ICES | **Institute for Clinical Evaluative Sciences**

Quality Assessment of Administrative Data (QuAAD): An Opportunity for Enhancing Ontario's Health Data

ICES Investigative Report

July 2007

Quality Assessment of Administrative Data (QuAAD): An Opportunity for Enhancing Ontario's Health Data

ICES Investigative Report

Authors

Karey Iron, MHSc

Douglas G. Manuel, MD, MSc, FRCPC

July 2007

Publication Information

Published by the Institute for Clinical Evaluative Sciences (ICES) © 2007

How to cite this publication

Iron K, Manuel DG. Quality assessment of administrative data (QuAAD): an opportunity for enhancing Ontario's health data. ICES Investigative Report. Toronto: Institute for Clinical Evaluative Sciences; 2007.

Authors' Affiliations

Karey Iron, MHSc
Health Information Officer, Institute for Clinical Evaluative Sciences

Douglas G. Manuel, MD, MSc, FRCPC
Senior Scientist, Institute for Clinical Evaluative Sciences

Associate Professor, Department of Public Health Sciences, University of Toronto

Associate, Manitoba Centre for Health Policy

Acknowledgments

About ICES

Ontario's resource for informed health care decision-making

The Institute for Clinical Evaluative Sciences (ICES) is an independent, non-profit organization that produces knowledge to enhance the effectiveness of health care for Ontarians. Internationally recognized for its innovative use of population-based health information, ICES' evidence supports health policy development and guides changes to the organization and delivery of health care services.

Key to our work is our ability to link population-based health information, at the patient-level, in a way that ensures the privacy and confidentiality of personal health information. Linked databases reflecting 12 million of 30 million Canadians allow us to follow patient populations through diagnosis and treatment, and to evaluate outcomes.

ICES brings together the best and the brightest talent under one roof. Many of our scientists are not only internationally recognized leaders in their fields, but are also practicing clinicians who understand the grassroots of health care delivery, making the knowledge produced at ICES clinically-focused and useful in changing practice. Other team members have statistical training, epidemiological backgrounds, project management or communications expertise. The variety of skill sets and educational backgrounds ensures a multi-disciplinary approach to issues and creates a real-world mosaic of perspectives that is vital to shaping Ontario's future health care system.

ICES receives core funding from the Ontario Ministry of Health and Long-Term Care. In addition, our faculty and staff compete for peer-reviewed grants from federal funding agencies, such as the Canadian Institutes of Health Research, and project-specific funds are received from provincial and national organizations. These combined sources enable ICES to have a large number of projects underway, covering a broad range of topics. The knowledge that arises from these efforts is always produced independent of our funding bodies, which is critical to our success as Ontario's objective, credible source of *Evidence Guiding Health Care.*

Executive Summary

Introduction

The administration of universal health care in Canada produces a wide variety of health information that is routinely collected and covers large segments of the population. The need for more robust and better quality health data has been recently highlighted by the Canadian Health Quality Council. Indicators of data quality typically include:

- Accuracy—do the data reflect the truth?
- Validity—do the data make sense?
- Completeness—do the data include all records that are collected?
- Comprehensiveness and coverage—do the data cover 100% of the intended population?
- Reliability—are the data reproducible?
- Timeliness—is there a short lag between data collection and use?
- Linkable—can the data be connected to other data to reflect health care system complexity?
- Anonymous—do the data adhere to jurisdictional privacy laws, procedures and practices?
- Usable—are the data organized, accessible, and provided in a format that can be easily used?
- Temporal consistency—are the data elements standardized to evaluate change over time?

Study

This report describes a quality assessment model to evaluate administrative data from the perspective of the data user, who is required to accurately analyze and interpret health data for service planning, decision support, performance measurement and research.

This report aims to answer the following questions:

- How is data quality defined?
- How is data quality evaluated in other jurisdictions?
- What is the scope of administrative data in Ontario, and how can the quality of these data be assessed?
- What is the context surrounding data quality evaluation?
- To what degree does data evaluation affect health planning, service delivery and, ultimately, the health of Ontarians?
- How can targeted data evaluation by users facilitate opportunities for data improvement in Ontario?

This is the first in a series of reports describing the development of the Quality Assessment of Administrative Data (QuAAD) model. QuAAD aims to provide data users with a comprehensive model to understand the full capacity of the data that are routinely used for health system planning, performance evaluation and research. The QuAAD model is not meant to be a panacea for evaluating data quality, but rather, an organic mechanism that can be shaped for particular uses and by targeted users. In the longer term, it is hoped that QuAAD will provide an opportunity for more focused dialogue about Ontario data and future system-wide data quality improvement as the health system evolves in Ontario.

Chapter 1 of this report is an environmental scan of international initiatives that attempt to define, evaluate and quantify the quality of routinely collected data for users. This section highlights data quality indicators and their various definitions which are used throughout the literature, as well as methods for quantification (where available). Chapter 2 highlights the data that are available in Ontario and their potential users. Chapter 3 illustrates the process of the QuAAD team and development of a data quality model for Ontario. Chapter 4 illuminates next steps on how the model can be systematically used to enhance data quality for Ontario users.

Key Findings

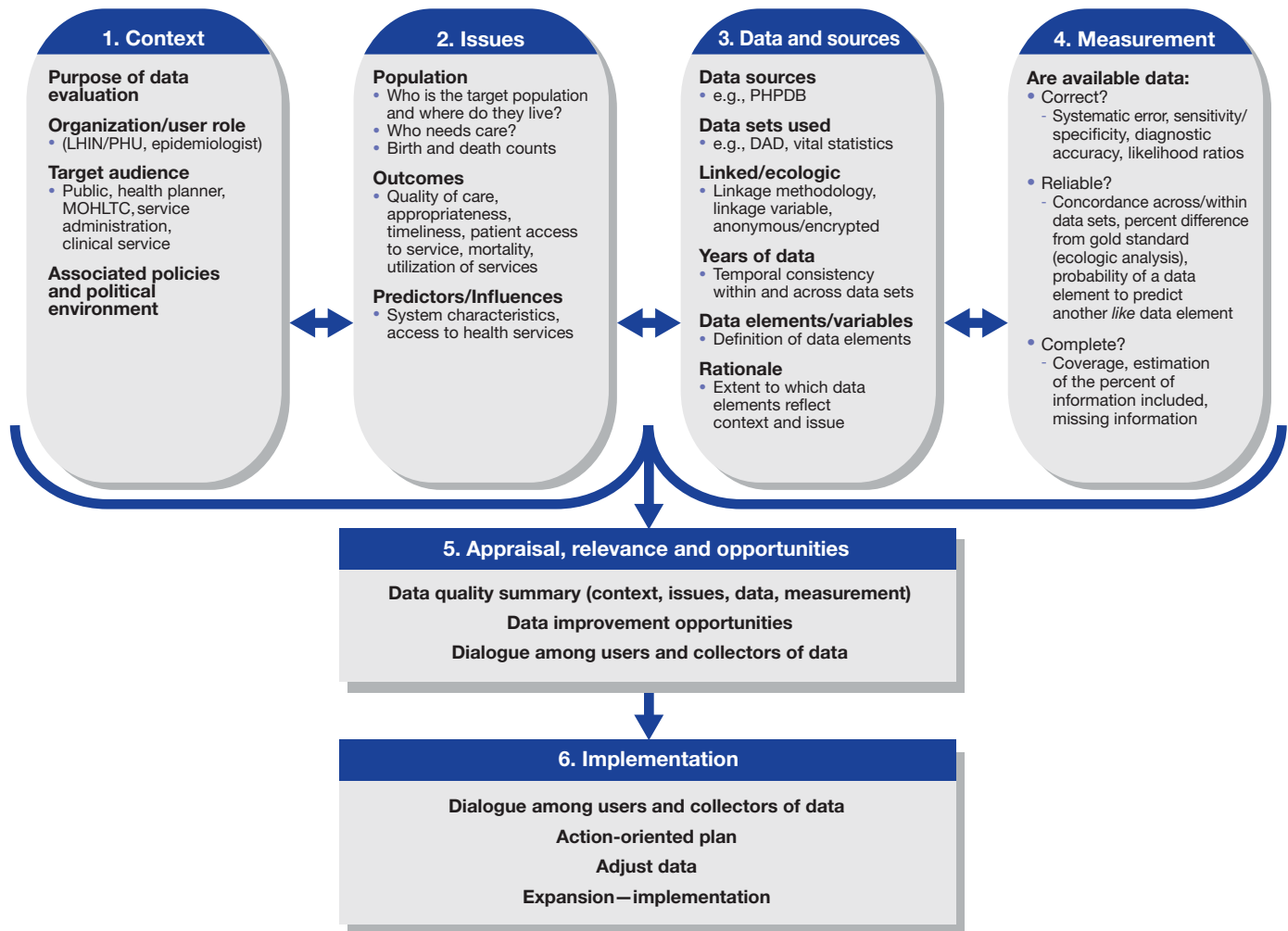The results of the environmental scan suggest that:

- Quality should be routinely and systematically evaluated for all generally-used data.

- Data quality must be user-defined.

- The evaluation and interpretation of data quality depends on the purpose for which the data are being used.

- The constructs of accuracy and validity are often confused.

- Accuracy (or truth) is an elusive construct and should not be expected.

- The most common ways to evaluate validity are concordance, comparability and inter-database reliability.

- Where available and where chart abstraction is not feasible, linked data should be made available and used to evaluate data quality.

- There are no standard methods for measuring data quality.

- There is a need for more investigation into evaluating data quality.

- The relevance of every data quality assessment requires full discussion.

The QuAAD team met over the course of six months to identify methods for developing a model. Ontario data user issues that impacted the development of an evaluative data assessment tool in the province were identified as follows:

- In general, the developers of the Ontario data have little discussion with data users; therefore, they have limited understanding of the relevance and quality of their data from the users' perspective.

- Data users include regional and Ministry of Health and Long-Term Care health system planners, hospital administrators, public health units, disease-based organizations, professional organizations, health services researchers and others whose mandate requires health information.

- There is a wide variation in data need among data users.

- Access to Ontario administrative data varies among users; this is constrained by a personal health information privacy law in Ontario.

- There are various platforms used and various versions of data that feed into these platforms; that is, there is currently no authoritative data source for all users.

- A variety of platforms exist to analyze the data, and this may be a limitation in developing a data quality tool that can be used across a variety of users.

- The calculation of some data quality indicators may only be possible using linked data.

Based on these issues and the results of the environmental scan, the QuAAD model was developed, as shown in Figure 1.

Figure 1 Quality Assessment of Administrative Data (QuAAD) model for the evaluation of health data quality in Ontario

1. Context

Purpose of data evaluation

Organization/user role
• (LHIN/PHU, epidemiologist)

Target audience
• Public, health planner, MOHLTC, service administration, clinical service

Associated policies and political environment

2. Issues

Population
• Who is the target population and where do they live?
• Who needs care?
• Birth and death counts

Outcomes
• Quality of care, appropriateness, timeliness, patient access to service, mortality, utilization of services

Predictors/Influences
• System characteristics, access to health services

3. Data and sources

Data sources
• e.g., PHPDB

Data sets used
• e.g., DAD, vital statistics

Linked/ecologic
• Linkage methodology, linkage variable, anonymous/encrypted

Years of data
• Temporal consistency within and across data sets

Data elements/variables
• Definition of data elements

Rationale
• Extent to which data elements reflect context and issue

4. Measurement

Are available data:
• Correct?
  - Systematic error, sensitivity/specificity, diagnostic accuracy, likelihood ratios

• Reliable?
  - Concordance across/within data sets, percent difference from gold standard (ecologic analysis), probability of a data element to predict another *like* data element

• Complete?
  - Coverage, estimation of the percent of information included, missing information

5. Appraisal, relevance and opportunities

Data quality summary (context, issues, data, measurement)

Data improvement opportunities

Dialogue among users and collectors of data

6. Implementation

Dialogue among users and collectors of data

Action-oriented plan

Adjust data

Expansion—implementation

DAD = Discharge Abstract Database; LHIN = Local Health Integration Network; MOHLTC = Ministry of Health and Long-Term Care; PHPDB = Provincial Health Planning Database; PHU = Public Health Unit

Implications

The QuAAD is a suggested model and may be used at local and provincial levels when evaluating administrative data. The model provides a framework to realize and develop opportunities for data enhancement in Ontario. It is intended to be used for a variety of issues, data sets and data sources.

The next step in the QuAAD project is to test the model across a variety of Ontario health care data sets, environments and users. It is hoped that various settings and user groups can be established to test the usability, flexibility and interpretability of the model and to provide an opportunity for users and data custodians to dialogue more formally about where further data quality initiatives are required at the data collection source.

Ontario is rich with data. As the health information system evolves, a more detailed and systematic dialogue for evaluating and enhancing public data is necessary so that health care system planning can accurately reflect and respond to the health care needs of Ontarians.

Contents

List of Exhibits

Exhibit 1.1   Environmental scan of peer-reviewed articles about data quality measurement and reporting, January 1996 to March 2006

Exhibit 1.2   Environmental scan of internet-based reports and commentary about data quality measurement and reporting, January 1996 to March 2006

Exhibit 2.1   Characteristics for examining health system performance

Exhibit 3.1   Quality Assessment of Administrative Data (QuAAD) model for the evaluation of health data quality in Ontario

Exhibit 3.2   Example of a general table used to describe the data source(s) and elements to be evaluated using the QuAAD model

Issue

The administration of universal health care in Canada produces a wide variety of health information that is routinely collected and covers large segments of the population. The Canadian Health Quality Council recently highlighted the need for more robust and better quality health data, stating that "without adequate data, it will be difficult to determine if additional public investments [in health care] are actually leading to better health—i.e., improved outcomes for the population overall".[2]

The concern about administrative health data quality is not new. In Ontario, as in other provinces, increased use and experience with these data over the past decade have led health service organizations, health planners and academic data users to think about methods to evaluate and describe data quality for particular purposes. Currently, there are few comprehensive and needs-based evaluation mechanisms for users to easily evaluate health information in Ontario.

This report describes *Quality Assessment of Administrative Data* (QuAAD), a proposed model to evaluate data quality. The perspective for this evaluation is from that of the data user, who is required to accurately analyze and interpret health data for service planning, decision support, performance measurement and research.

Introduction

Routinely-collected administrative health services data are useful for system-level health service planning, performance reporting and evaluation, clinical decision-making and research. Under the 1982 *Canada Health Act*,[3] all health services deemed essential by each province are universally covered by a publicly-funded single payer to all residents. As such, many of the provincial administrative health services databases cover much of the scope of health services provided to entire patient groups. In many provinces, including Ontario, these data are highly comprehensive and can be used for health services planning and health system performance measurement. Furthermore, these data are the basis for understanding information about:

- populations who use the health care system;
- the utilization and quality of health care services;
- general health outcomes, such as death and hospital readmissions; and,
- system characteristics, such as organization, funding, costing and efficiency.[4]

However, in some health services sectors, such as primary care, there are serious data inconsistencies and information gaps that make comprehensive performance measurement and accurate health system planning difficult.[4,5]

On December 1, 2005, the Government of Ontario released the *Information Management Strategy*, with the goal to "track and monitor how the health care system serves the public", and with a particular focus on "producing better data, supporting accountability and quality improvement through performance measurement, and supporting evidence-based decision-making".[6]

To this end, the Ministry of Health and Long-Term Care (MOHLTC) instituted a government-centred team to develop strategies aimed at improving the quality of routinely-collected health information, particularly those abstracted from medical charts and other types of written records.[7] These strategies included the creation of multi-disciplinary working groups to discuss:

- data collection practices within the standard medical chart coding mechanisms used in Ontario *(International Classification of Diseases, 10th Revision, Canada* for diagnoses and *Canadian Classification of Interventions* for procedures);
- a Ministry data quality framework; and,
- the enhancement of data user dialogue, using a voluntary interactive web-based program.[8–10]

At the same time, and in the decade preceding, various reports highlighted the major advantages, disadvantages and gaps in Ontario administrative data from the users' perspective. Several organizations have recommended the need to develop a simple model for data quality evaluation that is usable and relevant to health services planners, program developers and researchers, in an effort to better understand and improve the data that they use every day.[4,5,11]

According to Roos and others,[4,12–16] data quality assessment can be obtained through comparisons of linked information across data sets. This method is much less costly than chart abstraction validation and may be automated if defined evaluative measures are developed to assess the data quality from the users' perspective. According to many users of health data, data quality should be evaluated in a routine and *seamless* process,[17–19] but to date there are few comprehensive and systematic strategies available to users for this purpose.

Data quality indicators and definitions

Data quality has been defined as "the whole of planned and systematic procedures that take place before, during and after data collection to guarantee the quality of data in a database…*for its intended use*".[20] Several different data quality indicators are typically used (Appendix A), and can be grouped into four broad categories as illustrated in Figure 2.

Figure 2 Indicators of data quality grouped into four broad categories

Are the data correct?	Are the data complete?
• **Accuracy**—do the data reflect the truth?	• **Completeness**—do the data include all records that are collected? • **Comprehensiveness and coverage**—do the data cover 100% of the intended population?
Are the data reliable?	**Are the data usable?**
• **Reliability**—are the data reproducible? • **Validity**—do the data make sense?	• **Anonymity**—do the data adhere to jurisdictional privacy laws, procedures and practices? • **Linkability**—can the data be connected to other data to reflect health care system complexity? • **Timeliness**—is there a short lag between data collection and use? • **Usability**—are the data organized, accessible, and provided in a format that can be easily used? • **Temporal consistency**—are the data elements standardized to evaluate change over time?

Are the data correct and reliable?

The accuracy of data gathered from an external source compared to a primary source of information—deemed as the *truth* or *gold standard*—can be best evaluated using quantitative methods such as sensitivity (i.e., the proportion of people who are identified as having a characteristic according to the external data source, who truly have the characteristic), specificity (i.e., the proportion of people who do not have the characteristic according to the data, who really do not have the characteristic), and related measures such as positive and negative predictive values and likelihood ratios.

Most of the time, however, there is no way of defining the truth in administrative data. Various factors can affect the transfer of information that stems from a medical chart, such as "observer variation, poor documentation, illegibile charts, data loss, unavailability and timeliness of chart completion".[17] Therefore, the concept of validity, rather than accuracy, is more likely to be used when evaluating administrative data. Validity can be measured through blinded re-abstraction of data from medical charts compared to the original data abstract.[21] For example, measures such as inter-rater reliability (kappa statistic) or percent difference in counts between one data set and another, can be used to ecologically evaluate data. Although chart abstraction may be a necessary method for evaluating the validity of administrative data at the time of re-abstraction, it is costly, and, therefore, cannot be used for routine assessment of data quality. Linkage of like data elements from one data set to another can also be used to quantitatively evaluate the magnitude of concordance between two data sources. Stratified analyses can aid in determining whether the data are reliable across time, age, sex and geographic location, and to identify systematic bias within and across data sets. Other measures of reliability can be made, depending on the type of information being measured (discrete or continuous) and the purpose of the data evaluation. It is advisable to consult the epidemiologic literature to make sure that the correct measures are used.[22]

Are the data complete?

The quantification of data gaps is difficult to measure. For example, coverage can be evaluated by comparing two similar data sets such as population counts (using the MOHLTC central patient registry based on eligibility for universal health care) and census estimates (derived by Statistics Canada). Comparisons of administrative data to targeted survey information may also provide clues about the magnitude of coverage in administrative data. Completeness can be measured by evaluating the extent of missing or unknown data.

Are the data usable?

Evaluation of useful and usable data is more contextual than quantitative. A description of the data collection and processing mechanisms; who uses the data and for what purpose; how the data are organized and accessed by users; whether and how the data can be linked with other data sets; and the timeliness of the data, are all germane to understanding how well the data can be used and interpreted.[19]

About this report

This is the first in a series of reports describing the development of the Quality Assessment of Administrative Data (QuAAD) model. QuAAD is aimed at providing data users with a comprehensive model to understand the full capacity of the data that are routinely used for health system planning, performance evaluation and research. The QuAAD model is not meant to be the panacea for evaluating data quality, but rather, an organic mechanism that can be shaped for particular uses and by targeted users. In the longer term, it is hoped that QuAAD will provide an opportunity for more focused dialogue about Ontario data and future system-wide data quality improvement as the health system evolves in Ontario.

This report aims to answer the following questions:

- How is data quality defined?
- How is data quality evaluated in other jurisdictions?
- What is the scope of administrative data in Ontario, and how can the quality of these data be assessed?
- What is the context surrounding data quality evaluation?
- To what degree does data evaluation affect health planning, service delivery, and ultimately the health of Ontarians?
- How can targeted data evaluation by users facilitate opportunities for data improvement in Ontario?

Chapter 1 of this report is an environmental scan of international initiatives that attempt to define, evaluate and quantify the quality of routinely collected data for users. This section highlights data quality indicators and their various definitions used throughout the literature, as well as methods for quantification, where available. Chapter 2 highlights the data that are available in Ontario and their potential users. Chapter 3 illustrates the process of the QuAAD team and development of a data quality model for Ontario. Finally, Chapter 4 illuminates next steps on how the model can be systematically used to enhance data quality for Ontario users.

Chapter 1—Environmental Scan of Data Quality Measurement

An environmental scan of the peer-reviewed and internet-based literature was conducted to elucidate:

- automated processes for quantitative measurement of routinely collected data or registry quality in other jurisdictions;
- indicators and associated definitions of data quality assessment; and,
- methods for quantitatively measuring specific indicators of data quality.

How the scan was conducted

The peer-reviewed literature was searched (Medline and Embase) from January 1996 to March 2006 using the following key words:

- administrative data
- registry
- electronic data
- data quality
- data improvement

- framework
- model
- data quality reporting
- data quality benchmarking
- measurement indicators

Primary key words for exclusion were:

- primary data
- chart abstraction
- research data

Eighteen peer-reviewed articles were found using this strategy and eight were included in the analysis. Two articles were used as theoretical or background information.

In addition, data quality reports from select organization and association websites known to use administrative data were sought using *Google*, and *Scirus* for additional information. Six reports were included from the following searched websites:

- Manitoba Centre for Health Policy and Evaluation
- Centre for Health Services and Policy Research (British Columbia)
- Saskatchewan Health Quality Council
- Alberta Heritage Foundation for Medical Research
- Health Quality Council of Alberta
- Alberta Institute for Health Economics
- Rand Corporation (United States [US])
- Centers for Medicare & Medicaid Services (US)
- Ontario Ministry of Health and Long-Term Care

- Statistics Canada
- Canadian Health Quality Council
- Canadian Health Services Research Foundation
- Veterans Affairs (US)
- National Health Service (United Kingdom)
- Agency for Healthcare Research and Quality (US)
- Canadian Institute for Health Information
- Kaiser Permanente (US)
- Blue Cross and Blue Shield Association (US)

Findings from peer-reviewed literature

Exhibit 1.1 characterizes the peer-reviewed articles included in this analysis. A more detailed summary is available in Appendix B. In general, the literature suggested that the definition of data quality and the indicators to measure data quality need to be user-defined and flexible.[18] Stein et al. suggested that data quality can only be measured in the context of a targeted question or topic area, with careful evaluation of the data fields that can be used for measurement.[23]

Most of the articles attempted to find methods to measure data accuracy through, for example, comparability, concordance and inter-rater reliability.[12,13,23] Stein et al. suggested a staged approach to data quality evaluation by highlighting direct concordance of data elements, supportive or late matches, and no concordant data elements at all.[23] Roos et al. used linked population-based hospital discharge abstracts and physician claims to assess concordance and external validity for similar data elements across hospital discharge abstracts and physician claims data using time-sequenced matching techniques.[12] They concluded that this approach is easily generalizable and provides insight into the "strong and weak points" of administrative data. This method was also successfully used to examine the validity of hospital coding and physician claims submissions for surgery in previous reports which used Ontario data.[24,25]

In the examination of Saskatchewan hospital discharge abstracts, Rawson and D'Arcy suggested evaluating internal consistency and external validity through concordance of data elements, contextual consistency across time and generalizability across situations.[14]

Peabody et al. characterized three error types contributing to the assessment of accuracy:[15]

- beta error, as measured by the percent of occasions that an incorrect diagnosis or data entry was made by the physician completing the medical chart;

- gamma error, as measured by missing information in the medical chart; and,

- delta error, as characterized by incorrect data entry or interpretation of the medical chart.

However, these error types may not be measurable without a detailed chart review. Holt and Jones suggested that without measurement of variance and bias, accuracy cannot be reliably measured.[18]

Silcocks and Robinson developed a statistical bootstrapping method to evaluate the completeness of data based on a survival distribution for cancer patients. Incidence and death data were required in order to produce these completeness estimates;[26] however, a proxy measure of completeness can also be evaluated using comparisons across data sets.

Exhibit 1.1 Environmental scan of peer-reviewed articles about data quality measurement and reporting, January 1996 to March 2006

A. Peer-reviewed studies

Citation	Type of study/ data used	Are the data correct?	Are the data reliable?	Are the data complete?	Are the data usable?	Comments
• Byrne N et al.[27]	• Systematic review		• Percent level agreement between patient register and case note; kappa correlation and specificity			• Validity method should be relevant to data requirements.
• Peabody JW et al.[15]	• Comparison of outpatient physician claims and simulated patient encounters	• Diagnostic accuracy– three error types - beta: percent incorrect diagnosis by physician - gamma: missing data on form - delta: incorrect data entry				• Not generalizable since standardized simulated patient encounters were used for comparison with chart information.
• Roos LL et al.[13]	• Used linked hospital discharge abstracts and physician claims to assess concordance		• Concordance using kappa statistic		• Used linked provincial data sets	• Using linked data may be a way to reduce the necessity of primary data collection to check validity. • Test/re-test methods may be used more extensively.
• Silcocks PBS and Robinson D[26]	• Cancer incidence and survival data			• Developed a bootstrapping program to estimate completeness using capture/ recapture methods		
• Arts DG et al.[20]	• Literature review	• Types and causes of error				

Citation	Type of study/ data used	Are the data correct?	Are the data reliable?	Are the data complete?	Are the data usable?	Comments
• Stein HD et al.[23]	• Determination of how well particular data elements compare and answer questions		• Concept of concordance partitioned into: positive (confirmatory), supportive (complementary), contradictory, or absent altogether			• Truth is difficult to assess. Proxy measures can only assess validity.
• Rawson NS and D'Arcy C[14]	• Examination of concordance using like data elements across data sets		• Concordance using internal and contextual consistency measures	• Used linked hospital discharge abstracts, physician claims and drug claims data	• Used provincial data	
• Roos LL et al.[12]	• Examination of linking flow of information across continuity of care (physician claims, hospital discharge)		• Internal and external consistency	• Examination of: - perfect match - late perfect match (medical claim late) - match on two keys (date and MD/tariff) - late match on two keys (medical claim late, MD/tariff agree) - match on date only - late match only	• Approach relies on linked administrative data comparisons	

B. Peer-reviewed editorials/commentary

Citation	Type of study/data used	Report summary
• Brennan PF and Stead WW[30]	• Editorial	• Conceptual discussion about concordance, validity measures and assessment of accuracy or truth.
• Hierholzer WJ Jr.[17]	• Discussion of data quality issues in individual patient record data sets, public health and vital statistics, research data and bibliographic reviews	• Calls for defining context in data quality assessment; promotes linkage of data, standard methods for data collection. States that misclassification bias may result from large data collection initiatives. Proposes that all data sets should be consistently audited.

Findings from internet-based reports

Exhibit 1.2 describes the characteristics of the various internet-based reports and commentary that were retrieved from the environmental scan. In general, the reports illustrate the necessity and complexity of systematically evaluating administrative data quality for data custodians who develop databases and data user reports.

Exhibit 1.2 Environmental scan of internet-based reports and commentary about data quality measurement and reporting, January 1996 to March 2006

Citation	Type of study/data used	Report summary
• Canadian Institute for Health Information (CIHI) data quality framework 2005[28]	• Discharge abstracts and other institution-based information collected from patient charts	• Provides a data quality framework to assess the data collected by CIHI. The framework includes measures on: accuracy, timeliness, comparability, usability and relevance.
• Data quality framework in Ontario[10]	• To assess institutional, clinical and financial data quality for the Ministry of Health and Long-Term Care and other Ontario users	• Conceptual framework focusing on: timeliness (fresh and current), validity (accurate, complete, comprehensive), reliability (consistent) and usability (relevant, interpretable, accessible). Presents a proposal for a voluntary discussion and action system (Data Quality Issues and Action Management System [DIAMS]) for provincial users.
• Improving health care data in Ontario[5]	• Highlights uses and limitations of Ontario administrative data with recommendations for improvement	• Data quality indicators defined: completeness, accuracy, timeliness, linkability, anonymity, usability and consistency over time. Recommends that measurement should be contextual and that the development of a standardized quantitative model is needed.
• Moving toward a better health data system for Ontario[4]	• Highlights issues and Ontario data needs/gaps to measure health system performance	• Provides a framework to examine health system performance focused on population characteristics, quality of care, health care outcomes, health services utilization and health system/organization issues; and provides real-life scenarios to highlight Ontario data needs/gaps. Provides recommendations for a centralized data agency in Ontario, and systematic and quantitative data quality evaluation.
• Holt T and Jones T[18]	• Defines an authoritative data source and data quality measures for the United Kingdom	• Stresses that data quality is user-defined. Measure of accuracy could be obtained through bias and error measurement but it is a relative construct and therefore, should not be expected. Other quality measures discussed are relevance, coherence and consistency, continuity and timeliness.
• Veterans Affairs (VA) fiscal year performance and accountability report[29]	• Definition of data quality in VA	• Discussion of data quality definitions. Measures include: completeness, validity, consistency, accuracy and appropriateness of use.

Conclusions from environmental scan

The scan suggested that:

- Quality should be routinely and systematically evaluated for all generally-used data.

- Data quality must be user-defined.

- The evaluation and interpretation of data quality depends on the purpose for which the data are being used.

- The constructs of accuracy and validity are often confused.

- Accuracy (or truth) is an elusive construct and should not be expected.

- The most common ways to evaluate validity are concordance, comparability and inter-database reliability.

- Linked data, where available, should be used to evaluate data quality (when chart abstraction is not feasible).

- There are no standard methods for measuring data quality.

- There is a need for more investigation into evaluating data quality.

- The relevancy of every data quality assessment requires full discussion.

Finally, Holt and Jones suggested that "quality is not so much an absolute property of a statistical estimate but is related to the purpose for which the estimate is used."[18] This must be considered each time an estimate is derived.

Chapter 2—Uses of Ontario Administrative Data

What are administrative data in Ontario?

For the purposes of this report, Ontario administrative data are sets of individual records that are routinely collected for the purposes of payment and funding and are mandated and/or collected by the Ministry of Health and Long-Term Care (MOHLTC). These data are immense in breadth and scope, and provide a vast array of potential uses. Many reports have been published describing these data. Broadly, they can be categorized into three discrete topic areas:

- data about populations;
- data about the use of services within institutions; and,
- data about physician services.

Data about populations

The Registered Persons Database (RPDB), collected, maintained and operated by the MOHLTC, is a historical registry that logs demographic information about persons eligible for health care in Ontario using the unique eligibility health card number (HCN) as the individual reference. Although these data are maintained for the management of publicly-funded insurance in Ontario, they are also used anonymously by restricted users to gain residential, age and gender information about the Ontario population for health care planning purposes. Census data, provided by Statistics Canada, can also be used in aggregate form to count the number of people living in the province—by age, sex, socioeconomic status and area of residence.

The Ontario Registrar General (ORG), Ministry of Government Services, manages vital statistics for the province. At the time of death, a certificate must be completed by the attending physician, coroner or funeral home director. The information is collected and managed by the ORG and then sent to Statistics Canada. Out-of-province deaths are also included. The data are then sent back to the ORG and the MOHLTC. For tracking health insurance eligibility, the fact of death of an Ontario resident is obtained by the MOHLTC from the ORG and linked to the RPDB through matching techniques.

Information about who requires health care can be gleaned from a variety of health care data sources—the diagnosis and procedure information in health care utilization data can be used in many cases as a proxy for why people need health care.

Data about the use of services within institutions

These data are collected by expert health data information specialists in hospitals, emergency departments, community home care organizations and other institutions (e.g., long-term care facilities). The collection of these data is mandated by the MOHLTC and organized by standards and procedures facilitated by the Canadian Institute for Health Information (CIHI). The primary purposes of these data are costing and funding; however, they are often used for health system planning and health services research. Generally, these data are well managed and easily used for health system planning, evaluation and research.
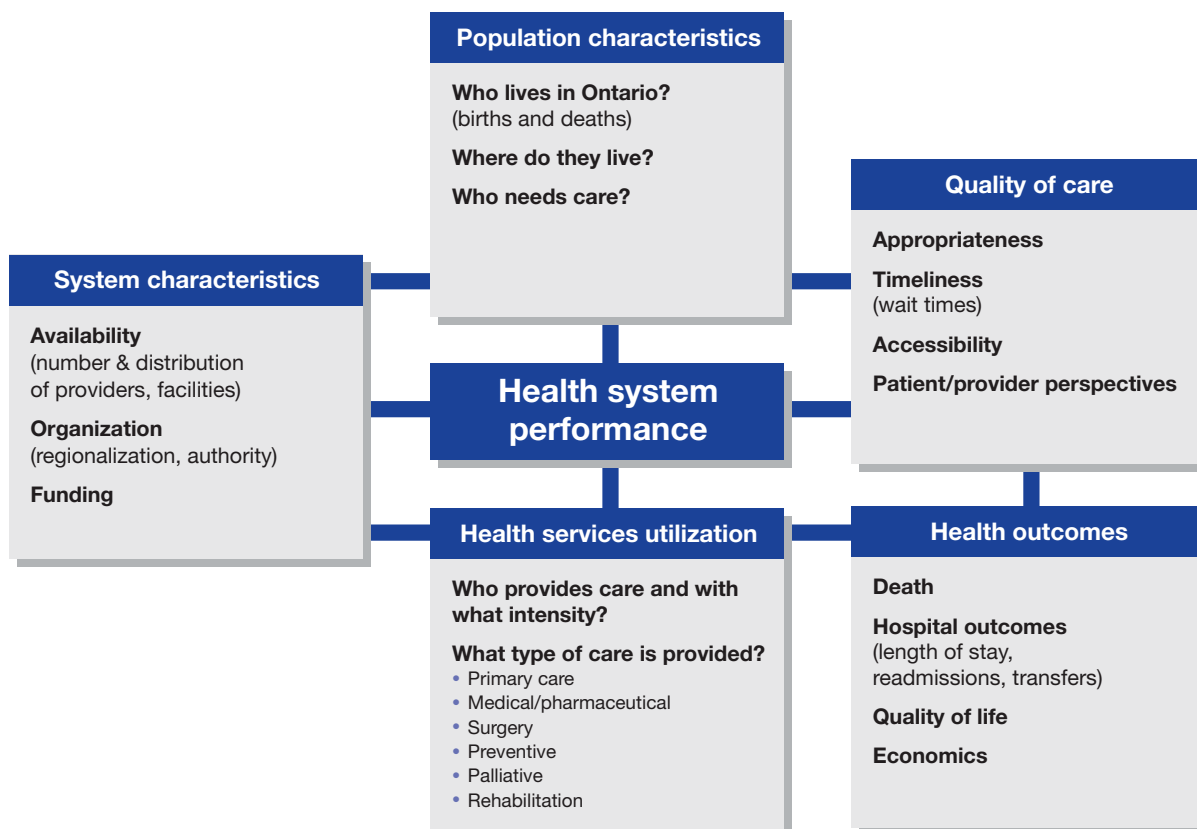
Data about physician services

Health services provided to patients by physicians who receive fee-for-service payment through the universal Ontario Health Insurance Plan (OHIP) are captured through the Medical Claims Database (operated by the MOHLTC). These data are used primarily for remuneration purposes, but are also used to understand health service utilization and physician service provision. Information about the medical specialty of physicians who are licensed to practice in Ontario is also used anonymously to understand physician practices and health system characteristics, such as health human resources. These data are very useful to health data users, but access is limited and the data often lack specificity for focused uses.

Characteristics for examining health system performance

A regularly updated compendium of Ontario's administrative data is available from the MOHLTC.[31] These data are used for many purposes and some can be anonymously linked to each other at the patient or physician level by organizations that have the privacy authority to do so.[1] Exhibit 2.1 provides an example of the types of issues that can be addressed for health system performance measurement through the use of Ontario administrative data.[4] For example, the data may be used to develop indicators about who requires care, whether the care provided was timely and based on evidence, and the historical utilization of certain health services, as well as to understand organizational aspects of the health care system for planning purposes.

Exhibit 2.1 Characteristics for examining health system performance

Population characteristics

Who lives in Ontario?
(births and deaths)

Where do they live?

Who needs care?

System characteristics

Availability
(number & distribution
of providers, facilities)

Organization
(regionalization, authority)

Funding

Health system performance

Quality of care

Appropriateness

Timeliness
(wait times)

Accessibility

Patient/provider perspectives

Health services utilization

Who provides care and with what intensity?

What type of care is provided?
• Primary care
• Medical/pharmaceutical
• Surgery
• Preventive
• Palliative
• Rehabilitation

Health outcomes

Death

Hospital outcomes
(length of stay,
readmissions, transfers)

Quality of life

Economics

Adapted from: Iron K. Health data in Ontario: taking stock and moving foward. *Healthcare Quarterly* 2006: 9(3); 24–6.

Chapter 3—Development of the QuADD Model

In December 2005, the Ontario government released a health information strategy to "track and monitor how the health care system serves the public", with a particular focus on "producing better data, supporting accountability and quality improvement through performance measurement, and supporting evidence-based decision-making".[6]

To this end, the Ministry of Health and Long-Term Care (MOHLTC) instituted a government-centred team to improve the quality of routinely collected health information, particularly those abstracted from medical charts and other types of written records.[7] The developed strategies included the creation of multi-disciplinary working groups to discuss data collection practices within the standard medical chart coding mechanisms used in Ontario (*International Classification of Diseases, 10th Revision, Canada* [ICD-10-CA] for diagnoses and *Canadian Classification of Interventions* [CCI] for procedures), a Ministry data quality framework and the enhancement of data user dialogue using a voluntary interactive web-based program.[8–10]

At the same time, and in the decade preceding, various reports highlighted the major advantages, disadvantages and gaps in Ontario administrative data from the users' perspective. Several organizations have recommended the need to develop a simple model for data quality evaluation that is usable and relevant to health services planners, program developers and researchers, in an effort to better understand and improve the data that they use.[4,5,11]

In the spring of 2006, a team with experience in using the array of Ontario administrative data for a multitude of purposes, met to discuss the development of a data quality model that could be routinely reported to Ontario users.

The project, called Quality Assessment of Administrative Data (QuAAD), aimed to:

- evaluate methods for examining administrative data in Ontario;
- assess the feasibility of standardized reporting of data quality for targeted uses and users;
- in the short term, inform health planners and researchers of the limitations of the currently available data for health system planning, measurement and evaluation and health services research; and,
- in the long term, act as a foundation and discussion forum for future system-wide data quality improvement.

Overall goals

The overarching goal of QuAAD was to develop a model that would evaluate data quality and would systematically become part of the discussions surrounding data dissemination and interpretation processes as the health information system evolves in Ontario. It was critical that dissemination of this information would work within existing reporting systems and would promote new educational opportunities towards enhancing the data at source, by identifying the uses and limitations of administrative health care data.

The QuAAD model was to take advantage of linked and/or unlinked administrative data to measure discrete data quality indicators as defined in the literature, and to create new methods where necessary. It was meant to assess data quality from the *top-down* (or users' perspective), while recognizing the *bottom-up* data collection improvement projects initiated by the MOHLTC and others in Ontario.

Development Process

The QuAAD team met over the course of six months to identify methods of developing a model. Several data user issues that impacted the development of an evaluative data assessment tool in Ontario were identified.

- In general, the developers of the Ontario data have little discussion with data users and, therefore, have limited understanding of the relevance and quality of their data from the users' perspective.

- Data users include regional and MOHLTC health system planners, hospital administrators, public health units, disease-based organizations, professional organizations, health services researchers and others whose mandate requires health information.

- A wide variation of data need exists among data users.

- Access to Ontario administrative data varies among users—constrained by a personal health information privacy law in Ontario.[1]

- There are various platforms used, and various versions of data that feed into these platforms; that is, there is currently no authoritative data source for all users.

- The variety of platforms to analyze the data may be a limitation to developing a data quality tool that can be used across a variety of users.

- It is unclear whether the data quality tool will be disseminated with data or whether data users will create their data quality report based on their needs.

- The calculation of some data quality indicators may only be possible using linked data which are not available to all users.

Based on these issues and the results of the environmental scan, the following key questions were identified for evaluating data:

- What is the context of the data evaluation?

- What are the issues, or foci, for which the data are being used?

- What are the data associated with the issues and context?

- Are the data correct? Reliable? Complete?

- How relevant is the data evaluation to the context?

- How will the data quality issues be disseminated and/or discussed?

- What are the mechanisms to improve the data?

Exhibit 3.1 integrates the concepts that stem from the above questions into a proposed model for evaluating data quality in Ontario. The model relies heavily on qualitative contextual information to support the quantitative analysis. In addition, there is an evaluative appraisal to support opportunities and implementation for adjustment, improvement and expansion of administrative data sources. The following six discrete areas for evaluation are presented in Exhibit 3.1 and will be discussed in turn:

1. Context

2. Issues

3. Data and sources

4. Measurement

5. Appraisal, relevance and opportunities

6. Implementation

1. Context

The context in data quality assessment refers to the multitude of factors that could potentially influence the uses of data, interpretation of data quality results and/or implementation of proposed changes to data collection practices. The articulation of context provides the foundation for the rest of the data quality assessment, in that it underscores the environment in which the rest of the analysis will occur. According to Dobrow, Goel and Upshur (2004),[32] internal factors are intrinsically tied to articulation of purpose including: temporal boundaries (years of evaluation); participants and target audience of the evaluation; political factors; environment in which the evaluation is occurring; and any other factor that may influence decision-making. External factors are fixed and describe the environment in which the evaluation is conducted and applied—including the political and social landscape.

The context can be formed as part of the background to the data quality evaluation, but at minimum should explicitly outline:

- The purpose of the evaluation
- The organization that is writing the evaluation
- The intended users of the information
- Policies or current environmental factors that may affect the data quality or the uptake of the data quality evaluation

Exhibit 3.1 Quality Assessment of Administrative Data (QuAAD) model for the evaluation of health data quality in Ontario
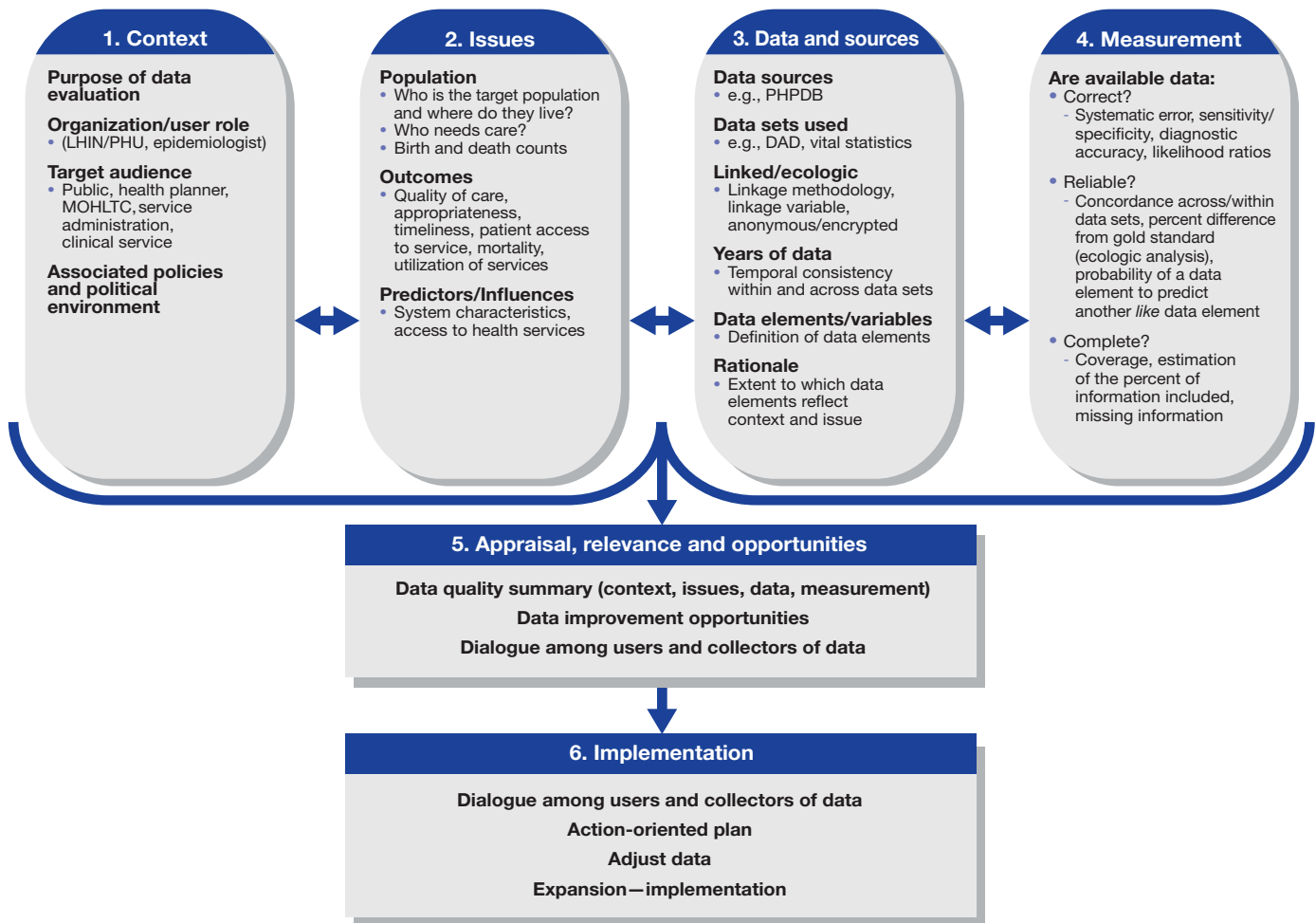
DAD = Discharge Abstract Database; LHIN = Local Health Integration Network; MOHLTC = Ministry of Health and Long-Term Care;
PHPDB = Provincial Health Planning Database; PHU = Public Health Unit

2. Issues

The description of issues highlights the scope of the data quality evaluation since the quality of data will vary according to the issues surrounding the evaluation. For example, the quality of the physician billing claims data will vary according to whether a surgical procedure or a consultation is being evaluated.[24,25] The characteristics for health system performance previously described can be further grouped to define the scope of the issues, and this will necessarily tie into the purpose and context of the data quality evaluation highlighted in the first evaluation area (context) presented above. For example, the scope of the issues pertaining to a combination of the following three areas will require identification:

- Population
  - Who is the target population and where do they live?
  - Who needs care?
  - How many people have died?
  - How has the population changed over time?
  - What is the geographical distribution of the population?

- Outcomes
  - Quality of care
  - Appropriateness
  - Timeliness
  - Patient access to services
  - Mortality rates
  - Utilization of services

- Predictors/influences
  - System characteristics
  - Access to health services

3. Data and sources

Information about the data source(s) and the data elements used for evaluation are a necessary component of QuAAD. A clear articulation is necessary of the administrative purpose of the data set, who manages it, how it is used for health system planning and performance, how it is manipulated from its original state to one that is used by health system planners and how it is organized. A general table such as the one presented in Exhibit 3.2 could efficiently organize the data information.

Exhibit 3.2 Example of a general table used to describe the data source(s) and elements to be evaluated using the QuAAD model

Data set	Source	Managing organization	Data use	Years available	Data elements	Rationale for use
• DAD	• PHPDB	• CIHI/MOHLTC	• Linked ecologically by age and sex	• 2000/01–2004/05	• ICD-10-CA diagnosis • Diagnosis type • CCI procedure code	• To create a disease-based cohort
• NACRS	• PHPDB	• CIHI/MOHLTC	• Linked ecologically by age and sex	• 2000/01–2004/05	• ICD-10-CA diagnosis • ICD-10-CA diagnosis type • CCI procedure code	• To create a disease-based cohort

CCI = Canadian Classification of Health Interventions; CIHI = Canadian Institute for Health Information; ICD-10-CA = International Classification of Diseases, 10th Revision, Canada; MOHLTC = Ministry of Health and Long-Term Care; NACRS = National Ambulatory Care Reporting System; PHPDB = Provincial Health Planning Database

4. Measurement

The QuAAD project emanated from questions about quantitatively measuring data quality. The environmental scan suggested that the quantitative measurement must be contextual in order to be of relevance to users, and that there may be more than one way of measuring data quality. The various data quality indicators can be broadly grouped based on their ability to answer three basic questions as below:

- Are the data correct? Do they reflect the truth?
  - Systematic error
  - Sensitivity/specificity
  - Diagnostic accuracy
  - Likelihood ratios

- Are the data reliable? Are they generalizable and reproducible?
  - Concordance across/within data sets
  - Percent difference from gold standard (ecologic analysis)
  - Probability of a data element to predict another *like* data element

- Are the data complete? What and how much information is missing?
  - Coverage
  - Estimation of the percent of information included
  - Missing information

As previously mentioned, quantitative evaluations are contingent on the organization of the data (ecologic or individual level), availability of and access to the data, as well as the purpose of the evaluation. Careful selection of quantitative measures and how they relate to the qualitative assessment is necessary.

5. Appraisal, relevance and opportunities

The appraisal is a distillation of all the components above. This information can best be presented in point form, highlighting the key context and issues, as well as data and measurement messages. The appraisal should focus on answering questions such as:

- How does this analysis reflect the initial data quality question?
  - Did the analysis reflect the defined purpose?
- How will the data quality analysis affect health system planning?
- Which stakeholders will be affected by the data quality analysis?
  - Planners?
  - Policy makers?
  - Administrators?
- How does data access impact the evaluation?
- How will this information be communicated to data users?
- What health system policies will be affected? How will they be affected?
- What are the data needs and where are the gaps?

6. Implementation

The implementation section of the analysis provides an action-oriented plan for data quality enhancement opportunities and would involve all named data users and stakeholders. Recommendations might include a temporary data set adjustment to reflect the results of the evaluation, so that data analysis activity can continue. This could be produced using data linkage or acquisition of data elements to enhance the interpretation of the data for users. At the same time data improvement opportunities would be identified, again with the input of all stakeholders, and would refer back to the context, issues, data, measurement and appraisal highlighted in the QuAAD model. The actions defined in this section need to be feasible and goal-oriented to ensure successful implementation. Specific groups of individuals who have the legislative authority to action change would need to be involved in this process. Finally, implementation plans are required, including funding and human resources for data enhancement and feedback.

Chapter 4—Next Steps in Using the QuAAD Model

The QuAAD is a suggested model and may be used at the local and provincial levels when evaluating administrative data. The model provides a framework to realize and develop opportunities for data enhancement in Ontario. It is intended to be used for a variety of issues, data sets and data sources.

The next step in the QuAAD project is to test the model across a variety of Ontario health care data sets, environments and for a variety of users. It is hoped that a variety of settings and user groups can be established to test the usability, flexibility and interpretability of the model, and to provide an opportunity for users and data custodians to dialogue more formally about where further data quality initiatives are required at the data collection source.

Ontario is *data rich*. As the health information system evolves, a more detailed and systematic dialogue for evaluating and enhancing our public data is necessary so that health care system planning can accurately reflect and respond to the health care needs of Ontarians.

Appendix A—Definitions Used Throughout the Literature

Term	Definition
• Data quality	• The whole of planned and systematic procedures that take place before, during and after data collection to guarantee the quality of data in a database…for its intended use.[20]
• Accessibility	• Users provided with information about quality of the statistics and methods used to derive figures.[33] • Ease with which data can be obtained from Agency. Suitability of the form that information can be accessed. Cost may be included.[33] • Affordability, multiple dissemination formats, and selectivity of data presentation.[34]
• Anonymity	• Adherence to strict privacy regulations (PHIPA, 2004).[5]
• Comparability	• Allowing comparisons over time, geographies and sub-populations.[18,19,35] • Extent to which databases are consistent over time using standard conventions.[28]
• Accuracy	• Distance between estimate and truth.[19,35] • How well information in (or derived from) the database or registry reflects the reality it was designed to measure.[28] • Extent to which the source information is correct.[27] • Reflect actual utilization and patient-provider characteristics.[5] • The information is correct.[4] • Extent to which the registered data are in conformity with the truth.[20] • Degree to which the information correctly describes the phenomena it was designed to measure.[33] • Degree to which data correctly estimates or describes the quantities or characteristics they are designed to measure.[34]
• Coherence	• Consistent standard.[19,35] • Degree to which data can be successfully brought together with other information, within a broad analytic framework over time. Standard concepts, classifications and target populations promote coherence as does the use of common methodology across data. Does not imply full numeric consistency.[33]
• Completeness	• Coverage reflecting user needs.[19,35] • Include all settings, providers and populations.[5] • Extent to which all necessary data have actually been registered.[20]
• Concordance	• The level of agreement between two data elements or two data sources; does not necessarily represent truth.[30] • Examining concordance at the diagnosis level may provide more granularity.[14]
• Consistency	• Temporal; consistent over time for temporal analyses.[5] • Data elements consistent to evaluate change over time.[4] • **Contextual**: logically, time-sequenced relationships; **external**: across data source; and, **internal**: within data source.[14]
• Error	• Amount of bias or systematic/random variance in evaluating accuracy; also described in coverage, sampling, non-response and response rate.[35] • Amount to describe variance in coverage, sampling, non-response, response, processing and dissemination.[34]
• Interpretability	• Ease with which user understands and properly uses and analyses the data; degree of documentation available.[34]
• Linkability	• Two or more separate sources of information to specify individuals present on both files.[13] • Connect different services that people receive to provide analyses that reflect the complexity of the health care system.[4,5]
• Relevance	• Degree to which a database or registry meets the current and potential needs of users.[28]
• Timeliness	• Punctuality in disseminating results to users[19,35] or use in real-time management and planning.[5] • Short lag between data collection and use.[4] • Delay between the reference point (or end of the reference period) to which the information pertains and the date on which the information becomes available—a trade-off with accuracy and relevance.[33]
• Usability	• The ease with which a database or registry may be understood and accessed.[28]

PHIPA = *Personal Health Information Protection Act.*

Appendix B—Detailed Summary of Environmental Scan of Peer-reviewed Articles on Data Quality Measurement and Reporting, January 1996 to March 2006

Citation	Organization/country	Type of data	Measures	Methods	Authors' comments
• Byrne N et al.[27]	• University College Medical School, King's College, London	• Systematic review via search of Medline (1966–2004) and EMBASE (1980–2004) • Search terms: 'validity or reliability', and 'register or databases' • Exclusion terms: psych, research, research register	• There was no gold standard found in the literature for register quality • 278 relevant articles were retrieved; 12 fit inclusion/exclusion criteria • Validity versus accuracy	• Developed indicators to assess validity of register studies since none existed previously • Measured percent level of agreement between case note and register, patient representativeness, kappa correlation and specificity	• Definitions were given of what was being measured, but validity measure was not explicit • Examination of primary relative to subsequent secondary diagnosis • Diagnosis stability unknown • Rater blinding/inter-rater reliability not captured • No assessment of the nature of validation for its quality and relevance to data requirements
• Peabody JW et al.[15]	• Veterans Administration (VA)	• Outpatient physician claims versus prospective data through standardized patient encounters	• Diagnostic accuracy– three error types - beta: percent incorrect diagnosis by physician - gamma: missing data on form - delta: incorrect data entry	• Used standardized patients to simulate patient/physician encounters • Abstracted associated charts • Examined associated administrative data • Examined primary diagnosis and comorbidity data	• Only three or four diagnosis conditions used • Only one visit assessed • No relationship between system incentives and error types • No distinction between incorrect chart coding and physician's diagnosis that may not have been recorded • Not generalizable
• Roos LL et al.[13]	• Manitoba Centre for Health Policy	• Provincial administrative data in Canada - Population-based registries - Hospital discharge - Physician claims	• Concordance (kappa)	• Record linkage and re-abstraction (reliability) • Coverage • Define gold standard • Concordance • Kappa	• Using linked data may be way to reduce necessity of primary data collection to check validity • Test/re-test methods may be used more extensively

Citation	Organization/country	Type of data	Measures	Methods	Authors' comments
• Silcocks PBS et al.[26]	• Trent Institute for Health Services Research, Nottingham, United Kingdom	• Cancer registry for one health region • Incidence data • Death data	• Development of program to estimate completeness using 'flow' method by Bullard et al.[36] compared to capture-mark-recapture • Probability that patient with cancer is still alive after diagnosis (survival distribution) • Probability that death certificate of patient who dies after diagnosis includes a mention of cancer • Probability that patient surviving after diagnosis is still unregistered	• Needs incident data with survival time and death data for cancer patients to produce estimates of completeness	• Estimation of coverage or completeness using bootstrapping methods

• Arts DG et al.[20]	• Department of Medical Informatics, Academic Medical Centre, Amsterdam	• National Institute for Health and Clinical Excellence (NICE) data set (registry for Intensive Care Unit treatment)	• Types and causes of errors
			• Procedure to minimize errors in registry
			• NICE versus re-abstracted data (20 random patients)
			• Underscores importance of definitions
			  - inaccurate
			  - incomplete
			  - measures
				• Automated literature search; key words: data quality, registries, data collection, validity, accuracy, quality control, 1990 to 2000
				• Clinical trial or registry
				• Procedures for assurance of data quality at start-up of registry
					• Literature review on framework identifies error types

• Stein HD et al.[23]	• Veterans Administration (VA)	• Clinical data repository	• Concordance: positive (confirmatory), supportive (complementary), contradictory, or absent altogether	• First identify question • Determine how well particular data elements answer the question • Comparison of data fields used to answer question	• Did not assess accuracy (i.e., truth) but how well data elements compare to come up with proxy of truth • Outlined limitations of administrative data

Citation	Organization/ country	Type of data	Measures	Methods	Authors' comments
• Rawson NS and D'Arcy C.[14]	• Memorial University, Newfoundland and College of Medicine, University of Saskatchewan	• Hospital discharge abstracts, physician claims data, drug claims data	• Consistency with external information (patient charts) • Consistency between administrative data files; hospital and physician claims for same event	• Concordance (exact match) • Internal consistency or agreement • Surrogate markers (drug data) • Contextual consistency (logical time-sequencing of events) • Guidelines • External validity–generalizability	• Want to determine how closely the information represents usual clinical practice • Recommended combination of all types of internal, external and contextual consistency
• Roos LL et al.[12]	• Manitoba Health Services Commission	• Patient registry • Hospital data • Physician claims data • Chart abstraction • More than one physician claim for the same patient, same diagnosis • Linking data sets	• How well data can be linked • How much agreement in linkage • Tracing of flow of information • Linking on data of procedure, surgical procedure and surgeon	• Identify perfect match • 'Late' perfect match (medical claim late) • Match on two keys (date and MD/tariff) • 'Late' match on two keys (medical claim late, MD/tariff agree) • Match on date only • 'Late' match only	• Reliance of linked administrative data comparisons • Approach easily generalizable • Reliance on these data suggest that we must know the "strong and weak points" • Defines a framework

References

1. *Personal Health Information Protection Act, 2004. S.O. 2004*, c.3, Schedule A. Accessed June 15, 2007, at http://www.e-laws.gov.on.ca/DBLaws/Statutes/English/04p03_e.htm.

2. Health care renewal in Canada: measuring up? Annual report to Canadians 2006. Toronto: Health Council of Canada; 2007. Accessed June 15, 2007, at http://www.healthcouncilcanada.ca/en/index.php?option=com_content&task=view&id=136&Itemid=115.

3. *Canada Health Act*. R.S.C. 1985, c.C-6. Accessed June 15, 2007, at http://laws.justice.gc.ca/en/C-6/index.html.

4. Iron K. Moving toward a better health data system for Ontario. ICES Investigative Report. Toronto: Institute for Clinical Evaluative Sciences; 2006. Accessed June 15, 2007, at http://www.ices.on.ca/file/Moving_towards_a_better_health_care_system_apr06.pdf.

5. Improving health care data in Ontario. ICES Investigative Report. Toronto: Institute for Clinical Evaluative Sciences; 2005.

6. Ontario Ministry of Health and Long-Term Care. The information management strategy. December 1, 2005. [News release.] Accessed June 15, 2007, at http://www.health.gov.on.ca/english/media/news_releases/archives/nr_05/bg_120105_2.pdf.

7. Ontario Ministry of Health and Long-Term Care. Health Results Team for Information Management. Update. August 31, 2006. Accessed June 15, 2007, at http://www.health.gov.on.ca/transformation/providers/information/updates/upd_083106.pdf.

8. Ontario Ministry of Health and Long-Term Care. Health Results Team for Information Management. Registration Data Entry Working Group: final report and recommendations. [2007].

9. Ontario Ministry of Health and Long-Term Care. Health Results Team for Information Management. Data quality management framework: data quality scorecard indicator companion document. March 2007.

10. Ontario Ministry of Health and Long-Term Care. Health Results Team for Information Management. Data quality management framework: orientation guide. March 2007.

11. Bains N. Population health indicators: mortality geographic data quality, 2000–2001. Kingston, ON: Health System Intelligence Project; 2005.

12. Roos LL, Sharp SM, Wajda A. Assessing data quality: a computerized approach. *Social Science and Medicine* 1989; 28(2):175-82.

13. Roos LL, Gupta S, Soodeen RA, Jebami L. Data quality in an information-rich environment: Canada as an example. *Canadian Journal of Aging* 2005; 24(suppl 1):153–70.

14. Rawson NS, D'Arcy C. Assessing the validity of diagnostic information in administrative health care utilization data: experience in Saskatchewan. *Pharmacoepidemiology and Drug Safety* 1998; 7(6):389-98.

15. Peabody JW, Luck J, Jain S, Bertenthal D, Glassman P. Assessing the accuracy of administrative data in health information systems. *Medical Care* 2004; 42(11):1066–72.

16. Rosendall JP, Williams J. Producing better data – thinking about gaps. A report on information management for the Ontario Ministry of Health and Long-Term Care. April 29, 2005.

17. Hierholzer WJ Jr. Health care data, the epidemiologist's sand: comments on the quantity and quality of data. *American Journal of Medicine* 1991; 91(suppl 3B):21S–26S.

18. Holt T, Jones T. Quality work and conflicting quality objectives. London: Office for National Statistics; 1998. Accessed June 15, 2007, at http://www.statistics.gov.uk/methods_quality/downloads/QualityWork.pdf.

19. Guidelines for measuring statistical quality. Version 3.0. London: Office for National Statistics; 2006. Accessed June 15, 2007, at www.opsi.gov.uk/click-use/index.htm.

20. Arts DG, de Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. *Journal of the American Medical Informatics Association* 2002; 9(6):600–11.

21. Juurlink D, Preyra C, Croxford R, Chong A, Austin P, Tu J, et al. Canadian Institute for Health Information Discharge Abstract Database: a validation study. Toronto: Institute for Clinical Evaluative Sciences; 2006.

22. Kelsey JL, Thompson WD, Evans AS. Methods in observational epidemiology. New York: Oxford University Press; 1986.

23. Stein HD, Nadkarni P, Erdos J, Miller PL. Exploring the degree of concordance of coded and textual data in answering clinical queries from a clinical data repository. *Journal of the American Medical Informatics Association* 2000; 7(1):42–54.

24. Iron KS, Goel V, Williams JI. Concordance of hospital discharge abstracts and physician claims for surgical procedures in Ontario. Working Paper 42. Toronto: Institute for Clinical Evaluative Sciences; 1995.

25. Pinfold SP, Goel V, Sawka C. Quality of hospital discharge and physician data for type of breast cancer surgery. *Medical Care* 2000; 38(1):99–107.

26. Silcocks PB, Robinson D. Completeness of ascertainment by cancer registries: putting bounds on the number of missing cases. *Journal of Public Health* 2004; 26(2):161–7.

27. Byrne N, Regan C, Howard L. Administrative registers in psychiatric research: a systematic review of validity studies. *Acta Psychiatrica Scandinavica* 2005; 112(6):409–14.

28. The CIHI data quality framework. June 2005 revision. Ottawa: Canadian Institute for Health Information; 2005. Accessed June 15, 2007, at http://secure.cihi.ca/cihiweb/en/downloads/Data_Quality_Framework_2004_e.pdf.

29. Fiscal year 2005 performance and accountability report. Part II. Assessment of data quality. Washington, DC: Department of Veterans Affairs; 2005. Accessed June 15, 2007, at http://www.va.gov/budget/report/2005/AssessmentDataQuality.pdf.

30. Brennan PF, Stead WW. Assessing data quality: from concordance, through correctness and completeness, to valid manipulatable representations. *Journal of the American Medical Association* 2000; 7(1):106–7.

31. Ontario health planning data guide. Release 3.0. Toronto: Ontario Ministry of Health and Long-Term Care; 2006. Accessed June 15, 2007, at http://www.health.gov.on.ca/transformation/providers/information/resources/healthplan_guide.pdf.

32. Dobrow MJ, Goel V, Upshur RE. Evidence-based health policy: context and utilisation. *Social Science and Medicine* 2004; 58(1):207–17.

33. Statistics Canada's quality assurance framework. Catalogue No. 12-586-XIE. Ottawa: Statistics Canada; 2002. Accessed June 15, 2007, at http://www.statcan.ca/english/freepub/12-586-XIE/12-586-XIE2002001.pdf.

34. Giovanni E, Colledge M, Ward D, Salou G. Quality framework and guidelines for OECD statistical activities. Geneva: Organisation for Economic Co-operation and Development; 2003. Accessed June 15, 2007, at http://www.oecd.org/dataoecd/26/42/21688835.pdf.

35. National Statistics Methodology Advisory Committee. Quality measurement and reporting. London: Office for National Statistics; 2007. Accessed June 15, 2007, at http://www.statistics.gov.uk/methods_quality/downloads/NSMAC(03)2.pdf.

36. Bullard J, Coleman MP, Robinson D, Lutz JM, Bell J, Peto J. Completeness of cancer registration: a new method for routine use. *British Journal of Cancer* 2000; 82(5):1111–6.