

**The Canadian Institute for Health Information (CIHI)
Data Quality Framework, Version 1:
A Meta-Evaluation and Future Directions**

(Practice Oriented Paper)

J.A. Long
Consultant, Data Quality
Canadian Institute for Health Information
Phone: 416-481-1616 Ext. 3579
Fax: 416-481-2950
Email: jlong@cihi.ca
www.cihi.ca

J.A. Richards
Manager, Data Quality
Canadian Institute for Health Information
Phone: 416-481-2002
Fax: 416-481-2950
Email: jrichards@cihi.ca
www.cihi.ca

C.E. Seko
Senior Analyst, Health and Outcomes Statistics
Health Statistics Division, Statistics Canada
Phone: 613-951-4931
Fax: 613-951-4251
Email: sekocet@statcan.ca
www.statcan.ca

Abstract. Information quality (IQ) problems can have severe consequences in the health care sector. Since its inception, the Canadian Institute for Health Information (CIHI) has recognized the importance of information quality and has implemented a framework designed to evaluate the data quality of the numerous CIHI data holdings. After one year of implementation, a meta-evaluation of the CIHI data quality framework evaluation process was conducted and the framework was found to be both relatively strong theoretically as well as practical. Despite a relatively favourable meta-evaluation, several aspects of the framework are scheduled for improvement. It is recommended that data quality framework and meta-evaluation development be recognized as crucial with the ultimate objective of improving information in the health care sector.

Keywords: Data quality, framework, information quality research, meta-evaluation

Introduction. Information quality (IQ) problems can have severe financial and operational consequences for organizationsⁱ, and in the health care sector, can impact life and death decisions. Since its inception, the Canadian Institute for Health Information (CIHI) has

recognized the importance of data quality. Although data quality has been a priority since the establishment of CIHI, a new framework, designed to evaluate the data quality of the numerous CIHI data holdings, has recently been implemented. The following is a meta-evaluation of the current CIHI data quality framework evaluation process.

Background. The CIHI Data Quality Framework, Version 1 (CIHI-DQF, v1) is a four-level conceptual model designed to standardize and facilitate the systematic quantification and measurement of data quality at CIHIⁱⁱ. Incorporated in 1993, CIHI is a federally chartered, yet independent, not-for-profit organization. The Institute was established through the amalgamation of two non-governmental organizations, i.e., the Hospital Medical Records Institute (HMRI) and the MIS Group, along with selected databases and functions from the Health Information Division of Health Canada and the Health Statistics Division of Statistics Canada (STC)^{iii,iv}.

CIHI has taken a central role in the development of Canada's health information system and is mandated to “serve as the national mechanism to coordinate the development and maintenance of a comprehensive and integrated health information system for Canada” and “to provide and co-ordinate the provision of accurate and timely information required for: the establishment of sound health policy, the effective management of the Canadian Health System, and for generating public awareness about factors affecting good health. Consistent with its mandate, an important role taken on at CIHI is the collection, processing, and maintenance of a growing number of clinical databases or registries, as well as, health human resources, health services, and health expenditures databases. To date, the CIHI data holdings include 22 databases and registries, many of which are national in scope^v.

As a result of CIHI being an amalgamation of several programs, the data quality methods were initially non-standardized and database or registry specific. In recognition of the vital importance of data quality, as well as, due to the responsibility of maintaining 22 databases or registries, the need for a standard strategy to identify data quality problems, to enable senior CIHI management to allocate finite resources across 22 data holdings, and to solve unforeseen problems is fast becoming imperative.

In response to the identified need for a standard, organized, and systematic approach to data quality, one of the authors (Seko) was seconded from STC in 1999 to develop a data quality strategy in collaboration with CIHI senior management. The cornerstone of the resulting strategy is the CIHI Data Quality Framework, Version 1 (CIHI-DQF, v1). The CIHI-DQF, v1 draws on the statistical literature^{vi,vii,viii,ix,x} the STC guidelines and methods^{xi}, the Information Quality literature^{xii}, the CIHI mandate, as well as the principle of Continuous Quality Improvement (CQI). The first version of the framework was ready and implemented in April 2000. To date, two major database evaluations have been conducted and another three major database evaluations are taking place at the time of this study.

Rationale. Consistent with the principle of CQI for ongoing data quality measurement, evaluation, and improvement, efforts must be made to ensure the integrity and relevance of the evaluation process itself. This can be achieved by requiring that the process itself be systematically assessed and improved on a continuing basis. That is, not only should the principle of CQI be applied to the data holdings, it should also be applied to the *methods* used to

evaluate the data. As CIHI-DQF, v1 has been successfully implemented for over one year now, in an ongoing endeavor to improve its effectiveness, it is an opportune time to conduct a meta-evaluation in the effort to assess its performance.

While the importance of applying CQI to any data quality framework or evaluation process should be apparent, there appears to be little evidence in the literature that this idea has been considered. In fact, it appears that the only relevant work available is a research-in-progress conducted by Eppler and Wittig^{xiii}.

Eppler and Wittig argue that an IQ framework should be practical in addition to being theoretically strong. They reason that an IQ framework should be theory driven and a theory based conceptual map upon which a framework is based should be available to the research community. Moreover, they suggest that a systematic and concise set of measurement and evaluation criteria, a scheme to analyze and solve quality problems, and a plan to facilitate proactive management should be available.

In order to assess whether some of the leading IQ frameworks are academically rigorous, as well as practical, Eppler and Wittig put forward a basic method for evaluation. The aim of this paper is to conduct a meta-evaluation of the CIHI-DQF, v1 according to the method developed by Eppler and Wittig. The meta-evaluation results integrated with the preliminary framework implementation experience will together direct future framework improvement.

Methods. A basic descriptive meta-evaluation was conducted of the CIHI Data Quality Framework, Version 1 (CIHI-DQF, v1). The evaluation methodology was simply a rating of whether the CIHI-DQF, v1 addressed the Eppler and Wittig categories or not; and if so, a simple qualitative statement was included as to how well the category was addressed. Applied findings based on the first year of the framework's implementation were integrated into the meta-evaluation.

The CIHI Data Quality Framework, Version 1 (CIHI-DQF, v1)

More specifically, the CIHI-DQF, v1 was designed to operationalize, measure, and evaluate the quality of the CIHI data holdings using a standard and systematic approach. The objectives of the CIHI-DQF, v1 are: 1) to standardize information on data quality, both for internal and external users; 2) to provide a common strategy for assessing data quality; and 3) to define a work process for CIHI's data holdings that identifies data quality priorities and produces continuous improvement in data quality.

According to the CIHI-DQF, v1, 'quality' is defined as 'fitness for use'² and 'data quality' is operationally defined and measured along five common and widely used quality dimensions. Consequently the framework was designed to facilitate the evaluation of these dimensions, as well as, to provide a single overall evaluation of a data holding based on the five dimensions.

Specifically, the CIHI-DQF, v1 is organized as a four-level conceptual model. At the foundation of the model are 86 basic unit items that are known as *criteria*. The 86 criteria can be collapsed into the second level of 24 *characteristics* (e.g., under-coverage, reliability, and interpretability)

that in turn, can be collapsed into 5 *dimensions* of data quality (i.e., i. accuracy, ii. timeliness, iii. comparability, iv. usability, and v. relevance). Finally, the 5 dimensions can be collapsed into one overall evaluation of the database. Figure 1 below provides a summary of the CIHI-DQF, v1.

Figure 1. The CIHI Data Quality Framework Evaluation Instrument, Version 1 (CIHI-DQF, v1)

Dimension	Characteristics	Criteria
I. Accuracy	i.1. Over-coverage	1-6
	i.2. Under-coverage	7-12
	i.3. Simple response bias	13
	i.4. Reliability	14-15
	i.5. Correlated response bias	16
	i.6. Collection and capture	17-24
	i.7. Unit non-response	25-26
	i.8. Item (partial) non-response	27-30
	i.9. Edit and imputation	31-37
	i.10. Processing	38
	i.11. Estimation	39-41
II. Timeliness	ii.1. Timeliness	actual release-planned release 42-45
III. Comparability	iii.1. Comprehensiveness	46-49
	iii.2. Integration	50-53
	iii.3. Standardization	54-57
	iii.4. Equivalency	58-59
	iii.5. Linkage-ability	60-64
	iii.6. Product comparability	65
	iii.7. Historical comparability	66-69
IV. Usability	iv.1. Accessibility	70-75
	iv.2. Documentation	77-78
	iv.3. Interpretability	79-81
V. Relevance	v.1. Adaptability	82-84
	v.2. Value	85-86

Whereas the CIHI-DQF, v1 provides standard definitions and a common strategy, the framework itself is designed to be a part of a work process that identifies data quality priorities and produces continuous improvement in data quality. Once the measurement of data quality is achieved, then it must be improved, then measured again, improved, and so on. While the framework quantifies the concepts of data quality and enables measurement, the evaluation process, based on the CIHI-DQF, v1, puts the principle of CQI into action.

The CIHI Data Quality Framework, Version 1 (CIHI-DQF, v1) Database Evaluation

While the objectives of the CIHI-DQF, v1 are essentially to standardize efforts and to define a work process, the objectives of the data holdings evaluation process based on the CIHI-DQF, v1 are: 1) to identify and rank aspects of data quality needing improvement; and 2) to produce information on data quality that feeds into the creation of data quality documentation for users^{xiv}. In fact, the facilitation of data quality documentation for users is key to the database evaluation process.

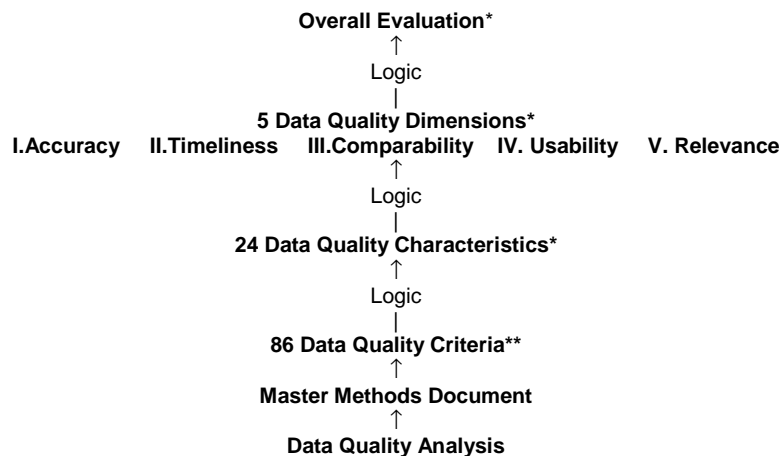
In addition to the CIHI-DQF, v1, a data quality manual that itemizes the step-by-step evaluation process is also provided to those responsible for CIHI data holdings. In fact, the manual was designed to complement, and is integral, to the framework. Part 1 of the manual provides instruction for analysing data quality, describing data quality, and for improving data quality, and

Parts 2 and 3 contain the CIHI-DQF, v1 and an instrument based on the framework, respectively. Direction on recommended analyses and detailed instruction on how to prepare an evaluation document are provided in the manual. Alongside assessments, evaluators are also instructed to include short and long term recommendations. Moreover, upon assessment completion evaluators are instructed to rank and summarize the recommendations at the beginning of the evaluation document. Assignments are made in consultation with the staff responsible for a given database and in tandem with the Data Quality Section. In addition, ongoing Data Quality Workshops designed to educate staff with respect to the database evaluation process are offered annually. In fact, training is considered to be central to the entire data quality strategy.

Besides the preparation of an evaluation document, the evaluation process includes the completion of the CIHI Data Quality Framework Evaluation Instrument, Version 1 which is designed to facilitate the computation of an evaluation and to enable the Data Quality Section track progress within and across databases. Specifically, the 86 criteria are scored in a consistent fashion so that a low value indicates a less favorable evaluation and a score of 0 indicates that a criterion is 'not applicable'. The criteria evaluations are: 0) not applicable; 1) unknown; 2) not met; and 3) met.

Likewise, the characteristic, dimension, and overall evaluations are scored such that low values indicate less favorable evaluations (not including 'not applicable'). The characteristic, dimension, and overall evaluations are: 1) unknown; 2) not acceptable; 3) marginal; and 4) appropriate. Once categories are assigned at the criteria level, the characteristic, dimension, and overall database evaluations can be easily computed based on the framework algorithm. SAS code, based on the CIHI-DQF, v1 algorithm for reading and scoring the data is currently under development. Figure 2 below provides an outline of how the CIHI Data Quality Framework Evaluation Instrument, Version 1 is scored.

Figure 2. The CIHI Data Quality Framework, Instrument, Version1 (CIHI-DQF, v1) Algorithm



*1. Appropriate, 2. Marginal, 3. Not acceptable, and 4. Unknown
 **0. Not Available, 1. Unknown, 2. Not Met, and 3. Met

In other words, at the bottom of the four-level model are 86 criteria that roll up to 24 data quality characteristics. Each of the 86 criteria, and hence the 24 data quality characteristics, can be regularly evaluated for a database or registry. Combined evaluations of constituent

characteristics define the assessment of a dimension. Combining dimensions gives an overall impression of the data quality for a database. All levels can also be combined across databases to summarize a dimension (or characteristic) for the entire set of CIHI data holdings. The aim is to identify and rank aspects of data quality in need of improvement in a comparable way such that resources can be optimally allocated across competing data holdings. The result is a comprehensive and integrated picture of the data quality within and across databases.

Again, the purpose of collecting and scoring evaluation data is to help identify and prioritize data quality improvement tasks^{xv}. Lastly, evaluators are instructed to think of the evaluation process as ongoing and to decide on a fixed time period (e.g., annually) for continuous data quality evaluations and improvement, with the ultimate objective of improving data quality and, by extension, critical health information based on the data. Lastly, the database evaluation process also includes a careful consideration of confidentiality, privacy, and security issues and all concerns are addressed.

The CIHI Data Quality Framework, Version 1 (CIHI-DQF, v1) Meta-Evaluation

The evaluation of the CIHI-DQF, v1 evaluation process, or meta-evaluation, was conducted according to Eppler and Wittig. Briefly, Eppler and Wittig put forward a basic method to assess the academic rigour as well as the practicality of an information quality framework. Their method entails assessing a framework according to the following six evaluation categories or ‘meta-criteria’: 1) definitions; 2) positioning; and 3) consistency, to assess theoretical robustness, as well as, 1) conciseness; 2) examples; and 3) tools, and to assess practicality. Each category or criterion is associated with a key question that is used to evaluate a framework. Figure 3 below illustrates the Eppler and Wittig meta-criteria and associated evaluation questions.

Figure 3: The Eppler and Wittig (2000) Meta-Criteria for the Evaluation of IQ Frameworks

Meta-Criteria	Evaluation Questions
I. Analytic	
I.1 Definitions	Are all individual information quality criteria clearly defined and explained? Are all the dimensions to which the individual criteria are grouped (if existing) defined and explained?
I.2 Positioning	Is the context of the framework’s application (and its limits) clear? Is the framework positioned within existing literature?
I.3 Consistency	Are the individual criteria mutually exclusive and collectively exhaustive? Is the framework overall divided into systematic dimensions that are also mutually exclusive and collectively exhaustive? Is it clear why a group of criteria belongs to the same dimension?
II. Practical	
II.1 Conciseness	Is the framework concise in the sense that it can be easily remembered? Are there (as a minimal rule of thumb) less than seven dimensions and less than seven criteria per dimension?
II.2 Examples	Are specific and illustrative examples given to explain the various criteria (e.g., case studies)?
II.3 Tools	Is the framework accompanied by a tool that can be used to put it into practice, such as a questionnaire, a software application, or a step-by-step implementation guide or methodology?

Hence, the CIHI-DQF, v1 was evaluated according to the six Eppler and Wittig theoretical and practical evaluation questions. The evaluation methodology was simply a brief description of if the meta-criteria were addressed, and if so, how each question was addressed by the CIHI-DQF, v1. Where applicable, findings from applied experience were included and possible future improvements were also itemized.

In terms of the applied experience, while the CIHI-DQF, v1 was officially released in April 2001, its test phase spanned April 2000-April 2001. Though applied experience involving all of the CIHI databases and registries was not available at the time of the study, two major database evaluations using the framework were conducted prior to the framework's official release and another three major evaluations were underway at the time of the study. A summary of evaluator feedback as well as notes from those involved in the development of version 1 was carried out and integrated with the meta-evaluation. Future directions were itemized based on the meta-evaluation results and the applied experience.

Results. Included for evaluation was the CIHI Data Quality Framework, Version 1 (CIHI-DQF, v1) which was developed in 1999 and tested from April 2000-April 2001. The following represents the results of a meta-evaluation of the CIHI-DQF, v1 conducted according to the method put forward by Eppler and Wittig. Applied findings and test phase notes were integrated along with the meta-evaluation results and future directions are presented.

Table 1 provides a summary of the CIHI-DQF, v1 meta-evaluation findings for the analytic or theoretical performance of the CIHI-DQF, v1. Regarding the first meta-criterion known as 'definitions', characteristic and dimension definitions are provided, however, applied findings suggest that the definitions are not detailed enough. Moreover, references are not provided. Certain dimensions (e.g., relevancy) and characteristics (e.g., equivalency) have been found to be unclear. The level 1 criteria, on the other hand, have been found to be clear. Improving the framework definitions and referencing have been targeted in version 2.

Table 1. A Meta-Evaluation of the Theoretical Foundation of the CIHI Data Quality Framework, Version 1 (CIHI-DQF, v1)

Meta-Criteria*	The CIHI-DQF, v1 response	Applied findings Year 1	Summary	Future Development
I. Analytic I.1. Definitions	Definitions of the dimensions and the characteristics exist and are provided.	Overall, definitions are not detailed enough and references are not provided. Certain dimensions (e.g., relevancy) and characteristics (e.g., equivalency) have been found to be particularly unclear.	Definitions exist and are provided but, similar to other frameworks, are in need of development.	Improve and develop definitions. As well, include examples and references.
I.2. Positioning	The framework is based on STC DQ Guidelines and the CQI literature as well as the CIHI mandate.	Not raised as an issue. Those applying the framework focused on the practical aspects of the framework.	More information regarding how the framework fits with respect to the field was nevertheless flagged by the authors as necessary. The context of the framework's application and limits could be clearer.	Despite not being raised as an issue and although a brief summary of the literature is provided, the position of the framework within the literature will be expanded and referenced in order to contribute to the field.
I.3. Consistency	Individual dimensions, characteristics, and criteria are reasonably mutually exclusive and collectively exhaustive.	While most dimensions were found to be reasonably distinct, the characteristics within Comparability (e.g., integration, standardization, and linkage) and Relevance were reported to be confusing. The framework is generic and was found to be applicable for several different data CIHI sources.	The concepts underpinning the Comparability and Relevance dimensions need further clarification. Interdependency across the dimensions (e.g., the tradeoff between accuracy and timeliness) is not available. The framework was found to be generic.	The concepts underpinning the Comparability and Relevance dimensions will be clarified for version2. The discussion of the Interdependency across dimensions or characteristics will be explored.

*Eppler and Wittig (2000)

In terms of positioning, while the CIHI-DQF, v1 is based on the Statistics Canada (STC) framework, the STC framework in turn is based on the literature. Nonetheless, the context of the framework's application and its limits have been found to be unclear in the documentation and, again, references are not included. Whereas those who applied the framework did not raise these issues as a problem, those involved with the design of version 1 flagged the background description of the framework as cursory and an area in need of expansion when time and resources permit. Version 2 will include an expanded background section that describes how the CIHI-DQF, v1 fits with the literature and all references will be included.

For consistency, the framework's individual components were, for the most part, found to be mutually exclusive and collectively exhaustive. The framework overall is generally thought to be divided into systematic dimensions that make sense and the numerous characteristics are thought to be logically assigned to their dimensions. While most dimensions were found to be reasonably distinct, the characteristics within 'comparability' (e.g., integration, standardization, and linkage)

were flagged as unclear. Although no comments were made by the test phase evaluators, consistent with Eppler and Wittig we also targeted the relevance dimension as a possible area for clarification. Furthermore, an explanation of the interdependency across the dimensions (e.g., the tradeoff between accuracy and timeliness or linkage-ability and privacy) was not well developed. These tradeoffs will need to be further explored in version 2. Also in terms of consistency, the framework was found to be applicable for different data sources (e.g., clinical data or health human resources data). The applicability of the framework outside of the collection of official health care sector statistics is unknown.

Table 2 summarizes how pragmatic the CIHI-DQF, v1 is. In terms of practicality, test phase participants initially felt overwhelmed and where to start was not clear to them. As well, they found that the framework evaluation document was not user-friendly. In response, an instrument based verbatim on the framework was developed and initial feedback indicates that user-friendliness may have been improved. Despite not knowing where to start, once underway evaluators indicated that the framework was concise and practical. Also of note, the framework contains less than seven dimensions and, for the most part, less than seven characteristics per dimension. Version 2 will include clearer step-by-step directions in the manual and the manual will become more centered on the framework.

Table 2. A Meta-Evaluation of the Pragmatic Aspects of the CIHI Data Quality Framework, Version 1 (CIHI-DQF, v1)

Meta-Criteria*	The CIHI-DQF, v1 response	Applied findings Year 1	Summary	Future Development
II. Pragmatic II.1. Conciseness	The framework has less than 7 dimensions and, for the most part, less than 7 characteristics per dimension. For the most part, each characteristic is based on fewer than 7 criteria.	Length has not been raised as an issue. Where to start and user-friendliness have been flagged as areas in need of improvement.	Good start.	Conciseness has not been targeted in version2. The manual will become more framework centered and the step-by-step evaluation instructions will be expanded.
II.2. Examples	Of the 86 criteria, only 6 specific examples are provided. No case studies are provided.	Examples were requested for all of the components as well as for the entire process of the data quality evaluation.	More criteria or characteristic examples should be included as well as entire evaluations as they become available.	Available examples will be incorporated and, as they become available, entire hardcopy and online evaluations will be included.
II.3. Tools	Tools include ongoing training, a manual, an instrument version of the framework, and an evaluation algorithm.	Although several tools exist, where to start has been raised as a consistent concern. The connection between the manual and the framework was found to be unclear.	The value and number of tools is sufficient however the tie between them is not clear enough.	The connection between the framework and the manual will be improved. Clearer step-by-step instructions will be provided. Framework algorithm SAS code will be developed. A software application has been suggested, however, this suggestion may not be addressed in version2.

Of the 86 criteria included in the CIHI-DQF, v1, only 6 specific examples are provided and no case studies are provided. Congruent with the poor evaluation for the ‘examples’ meta-criterion, the authors, as well as the test phase participants, indicated that more examples were necessary in version 2. Available examples will be included and as new examples become available they will also be incorporated.

The data quality training, manual, framework, and instrument were found to be valuable tools and the number of tools was found to be sufficient. A software version of the instrument was suggested however operational restraints might prevent software development in the near future. Although the existence and number of tools was found to be sufficient, how they related was identified as an area for improvement. Of specific concern, was the relation between the manual and framework documents. The connection between these documents will be clarified in version 2.

Discussion. After one year of implementation, the time had come to evaluate the performance of the CIHI Data Quality Framework, Version 1 (CIHI-DQF, v1). Although the importance of quality data is unchallenged and frameworks designed to operationalize or quantify data quality, with the ultimate objective of improving it, are crucial, few meta-evaluation protocols are readily available. This is troubling for those who maintain health data, as poor quality data can be dangerous and has the potential to cause great harm. The approach used by Eppler and Wittig in their review of seven principal frameworks was applied to the CIHI-DQF, v1 and integrated with preliminary implementation findings in order to guide improvement efforts.

The CIHI Data Quality Framework Evaluation, Version 1 (CIHI-DQF, v1) Meta-Evaluation Findings

In terms of theoretical robustness, the CIHI-DQF, v1 concepts were found overall to be reasonably well defined, well positioned in the literature, and consistent. However, certain weaknesses were detected. While definitions of the framework characteristics and dimensions are provided in the CIHI-DQF, v1, the definitions may not be detailed enough and references need to be provided. Consistent with the findings of Eppler and Wittig, certain dimensions (e.g., relevancy) or characteristics have been found to be unclear.

While the framework is based on the literature, the context of the framework’s application and its limits are not clearly stated in the documentation. The framework was also found to be consistent within its dimensions and characteristics, as well as, with other frameworks. Similar to the seven other frameworks reviewed by Eppler and Wittig, the CIHI-DQF, v1 includes the concepts of timeliness, accessibility (as a characteristic within usability), relevance, and accuracy. While no dimensions, characteristics, or criteria in the CIHI-DQF, v1 were specifically referred to as ‘objectivity’, ‘consistency’, or ‘completeness’, there may be significant overlap with some of the categories included in the CIHI-DQF, v1. For example, ‘completeness’ might be similar to the framework concept of ‘comprehensiveness’.

Consistent with other work, however, the interdependency across the dimensions (e.g., the tradeoff between accuracy and timeliness or the tradeoff between linkage-ability and privacy) was not well developed and was identified as an area for expansion in version 2. While Eppler

and Wittig found that many of the frameworks they reviewed were domain-specific, it's believed that the CIHI-DQF, v1 is also fairly generic within the realm of official health statistics collection and was found to be applicable across several diverse data sources at CIHI. In fact, given that much of the health care sector population data is 'administrative data', it is possible that the framework has not considered fully this type of specific data, along with its unique challenges^{xvi}. One solution might be to maintain applicability as well as to include more components specific to administrative data.

In terms of practicality, test phase participants tended to feel overwhelmed. Once underway, however, evaluators indicated that the framework was concise and practical. The inclusion of illustrative examples on the other hand was flagged as an area for improvement. Lastly, unlike other frameworks the CIHI-DQF, v1 is accompanied by several tools including a workshop, a manual, framework documentation, and an evaluation instrument. Although the value and the number of tools were found to be sufficient, how the tools related was found to be unclear.

CIHI Data Quality Framework, Version 2 (CIHI-DQF, v2) Development Plans

Based on both the meta-evaluation results and the applied experience, improvement plans for the framework include:

- expanding the framework definitions;
- expanding the background sections of the CIHI-DQF, v1 documentation so that the conceptual map is clear for evaluators as well as for the research community;
- including all references;
- making the documentation more framework centered and user-friendly (e.g., clarifying the step-by-step instructions as well as the relationship between the manual and framework);
- redesigning the manual so that it will better support the framework and instrument (e.g., an analogy could be a tax guide and tax form)
- expanding the explanation of the interdependency across the dimensions (e.g., the tradeoff between accuracy and timeliness); and
- examples will be included.

Although the CIHI-DQF, v1 appears to compare favorably in many respects to other frameworks, certain 'big picture' changes, consistent with Eppler and Wittig's recommended future directions, are being considered. First, at present an evaluation based on the CIHI-DQF, v1 is more representative of whether an important list of criteria have been considered rather than an actual qualitative evaluation of the data quality or, by extension information quality, of a database. More qualitative versions are currently under development.

Second, future development could also involve the incorporation of the entire error model. For example, the CIHI-DQF, v1 does not cover possible errors made between the time of the clinical action (e.g., a clinical intervention) and the chart documentation, nor does it consider the critical importance of system design. While the CIHI-DQF, v1 has been found to be systematic and concise, it does not provide a scheme to solve detected problems and its ability to facilitate proactive data quality management needs enhancement. Suggested improvement schemes and proactive management approaches will be explored.

Finally, the cost of framework implementation must be addressed. Outside of the cost of a new data quality unit responsible for the development and support of the framework, within CIHI experience to date suggests that most of the resources, e.g., database managers, analysts, technical support, and documentation, required for framework implementation are already in place. Other than the initial study time required to learn the revised standard approach, it appears that no additional resources have been necessary. One preliminary observation, however, might be that the cost of implementation, as measured by the time required for existing staff to complete an evaluation, varies with the level of methodological expertise available.

The importance of establishing the cost of implementation within CIHI has resulted in an effort to track relevant time and resource use. Future plans include the development of an external version of the framework, hence an understanding of the issues and costs involved may be of interest to other health care sector settings, e.g., hospitals or clinical research institutions.

In addition to existing data quality efforts, a new CIHI data quality section (three methodologists, a classification expert, some administrative support, and a manager) has been put in place and is devoted to studying data quality issues, framework and methodological development, and support. While replicating such a unit in many external health care settings would be costly there may be no need to do so. Although the framework has not yet been implemented externally, some additional observations based on our experience can be made. Primarily what is required for implementing such a framework is methodological (i.e., statistical or epidemiological) expertise. However, what is fundamental, in addition to readily available methodological expertise, as well as the basic infrastructure required for a clinical or administrative database, is senior management commitment, active sponsorship for the idea, and an assurance of commitment and resources in all operational plans. Such factors have proven crucial for successful framework implementation at CIHI and thus should be considered for any external implementation.

Potential Limitations of the Study

It is recognized that the main weakness of this study is that the meta-evaluation methodology was based on only one article (i.e., Eppler and Wittig (2000)). While the findings of the meta-evaluation were congruent with our applied experience, we acknowledge that the field of meta-evaluation is in its infancy. In fact, other than Eppler and Wittig, prior work in the field of data quality framework meta-evaluation appears to be unavailable. To interpret results with confidence, a meta-evaluation must be based on a solid body of literature.

In addition to a call for more meta-evaluation work, data quality practitioners would benefit from comparative analyses of the different types of meta-evaluation methodologies, i.e., meta-meta-evaluations or meta²-evaluations. Nevertheless, conducting the Eppler and Wittig meta-evaluation combined with applied results is a good start. It is hoped that this paper will not only elicit feedback regarding the CIHI-DQF, v1, but will also help to stimulate the field of framework meta-evaluation as well as meta²-evaluation. The need for more in-depth, rigorous, and complete meta-evaluation methodologies is obvious, especially in health care where quality information is crucial.

Another limitation of this study might be that the literature search was not comprehensive enough. That said, data quality research seems to be spread across numerous fields and the search and acquisition process was found to be challenging. An additional recommendation based on this study could be a reiteration of the importance of *The Data Quality Journal* and the annual MIT IQ conference and proceedings as centers of excellence for practitioners of data quality. Lastly, the omission of an explanation of the difference between the concepts of ‘information quality’ and ‘data quality’ might be interpreted as another limitation. Due to time and space constraints, and for the purposes of this study, no definitions were provided except to state that information quality follows from data quality.

Recommendations for Future Research

Other recommendations based on this study echo the call sounded by Huang, Lee, and Wang (1999), as well as, by Eppler and Wittig (2000) for improved concept definition and standardization for the field. Even the definitions of framework components, such as the difference between a criterion and a characteristic, could benefit from standard definitions. One suggestion could be a dictionary for the field of data quality much like the *Dictionary of Epidemiology, 4th Edition*^{xvii} is for the field of Epidemiology. Such a dictionary might draw on and collate several disciplines where work on the important concepts (e.g., accuracy) has also been conducted. Moreover, while Eppler and Wittig provide an excellent start, given the vital importance of quality information, especially in health care where lives can be affected, and consistent with the principle of CQI, meta-evaluation methodology development is necessary.

As a final point, whereas senior management at CIHI understands the importance of data quality, framework development, and meta-evaluation research, some in the health care field may not fully recognized the impact of adequate data quality or conversely the impact of poor quality data and by extension, information. The study and communication of the extent, impact, and resolution of data quality, and hence information quality, must be more forcefully pursued. Nowhere else may this be more pertinent than in the health care sector where critical decisions are being made and lives may be in the balance.

Conclusion. In summary, the CIHI Data Quality Framework, Version 1 (CIHI-DQF, v1) is a new framework and its evolution was anticipated. The CIHI-DQF, v1 was evaluated and found to rank relatively well when compared to other frameworks. The framework was found to be both relatively strong theoretically as well as practical and reasonably generic. Despite a relatively favorable meta-evaluation, several aspects of the CIHI-DQF, v1 are slated for improvement (e.g., more explanation of the trade-offs involved across quality dimensions). Given the importance of quality information, especially in health care where life and death decisions may be involved, it is surprising that so few generic frameworks seem to exist and meta-evaluation methodologies seem almost nonexistent. It is also recommended that framework and meta-evaluation development be flagged as crucial with the ultimate objective of improving information especially in the health care sector.

References

- ⁱ Strong, D.M., Lee Y.W., & Wang, R.Y. (1997) 10 Potholes in the Road to Information quality, in: *Computer IEEE*, pp. 38-46.
- ⁱⁱ CIHI Data Quality Framework, Version 1, April 2001, CIHI.
- ⁱⁱⁱ National Consensus Conference on Population Health Indicators, 1999, CIHI.
- ^{iv} www.cihi.ca
- ^v CIHI Products and Services Catalogue 2001, CIHI.
- ^{vi} Deming, W.E., *The New Economics for Industry, Government, and Education*. W. Edwards Deming Institute, 1994.
- ^{vii} Deming, W.E., *Out of the Crisis*, W. Edwards Deming Institute, 1986.
- ^{viii} Brackstone, G. J. (1987) Issues in the Use of Administrative Records for Statistical Purposes, *Survey Methodology* 13.
- ^{ix} Brackstone, G. (1999) Managing Data Quality in a Statistical Agency. *Survey Methodology* Vol 25(2) pp, 139-149.
- ^x U.S Federal Committee on Statistical Methodology, *Statistical Policy Working Paper 4 – Glossary of Nonsampling Error Terms: An Illustration of a Semantic Problem in Statistics*, 1987.
- ^{xi} Statistics Canada Quality Guidelines, 3rd Edition, October 1998, Statistics Canada.
- ^{xii} Huang, K.T., Lee, W. L., and Wang, R. Y., *Quality Information and Knowledge*, Prentice-Hall, October 1998.
- ^{xiii} Eppler, M.J. and Wittig, D. (2000) Conceptualizing Information Quality: A Review of Information Quality Frameworks from the last Ten Years. *Proceedings of the 2000 Conference on Information Quality*. Eds Klein D. and Rossin D. F. IQ-2000 MIT Cambridge, Massachusetts, USA. pp 83-91.
- ^{xiv} CIHI Data Quality Manual, Version 1, April 2001 p 48.
- ^{xv} CIHI Data Quality Framework Evaluation Instrument, Version 1 Users' Guide, CIHI 2001.
- ^{xvi} Iezzoni, L. I. (1997) Assessing Quality Using Administrative Data. *Annals of Internal Medicine*. Vol 127(8 part2), pp 666-674.
- ^{xvii} Last, J. M. *A Dictionary of Epidemiology*, 4th Edition. Oxford University Press, 2001.