Quality Assurance Processes Applied to the Discharge Abstract and Hospital Morbidity Databases

December 2007

Canadian Institute for Health Information
Institut canadien d'information sur la santé

Canadian Institute for Health Information
495 Richmond Road
Suite 600
Ottawa, Ontario
K2A 4H6

Phone: 613-241-7860
Fax: 613-241-8120
www.cihi.ca

ISBN 978-1-55465-222-8 (PDF)

How to cite this document:
Canadian Institute for Health Information, *Quality Assurance Processes Applied to the Discharge Abstract and Hospital Morbidity Databases* (Ottawa: CIHI, 2008).

Quality Assurance Processes Applied to the Discharge Abstract and Hospital Morbidity Databases

Table of Contents

Introduction

This document reviews the quality assurance activities of the Discharge Abstract Database (DAD) and Hospital Morbidity Database (HMDB) in accordance with the dimensions outlined in CIHI's Data Quality Framework: accuracy, timeliness, comparability, relevance and usability.

Good data quality is the outcome of a solid quality assurance process used to manage a database. But what is "good" data quality? Where does a quality assurance process begin and end? For the DAD, the production of ***accurate*** and valid data begins with the ***timely*** submission of data according to standardized data elements, code sets and coding rules. Once data are submitted to CIHI, systematic quality assurance practices ensure ***comparability*** and ***usability***. ***Relevance*** is maintained through consultation with advisory committees and via the dissemination of comparative and special topic reports.

The Purpose of This Document

This document is intended to serve as a single reference source for information about the data quality and quality assurance processes applied to both the DAD and HMDB databases. It is an update to a document of the same name that was first published in 2002.

What Is the DAD?

The DAD is a national database for information on all acute care hospital separations (that is, discharges, deaths, sign-outs, transfers). Over time, the DAD has also been used to capture day surgeries, long-term care, rehabilitation and other data. The DAD reports on patient discharges within a fiscal year (April 1 to March 31). Over 3.0 million abstracts were submitted to the DAD in 2006–2007, representing approximately 75% of all acute inpatient discharges in Canada. Quebec does not submit data to the DAD; its inpatient discharges are reported to the HMDB and account for 25% of the total inpatient discharges in Canada.

The DAD was originally developed in 1963 to collect data on hospital discharges in Ontario. Over time, it has expanded to include all provinces and territories, with the exception of Quebec. DAD data have been available since 1979–1980, but the comprehensiveness of submissions varied as provinces and territories subscribed to the DAD; therefore, data from these years are not normally used for analysis and are not easily accessible.

What Is HMDB?

Like the DAD, the Hospital Morbidity Database (HMDB) provides a count of inpatients separated, through discharge or death, from a hospital, listed by the primary (most responsible) morbidity (disease) diagnosis.

Management and responsibility of the HMDB was assumed by CIHI from Statistics Canada in 1995 during a transfer of several health services databases. Data for HMDB are downloaded from the DAD for participating provinces. Data files for hospitals in provinces not submitting to the DAD (that is, Quebec) are submitted annually to CIHI by the ministry of health for inclusion in the HMDB. Data included in HMDB are from general and allied specialty hospitals, including acute care, convalescence and chronic care facilities (except in Ontario). Data do not include any outpatient services in any hospital or services in psychiatric hospitals.

CIHI's Data Quality Framework

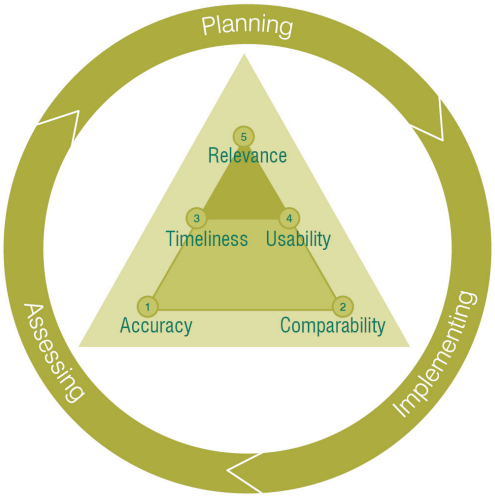
Good data quality is intrinsic to CIHI's mandate to inform public policy, support health care management and build public awareness about the factors that affect health. CIHI engages in rigorous activities to ensure that the data collected and provided are of the highest standard. CIHI has implemented a complete data quality program that includes processes and policies to improve data quality continuously, both within CIHI and in the broader health sector. Our corporate strategy includes a number of initiatives aimed at prevention, early detection and resolution of data quality issues. CIHI's Data Quality Framework is one of the tools that CIHI has developed to assist its data holdings in the assessment of its data quality. This tool was first developed in 2001 and is reviewed and refined every two years. This document will explore the dimensions, characteristics and criteria for the 2005 version of the Data Quality Framework, which is the most recent version.

The purpose of CIHI's Data Quality Framework is to aid in the systematic assessment, improvement and documentation of data quality for all of the databases and registries at CIHI. There are three components to this tool.

- The first sets out the **data quality work cycle**, which describes a work process that is cyclical and iterative in nature, where a data quality activity is planned, implemented and then evaluated. This work cycle also sets out the roles and responsibilities for each of these data quality activities within the data holding program area, in the Data Quality department and across CIHI.
- The second component includes the **assessment tool**, which describes the five dimensions, 19 characteristics and 58 criteria that capture CIHI's data quality principles. Table 1 in the Dimensions of Data Quality section below details these aspects of the assessment tool.
- The third component of the Data Quality Framework describes the various **documents** that should be produced annually to highlight the data quality limitations for each data holding, its data quality strengths and the action plan to address any data quality issues.

Dimensions of Data Quality

CIHI has adopted five dimensions that comprise CIHI's strategy for an ongoing approach to maintaining data quality.



Accuracy: how well the information in, or derived from, the data holding reflects the reality it was designed to measure.

Timeliness: how current or up to date the data are at the time of release.

Comparability: how consistent the data holding is over time and in its use of standard conventions.

Relevance: how well the data holding meets current and potential needs of users.

Usability: how easily understood and accessed a data holding is.

Table 1 Summary of Dimensions, Characteristics and Criteria

Accuracy	
Coverage	1 The population of reference is explicitly stated in all releases
	2 Known sources of under- or overcoverage have been documented
	3 The frame has been validated by comparison with external and independent sources
	4 The rate of under- or overcoverage falls into one of the predefined categories
Capture and collection	5 Practices exist that minimize response burden
	6 Practices exist that encourage cooperation
	7 Practices exist that give support to data suppliers
	8 Standard data submission forms and procedures exist
	9 Data capture quality control measures exist
Unit non-response	10 The magnitude of unit non-response is mentioned in the data quality documentation
	11 The number of records received is monitored to detect for unusual values
	12 The magnitude of unit non-response falls into one of the predetermined categories
Item (partial) non-response	13 Item non-response is identified
	14 The magnitude of item non-response falls into one of the predetermined categories
Measurement error	15 The level of overall measurement error falls into one of the predetermined categories
	16 The level of bias is not significant
	17 The degree of problems with consistency falls into one of the predetermined categories
Edit and imputation	18 Validity checks are done for each data element
	19 Rules for editing and imputation are logical and consistent
	20 Edit reports for users are easy to use and understand
	21 Imputation is automatically derived from editing
Processing and estimation	22 Documentation for all data processes is maintained
	23 Documentation for all systems, programs or applications is maintained
	24 The processing system has been tested after the last revision
	25 Raw data are saved in a secure location
	26 The estimation bias and variance of the estimates are at acceptable levels

Timeliness	
Data currency at the time of release	27 The difference between the actual date of release and the end of the reference period is reasonably brief
	28 The official date of release was announced in advance of the release
	29 The official date of release was met
	30 Database or registry methods are regularly reviewed for efficiency
Documentation currency	31 The recommended data quality documentation was available at the time of data or report release
	32 Major database or registry reports were released on schedule
Comparability	
Data Dictionary standards	33 Data elements are evaluated in comparison to the CIHI Data Dictionary
	34 Data elements conform to the CIHI Data Dictionary
Standardization	35 Data are captured at the finest level of detail as is practical
	36 The original data element is maintained on the main database for any derived data element
Linkage	37 Standard Geographical Classifications (SGC) can be used
	38 Data are collected using a consistent time frame
	39 Codes are used to uniquely identify institutions
	40 Codes are used to uniquely identify persons
Equivalency	41 The impact of problems related to crosswalks or conversions falls into one of the predetermined categories
	42 Methodology and limitations of crosswalks or conversions are documented
Historical comparability	43 Trend analysis is used to examine changes in core data elements over time
	44 The extent of problems in comparing data over time falls into one of the predetermined categories
	45 Accessible documentation on historical changes to the database exists

Usability	
Accessibility	46 An official subset of microdata is defined, created, made available and frozen per release for users where appropriate
	47 Standard tables and analyses are produced per release
	48 Products are defined, catalogued and/or publicized
Documentation	49 Data quality documentation exists per annual subset release
	50 Database or registry methods documentation exists for internal purposes per annual subset release
	51 A caveat accompanies any official preliminary release
Interpretability	52 A mechanism is in place whereby key users can provide feedback to, and receive notice from, the product area
	53 Revision guidelines are available and applied per annual subset release
Relevance	
Adaptability	54 Mechanisms are in place to keep clients and stakeholders informed of developments in the field
	55 The database or registry can adapt to change
Value	56 The mandate of the data holding fills a health information gap
	57 The level of usage of the data holding is monitored
	58 User satisfaction is periodically solicited

1 Accuracy

Accuracy refers to how well information in, or derived from, the data holding reflects the reality it was designed to measure.

CIHI enhances the accuracy of data contained in the DAD/HMDB in a number of ways. These include:

- Standardized classification systems
- Abstracting manuals and software
- Client support and education
- Processing and corrections activities
- In-year data quality assessments
- Annual application of the Data Quality Framework
- Reabstraction and special studies

1.1 Standardized Classification Systems

Classification systems in health care provide a standard mechanism for the capture and coding of diagnoses and interventions. The coding classification schemes supported by CIHI for the purposes of coding diagnoses and interventions in the DAD/HMBD include:

- ICD-10-CA—Enhanced Canadian version of the 10th revision of the International Statistical Classification of Diseases and Related Health Problems. ICD-10-CA replaces the earlier ICD-9/ICD-9-CM classification.
- CCI—Canadian Classification of Health Interventions, developed to accompany ICD-10-CA. CCI replaces the earlier CCP classification and the procedures component of ICD-9-CM.

1.1.1 What Is ICD-10-CA?

The International Statistical Classification of Diseases and Related Health Problems—10th Revision (ICD-10) was adopted by the World Health Assembly, the decision-making body of the World Health Organization (WHO) in 1990. It is the most recent revision of an international classification of mortality and morbidity statistics.

Recently it has been recognized that a single classification is incapable of 1) meeting the needs of the increasing number and variety of professional groups in the health field; and 2) satisfying the demands for uniform assessment of health problems for decision-making in disease prevention, provision of health care and research on particular health problems. Instead, the development of a family of classifications has been proposed. The International Classification of Diseases is the core classification.

ICD-10-CA, Canada's enhancement to the ICD-10, expanded the codes in ICD-10 at the 4th, 5th and 6th character levels. More than 4,000 extensions were added to the ICD-10 codes. These additions were built in consultation with clinicians and users of morbidity data to ensure the classification met clinical, epidemiological and administrative needs.

Ongoing Maintenance and Updating of ICD-10 and ICD-10-CA
Adaptability, maintenance and updating are critical if a classification system is to be dynamic enough to be used in our rapidly changing world. Unlike previous revisions, ICD-10 allows for enhancements to accommodate newly discovered diseases such as Severe Acute Respiratory Syndrome (SARS). The WHO has established an ongoing maintenance and updating process that ensures input from member states (including Canada) and from interested professional bodies. CIHI has provided the Chair and Secretariat support to the WHO Update and Revision Committee (URC) since 2005. Cumulative updates to the ICD have been available since 1996 and will continue until ICD-11 becomes available. In preparation for ICD-11, countries with clinical modifications are providing electronic versions of their databases to the WHO as a starting point for areas of enhancement to the base classification.

The ICD-10-CA is updated on a regular basis. It incorporates all updates from the WHO Update and Revision Committee and includes recommendations brought forward to CIHI and approved by the WHO and CIHI's Classification Advisory Committee.

1.1.2 What Is CCI?

The Canadian Classification of Health Interventions, referred to as CCI, is a multi-axial classification of health-related interventions that is developed and maintained by CIHI. It contains a comprehensive list of diagnostic, therapeutic and support interventions (>18,000 codes) and includes a tabular listing (listing of codes in alphanumeric order), an alphabetical index and relevant coding standards/recording instructions.

Key Features of CCI
A number of guiding principles were used to develop the new classification. These principles form the basis for some of CCI's key features.

Service Provider and Service Setting Neutral
One of the key features of the classification is its service provider and service setting neutrality. The classification has been developed in such a way that modes of practice are not reflected in the code structure. Therefore, the same codes are intended to be applicable regardless of whether a physician, a nurse or a respiratory technologist performs the interventions, or whether the intervention is performed in an operating room, emergency department, clinic or physician's office.

Multi-Axial and Hierarchical Structure
The code structure design uses a multi-axial approach to identify, for example, the body system/anatomy involved and the intervention performed, including the approach and technique or device used. The coding structure is also designed to be hierarchical to facilitate data analysis by providing roll-up and roll-down capabilities at various levels (for example, anatomy, intervention and qualifiers).

Comprehensive
The classification has a significantly expanded scope to meet the needs of organizations across the continuum of health services. The comprehensive range of interventions reflects the broad spectrum of providers and variety of applications beyond traditional classifications.

The classification provides NOS (Not Otherwise Specified) and NEC (Not Elsewhere Classified) categories, where required, to ensure that all interventions can be classified. By using inclusion and exclusion terms, the classification also clearly identifies interventions and their common synonyms and maps them to each conceptual term.

Relevant
The classification has been developed to ensure that the meaning of each conceptual term is unique and clinically significant. Complex and multi-component interventions are identified, where possible, by a conceptual term that recognizes the various levels of complexity. Furthermore, the developmental process included input from selected experts from various clinical specialties, to ensure that the classification is clinically relevant.

Simple
The classification's code-building logic has been made apparent, where possible, to users in order to enhance comprehension and improve coding practices and data quality. As well, it should facilitate the capturing of data through controlled or natural language interfaces as part of the electronic health record. Each code description has been kept as short and simple as possible, while still providing maximum detail supported by clinical documentation to meet the needs of users.

Dynamic and Expandable
Ongoing maintenance and updating of the classification is facilitated by reserving blocks of codes, which allows for future growth or changes.

Restricted to Procedure-Related Information
For the most part, diagnostic or other non-procedure variables are not included in the intervention code. This information will be collected elsewhere in the diagnosis portion of the abstract.

Ongoing Maintenance and Updating of CCI
CCI is revised/updated on a regular basis to reflect changes in practice and technologies used to perform the various interventions.

1.1.3 Canadian Coding Standards

The initial version of the Canadian Coding Standards for ICD-10-CA and CCI was released in 2001. These standards are amended, enhanced and annually reviewed and tested by representatives from all provinces and territories on the National Coding Advisory Committee. The standards are available in printable and searchable document format (PDF) on the CIHI website and may be downloaded free of charge. The standards are developed:

- to support a consistent and national approach to data capture and data quality
- to decrease subjectivity in clinical coding
- to reinforce the coding rules embedded in the ICD-10-CA
- to clarify the Canadian enhancements to the WHO's ICD-10
- to reinforce the coding rules embedded in the CCI
- to accurately reflect case complexity
- to aid in future grouper development
- to enhance CIHI data user confidence

1.2 Abstracting Manual and Software

1.2.1 The DAD Abstracting Manual

The DAD Abstracting Manual is the tool provided to clients, in PDF, to guide them with the data collection process for the demographic, administrative and clinical data elements. The manual is also used by researchers and by external software vendors who design abstracting (data collection) systems for use in hospitals. The manual is divided into two sections: the core section, with requirements applicable at a national level; and the provincial variation section, related to specific requirements for a province/territory. The data elements are defined as mandatory or optional. For each data element, a description of the data element with the corresponding valid values, an example of how the field is to be used and applicable edits to the field are documented in the DAD Manual.

The DAD Abstracting Manual has been designed around the group (19 groups) and field (161 fields) concept. Each data element has been assigned a group and field number for easy reference. Many data elements need to be collected only once on each abstract. However, the abstract provides flexibility in instances where a data element needs to be repeated. For example, the 2007–2008 DAD abstract can accommodate data collection on 8 provider numbers, 25 diagnoses, 20 interventions and 6 special care units.

1.2.2 Abstracting Software and the Role of External Software Developers (Vendors)

In order to standardize and ensure accurate data collection, CIHI's respective data suppliers purchase the necessary software infrastructure from external software vendors to enable data collection and submission. Data collection or abstracting software is generally developed by external companies or by hospital information systems staff. At CIHI, the external companies are commonly known as "vendors."

The abstracting software is developed according to CIHI electronic data submission specifications, standards and data holding user documentation (that is, DAD Abstracting Manual). Data providers (for example, hospitals) select and contract vendor systems that best meet their information and internal reporting needs. The abstracting software is either "stand-alone" or interfaced with hospital systems such as the Patient Registration or Admission/Discharge/Transfer (ADT) system. Where the abstracting system is interfaced, many of the demographic data elements are downloaded into the software.

Starting with the 2002–2003 DAD, vendors have been provided detailed edit (validation) specifications that matched the CIHI production system for the DAD. Prior to this, vendors developed their edits by using the DAD Abstracting Manual abstracting rules and error message descriptions. As of 2002–2003 DAD (for use with ICD-10-CA/CCI) CIHI expects that the vendor abstracting systems will incorporate all CIHI edits, which should improve data quality. If the vendor software is developed appropriately, it should edit each data element at the time that data are entered by the coder. Inter-field relationship edits should occur at some point before the abstract is sent to CIHI. An example of a single data element edit is the discharge date, where it must be in the valid format of year-month-day (YYYYMMDD). An example of inter-field edits is on the discharge date check to see if the discharge date falls within the submission period, that it occurs after the admission date, etc. The use of this software allows the data to be edited "at source" and corrected prior to being submitted to CIHI where they are edited again in a "batch mode."

Vendors are expected to submit test files to CIHI for each province/territory where they have clients. Facilities cannot submit "live" data until both the vendor and facility have submitted a successful test submission. There should be a sufficient number of test records on this file to provide good feedback to the vendor. A minimum of 250 records is expected. This process tests the control record, file format and file size, and performs a full edit test on the file. Feedback (that is, submission summary report, submission error detail file) is provided to the vendors, identifying problem areas where changes are required. CIHI staff analyzes the reports and supports the vendor through CIHI's system development process. Vendors are expected to correct all of their errors before submitting a second or subsequent test. Vendors are expected to test the various record types (for example, original abstracts, corrections, additions and deletions) within their test submissions. Starting with 2007–2008, vendors are expected to be 100% error-free on specified fields in order for CIHI to classify the test as "successful" according to pre-set criteria.

For 2007–2008, CIHI implemented an additional vendor testing phase for the CMG+ and DPG grouping methodologies. Licensed grouper vendors were required to group the data in a file provided by CIHI and return the file to CIHI for verification of the vendor-assigned grouper variables.

A list of vendors that have successfully passed CIHI's testing requirements is posted on the CIHI website. Clients are able to submit their facility test only after their vendor's name is posted on the website.

CIHI does not certify vendors but will post the vendor test results when they are deemed successful.

CIHI/Vendor Relationship
The data provider contracts with the vendor of their choice for products and services. CIHI sets the standards and receives the data from the data providers. CIHI provides support to the vendors for the interpretation of the standards, electronic data submission and edit specifications. CIHI liaises with vendors on an ongoing basis to provide fiscal updates to specifications, to identify issues and to provide feedback encountered during the transmission of data from client sites.

1.3 Client Support and Education

1.3.1 DAD Client Services Representatives

In the context of supporting accuracy through quality assurance practices for consistent coding and abstracting, CIHI's client services representatives (CSRs) serve as liaisons to data suppliers related to issues with the DAD. The CSR is responsible for providing direct client support related to the DAD products, assisting in the development and delivery of educational programs, completing annual updates to the DAD Abstracting Manual, testing the DAD production system, providing site visits to clients, providing data quality expertise and building relationships with provincial/territorial data consultants, health organizations and data users. CSRs have been assigned specific provinces/territories to ensure that effective and efficient support and expertise is provided. The CSRs are certified with the Canadian College of Health Information Management and have a minimum of five years' experience in the health information management field. The CSRs respond to client queries related to the DAD, usually within 48–72 hours. Many of these queries are answered using CIHI's eQuery tool.

1.3.2 CIHI's eQuery Tool

In December 2006, CIHI launched a bilingual eQuery tool. The eQuery tool replaces and expands on the Coding Query service that was launched in 2001 to support the implementation of ICD-10-CA and CCI. It allows clients to submit queries to the following programs at CIHI:

- Classifications
- Case Mix
- Discharge Abstract Database (DAD)
- National Ambulatory Care Reporting System (NACRS)
- National Rehabilitation Reporting System (NRS)
- Ontario Mental Health Reporting System (OMHRS)
- Home Care Reporting System (HCRS)
- Continuing Care Reporting System (CCRS)
- Portal Services

The eQuery tool allows users to search through a knowledge base of previously submitted questions and answers. Users can also submit queries through a common point of entry. Query responses are sent via email in the language of the query.

The eQuery tool has collaborative features that allow CIHI staff to seek and incorporate input from other program specialists within CIHI, in order to provide a complete and efficient response to client requests for information. The eQuery tool knowledge base is used to develop and improve existing manuals, coding standards and education sessions.

1.3.3 CIHI Education Program

The Education Program is a core function of CIHI. In the context of data accuracy, education programs facilitate the understanding of health information and CIHI products and services. This program offers its clients a series of workshops and distance learning courses (paper-based, elearning and/or teleconference) to support the implementation and maintenance of national standards and reporting systems.

Sessions focusing on the interpretation, uses and application of data, indicators and other information tools are also offered. Through these initiatives, the Education Program enhances the quality of data submissions to national databases and registries as well as ensures the correct interpretation and application of data.

For the DAD, the Education Program is developed with our internal partners in Classifications and Case Mix. The program goes a long way to promote and maintain the integrity of this CIHI database. The Education Program focuses on data collection (for example, coding and abstracting), data processing (for example, submissions, errors and corrections) and outputs (for example, report interpretation) with data quality being emphasized in all three areas. Refer to the following table for the types of education provided by CIHI related to DAD/HMDB.

Table 2 Overview of Educational Programs for DAD/HMDB Data Providers

Focus	Session	Highlights
Annual update	What's New for DAD in 2008–2009 What's New for Classifications in 2008–2009 What's New for Case Mix in 2008–2009	• Introduces and reviews the annual revisions to the DAD, Classifications and Case Mix. • Done for both internal and external clients.
Input and data quality	Applied Diagnosis Typing, Main Problem/Other Problem Assignment (2008–2009)	• This elearning course provides a general review of diagnosis types and main/other problem assignment with emphasis on the importance of these data, uses and the consequences of incorrect assignment. There are case studies to test comprehension.
Input and data quality	Applied ICD-10-CA and CCI: Series 2	• This elearning program contains a set of advanced-level inpatient case studies to assess coders' understanding of classifications, coding conventions and coding standards. • Provides an opportunity to reinforce and evaluate coding knowledge.
Input and data quality	Search Techniques for ICD-10-CA/CCI Using Folio Views (2008–2009)	• This elearning course outlines the basic Folio (electronic book) navigation functions for ICD-10-CA and CCI classifications.
Input and data quality	Specialized Coding Workshops • Coding for Diabetes, Part 2 • Acute Coronary Syndrome: Understanding the Spectrum • Obstetrical Coding—Moving Beyond the Basics • Classifying Complications of Care—Part 1 (2008–2009) Knee Replacement Surgery (2008–2009) • Acute Coronary Syndrome—SLP follow up (2008–2009)	• Provides a depth of knowledge on specialized areas of coding (diabetes, obstetrics, acute coronary syndrome, etc.). • Promotes accurate and consistent national coding of diagnoses and interventions.

Focus	Session	Highlights
Data collection and data quality	Basic DAD Abstracting in 2008–2009	• Provides a detailed overview of the DAD abstracting guidelines for each data element. • Highlights changes for this data year.
Data collection and data quality	Managing DAD Submissions and Corrections (2008–2009)	• Introduces and reviews the CIHI data correction system. • Ensures the integrity of the database.
Data collection and data quality	Improving the Quality of Admitting and Registration Data	• Self-learning interactive tool that ensures accuracy of DAD data elements collected at registration.
Output and application	Understanding the Changes to the DAD Reports in 2007–2008	• Self-learning tool to provide overview of changes made to DAD Reports in 2007–2008.
Output and application	Understanding 2007–2008 eCHAP Reports	• Promotes standardized interpretation and application of information from quarterly/annual comparative reports. • Enhances the overall understanding of ELOS reports and CHAP 1, 2 and 3 reports.
Output and application	Trending and ICD-10-CA and CCI (2008–2009)	• An elearning course that provides background information on the factors critical to trending in ICD-10-CA and CCI. • It also identifies some of the tools and resources used for trending.
Output and application	CMG+ and Resource Indicators 2008	• An elearning course that provides significant detail regarding the evolution of CMG+ methodology for 2008 and the associated indicators (RIW and ELOS).

Focus	Session	Highlights
Output and application	Introduction to the CMG+	• Five elearning modules that assist stakeholders in their understanding of the changes to case mix grouping and in applying the new methodology in their analyses. The five modules are offered in both English and in French. The modules are: – Module 1: Introduction to Grouping Methodologies and CMG+ High-Level Business Rules – Module 2: CMG+ Major Clinical Categories – Module 3: CMG+ Five-Factor Methodology – Module 4: CMG+ Clinical and Statistical Performance – Module 5: CMG+ Resource Indicators—RIW and ELOS Methodologies
Output and application	An Advanced Look at CMG+	• A workshop targeted to audiences who are looking for a more in-depth knowledge of the CMG+ methodology and associated health resource indicators (ELOS, RIW).
Output and application	Introduction to Case Mix for Acute Facilities (2008–2009)	• An elearning course that provides general information about the grouping methodologies that are used in acute care settings.
Output and application	Understanding CACS and DPG Grouping Methodologies (2008–2009)	• An elearning course that provides basic information regarding the acute ambulatory care grouping methodologies and associated resource indicators. The course will also demonstrate the similarities and differences between the CACS and DPG methodologies.

Focus	Session	Highlights
Output and application	Executive Summary of the New CMG+ Grouping Methodology	• A tool that provides a concise summary of the new CMG+ methodology. This summary is intended for executives and/or other decision-makers who would not normally be involved in the analysis of data.
Output and application	Introduction to HSMR (2008–2009)	• An elearning course that offers an opportunity for participants to learn about the use and interpretation of the hospital standardized mortality ratio (HSMR). The intended audience is senior leadership staff.
Output and application	HSMR for the Analyst (2008–2009)	• An elearning course that describes the HSMR in detail and provides more information about the technical aspects of the tool, interpretation of data and some suggestions for analysis.

Education is a bridge between CIHI products and the successful application of the content of these products. Through education initiatives, CIHI is able to work with users (for example, health record personnel) to capture data in an accurate and acceptable manner and provide an opportunity to support the correct interpretation and application of the data (that is, to prevent users from making incorrect assumptions or from drawing incorrect conclusions). Both initiatives (data capture and data use) contribute to data quality and the appropriate application of CIHI's various data holdings, the DAD in particular.

1.3.4 Bulletins

Bulletins provide periodic updates for data collectors, ministries of health and vendors. Bulletins are generally topic-specific and directed to a specific data holding. This communication medium is used to identify issues, provide updates and circulate general data quality information. Bulletins are distributed via regular email channels and are posted on the CIHI website. The distribution list for client groups is identified at the bottom of each bulletin. The authoring unit and bulletin number are identified under the CIHI address on the top left side of the bulletin.

1.4 Processing and Corrections Activities

1.4.1 CIHI DAD Production System Edits

The CIHI production system edits (validations) are developed to ensure validity and integrity of the data submitted to a data holding. The DAD edits contain hard errors (response is impossible) and data quality warnings (response is improbable). There are approximately 780 production system edits in the DAD. The DAD accepts erroneous data on submission from the hospital; however, during data processing these erroneous data are defaulted to "Z" for all hard errors.

Types of Production System Edits are as follows:

a) individual field (data element) edit
   - mandatory or optional (hard error or data quality warning)
   - valid values in appropriate use (specific values, range of values)
   - format (that is, justification, numeric, alpha or alphanumeric fields)

b) inter-field (data element) edits (2 or more data elements)

c) provincial/territorial variations in correct usage

d) abstract against institution file

e) control and batch integrity testing

f) post-grouping methodology edits

CIHI has developed a standard layout for the types of edits listed in a), b) and c) above. These are contained in an MS Excel workbook format and list the core edits (those that apply to all provinces/territories) and provincial/territorial edits as separate worksheets within the workbook. This workbook is maintained for each fiscal year. Enhancements or modifications may be made during a fiscal year.

CIHI's continuous quality improvement activities identify additional validations to be added at the beginning of a new fiscal year. Since the implementation of ICD-10-CA/CCI, CIHI has developed edits that correspond to the ICD-10-CA/CCI Coding Standards. The general "rule of thumb" is to identify these edits as warnings for at least one fiscal year and, through analysis, change them to hard errors in subsequent years. CIHI staff analyzes the open year of data for inconsistencies that the edit programs do not identify and notifies the clients of any issues that can be corrected prior to the database year-end closure.

1.4.2 Corrections Submitted to the DAD

Before the information on the abstract is accepted for reporting and storage, it is edited for validity and consistency against the CIHI-defined edits. The purpose of the edit program is to identify the erroneous data, to flag the errors to data collectors in order to facilitate correction, and to guard against both the inclusion of the erroneous data in reports and the storage of erroneous data in the database.

To facilitate the correction process, an error message appears on the Submission Detailed Error File for each error. There are two types of error messages: hard errors and soft, or warning, errors. When a data element is mandatory to record but is missing or invalid, the data element is replaced by "Z," meaning that a correction must be applied to this data element. Soft/warning edits require verification by the abstractor/coder to ensure the accuracy of the information. Each error message consists of three sections: error identification, error description and data as submitted. All error messages are preceded by the following identifiers: group number, field number, occurrence number (where applicable) and edit code number (number identifying the specific error message). The group and field identify the data element to be corrected and the edit code explains the cause of the error and the action to rectify it.

The correction abstract method is designed to allow the hospital to correct errors in the abstract and to create a Correction Abstract in the hospital's abstracting system for submission to CIHI. Detailed vendor specifications describing the requirements are provided to abstracting vendors.

Errors detected by the edit system and reported on the Submission Detailed Error File, or additional changes requested by the client, can be applied at any time after receipt of the initial monthly reports but prior to the production of the closure of the annual database. During the course of a fiscal year, clients can submit additional abstracts that were missing at the time of original submission of a period or delete duplicate abstracts detected in subsequent analyses.

1.4.3 Corrections Applied to the DAD

The edit program processes data as the abstracts are received by CIHI. All errors detected in the abstracts, for the same period, are reported together. After each editing process, the system ensures that the appropriate added values such as Case Mix Group (CMG+) or Day Procedure Group (DPG), expected length of stay (ELOS) and Resource Intensity Weight (RIW) are re-calculated when corrections are applied. The corrections and editing steps are repeated until the client (that is, hospital) successfully corrects the abstracts or the database closes as per its year-end deadline (July 31 for the previous fiscal year).

1.5 In-Year Data Quality Assessments

Trend analyses on specific data elements, as well as data quality tests, are performed during the open year and at the database closure as part of the data quality evaluation process. Significant deviations are promptly reviewed and followed up with clients to allow corrections before the closure of the database.

1.6 Quality Assurance in the HMDB

Since April 1, 2004, Quebec has been the only province whose ministry of health submits data directly to the HMDB rather than through the DAD.

CIHI sends a formal request to the Quebec Ministry of Health to acquire Quebec data, which are typically received by CIHI in November of that year or early in the following year. The data elements from the Quebec file, which contribute to the HMDB, are then extracted for verification. The source data are checked for counts of records and any obvious anomalies. Following validation of these source data with the Quebec Ministry of Health, including resolving any anomalies found, the data are mapped (that is, transformed) to DAD data element names and code values. Record counts on the mapped data file and source data file are compared, at data element level, to ensure agreement between the two data files. The edits that are applied to the DAD data are then applied to the mapped Quebec data where applicable. For example, if a data element in the DAD was not provided by Quebec, then edits related to that particular data element would not be applied to the mapped Quebec data. The diagnosis and procedure codes are validated at this stage and sequence numbers are assigned to them in accordance with editing rules. These data are then checked to ensure that the edits were appropriately applied. The results of this investigation are documented and signed off by the program area. These edited data are then appended to the DAD data for that fiscal year to create the merged DAD/HMDB data file for that fiscal year. Several derived data elements are added to allow isolation of the DAD and HMDB populations from the merged data file. A comprehensive data quality test plan is then applied to this file. This plan consists of a series of logic checks on the data elements (alone or in combination with other data elements) as well as a trend analysis on select data elements to determine if there are any unusual patterns observed over time. The findings from this test plan feed into the various data quality documents that accompany the release of the file (for example, data quality assessment reports, field-specific data limitations).

From this merged data file, an Oracle table dump is prepared and sent to Statistics Canada, along with the above-noted data quality documentation. Statistics Canada then begins its own data quality verification activities. This initiates an iterative process between CIHI and Statistics Canada to discuss anomalies and/or questionable trends that may require further updates to the data quality documentation. Statistics Canada reviews the data file and reports any questions to CIHI. CIHI takes appropriate steps, which may include re-processing of a file, depending on the size or impact of findings. Any findings are also corrected on CIHI's internal files. Following Statistics Canada's completed review of the file, the HMDB data are made available for internal and external reporting purposes. SAS data sets are made available for the internal CIHI analytic community to query, and the data quality documentation is loaded onto the CAD Essential Documents intranet page.

1.6.1 HMDB Specifications Manual

The HMDB Specifications are compiled to serve as a source of information on the steps involved in the processing and editing of data used in the creation of the Morbidity Database. The specifications also serve as a source for the definition, description and interpretation (including code values) of the data elements in the database. The main body of the original specifications document reflects the situation as of 1996–1997, with an addendum written for 1997–1998 to reflect any changes in processing and editing that were made in that year. A third document was prepared to capture processing information and changes for the years 1998–1999 through 2000–2001. Since 2001–2002, the year of the technical merge of the DAD and the HMDB, a separate HMDB processing specifications document has been prepared for each fiscal year. While the information in the specifications documents prior to 2001–2002 is still captured from 2001–2002 onward, some of this information is no longer included in the specifications documents, as summarized in the following table.

Table 3 Guide to HMDB Supporting Documents

Information	Location		
	1994–1995 to 1997–1998	1998–1999 to 2000–2001	2001–2002 Onward
Statistics Canada file layout	Processing specifications (Section 1)	Processing specifications (Section 4)	Processing specifications (STC extraction specifications)
Data element definitions, descriptions, code values	Processing specifications (Section 2)	Processing specifications (Section 1)	DAD Abstracting Manuals (fiscal-year specific)
Steps to create file	Processing specifications (sections 3 and 4)	Processing specifications (sections 2, 3, 4)	Processing specifications (process description)
Data flow diagram	Processing specifications (Section 5)	Processing specifications (Section 2)	CAD essential documents web page (DAD/HMDB data flow diagram)
Provincial/territorial record layouts	Processing specifications (Section 6)	Processing specifications (Section 5)	Processing specifications (Man. and Que.; Que. only as of 2004–2005)
Institution information	Processing specifications (Appendix A)	Processing specifications (Appendix A)	CAD essential documents web page (submitting institution numbers)
Crosswalk between provincial/territorial code values and CIHI code values	Processing specifications (Appendix B)	Processing specifications (Appendix D)	Processing specifications (Man. and Que.; Que. only as of 2004–2005)
CDL and CPL tables	Processing specifications (Appendix C)	Processing specifications (Appendix C)	Program area folders
Database schema diagram	Processing specifications (Appendix D)	Processing specifications (Section 4)	CAD essential documents web page (database schema)

1.7 Annual Application of the Data Quality Framework

The Data Quality Framework has been applied on an annual basis to the DAD and the HMDB. Through this process, data quality strengths and weaknesses are identified and monitored. The findings from the application of the Data Quality Framework are fed into the ongoing data quality enhancement activities.

1.8 Reabstraction: Special Studies in Data Quality

CIHI's corporate data quality program includes a reabstraction special study program aimed at detecting and resolving data quality issues. This program first saw the implementation of a three-year series of reabstraction studies on the DAD for data spanning a four-year period from 1999–2000 to 2002–2003. The goal of this first set of studies was to establish a baseline of information on a pan-Canadian basis on data quality issues pertaining to the clinical and non-clinical data in the DAD.

The first year of this study examined the coding of specific conditions that were used in CIHI's Health Indicator Framework. These included:

- Ambulatory care sensitive conditions
- Caesarean sections
- Coronary bypass grafts
- Hospitalization due to pneumonia or influenza
- Injury hospitalizations
- Total hip replacements
- Vaginal births after a Caesarean section

Additional medical conditions were studied as part of a collaboration with Health Canada's Canadian Perinatal Surveillance System, whose long-term goal was the creation of a national database to monitor a comprehensive set of perinatal indicators. The following additional medical conditions were studied in this first year:

- Rare congenital anomalies
- Rare maternal conditions
- Rare neonatal conditions
- Respiratory distress syndrome
- Third-degree perineal laceration
- Other non-rare maternal and neonatal conditions

In the second year of these studies, the following new conditions were examined:

- Acute myocardial infarction
- Hip fracture
- Hysterectomy
- Total knee replacement

When collecting information from these earlier reabstraction studies, CIHI determined that it needed to collect more information on more complex hospital cases to assist with preparation for the Grouper Redevelopment Project. This project saw the redevelopment of the acute care grouping methodologies to accommodate the coding of the new classification systems that were being phased in across the country. With the advent of ICD-10-CA and CCI, CIHI recognized this was just one of the changes that would need to take place. To assist with this project, CIHI performed a reabstraction study on these more complex and resource-intensive cases.

The final study in this three-year plan was conducted in 2003. Its objective was to collect data from across Canada on the quality of the acute inpatient care data submitted for the first time using the new ICD-10-CA/CCI classification system.

Since then, CIHI has conducted many other reabstraction studies on the DAD for specific provincial ministries. Both the Ontario Ministry of Health and Long-Term Care and Alberta Health and Wellness have approached CIHI on numerous occasions to assist them with their own data quality initiatives. The Alberta government conducted a reabstraction study for the 2002–2003 data year involving ICD-10-CA/CCI coding. The Ontario government has conducted two studies in the last several years: an audit of 2001–2002 data using the old classification systems; and a more recent study of its case-costing hospitals for 2002–2003 and 2003–2004 data. These reports are all available on CIHI's website.

CIHI has recently established a new data quality plan for the reabstraction of the DAD to span the next five years (2005–2010). The first-year study, which examines 2005–2006 data, will provide comparable rates across the nine provinces that submitted to the DAD during the year. This study will be examining the following health conditions:

- Angina and other acute ischemic heart disease
- Hypertension and hypertensive heart disease
- Chronic obstructive pulmonary disease
- Percutaneous coronary intervention
- Heart failure and pulmonary edema
- Diabetes mellitus
- Influenza and pneumonia
- Epilepsy and status epilepticus
- Total hip replacement
- Asthma

Over the next four years, CIHI will be conducting targeted studies to examine the accuracy of clinical coding using ICD-10-CA/CCI, the adoption of coding standards and the implementation of CIHI's new grouping methodology, CMG+. Once completed, these reports will be available on CIHI's website.

A reabstraction study involves the collection, by expert reabstractors, of clinical and non-clinical data elements from hospitals, which are then compared to what was originally submitted to CIHI by these same hospitals. These comparisons allow CIHI to comment on the accuracy of coding, as well as on the impact that variation observed in the reabstracted data has on the measures of hospital output and resource utilization produced from the grouping methodology. It also allows CIHI to assess the adoption of specific coding standards. More recent reabstraction studies have included inter-rater and intra-rater reliability components that facilitate the understanding of the study results. The sampling methodology for these studies usually consists of a two-stage study: facilities are selected in the first stage, and a sample of charts from each facility is selected in the second. For each of these studies, the frame has depended upon which provinces and facilities were submitting to the DAD at the time of the study, and also on the parameters (data elements) being studied.

CIHI is very appreciative of the response received from facilities that have agreed to participate in the studies and of those who have contributed (new) one or two coders to be seconded to these studies as reabstractors. Facilities that participate in these studies receive a customized report with data files. These reports are intended to provide information to each facility on the status of the quality of its data, an explanation of any data quality issues found and recommendations for addressing these issues.

2 Timeliness

Timeliness measures how current or up to date the data are at the time of release by measuring the gap between the end of the reference period to which the data pertain and the date on which the data become available to users.

2.1 CIHI Submission and Correction Deadlines

Timeliness of the DAD and HMDB is monitored regularly according to a production schedule that coordinates all submissions and corrections to data submitted. At the beginning of each fiscal year, CIHI determines fiscal year-end submission and correction deadlines. These deadlines are established well in advance of the fiscal year, and all facilities and provincial and territorial ministries of health are notified. CIHI sets quarterly deadlines for inclusion of data in comparative electronic reports as well as year-end deadlines. Provincial/territorial ministries of health, health authorities or hospitals may set submission deadlines that are earlier than those established by CIHI to accommodate their own jurisdiction's reporting needs.

2.2 Release Dates

2.2.1 Annual File

Typically, the annual DAD data are made available to internal users and external clients in October following the end of the fiscal year.

2.2.2 Periodic Reports

CIHI is responsible for supporting the design and production of hundreds of reports and data sets, based on open and closed DAD data, which are sent to ministries/departments of health, institutions, researchers and policy-makers both internal and external to CIHI.

The output includes:

- eHSR (Hospital Specific Reports): Hospitals receive access to their own data via eHSR two to three days after submitting data to CIHI. These reports are produced on a monthly, quarterly, cumulative quarterly and annual basis.
- eCHAP: Hospitals receive open year, comparative, cumulative, aggregate quarterly data accessed securely online via pre-generated standard reports or user-defined reports. These are produced quarterly and the extraction date is 60 days after the end of the quarter.
- Monthly data cuts to ministries/departments of health: Ministries and departments receive monthly cumulative cuts of open-year data for their jurisdiction as well as a final closed year file.
- SAS data sets: These data sets are used by internal CIHI users for research and provide record-level access to open- and closed-year data.
- CIHI Portal: DAD data from fiscal 2001 to the current open year are loaded in the data warehouse and are accessible on the Portal to authorized users.

3 Comparability

Comparability refers to the extent to which a data holding is consistent over time and uses standard conventions to make it similar to other data holdings.

The DAD and HMDB fulfill the criteria of comparability according to CIHI's Data Quality Framework. The DAD is the central database at CIHI, as it includes detailed information on all hospital discharges in Canada (with the exception of Quebec). Consequently, it is the prime resource for investigating issues around acute care and hospital activity. The DAD also contributes to the following CIHI databases: the HMDB; the Hospital Mental Health Database; the National Trauma Registry; the Ontario Trauma Registry; and the Therapeutic Abortions Database.

3.1 Standardization

The DAD strives to capture comprehensive, detailed information on all hospital discharges. For example, the DAD abstract captures admission/discharge time in hours and minutes (HHMM), which provides the flexibility to analyze the data in any time frame and also facilitates comparison across databases and registries.

The following table highlights the level of detail for select data elements.

Table 4 Level of Detail of Select Data Elements, 2005–2006 DAD

Type of Information	Level of Detail
Patient demographics	• Health card number (unique per individual) • Province issuing health card number • Responsibility for payment • Gender • Birth date: full birth date (YYYYMMDD) • Age (derived from birth date; in years, months and days) • Weight (for newborns and neonates less than 29 days of age at admission)
Patient/institution geography	• Patient province/territory of residence (where the patient lives) • Province/territory of report (where the institution is located) • Patient postal code: submitted at the 6-digit level for all provinces/territories • Postal code of the institution of report: available at the 6-digit level
Admission/discharge information	• Admission date: full admission date (YYYYMMDD) • Admission time: in hours and minutes (HHMM) • Discharge date: full discharge date (YYYYMMDD) • Discharge time: in hours and minutes (HHMM)

Type of Information	Level of Detail
Institution information	• Submitting institution number: unique number (within a province/territory) assigned to each institution that provides data • Institution type: assigned to each institution by provincial/territorial ministries of health • Analytical institution type: assigned to each institution by CIHI based on the "true" type of care the institution provided • Transfer to and transfer from institution numbers and types
Clinical information	• Up to 25 diagnosis codes (ICD-10-CA or ICD-9) • Coding system used (ICD-9 versus ICD-10-CA) • Most responsible diagnosis • Up to 20 intervention codes (CCI, ICD-9-CM or CCP) • Principal intervention • Special care unit information
Provider information	• Up to 8 provider services and provider numbers • Intervention provider services and provider numbers

Source: Discharge Abstract Database, 2005–2006, Canadian Institute for Health Information.

3.2 Linkage

Overall, the various data elements for geography, institution, time and person are sufficiently detailed to create equivalent concepts with other CIHI databases.

The mandatory postal code field and use of the Postal Code Conversion File (PCCF) ensures that any Standard Geographic Code (SGC) used in other CIHI databases can be obtained for the DAD. Province codes may not be consistent, but there is sufficient detail to create standard codes across different databases if required.

The institution number is the widely used standard and is unique within the DAD, which uses provincial institution numbers with an added provincial prefix to make the number. With minor modifications, it can be made equivalent to the institution numbers used in other databases such as the Ontario Trauma Registry.

Although the data are collected on a fiscal-year basis, the range of date variables in the DAD (admission date, discharge date, etc.) allows the user to examine any time frame. This flexibility is especially useful when comparing DAD data to data in the registries. Registries contain cumulative visit data; they do not separate the data by discrete years.

A person dimension can be identified using the encrypted health card number. This is a data element common to most other CIHI databases. Other important data elements that support the person dimension are gender, birth date and postal code. Although variables like birth date and postal code are collected, access to these data elements is restricted.

3.3 Historical Comparability

The DAD program area maintains documents detailing the evolution of DAD data elements (including key provincial abstracting differences) as well as the DAD reporting frame from 1995–1996 to present. The DAD contains a mixture of data submitted to CIHI in ICD-10-CA/CCI and CCP and ICD-9-CM. To assist in comparing data over time, CIHI created conversion tables to map ICD-10-CA diagnoses and CCI interventions to ICD-9 and CCP. Where direct one-to-one conversion was not possible, the conversion tables use a best "force fit" of codes, which compromises comparability. Users are reminded in all data releases to first analyze data in their original classification scheme (ICD-10-CA/CCI) before using the conversion tables. The following table shows the adoption of ICD-10-CA/CCI since 2001–2002.

Table 5 Acute Care and Day Surgery Data Submission in ICD-10-CA and CCI, by Province and Fiscal Year

Province/ Territory	Classification Used Prior to ICD-10-CA and CCI Adoption	2001–2002	2002–2003	2003–2004	2004–2005	2005–2006	2006–2007
B.C.	ICD-9 and CCP and ICD-9-CM	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓
N.L.	ICD-9 and CCP and ICD-9-CM	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓
N.S.	ICD-9-CM	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓
P.E.I.	ICD-9 and CCP and ICD-9-CM	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓
Y.T.	ICD-9-CM	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓
Sask.	ICD-9 and CCP and ICD-9-CM	✓	✓✓	✓✓	✓✓	✓✓	✓✓
Ont.	ICD-9 and CCP and ICD-9-CM		✓✓	✓✓	✓✓	✓✓	✓✓
Alta.	ICD-9-CM		✓✓	✓✓	✓✓	✓✓	✓✓
N.W.T.	ICD-9-CM		✓✓	✓✓	✓✓	✓✓	✓✓
Nun.	ICD-9-CM		✓✓	✓✓	✓✓	✓✓	✓✓
N.B.	ICD-9-CM			✓✓	✓✓	✓✓	✓✓
Man.	ICD-9-CM				✓✓	✓✓	✓✓
Que.	ICD-9-CM						✓✓

Notes
✓ Partially adopted ICD-10-CA and CCI.
✓✓ Fully adopted ICD-10-CA and CCI.
Blank did not adopt ICD-10-CA and CCI.

The staggered adoption has created some challenges in comparing over time and between jurisdictions.

4 Relevance

Relevance reflects the degree to which a data holding meets the current and potential future needs of users.

4.1 How a Data Element Appears in the DAD

There are several sources for identifying refinements, enhancements and new data elements. These include (1) routine communication between clients and CIHI client services representatives; (2) input from advisory committees; (3) internal CIHI requests for change; (4) program area data analysis; and (5) reabstraction findings. Each of these sources will be described briefly below.

(1) One of the best sources of input to database enhancements is suggestions received directly from the database users and data suppliers. Health information management professionals represent the majority of individuals charged with the task of collecting and verifying information reported by hospitals. CIHI's client services representatives (described in section 1.3.1) serve as liaisons for these individuals in interpreting and communicating issues related to the DAD. Suggestions are often made for the addition or modification of data elements that would enhance the utility of the DAD to hospitals. Informal logs are maintained by CIHI client service representatives, and suggestions are discussed periodically with the program area leads and managers. These suggestions are reviewed for practicality and appropriateness and then are aggregated to a list that is discussed at the National Clinical Administrative Databases (NCAD) Steering Committee (described in section 4.2 below).

(2) CIHI supports and benefits from many advisory committees that are multidisciplinary in membership and geographically representative. These committees are often created to provide specific advice on issues related to the activities of CIHI. In turn, CIHI staff are often invited to participate on external advisory committees and are given opportunities to discuss health information issues as well as epidemiological surveillance issues. These groups provide an excellent forum for the generation of new data elements and coding legends that eventually find their way into the DAD. Of particular importance in this regard is the NCAD Steering Committee (see section 4.2).

(3) A variety of internal CIHI departments provide valuable feedback and suggestions for database enhancements. These include the Health Services Research, Health Indicators, Case Mix and Classifications departments.

(4) The DAD and HMDB program areas perform a variety of tests to identify data quality issues. The annual assessment of the database using the CIHI data quality framework results in areas for improvement.

(5) Reabstraction findings (previously described).

4.2 The National Clinical Administrative Databases Steering Committee (NCAD)

Once all suggestions for database enhancements are tallied from each of the sources described above, semi-annual meetings are organized by CIHI for its NCAD. The mandate of NCAD is to provide CIHI with guidance on the impact of strategic decisions concerning the DAD and HMDB. This committee advises CIHI on how to operationalize decisions in the provinces/territories and how to focus on continued improvement of data quality. Each province or territory appoints the members of the committee, with a requirement that each member possess decision-making authority on matters related to the DAD and/or HMDB in their province or territory. On behalf of their jurisdictions, the provincial/territorial representatives bring suggestions related to possible database improvements. They also support committee initiatives and facilitate communication between clients in their jurisdictions and CIHI. Membership on NCAD includes representatives from the Public Health Agency of Canada and Statistics Canada, whose role is to identify topics for consideration that will improve their ability to fulfill their national mandates. Through these processes, new data elements appear in the DAD or new edits are introduced. In 1999–2000, a significant re-development of the DAD abstract occurred in which new elements were added and others were deleted. A major revision to the 2001–2002 DAD abstract was required in order to accommodate the ICD-10-CA and CCI classifications.

4.3 Production System Processes for Change Control

Enhancements and modifications are a necessary measure to ensure continued utility and relevance to any database. In the context of database management, it is important to ensure that a documented and standardized process is in place for change control. For this reason, enhancements or modifications to the DAD or HMDB, while logged throughout the fiscal year, are applied to the databases' production systems only at the beginning of a fiscal year.

There are two main reasons for applying enhancements or modifications only at the beginning of the fiscal year. First, sufficient time is needed to notify software developers (vendors) and provide them the opportunity to make changes to their software and implement this updated software at their client sites; the general rule of thumb is a minimum of six months for existing systems and a minimum of nine months for new or redeveloped systems. Second, a scheduled time for changes in the production cycle ensures consistent data content for a given fiscal year, for internal and external analysis purposes. There are two exceptions to the schedule of changes:

1) To fix system problems (for example,"bugs") to enable the receipt of error-free data.
2) To add or modify edits not initially identified but that are required to improve data quality.

With the above two exceptions noted, the following six steps define the change management policy for the DAD and HMDB:

i. All fiscal updates to existing systems should be determined by the program area in consultation with ITS Operations and finalized prior to September 30.

ii. The ITS Operations Program Coordinator or Program Lead completes the ”request for change” portion of the system request forms for all of the agreed-upon changes.

iii. ITS Operations updates and distributes the corresponding Electronic Submission Requirements Document (that is, ”vendor specifications”) no later than October 30.

iv. The Program Area (user) develops test scenarios and creates test data for all of the agreed-upon changes.

v. ITS Operations programming staff makes the modifications, tests the system (requires user sign-off on the test results) and implements the “new” fiscal year system prior to March 1.

vi. Vendors are notified that the “vendor test system” is operational and available for them to submit vendor test files.

The following seven steps define the procedures for change management:

i. The Program Area completes a ”system request” form detailing the exact requirement.

ii. The Program Area submits the system request log to the Team Lead, ITS Operations (ideally, the program area should provide ”test data” along with the system request form).

iii. The ITS Operations Team Lead reviews the system request to determine extent and assigns the task to the appropriate programmer.

iv. The programmer makes the modification and performs preliminary system testing.

v. The programmer completes the system request form and returns the form to the ITS Operations Team Lead.

vi. The ITS Operations Team Lead reviews the code and preliminary test results and implements the changes in the production systems environment. The testing process requires user testing and sign-off on the test results.

vii. The ITS Operations Team Lead informs the requestor that the change has been completed and closed.

5 Usability

Usability reflects the ease with which a data holding may be understood and accessed.

5.1 Accessibility

5.1.1 Micro-Data

The official subset of micro-data for the DAD is well-defined on both the Oracle production server and the QnA server. The SAS data sets created from both servers are available to users. The DAD closure date is announced to all institutions well in advance. This closure date marks the date the database is frozen and after which no changes or additions are made, except under special circumstances. The DAD is then made available to both internal and external users once the executive summary and Database Limitations document are completed.

5.1.2 Products for DAD Submitting Facilities

DAD submitting facilities receive a series of reporting products that are generated from the DAD. These include the eHSR and eCHAP reports described in Section 2.2.2.

In addition to the above products, many others previously described in this report (the DAD Manual, the CMG Directory and others) can assist in interpreting DAD/HMDB information.

5.1.3 Other External Reports

External reports using data from the DAD are also well-publicized and easily obtained. The annual *Health Indicators* reports, for example, draw heavily from the DAD/HMDB and receive considerable media coverage.

5.2 Documentation

A wide variety of documents exist for the DAD, including the abstracting manual, abstract layout, documentation pertaining to DAD data quality, submission statistics, historical changes and reports production. Processing and data management documents are available in a central location.

All external releases of the DAD are accompanied by the *Executive Summary—Data Quality Documentation: DAD.*

5.3 Interpretability

Internal analysts are notified of DAD-related issues by the DAD analysts via continuous updates to the *Field Specific Data Limitations: DAD* and via the Data Quality Email List. Contact information (phone number and email address) for the program area is provided on the CIHI website, in data quality documentation and in all bulletins for external clients.

The program area evaluates all issues that are identified during the fiscal year. Clients and users are notified promptly if the issues create significant impacts on the data and necessary revisions will be made for the future years. Database changes after the beginning of the fiscal year are kept to a minimum to prevent problems with interpreting and comparing data resulting from changes within the open year of data.

6 Quality Assurance for Specific Analytical Products

The analysis process begins with the identification of the question to be addressed and methodology to be used for analysis. This step is designed to ensure the relevance, utility and feasibility of the analysis for the intended target audience. It is typically undertaken with the guidance of external advisory groups, clinical experts, methodological advisors, biostatisticians and data quality and classification specialists where appropriate. This step draws upon the rich Canadian and international experience with, and literature on, clinical utilization and outcomes analysis.

The resultant analytical plans include well-defined quality assurance processes. Appropriate quality assurance strategies vary, depending on the type of analysis, complexity of the methodologies, data source(s), intended audience and other factors.

For example, where methodologies are being adapted from the literature or from previous Canadian research, their appropriateness in the context of the characteristics of the relevant data set is assessed and methodologies are revised as required. Newly developed methods and approaches are likewise reviewed with internal and/or external expert advisors, and in the context of the data set being used.

Additional assessments of the underlying quality of the data being used for specific analyses may also be undertaken. For example, the estimation of false positive/false negative rates for specific health indicators is an integral part of the multi-year reabstraction process currently underway for the DAD.

When numerical analysis begins, two separate, independently prepared sets of computer codes are typically developed and run to extract data and perform statistical analysis. The results are then reviewed to identify and resolve any discrepancies.

An internal verification process is then undertaken. This step includes, where possible or relevant, a comparison of results with historical trends and/or other data sources, careful review of potential anomalies or outliers, verification of results against control totals established for the database and much more. In some cases, automated quality assurance programs have been designed to assist with this process. Senior analytical staff also review preliminary results, and confidentiality checks and rate stability checks are typically applied at this point. CIHI privacy and confidentiality policies limit the disclosure of potentially identifiable data.

An external verification process typically follows. This involves sending preliminary results (often at a greater level of detail than will ultimately be published in order to facilitate verification) back to the original data sources/subjects. For example, preliminary regional health indicators data are shared with health regions and ministries of health. Data are accompanied by definitions, technical notes and a request to review the results and advise CIHI of any potential issues. Through this process, a number of regions and/or ministries typically replicate the results of the analysis from independent data sources. CIHI analysts are available to provide further information and assist throughout the verification process.

Based on the results of the internal and external verification processes, CIHI analytical staff, in consultation with expert advisors as required, determines whether the information to be presented meets the organization's data quality protocols. They also review the definitions, detailed technical notes and related materials that will accompany the release of data/analytical results to clarify any questions or issues that may have arisen during the verification processes.

Final results and documentation are then produced. These results, as with those circulated in the external verification process, are the subject of an extensive fact-checking process. This iterative process involves pairs of analysts cross-checking draft and final reports against original sources based on standardized checklists to ensure that the results and associated documentation accurately reflect the results of the analysis. Dedicated staff time is assigned to this process.

Summary

As stated in the introduction, the primary purpose of this document is to serve as a single-source reference for the quality assurance processes applied to the DAD and HMDB.

CIHI is committed to ensuring quality data. While there is no standard definition of data quality, there are a number of dimensions of quality that can be consistently applied to the maintenance of data quality. These include accuracy, timeliness, comparability, usability and relevance. Policy-makers, health care leaders and the general public depend on quality data for decisions that affect the Canadian health care system. By conducting ongoing data quality evaluations of CIHI's data holdings and by performing special data quality studies, CIHI facilitates the continuous production of quality information. CIHI already has an established reputation for producing high-quality information. The ongoing challenge is to build on that reputation by continually enhancing the quality of the underlying data.