THE DATA QUALITY STUDY OF THE
CANADIAN DISCHARGE ABSTRACT DATABASE

Julie Richards, Ann Brown, Craig Homan [1]

ABSTRACT

The Discharge Abstract Database (DAD) is one of the key data holdings held by the Canadian Institute for Health Information (CIHI). The Institute is a national, not-for-profit organization, which plays a critical role in the development of Canada's health information system. The Discharge Abstract Database contains acute care discharge data from most Canadian hospitals. The data generated are essential for determining, for example, the number and types of procedures, and the length of hospital stays. CIHI is conducting the first national data quality study of selected clinical and administrative data from the DAD. This study is evaluating and measuring the accuracy of the DAD by returning to the original data sources and comparing this information with what exists in the CIHI database in order to identify any discrepancies and their associated reasons. This paper describes the DAD data quality study and some preliminary findings. The findings are also briefly compared to another similar study. In conclusion, the paper discusses the next steps for the study and how the findings from the first year of the study are contributing to improvements in the quality of the DAD.

1. INTRODUCTION

The Canadian Institute for Health Information (CIHI) is conducting the first national study of the accuracy of the data contained in its Discharge Abstract Database (DAD). The study is aimed at measuring discrepancies and their associated reasons, and providing users of the study with statistically reliable information about the accuracy of the DAD. The study is being conducted on an annual basis for three years. The report first describes the background and objectives of this study, summarizes the study approach and methodology, and presents some preliminary findings. These findings are also compared to another similar study. The paper concludes by discussing the next steps for the study.

1.1 Background

Established in 1994, CIHI is a national, not-for-profit organization that plays a critical role in the development of Canada's health information system. CIHI's mandate is to co-ordinate the development and maintenance of a comprehensive and integrated approach to health information in Canada. CIHI's diverse data holdings are playing an increasingly important role in supporting public debate and decision-making about the Canadian health system.

In order to carry out this mandate, CIHI established a comprehensive and systematic data quality program. The purpose is to enhance the quality of existing data holdings and ensure that new data holdings and information products meet the standards of quality consistent with CIHI's program objectives and CIHI's commitment to excellence. The data quality program involves the implementation of a data quality framework and special studies focusing on specific data quality issues. The framework provides criteria for the analysis and documentation of the data quality dimensions (and their specific characteristics) of

[1] Julie Richards, Ann Brown and Craig Homan, Canadian Institute for Health Information, 377 Dalhousie St, Suite 200, Ottawa, Ontario, CANADA, K1N 9N8. Ann Brown is on corporate assignment to CIHI from Statistics Canada.

accuracy, timeliness, usability, relevance and comparability. Given its size, coverage and importance, the first special data quality study was designed to evaluate the accuracy of the Discharge Abstract Database. The DAD study aimed at measuring the reliability of approximately 50 different data elements contained in the discharge abstract.

1.2 The Discharge Abstract Database

The Discharge Abstract Database is a key data holding at CIHI. It contains clinical and administrative data relating to health care services provided to patients. Using a standard data set, hospitals prepare a discharge summary that contains information retrieved from patient charts. This information is subsequently forwarded to CIHI where it goes through extensive edits prior to being included in the database. The target population of the DAD includes inpatient hospital discharges (acute care, chronic care and rehabilitation) and same-day surgeries. About 75%[2] of all hospital discharges are submitted to CIHI and are included in the DAD. The DAD receives approximately 4.3 million records annually for inpatient and same-day surgeries in hospitals. Using the diagnostic and procedure data contained on each abstract record, health indicators such as the rates of reported pneumonia and coronary artery bypass graft surgery are derived.

The DAD contains demographic (e.g. postal code, date of birth, etc.), non-medical administrative (e.g. health care number, date of admission, etc.), clinical (e.g. diagnosis, procedure, etc.) and value added derived data elements. During the fiscal year studied (FY99/00), two classification systems were in use for diagnosis (e.g. International Statistical Classification of Diseases, Injuries and Causes of Death, Ninth Revision (ICD-9) and the ICD-9-Clinical Modification (ICD-9-CM)) and for procedures (e.g. Canadian Classification of Diagnostic, Therapeutic, and Surgical Procedures (CCP) and Volume 3 of ICD-9-CM). Starting in fiscal year 2001/2002, Canada introduced the new classifications for diagnosis and interventions, the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Canada – Canadian Modification (ICD-10-CA) and the Canadian Classification of Health Interventions (CCI). At the same time a new DAD abstract is being implemented to accommodate these new national classification systems and to adapt to the evolving health information needs of stakeholders. This new DAD abstract is designed to improve inter-provincial standardization, improve linkages among databases and registries, improve specific data elements, add new data elements and delete irrelevant data elements.

The DAD data are used by a variety of organizations in many ways. These include the monitoring of hospital utilization practices, for analyses of health conditions and injuries, and tracking patient outcomes. The DAD is a major source for CIHI's annual Health Report regarding the various health indicators that reflect the key strategic directions and endorsed by the Conference of Deputy Ministers of Health.

1.3 Privacy, Confidentiality and Security

CIHI has developed very strict policies on matters of data privacy, confidentiality, and security. In order to respect personal privacy and safeguard the confidentiality of individual records and facilities, a number of procedures were developed and adhered to during the course of this study. CIHI staff signed confidentiality agreements with the participating hospitals and CIHI agreed not to release the names of the participating hospitals. All results, other than the reports provided to the participating hospitals, are being disseminated in an aggregate form only, so that it is not possible to identify individual patients, physicians or institutions included in the study.

---

[2] Facilities in Quebec and some in Manitoba do not currently submit data to the DAD.

2. STUDY DESIGN AND IMPLEMENTATION

2.1 Objectives of the Study

The goal of the study is to evaluate the accuracy of selected administrative data, at the national level, from CIHI's Discharge Abstract Database. The proposed objectives of the study over three years include: evaluating and measuring the overall accuracy of the DAD; evaluating and measuring the impact of data collection from incomplete charts; evaluating and measuring the coding quality of diagnoses and procedures relevant to CIHI Indicators represented in the Health Indicators Framework; and evaluating and measuring how often diagnoses and procedures are not coded to CIHI guidelines and where additional guidelines may be required.

The proposed data elements and the associated health indicators to be evaluated are identified for each year of the three-year study. In year one, the study focused on the following CIHI health indicators: Injury Hospitalizations; Vaginal Births After Cesarean Sections; Cesarean Sections; Hospitalization due to Pneumonia and Influenza; Ambulatory Care Sensitive Conditions; Coronary Artery Bypass Graft; and Total Hip Replacement.

2.3 Sample Design

The target population consists of all inpatient stays at provincial acute care hospitals reporting to the DAD. The study population consisted of approximately 2.5 million patient stays (charts) in about 550 acute care hospitals.

The study featured a multi-stage stratified sample design (Brown and Richards, 2001). The first stage randomly selected hospitals. The second stage randomly selected inpatient discharge records stratified by health indicators within each participating facility. A representative sample of patient charts was randomly selected within each sample stratum. Additional charts not containing any of these indicators were randomly selected.

Due to operational constraints, the study was conducted at a maximum of 18 hospitals in the first year. In order to get an optimal sample design, it was assumed (at the national level) that the proportion of charts for each indicator that contained a discrepancy was 15% and that the reliability required for the sample was a coefficient of variation (C.V.) of 16.5% (that is, a standard error of 2.5%). Using this assumption and reliability requirement, a minimum sample size was then determined for each indicator. This sample size was increased by 10% to account for chart non-response (unavailability) and a further 10% for possible situations of better than expected productivity by the re-abstractors. The initial sample size for the CIHI health indicators was 2271 with an overall target sample size of 1950 charts.

2.4 Data Collection and Data Processing

CIHI classification specialists re-abstracted the data for the study by returning to the original source of the data at each facility for a one-week period during September to November 2000. For this study, the CIHI classification specialists were given specific training on the collection instrument, exercises and opportunities to standardise outstanding coding issues.

A laptop application was designed and developed by CIHI to facilitate the on-site re-abstraction and identification of possible discrepancies between the DAD and the re-abstracted data. All the original information from the DAD was downloaded to the application immediately prior to the collection week. The application featured the use of pull-down lists of discrepancy codes and reasons, and a comment field that allowed entry of additional information pertaining to the discrepancy. Additional reference material that would ordinarily be available was also loaded onto the laptop.

A processing system was designed and developed for possible re-use in the subsequent years of the study. The separate files in the application were merged and re-structured for analytic purposes.

The classification specialist from the data quality team manually reviewed re-abstracted clinical diagnosis and procedural discrepancy data and ensured that the correct discrepancy code was assigned. This resulted in less than 2% of the discrepancy data for the diagnosis and procedures being revised.

2.5 Identification of Discrepancies and Reasons

The data elements that were re-abstracted included demographic, non-medical, diagnosis, procedure and blood product data elements. All clinical data for the diagnoses, procedures and blood products were re-abstracted blindly. Upon completion of the entry of the data, discrepancies were identified and the discrepancy and reason codes were selected. Non-medical data such as birth date and discharge data were not re-abstracted these were compared with the original data and identified as a match or a discrepancy. If a discrepancy occurred then the non-medical data was re-abstracted and the discrepancy and reason codes were selected from the pull down lists. If clarification was needed, additional text was written in the comments field.

Discrepant data occurs when the re-abstracted data element is different to the original DAD data element. The discrepancy code indicates the type of difference between the DAD and the re-abstracted data. Generally this consisted of:
- Entry missing (the re-abstractor entered data not contained in the original DAD);
- Entry not re-abstracted (the re-abstractor did not code data that was present in the original DAD data); and
- Entry different (the value for the data element was present but differed between the original and re-abstracted data).

For the diagnosis and procedure fields, additional discrepancy codes indicated, first of all, whether the discrepancy occurred in the determination of the most responsible diagnosis or principal procedure. For other diagnoses or procedure codes, discrepancy codes were used to indicate whether or not: the re-abstractor agreed with how significantly each diagnosis affected treatment and/or length of stay (diagnosis type); whether the diagnoses appeared pre- or post- admit; and if there was agreement on the actual diagnosis and procedure code selections used to represent the conditions and procedures.

Reasons were also assigned to each discrepancy. Reasons included: transcription error; inconsistent or conflicting information on the chart; different interpretation of documentation; and code specificity not supported by documentation. Note that more than one reason was possible for each discrepancy. In some cases the reason for the discrepancy between the original data and the re-abstracted data may be due to the unavailability of data; because it was an optional data element for collection; or for other reasons beyond the control of the re-abstractor. Discrepancies such as these were not included in the findings presented in this paper.

2.6 Weighting and Estimation

The sampling weight that indicated the number of charts being represented by each sampled chart was calculated by using the probability of selection. The sampling weight was used to obtain estimates of the percentage of discrepancies for all data elements in the study.

2.7 Non-Sampling Error and Sampling Error

Considerable time and effort was taken to measure and minimize non-sampling errors, since errors that are not related to sampling may occur at almost any phase of a study. Quality assurance procedures were implemented at each step of the data collection and processing to monitor the quality of the data. These procedures included the use of highly skilled re-abstractors, specific training of the re-abstractors with respect to the study procedures, and quality checks to verify the logic of the data processing. (Brown, A. and Richards, J., 2001)

As non-response is a major source of non-sampling error that can potentially bias the results of the study, it was carefully monitored throughout collection. For this study, there were two possible types of non-response. The first was non-response by the randomly selected hospitals for the first stage in the sample design. Of the 26 facilities that were randomly selected, 22 initially agreed to participate, from these 18 facilities were randomly selected. The second type of non-response was chart non-response. The targeted number of charts for re-abstraction was 1950. Of the initial sample of 2271 charts for the CIHI health indicators, 1936 were re-abstracted giving a chart response rate of 84.5%. It is to be noted that 99.3% of the target sample size was achieved.

Edit rules were applied to systematically identify and correct invalid and inconsistent data. Approximately 5% of the discrepancy codes in the non-medical administrative data were corrected and less than 1% of diagnosis codes.

As the findings of the study are based on a sample, the estimates are subject to sampling error. The reliability of the estimates of the percentage of discrepancies is being measured by calculating their coefficient of variation. As these are preliminary findings only, coefficient of variations have not been produced for all of the findings. After the second year of the study has been completed, final findings will include information about the reliability of the estimates.

3. PRELIMINARY FINDINGS FROM THE FIRST YEAR OF THE STUDY

This section presents the preliminary findings from the first year of the re-abstraction by health indicator and by demographic, non-medical, diagnostic and procedural discrepancies.

3.1 Health Indicator Findings

Table 1 shows the estimated percentage of the false positive and false negatives. Since indicators are value-added variables that are derived from diagnosis and procedures, the estimated percentage of false positives and negatives is due to discrepancies occurring in the diagnoses and procedures.

Table 1: Percent Re-abstracted to a Different Indicator (False positive and False Negative)

Health Indicator	Estimated % False Positive[1]	Estimated % False Negatives[2]
Ambulatory care sensitive condition (diabetes, asthma, alcohol/drug psychoses, non-dependant abuse of drugs, depression and hypertension)	10.9	6.5
Cesarean section	0.1	0
Coronary artery bypass graft	1.1	0
Hospitalization pneumonia	7.1	12.8
Injury hospitalization	5.5	7.1
Total hip replacement	0.8	0
Vaginal births after cesarean	0.7	10.5
Not assigned to any of above conditions	1.6	1.4

[1] A false positive occurs when the diagnosis and procedure codes in the original abstract met the criteria for inclusion in the indicator but the re-abstracted codes did not.
[2] A false negative occurs when the diagnosis and procedure codes in the original abstract did not meet the criteria for inclusion in the indicator but the re-abstracted codes did.

This table shows that the indicators related to elective procedures such as coronary artery bypass grafts or hip replacements, and those where the treatment is not as complex as in cesarean sections and vaginal births after cesarean section, were the most accurately coded. Diagnoses with more complex treatment protocols and those that are less easily defined such as pneumonia, injuries and ambulatory care conditions showed a higher degree of discrepancies. It should also be noted that the percentage of false negatives in vaginal births after cesarean section was much higher than the percentage of false positives. This is in

keeping with the assumption that it is much easier to miss collecting the previous cesarean section than it is to capture it when it doesn't exist.

3.2 Demographic and Non-medical Findings

Table 2 presents the estimated percentage error for the demographic and non-medical data elements where the data elements are in descending order of the size of the estimated percentage of the total discrepancy. The non-medical discrepancies are sub-divided into three categories: Entry missing; Entry not re-abstracted; and Entry different (as defined earlier). Table 2 shows that for the field "wait time in emergency", the information was either missing from the original DAD submission (5.5%) or was different from that re-abstracted (5.1%); resulting in an estimated error of 10.6% for this data element.

Table 2: Estimated Demographic and Non-Medical Discrepancies by Data Element[1]

Data Element	Entry Missing % Estimated Error[2]	Entry Not Re-abs % Estimated Error[2]	Entry Different % Estimated Error[2]	Total % Estimated Error[3]
Admission Category	0.1	0.0	13.8	13.9
Wait Time in Emergency	5.5	0.0	5.1	10.6
Discharge Hour	0.0	0.0	9.8	9.8
Postal Code	0.0	0.0	9.0	9.0
Entry Code	0.0	0.0	6.5	6.5
Readmission Code	0.0	0.0	5.0	5.0
Admit by Ambulance	3.2	0.5	0.6	4.3
Unplanned Readmission	3.4	0.0	0.0	3.4
Institution To	0.4	1.0	0.3	1.7
Institution From	0.7	0.3	0.1	1.1
Weight	0.3	0.1	0.7	1.1

The following data elements yielded <1% total estimated error: Autotransfusion Indicator, Blood Transfusion, SCU Days, Admit Hour, Albumin, SCU Unit Number, Other Blood Product, Health Care Number, Supplemental Death Code, Abstract Overflow, Birth Date, Exit Alive, and Red Blood Cells,

The following data elements yielded a 0% total estimated error: 2nd Chart/Registry Number; Admit Date; Date of Last Menses; Death Code; Discharge Date; Estimated Birth Date; Gender; Plasma; Platelets; Prov.Issuing HCN

[1]The population (N) is 2,391,440.
[2]The estimated percentage where this discrepancy was identified for this data element.
[3]The estimated percentage that had any discrepancy identified for this data element.

Table 3 presents the distribution by percentage of the reasons for these discrepancies, where the reasons are in descending order of the size of the estimated percentage. For example, Table 3 indicates that "the original abstractor missed information on the chart" (reason code P) is the most common reason (27.6%) for non-medical discrepancies.

Tables 2 and 3 provided information on the demographic and non-medical discrepancies. Table 2 illustrated that there were only two data elements where discrepancies appeared in more than 10% of the charts, and a further three that occurred between 5% and 10%. Many of the data elements, however, had discrepancies in less than 1% of the charts.

The two data elements where discrepancies occurred in over 10% of the charts were Admission Category (13.9%) and Wait Time in Emergency (10.6%). Discharge Hour (9.8%), Postal Code (9.0%), and Entry Code (6.5%) were the next most frequent. Many of the Admission Category discrepancies arose where hospitals were identifying all patients admitted through the Emergency Department as "Emergent" when only patients with life-threatening conditions should be designated as such. In addition, there was some difficulty in identifying proper admission codes for Obstetrical patients.

Table 3: Estimated Demographic and Non-Medical Discrepancies by Reason Code

Reason Code	% reason Entry Missing[1]	% reason Entry Not Re-abstracted[2]	% reason Entry Different[3]	% reason All Discrepancy Types[4]
A Transcription error	0.0	0.0	0.7	0.5
B Incomplete documentation	0.0	0.2	0.0	0.0
E Code specificity not supported	0.0	0.7	0.1	0.1
F Different interpretation	0.4	16.0	4.6	4.2
K Other grey area coding	0.0	0.0	0.4	0.3
L Inconsistent or conflicting information	0.0	27.9	1.4	1.8
M Coding contrary to CIHI guidelines	5.4	0.0	10.9	9.7
N Hospital policy	36.1	15.3	10.4	14.9
O Coding error	0.0	0.0	0.0	0.0
P Information on chart missed	47.1	1.9	24.3	27.6
Q Mathematical/counting error	0.0	0.0	4.6	3.7
R Downloaded incorrectly	0.1	0.7	27.0	21.8
V Other	0.6	15.2	2.6	2.5
W No apparent reason	10.2	22.1	13.1	12.9
Total	**100.0**	**100.0**	**100.0**	**100.0**

[1]The estimated population (N) of reasons for 1-entry missing discrepancies was 366,390.
[2]The estimated population (N) of reasons for 2-entry not re-abstracted discrepancies was 55,817.
[3]The estimated population (N) of reasons for 3-entry different discrepancies was 1,735,497.
[4]The estimated population (N) of reasons for total non-medical discrepancies was 2,157,704.

It is to be noted from Table 3 that 27.6% of all non-medical discrepancies were due to coders missing information (reason code P) that was available on the chart. In addition 21.8% of all non-medical discrepancies occurred because of incorrect data downloads (reason code R)- Admission Discharge Transfer (ADT) download inconsistent with the rest of the chart.

3.3 Diagnosis Code Findings

The diagnosis codes in this study were compared using four different elements: the prefix, the actual code, the suffix, and the diagnosis type. In Table 4, all possible discrepancies for any of these elements are shown sub-and divided into categories. The percentage discrepancies found for each of these categories were: Most Responsible Diagnosis (MRDx) 13.4%, Comorbidity and Complication (CC) typing issues 11.0%, other diagnosis discrepancies 31.2%, and discrepancies associated with the code itself (regardless of whether Mrdx or CC) 6.5%.

Table 5 shows the distribution by percentage, of the reasons for the MRDx discrepancies and all other diagnosis discrepancies. For example, Table 5 indicates that "different interpretation of the documentation" (reason code F) is the most common reason (40.1%) for MRDx discrepancies.

Table 4: Estimated Diagnosis Discrepancies

Diagnosis Discrepancies	% Estimated Error of Estimated Population
MRDx Discrepancies[1]	
6 – MRDx coded as different type in DAD	6.3
7-MRDx missing in DAD	3.2
11-Post admit typed as MRDx in DAD	0.0
13-Secondary dx coded as MRDx in DAD	0.1
15-Dx not coded, typed as MRdx in DAD	3.9
Total MRDx discrepancies	**13.4**
CC Diagnosis Typing Discrepancies[2]	
8-CCdx coded as type 3 in DAD	2.1
10-Pre admit typed as post admit in DAD	0.2
12-Post admit typed as pre admit in DAD	0.6
14-Secondary dx coded as CCdx in DAD	8.1
Total diagnosis typing discrepancies	**11.0**
Other Diagnosis Discrepancies[2]	
9-CCdx missing in DAD	14.4
16-Dx not re-abstracted, typed as CC in DAD	16.6
18- Transfer Dx missing in DAD	0.2
19-Dx not re-abstracted typed as trans Dx in DAD	0.0
Total Other Dx discrepancies	**31.2**
4-Dx prefix/suffix different[3]	**0.1**
5-Different Dx code for same condition[3]	**6.5**

[1]The denominator for % error of MRDx was 2,504,600 for the estimated population.
[2]The denominator for % error of other Dx was 2,631,600 for the estimated population.
[3]The denominator for % error of all Dx was 5,870,200 for the estimated population.

Table 5: Estimated Diagnosis Discrepancies by Reason Code

Reason Code	% of reasons for MRDx discrepancies[1]	% of reasons for other Dx discrepancies[2]
A Transcription error	0.0	0.3
B Incomplete documentation	0.0	0.0
D Lack of Code specificity	0.3	3.1
E Code specificity not supported	1.5	9.3
F Different interpretation	40.1	12.2
I No significant impact on treatment	4.8	24.3
K Other grey area coding	8.5	6.9
L Inconsistent or conflicting information	8.0	1.8
M Coding contrary to CIHI guidelines	15.9	5.6
N Hospital policy	0.3	0.4
O Coding error	3.6	8.2
P Information on chart missed	15.2	25.5
Q Mathematical/counting error	0.0	0.0
R Downloaded incorrectly	0.0	0.0
V Other	1.6	2.1
W No apparent reason	0.1	0.3
Total	**100.0**	**100.0**

[1]The estimated population (N) of reasons for MRDx discrepancies was 363,900.
[2]The estimated population (N) of reasons for other Dx discrepancies was 1,672,400.

The most common Mrdx discrepancy identified was where the re-abstractor coded a diagnosis as the Mrdx when it had been coded as another diagnosis type in the original abstract (code 6).

For the other categories, the majority of the discrepancies fell into one of three areas:
- the original coder captured a code that the re-abstractor did not feel was significant (code 16);

- the re-abstractor coded a significant condition that the original coder did not (code 9); and
- the re-abstractor and original coder used a different code to represent the same condition (code 5).

The most common reasons for these discrepancies included:
- the re-abstractor disagreeing that the diagnosis significantly impacted on the treatment (Reason I);
- the original coder missing information that was documented on the chart (Reason P); and
- different interpretations of the documentation (Reason F).

3.4 Procedure Code Findings

Table 6 presents the estimated percentage of procedure discrepancies. These are presented by principal procedure, other procedures and discrepancies related to anaesthetic technique. For example, a total of 10% of all principal procedures re-abstracted were identified as having a discrepancy with the original data. Table 7 presents the individual reason codes for the procedure discrepancies as percentages of total reasons. For example, Table 7 indicates that the original abstractor missed information on the chart (reason code P) is the most common reason (56.0%) for principal procedure discrepancies.

Table 6: Estimated Procedure Discrepancies

Procedure Discrepancies	% Error of Estimated Population
Principal Procedure Discrepancies[1]	
22- Principal proc as other proc	0.3
23- Principal proc missing	4.9
25- Proc not coded, orig as P.P.	4.8
Total Principal Procedure discrepancies	**10.0**
Other Procedure Discrepancies[2]	
24- Other procedure missing	10.6
26- Proc not coded, orig as other	12.7
Total other procedure discrepancies	**23.3**
21- Procedure code different[3]	**5.3**
Anaesthetic discrepancies[4]	
27- Anaesthetic type different	6.2
28- Anaesthetic type missing	2.1
29- Anaesthetic type not coded	5.6
Total anaesthetic discrepancies	**13.9**

[1]The denominator for % error of principal procedure was 1,167,000 for the estimated population.
[2]The denominator for % error of other procedures was 1,130,900 for the estimated population.
[3]The denominator for % error of all procedures was 2,300,900 for the estimated population.
[4]The denominator for % error of anaesthetic techniques was 930,600 for the estimated population.

Table 7: Estimated Procedure Discrepancies by Reason Code

Reason Code	% of reason for PP Discrepancies[1]	% of reason for other procedure Discrepancies[2]	% of reason for Anaesthetic Discrepancies[3]
A Transcription error	0.0	0.5	0.0
B Incomplete documentation	0.0	0.0	0.0
D Lack of Code specificity	0.0	0.9	0.0
E Code specificity not supported	0.3	5.7	0.3
F Different interpretation	35.6	20.6	0.8
I No significant impact on treatment	0.0	1.1	0.0
K Other grey area coding	1.3	7.8	0.7
L Inconsistent or conflicting information	0.3	4.0	0.9
M Coding contrary to CIHI guidelines	2.1	6.2	38.6
N Hospital policy	0.1	5.4	10.2
O Coding error	0.5	15.6	0.0
P Information on chart missed	56.0	28.2	45.9
Q Mathematical/counting error	0.0	0.0	0.0
R Downloaded incorrectly	0.0	0.0	0.0
V Other	1.2	3.7	0.5
W No apparent reason	2.6	0.3	2.0
Total	100.0	100.0	100.0

[1]The estimated population (N) of reasons for Principal Procedures discrepancies was 154,000.
[2]The estimated population (N) of reasons for other procedures discrepancies was 502,000.
[3]The estimated population (N) of reasons for anaesthetic discrepancies was 130,700.

Table 6 illustrated the following discrepancies: Principal Procedure 10.0%, other procedures 23.3%, differences in the procedure code itself 5.3%, and anaesthetic types 13.9%.

The most common procedure discrepancies were:
- the original coder captured a procedure that the re-abstractor did not (code 26);
- the re-abstractor coded a procedure that the original coder did not (code 24); and
- the re-abstractor and original coder used a different code to represent the same procedure (code 21).

The most common reasons for these discrepancies included:
- the original coder missing information that was documented on the chart (Reason P) and different interpretation of the documentation (Reason F) for procedures; and
- the original coder missing information that was documented on the chart (Reason P), and coding contrary to guidelines (Reason M) for anaesthetic technique.

4. STUDY COMPARISONS

Although we are unaware of any re-abstraction studies conducted at the national level in Canada, the Ontario Hospital Association, the Ontario Ministry of Health and the Hospital Medical Records Institute (HMRI, a predecessor of CIHI) conducted a provincial re-abstraction study in 1991. The Baseline (Re-abstracting) Study was undertaken to determine the quality of Ontario data that was submitted to HMRI in the fiscal year 1988/99 (OHA, OMH, HMRI, April 1991). While there are some similarities in the sampling and re-abstraction methodology, note that the current CIHI study is of national scope while the Baseline study assessed Ontario data. Further, the Baseline Study does not appear to address the following discrepancies: (code 11) where the post admit Dx was typed as MRDx in DAD; (code 13) where the secondary Dx was coded as the MRDx in DAD; (code 15) where the Dx not coded, typed as Mrdx in DAD; and discrepancy (code 25) Proc not coded, orig as PP. In addition, changes in coding practises such as the change in classification systems in Ontario from ICD-9 to ICD-9-CM; changes to the data elements collected; and increased access to education and coding query services, may limit the comparability of these studies. Therefore, caution should be used when comparing the two studies. Comparison of results, where we could identify common elements is presented in Table 8.

Table 8: Comparison of Study Findings

Data Element	CIHI DAD 1999/00[1] % discrepancy	Ontario 1988/89[2] % discrepancy
Admit Date	0.0	0.5
Discharge Date	0.0	0.9
Health Care Number	0.2	1.3
Postal Code	9.0	7.0
Gender	0.0	0.8
Weight	1.1	2.5
Birth Date	0.1	1.5
Entry Code	6.5	1.9
Institution From	1.1	2.2
Institution To	1.7	2.7
Exit Alive Code	0.1	1.0
Death Code	0.0	0.0
Supplemental Death Code	0.2	0.1
Most Responsible Diagnosis	13.4	19.3
Principal Procedure	10.0	12.3

[1] percent discrepancy of the estimated population
[2] percent discrepancy of the sample. The paper presents % match. This was subtracted from 100 to determine % discrepancy.

5. NEXT STEPS

5.1 ICD-10-CA/CCI Guidelines

One of the major objectives of the study is to support the development and enhancement of coding guidelines as Canada moves into the use of ICD-10-CA/CCI. Results from the study relating to diagnosis and procedure discrepancies have been summarized and are in the process of being extensively reviewed by CIHI classification specialists. This analysis will feed directly into the ongoing coding guideline development process and into education sessions provided by CIHI. This will help to ensure that coding issues of national significance will receive appropriate consideration as the guidelines are built.

5.2 Other Uses of the Findings to Improve the Discharge Abstract Database

While it is necessary to further analyze and examine the findings to fully understand the data quality issues, the findings may be used to identify priorities for improvement. After the implementation of the new DAD abstract, should the discrepancy levels continue then non-medical data elements having a discrepancy estimate in excess of 10% should be investigated. This may involve collaboration with the facilities and/or system vendors to fully address specific quality issues. Follow-up processes include the use of appropriate CIHI mechanisms (such as revising the Discharge Abstract Manual documentation, holding education sessions and teleconferences) to support improvements for this as well as for coding practices (as noted in the previous section). For the first year of the study coders missing information that was available on the chart (reason code P) accounted for a substantial number of discrepancies for all types of data (both non-medical and clinical).

5.3 Year 2 & 3

For the second year of the study, the expertise from the first year of the study was used to enhance the software application. This included online validation of diagnosis and procedure codes. While the same study approach is being used, the focus of the second year is on four CIHI health indicators: Acute

Myocardial Infarction; Hip Fracture; Hysterectomy; and Knee Replacement. The second year of the study also includes a sample of charts not assigned to these indicators. After its completion, the two years will be combined to provide a baseline of the accuracy of the DAD as it existed prior to the use of the new classification systems and the new abstract.

The major objective of the third year of the study will be to review the quality of the new DAD, i.e. evaluate the quality of the coding of ICD-10-CA/CCI and the new abstract. Comparing these results with those of the first two years will provide an indication of the effectiveness of the national ICD-10-CA/CCI education program and the implementation of the new standards.

REFERENCES

Brown A., and Richards, J., *"Quality Measurements of the Canadian Discharge Database",* Statistics Sweden International Conference on Quality in Official Statistics, (May 2001).

Canadian Institute for Health Information (June 1, 2000), *Bulletin, "DAD Data Quality Study"*, www.cihi.ca

Canadian Institute for Health Information (2000), "*Improving Timeliness of the Discharge Abstract Database Data Quality Study"*, www.cihi.ca

Canadian Institute for Health Information (2001), "*Health Care in Canada",* www.cihi.ca

Canadian Institute for Health Information (2001), "*Interim Report for the Data Quality Study of the Discharge Abstract Database: First Year National Findings",* www.cihi.ca

Canadian Institute for Health Information (2001), "*Products and Services Catalogue",* www.cihi.ca

Ontario Hospital Association, Ontario Ministry of Health, Hospital Medical Records Institute (April 1991), "*Report of the Ontario Data Quality Re-abstracting Study".*