
Methods

Assessing Validity of ICD-9-CM and ICD-10 Administrative Data in Recording Clinical Conditions in a Unique Dually Coded Database

Hude Quan, Bing Li, L. Duncan Saunders, Gerry A. Parsons, Carolyn I. Nilsson, Arif Alibhai, and William A. Ghali for the IMECCHI Investigators

Objective. The goal of this study was to assess the validity of the International Classification of Disease, 10th Version (ICD-10) administrative hospital discharge data and to determine whether there were improvements in the validity of coding for clinical conditions compared with ICD-9 Clinical Modification (ICD-9-CM) data.

Methods. We reviewed 4,008 randomly selected charts for patients admitted from January 1 to June 30, 2003 at four teaching hospitals in Alberta, Canada to determine the presence or absence of 32 clinical conditions and to assess the agreement between ICD-10 data and chart data. We then recoded the same charts using ICD-9-CM and determined the agreement between the ICD-9-CM data and chart data for recording those same conditions. The accuracy of ICD-10 data relative to chart data was compared with the accuracy of ICD-9-CM data relative to chart data.

Results. Sensitivity values ranged from 9.3 to 83.1 percent for ICD-9-CM and from 12.7 to 80.8 percent for ICD-10 data. Positive predictive values ranged from 23.1 to 100 percent for ICD-9-CM and from 32.0 to 100 percent for ICD-10 data. Specificity and negative predictive values were consistently high for both ICD-9-CM and ICD-10 databases. Of the 32 conditions assessed, ICD-10 data had significantly higher sensitivity for one condition and lower sensitivity for seven conditions relative to ICD-9-CM data. The two databases had similar sensitivity values for the remaining 24 conditions.

Conclusions. The validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions was generally similar though validity differed between coding versions for some conditions. The implementation of ICD-10 coding has not significantly improved the quality of administrative data relative to ICD-9-CM. Future assessments like this one are needed because the validity of ICD-10 data may get better as coders gain experience with the new coding system.

Key Words. ICD-9-CM, ICD-10, chart data, validity, Canada

The World Health Organization adopted the first version of the International Classification of Diseases (ICD) in 1900 to internationally monitor and compare mortality statistics and causes of death. Since then, the classification has been revised periodically to accommodate new knowledge of disease and health. The sixth revision, published in 1949, was more radical than the previous five revisions because this edition made it possible to record information from patient charts to compile morbidity statistics. Subsequent revisions were made in 1958 (7th Edition), in 1968 (8th Edition), and in 1979 (9th Edition). The United States modified ICD-9 by specifying many categories and extending coding rubrics to describe the clinical picture in more detail. These modifications resulted in the publication of ICD-9 Clinical Modification (ICD-9-CM) in 1979 for coding diagnoses in patient charts (Commission on Professional and Hospital Activities 1986). The latest version, ICD-10, was introduced in 1992 (World Health Organization 1992).

The major differences between the ICD-10 and ICD-9-CM coding systems are: (1) the tabular list in ICD-10 has 21 categories of disease compared with 19 categories in ICD-9-CM and the category of diseases of the nervous system and sense organs in ICD-9-CM is divided into three categories in ICD-10, including diseases of the nervous system, diseases of the eye and adnexa, and diseases of the ear and mastoid process; and (2) the codes in ICD-10 are alphanumeric while codes in ICD-9-CM are numeric. Each code in ICD-10 starts with a letter (i.e., A–Z), followed by two numeric digits, a decimal, and a digit (e.g., acute bronchiolitis due to respiratory syncytial virus is J21.0). In contrast, codes in ICD-9-CM begin with three digit numbers (i.e., 001–999), that are followed by a decimal and up to two digits (e.g., acute bronchiolitis due to respiratory syncytial virus is 466.11).

Canada, Australia, Germany, and other countries have enhanced ICD-10 by adding more specific codes and released country-specific ICD-10 versions, such as ICD-10-Canada (ICD-10-CA; Canadian Institute for Health

Address correspondence to Hude Quan, M.D., Ph.D., Department of Community Health Sciences and Centre for Health and Policy Studies, University of Calgary, 3330 Hospital Dr. NW, Calgary, AB, Canada T2N 4N1. Bing Li, M.A., is with the Calgary Health Region, Calgary, AB, Canada. L. Duncan Saunders, M.B.B.Ch., Ph.D., and Arif Alibhai, M.H.S.A., are with the Department of Public Health Sciences, University of Alberta, Edmonton, AB, Canada. Gerry A. Parsons, R.N. (Ret), is with The Centre for Health and Policy Studies, University of Calgary, Calgary, AB, Canada. Carolyn I. Nilsson, C.C.H.R.A. (c), is with the EPICORE Centre, University of Alberta, Edmonton, AB, Canada. William A. Ghali, M.D., M.P.H., Departments of Medicine and Community Health Sciences, and Centre for Health and Policy Studies, University of Calgary, Calgary, AB, Canada.

Information 2003). However, ICD-10-CA has maintained its comparability with ICD-10. The basic ICD-10 structure, scope, content, and definition of existing codes are not altered in ICD-10-CA. This means that none of the ICD-10 codes are relocated or deleted. ICD-10-CA mainly extends code character levels, from third and fourth levels of ICD-10 to fourth, fifth, or sixth character levels (e.g., from I15.0 for renovascular hypertension to I15.00 for benign renovascular hypertension and I15.01 for malignant renovascular hypertension). A few additions of third- and fourth-level codes were also included in ICD-10-CA in a manner consistent with the existing classification. All of these additional codes are indicated with red maple leaf symbols in ICD-10-CA coding manuals.

To continuously study the health care system and investigate or monitor population health status with ICD-10 data, it is imperative to assess errors that could occur in the process of creating administrative data due to the introduction of the new coding system, ICD-10. We conducted this study to evaluate the validity of ICD-10 administrative hospital discharge data and to determine whether there were improvements in the validity compared with the validity of ICD-9-CM data. To achieve this aim, we reviewed randomly selected charts coded using ICD-10 at four Canadian teaching hospitals, determined the presence or absence of recorded conditions, and then separately recoded the same charts using ICD-9-CM. Then we assessed the agreement between originally coded ICD-10 administrative and chart review data, and the recoded ICD-9-CM administrative data and chart review data for recording the same conditions. This permitted us to compare the accuracy of ICD-10 data relative to the chart review data, with the accuracy of ICD-9-CM data relative to the chart review data for these conditions.

METHODS

Original ICD-10-CA Hospital Discharge Abstract Administrative Data

At each of the four adult teaching hospitals in Alberta, Canada, professionally trained health record coders read through the patients' medical charts to assign ICD-10-CA diagnoses that appropriately described the patient's hospitalization. Each discharge record contained a unique identification number for each admission, a patient chart number, and up to 16 diagnoses. Alberta hospital discharge records have been coded with ICD-10-CA since April 1, 2002. To avoid quality issues in coding during the transition period between ICD-9-CM and ICD-10-CA, we obtained all records for patients with ages

≥ 18 and discharged from January 1, 2003 through June 30, 2003 (i.e., 9 months after the implementation of ICD-10-CA) from the four study hospitals. After stratifying records by hospital, and assigning a random number to each record, we sorted them by ascendance of the random number and assigned a sequence number to each record within hospital. With the aim of having a final sample size of at least 1,000 records from each hospital, we located charts sequentially using a combination of patient chart number and admission identification number unique to admission at each hospital. We ended up reviewing 4,008 charts and did not locate 26 charts (i.e., a 99 percent success rate in locating charts).

Recoded ICD-9-CM Hospital Discharge Abstract Data (Simulating Real-World Coding)

Before April 1, 2002, discharge data were coded with ICD-9-CM and therefore, in our sampling period of January 1 to June 30, 2003, ICD-9-CM data were not available in Alberta. To create a new ICD-9-CM database, we attempted to simulate hospital coders' coding in ICD-9-CM (i.e., "real-world coding"). Four coders who had ICD-9-CM coding experience at these hospitals recoded the 4,008 charts following the ICD-9-CM coding guidelines used at the four hospitals at the average speed of coding staff, spending about 15–20 minutes per chart. These coders were blinded to the ICD-10-CA codes assigned to each record.

Defining Clinical Conditions in ICD-9-CM and ICD-10-CA Data

Through multiple steps, we developed ICD-10 coding algorithms and enhanced the Deyo and Elixhauser ICD-9-CM coding algorithms for adaptation of the Charlson and Elixhauser clinical conditions in ICD-9-CM and ICD-10 administrative data. Our multistep process for doing this is described in detail in a previously published paper (Quan et al. 2005). The ICD-10 coding algorithms used for this study did not contain country-specific ICD-10 codes. When the coding algorithms were used to define 32 conditions in ICD-9-CM and ICD-10-CA databases, respectively, using up to 16 diagnosis coding fields, we utilized the *SAS* functional command of "substr" to truncate the length of ICD-10 codes in the ICD-10-CA database. Therefore we defined the 32 conditions using the ICD-10 codes rather than ICD-10-CA codes and avoided influence of Canadian extended digits or additional codes on these conditions. This methodological approach is intentional, to increase the international relevance of our findings. We chose the Charlson index (Charlson et al. 1987)

and Elixhauser measures (Elixhauser et al. 1998) because they have been widely used by health researchers to measure burden of disease or case mix with administrative data (Southern, Quan, and Ghali 2004; Sundararajan et al. 2004; Needham et al. 2005).

Chart Review Data

Two reviewers who have nursing backgrounds and health records coding training, as well as extensive chart review experience, reviewed the randomly selected charts to determine the presence or absence of 32 conditions. The chart reviewers followed the definitions described by Charlson et al. (1987) to determine the presence or absence of the 14 conditions that constitute the Charlson index. To determine the presence or absence of the remaining 18 Elixhauser clinical conditions in the charts, we developed explicit definitions by describing all of the ICD-10 codes that were used to define the 18 conditions, with the clinical terms used in the ICD-10 manuals.

Two reviewers underwent training in data extraction with the lead investigator (H. Q.). In the training session, the definition of study variables was discussed and eight charts were reviewed. Any discrepancies between the two reviewers in reviewing these eight charts were discussed and resolved by consensus involving a third party. The agreement between the two reviewers was then evaluated. Both of the reviewers independently extracted clinical conditions from 70 charts using a predesigned standard form from one of the teaching hospitals. Of the 32 conditions extracted from these 70 charts, 17 conditions had near perfect agreement (κ : 0.81–1.0), 10 had substantial agreement (κ : 0.61–0.80), and four had moderate agreement (κ : 0.41–0.60) according to Landis and Koch (1977) criteria. κ could not be calculated for the remaining one condition (i.e., psychosis) due to its low frequency in the sample. After the agreement study, two reviewers started chart reviews. In the period of data collection, they discussed cases with uncertainty in determining conditions to ensure the consistency between them.

The two reviewers examined the entire chart, including the cover page, discharge summaries, narrative summaries, pathology reports (including autopsy reports), trauma and resuscitation records, admission notes, consultation reports, surgery/operative reports, anesthesia reports, physician daily progress notes (nursing notes excluded), physician orders, diagnostic reports, and transfer notes for evidence of any of the 32 conditions. This detailed chart review process took approximately 1 hour per chart.

Aside from the difference in the average length of time per chart between reviewers (1 hour) and coders (15–20 minutes), reviewers focused on determining presence or absence of medical conditions based on all documented information in the chart, including diagnostic imaging and laboratory results. This is in contrast to general coding guidelines (Canadian Institute of Health Information 2007) that instruct coders to confine their coding to clinical problems, conditions, or circumstances that are identified in the record by the treating physicians as the clinically significant reason for the patient's admission, or that require or influence evaluation, treatment, management, or care. Coders do not typically code problems that do not meet these requirements, whereas the reviewers who conducted our “reference standard” chart review included them regardless of the significance of the condition on resource use during hospitalization. Coders are instructed that when a condition is suggested by diagnostic test results, they should only code the condition if it has been confirmed by physician documentation.

Statistical Analysis

Three databases were thus created for the same hospital discharges: (1) ICD-10 discharge abstract data, (2) ICD-9-CM discharge abstract data, and (3) chart review data. The databases allowed us to calculate sensitivity, specificity, positive predictive value, and negative predictive value for each condition recorded in ICD-10 hospital discharge data and then in ICD-9-CM discharge data, accepting the chart review data as a “reference standard.” Recognizing that some might question the use of chart review data as a reference standard, the κ statistic was also used to assess the agreement between the two databases for individual conditions. For each condition identified in the chart data, McNemar's test was used to compare the sensitivity and specificity of ICD-10 versus ICD-9-CM data relative to the chart review data for detecting the conditions. To implement McNemar's statistical test for estimates of sensitivity and specificity, records with and then without a given condition present, respectively, based on chart data, were selected and agreement between ICD-9-CM and ICD-10 was tested in the subsample.

RESULTS

Table 1 presents the frequency of the 32 conditions by data source among 4,008 records. Compared with the chart review data, the ICD-9-CM data underreported 29 conditions, slightly overreported two conditions (diabetes

Table 1: Frequency of Clinical Condition by Data Source (%)

<i>Conditions</i>	<i>Chart Data</i>	<i>ICD-9-CM Data</i>	<i>ICD-10 Data</i>	<i>Difference Chart—ICD-9-CM</i>	<i>Difference Chart—ICD-10</i>	<i>p-Value ICD-9-CM versus ICD-10</i>
<i>In Charlson Index</i>						
Myocardial infarction	12.8	9.6	8.4	3.2	4.4	<.001
Cerebrovascular disease	8.1	4.6	4.5	3.5	3.6	.642
Rheumatic disease	2.6	1.5	1.4	1.1	1.2	.683
Dementia	3.3	1.1	2.4	2.2	0.9	<.001
<i>In Elixhauser Index</i>						
Cardiac arrhythmias	21.8	9.4	9.1	12.4	12.7	.241
Pulmonary circulation disorders	2.7	1.6	1.6	1.1	1.1	.578
Valvular disease	7.0	3.2	3.5	3.8	3.5	.134
Hypertension	30.2	25.2	22.2	5.0	8.0	<.001
Hypothyroidism	8.8	6.2	3.7	2.6	5.1	<.001
Lymphoma	1.0	0.9	0.8	0.1	0.2	.157
Solid tumor without metastasis	9.5	7.4	7.4	2.1	2.1	.736
Renal failure	4.0	4.6	4.9	-0.6	-0.9	.180
Blood loss anemia	1.1	0.7	0.6	0.4	0.5	.858
Deficiency anemia	1.9	1.9	1.4	0	0.5	.011
Coagulopathy	7.7	1.8	1.8	5.9	5.9	1.000
Fluid and electrolyte disorders	11.1	6.1	5.6	5.0	5.5	.089
Weight loss	3.7	0.5	0.9	3.2	2.8	.016
Obesity	8.3	2.7	1.9	5.6	6.4	<.001
Alcohol abuse	7.4	4.8	4.6	2.6	2.8	.477
Drug abuse	4.9	3.7	2.8	1.2	2.1	<.001
Psychoses	2.9	2.1	1.8	0.8	1.1	.048
Depression	11.9	7.3	5.8	4.6	6.1	<.001
<i>In Both Charlson and Elixhauser Indices</i>						
Congestive heart failure	8.3	6.6	6.3	1.7	2.0	.281
Peripheral vascular disease	4.3	2.9	2.8	1.4	1.5	.574
Hemiplegia or paraplegia	1.6	1.1	1.4	0.5	0.2	.028
Chronic pulmonary disease	15.0	9.0	8.7	6.0	6.3	.440
Diabetes with complication	2.7	2.8	2.6	-0.1	0.1	.292
Diabetes without complication	11.9	10.7	10.2	1.2	1.7	.114
Peptic ulcer disease	2.5	1.1	1.3	1.4	1.2	.088
Metastatic cancer	4.4	4.1	4.1	0.3	0.3	1.000
Liver disease	5.0	2.4	2.4	2.6	2.6	1.000
AIDS/HIV	0.6	0.2	0.3	0.4	0.3	.103

ICD-9-CM, ICD-9 Clinical Modification; ICD-10, International Classification of Disease, 10th Version.

with complications and renal failure), and equivalently reported one condition (deficiency anemia). The ICD-10 data underreported 31 conditions and slightly over-reported one condition (renal failure). ICD-10 data had a significantly lower frequency for eight conditions and higher frequency for three conditions compared with ICD-9-CM data.

Table 2 presents five quantitative indices to assess whether the administrative data accurately reproduced what was recorded in the patient charts by data source. Sensitivity was calculated to measure the extent of recording the presence of conditions in administrative data when these were present in the chart review data. Sensitivity for ICD-9-CM and ICD-10 data varied greatly by condition. Metastatic cancer had the highest sensitivity (83.1 percent in ICD-9-CM and 80.8 percent in ICD-10) and weight loss had the lowest sensitivity (9.3 percent in ICD-9-CM and 12.7 percent in ICD-10). Compared with ICD-10 data, ICD-9-CM data had significantly higher sensitivity for seven conditions and lower sensitivity for one condition. Sensitivity for the remaining 24 conditions was similar between ICD-9-CM and ICD-10 (see Table 2 and Figure 1). Positive predictive value, which determines the extent to which a condition present in the administrative data was also present in the chart review data, was higher than 75 percent for 20 conditions in ICD-9-CM and for 18 conditions in ICD-10 data. Specificity was used to determine the extent of reporting absence of these conditions in the administrative data when these diseases were absent in the charts. Negative predictive value was also used to determine the extent to which a condition absent in the administrative data was truly absent according to the chart review data. Specificity was higher than 98 percent for 29 conditions in ICD-9-CM (96.5 percent for solid tumor without metastasis, 97.7 percent for drug abuse, and 94.4 percent for depression) and for all 32 conditions in ICD-10. Negative predictive value was higher than 98 percent for 12 conditions in ICD-9-CM and 13 conditions in ICD-10. Cardiac arrhythmias had the lowest negative predictive value in both datasets (85.8 percent in ICD-9-CM and 85.3 percent in ICD-10).

The κ value indicates that a near perfect agreement (κ : 0.81–1.0 between coded data and chart review data) was found for two conditions in ICD-9-CM and one in ICD-10 data, substantial agreement (κ : 0.61–0.80) for 13 conditions in ICD-9-CM and 11 conditions in ICD-10, moderate agreement (κ : 0.41–0.60) for 10 conditions in ICD-9-CM and 15 conditions in ICD-10 and fair agreement (κ : 0.21–0.40) for six conditions in ICD-9-CM and five conditions in ICD-10. κ values relative to chart review data were generally similar for the ICD-9-CM and ICD-10 data for 29 conditions, but were discrepant for HIV/AIDS, hypothyroidism, and dementia (see Table 2 and Figure 2).

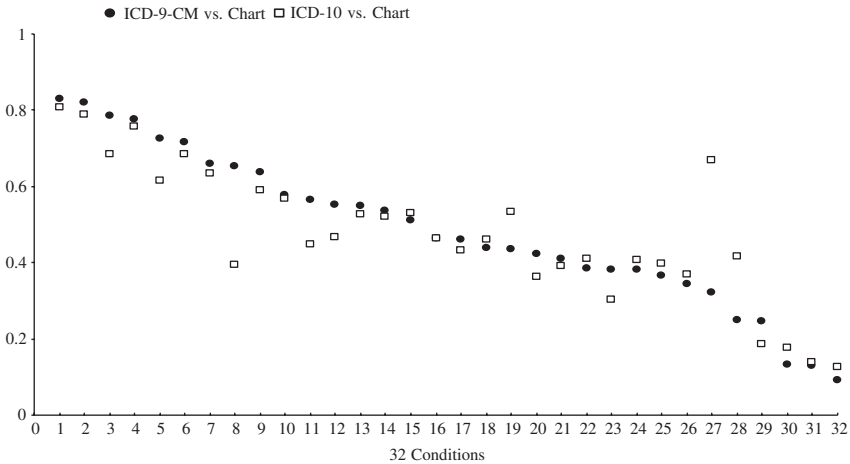
Table 2: Agreement between Chart and Administrative Data (%)

Conditions	ICD-9-CM Data				ICD-10 Data				p-Value ICD-9-CM versus ICD-10			
	Sensitivity	PPV	Specificity	NPV	κ	Sensitivity	PPV	Specificity	NPV	κ	Sensitivity	Specificity
<i>In Charlson Index</i>												
Myocardial infarction	72.4	95.9	99.5	96.1	0.80	61.5	93.5	99.4	94.6	0.71	<0.001	.221
Cerebrovascular disease	46.3	81.2	99.1	95.4	0.56	46.3	83.0	99.2	95.4	0.57	1.000	.433
Rheumatic disease	51.0	89.8	99.9	98.7	0.64	52.9	96.5	100	98.8	0.68	0.637	.103
Dementia	32.3	95.6	100	97.7	0.47	66.9	92.7	99.8	98.9	0.77	<0.001	.059
<i>In Elixhauser Index</i>												
Cardiac arrhythmias	41.1	95.2	99.4	85.8	0.51	39.0	93.4	99.2	85.3	0.48	0.056	.221
Pulmonary circulation disorders	34.3	59.7	99.4	98.2	0.42	37.0	61.5	99.4	98.3	0.45	0.439	1.000
Valvular disease	38.4	82.3	99.4	95.6	0.50	40.9	80.3	99.3	95.7	0.52	0.307	.225
Hypertension	78.6	94.5	98.0	91.4	0.81	68.3	93.1	97.8	87.7	0.72	<0.001	.414
Hypothyroidism	65.3	92.8	99.5	96.7	0.75	39.3	93.3	99.7	94.4	0.53	<0.001	.021
Lymphoma	65.9	73.0	99.8	99.7	0.69	63.4	78.8	99.8	99.6	0.70	0.564	.180
Solid tumor without metastasis	43.8	56.6	96.5	94.2	0.45	45.9	58.7	96.6	94.5	0.47	0.228	.398
Renal failure	81.9	71.2	98.6	99.2	0.75	78.8	64.3	98.2	99.1	0.69	0.411	.010
Blood loss anemia	13.3	23.1	99.5	99.0	0.16	17.8	32.0	99.6	99.1	0.22	0.414	.549
Deficiency anemia	38.2	39.2	98.9	98.8	0.38	30.3	40.4	99.1	98.7	0.34	0.083	.056
Coagulopathy	12.9	55.6	99.1	93.2	0.19	13.9	59.7	99.2	93.2	0.20	0.532	.549
Fluid and electrolyte disorders	42.4	76.7	98.4	93.2	0.51	36.3	71.6	98.2	92.6	0.44	0.005	.297
Weight loss	9.3	66.7	99.8	96.6	0.16	12.7	55.9	99.6	96.7	0.20	0.197	.033
Obesity	24.6	75.9	99.3	93.6	0.35	18.6	83.8	99.7	93.1	0.28	0.006	.003
Alcohol abuse	53.6	82.7	99.1	96.4	0.63	52.2	83.7	99.2	96.3	0.62	0.623	.590
Drug abuse	55.3	73.7	99.0	97.7	0.62	46.7	81.4	99.5	97.3	0.58	0.001	.002

Psychoses	57.8	79.8	99.6	98.8	0.66	56.9	90.4	99.8	98.7	0.69	0.763	.025
Depression	56.6	92.8	99.4	94.4	0.67	44.9	91.5	99.4	93.0	0.59	<0.001	.808
<i>In Both Charlson and Elixhauser Indices</i>												
Congestive heart failure	71.6	90.5	99.3	97.5	0.78	68.6	90.2	99.3	97.2	0.76	0.197	1.000
Peripheral vascular disease	46.2	67.0	99.0	97.6	0.53	43.3	65.5	99.0	97.5	0.50	0.423	1.000
Hemiplegia or paraplegia	43.6	62.8	99.6	99.1	0.51	53.2	58.9	99.4	99.3	0.55	0.109	.127
Chronic pulmonary disease	54.9	91.9	99.2	92.6	0.65	52.8	90.8	99.1	92.2	0.63	0.267	.590
Diabetes with chronic complication	63.6	62.5	98.9	99.0	0.62	59.1	63.1	99.0	98.9	0.60	0.384	.527
Diabetes without chronic complication	77.7	86.5	98.4	97.0	0.80	75.8	88.5	98.7	96.8	0.79	0.389	.124
Peptic ulcer disease	36.6	84.1	99.8	98.4	0.50	39.6	76.9	99.7	98.5	0.52	0.467	.025
Metastatic cancer	83.1	89.1	99.5	99.2	0.85	80.8	86.7	99.4	99.1	0.83	0.433	.655
Liver disease	38.1	80.2	99.5	96.8	0.50	40.6	85.4	99.6	96.9	0.54	0.384	.275
AIDS/HIV	25.0	100	100	99.6	0.40	41.7	100	100	99.7	0.59	0.103	1.000

Note: PPV, positive predictive value; NPV, negative predictive value; ICD-9-CM, ICD-9 Clinical Modification; ICD-10, International Classification of Disease, 10th Version.

Figure 1: Sensitivity Values for 32 Conditions Derived from ICD-9-CM and ICD-10 Hospital Discharge Abstract Data Relative to Chart Review Data as the “Reference Standard.”



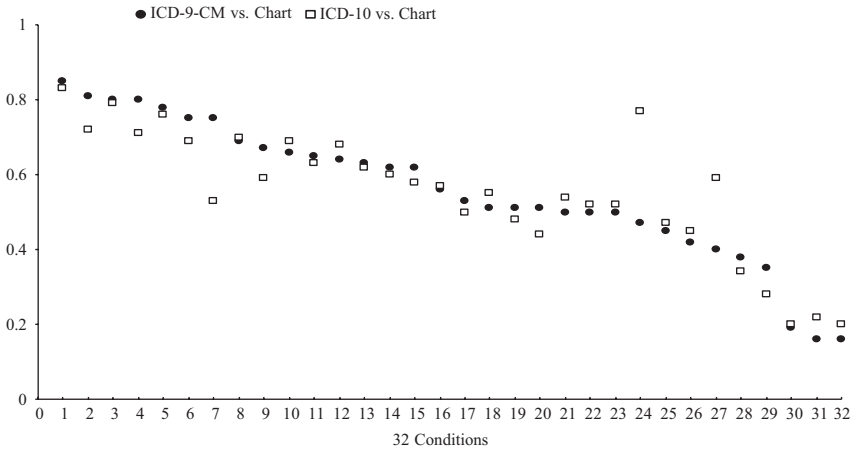
Notes: 1, Metastatic Cancer; 2, Renal Failure; 3, Hypertension; 4, Diabetes without Chronic Complication; 5, Myocardial Infarction; 6, Congestive Heart Failure; 7, Lymphoma; 8, Hypothyroidism; 9, Diabetes with Chronic Complication; 10, Psychoses; 11, Depression; 12, Drug Abuse; 13, Chronic Pulmonary Disease; 14, Alcohol Abuse; 15, Rheumatic Disease; 16, Cerebrovascular Disease; 17, Peripheral Vascular Disease; 18, Solid Tumor without Metastasis; 19, Hemiplegia or Paraplegia; 20, Fluid and Electrolyte Disorders; 21, Cardiac Arrhythmias; 22, Valvular Disease; 23, Deficiency Anemia; 24, Liver Disease; 25, Peptic Ulcer Disease; 26, Pulmonary Circulation Disorders; 27, Dementia; 28, AIDS/HIV; 29, Obesity; 30, Blood Loss Anemia; 31, Coagulopathy; 32, Weight Loss. ICD-9-CM, ICD-9 Clinical Modification; ICD-10, International Classification of Disease, 10th Version.

DISCUSSION

Our study documented the validity of ICD-9-CM and ICD-10 coding systems in coding clinical information. We found that ICD-10 administrative data were coded reasonably well on 32 conditions but that some conditions tended to be underdetected in ICD-10 data and had low validity relative to chart review data. The validity of ICD-10 data was generally comparable with that of ICD-9-CM data in recording clinical information, although ICD-9-CM coding demonstrated better sensitivity for a few conditions.

We anticipated that the new coding system had the potential to produce better validity relative to ICD-9-CM due to the new structure of codes in ICD-

Figure 2: κ Values for 32 Conditions Derived from ICD-9-CM and ICD-10 Hospital Discharge Abstract Data Relative to Chart Review Data as the “Reference Standard.”



Notes: 1, Metastatic Cancer; 2, Hypertension; 3, Diabetes without Chronic Complication; 4, Myocardial Infarction; 5, Congestive Heart Failure; 6, Renal Failure; 7, Hypothyroidism; 8, Lymphoma; 9, Depression; 10, Psychoses; 11, Chronic Pulmonary Disease; 12, Rheumatic Disease; 13, Alcohol Abuse; 14, Diabetes with Chronic Complication; 15, Drug Abuse; 16, Cerebrovascular Disease; 17, Peripheral Vascular Disease; 18, Hemiplegia or Paraplegia; 19, Cardiac Arrhythmias; 20, Fluid and Electrolyte Disorders; 21, Liver Disease; 22, Valvular Disease; 23, Peptic Ulcer Disease; 24, Dementia; 25, Solid Tumor without Metastasis; 26, Pulmonary Circulation Disorders; 27, AIDS/HIV; 28, Deficiency Anemia; 29, Obesity; 30, Coagulopathy; 31, Blood Loss Anemia; 32, Weight Loss. ICD-9-CM, ICD-9 Clinical Modification; ICD-10, International Classification of Disease, 10th Version.

10 that may enhance the accuracy and specificity of code identification. In this regard, ICD-10 partially reflects the advancement of medical knowledge of the past two decades. Yet, despite this potential for greater validity, our early validity assessment (performed 9 months after the implementation of ICD-10 coding) shows that sensitivity in ICD-10 was significantly lower than that in ICD-9-CM for myocardial infarction, hypertension, hypothyroidism, fluid and electrolyte disorders, obesity, drug abuse, and depression but higher in ICD-10 than in ICD-9-CM for dementia. The first possible explanation for the lower sensitivity in ICD-10 for several of the conditions is that coders were still in the early portion of an ICD-10 learning curve. The high sensitivity for dementia in ICD-10, meanwhile, may be related to the fact that ICD-10

groups dementias together as dementia in Alzheimer's disease (F00), vascular dementia (F01), dementia in other diseases classified elsewhere (F02), and unspecified dementia (F03). In contrast, ICD-9-CM does not group dementias together in the coding system as is done in ICD-10. The detailed grouping of "dementia" in ICD-10 may thus facilitate the work of coders in locating dementia codes, with the downstream result being an increase in the accuracy of coding. In contrast, there are no substantial enhancements in ICD-10 relative to ICD-9-CM in disease grouping and/or code descriptions for myocardial infarction and hypertension. For example, ICD-10 and ICD-9-CM were perfectly matched for hypertension codes I10.x/401.x-I15.x/405.x. The second possible explanation is that our coders who recoded charts in ICD-9-CM performed better than regular coders who coded ICD-10. About 16,000 charts were coded per year in Alberta. Coders rotate among hospital sites and are supervised under one manager within a health region. We recruited four coders who were working in the Health Records departments of the teaching hospitals studied and instructed them to code charts as they routinely do, following usual coding guidelines. Our coders coded 5.3 diagnoses per chart on average with median of four diagnoses in ICD-9-CM, which is very similar to the provincial average of 5.1 diagnoses per chart and median of four diagnoses in fiscal year 2001/2002 ICD-9-CM data. It therefore seems unlikely that the study coders performed better than regular coders. The third possible explanation is that our coders may have been randomly assigned to recode in ICD-9-CM some of the same charts that they had earlier coded in ICD-10 through their primary employment, thereby inflating the apparent similarity in performance between the two coding systems. While possible, we consider such a scenario to be infrequent, and also unlikely to have a major effect on the quality of our recoding. We randomly selected only 4,008 charts out of a total of about 70,000 (5.7 percent). Bearing in mind these numbers, it is quite unlikely for one of our coders to code the same randomly selected chart in the both ICD-9-CM and ICD-10. And even if this did occur on a few occasions, it would be quite difficult for a coder to remember much about the first time they coded a given chart. We therefore doubt that this scenario has occurred much and/or affected our results and conclusions significantly.

ICD-9-CM administrative data have been validated using various methodologies for various purposes. Hsia et al. (1992) assessed the accuracy of claims data by measuring incorrect grouping of clinically interrelated diagnostic codes with diagnosis-related groups (DRGs) and found that incorrect assignment of DRGs decreased significantly from 21 percent in 1985 to 15 percent in 1988. Many other investigators (Iezzoni et al. 1988; Jollis et al. 1993;

Romano and Mark 1994; Geraci et al. 1997; Muhajarine et al. 1997; Weingart et al. 2000; Best et al. 2002; Quan, Parson, and Ghali 2002; Romano et al. 2002; Lee et al. 2005; Yasmeen et al. 2006) conducted validation studies focusing on comorbidities, clinical conditions, and complications of substandard care, and found that administrative data are accurately coded for many severe or life-threatening conditions such as myocardial infarction and cancer, but that some clinically nonspecific and symptomatic conditions such as rheumatologic disease, are less accurately coded.

The introduction of the new coding system, ICD-10, raises new questions about the coding accuracy and completeness of clinical information recorded in administrative data and whether there have been changes in the magnitude of coders' errors between ICD-9-CM and ICD-10 coding systems. Anderson and Robenberg (2003) analyzed cause of death before and after implementation of ICD-10 in the United States. They found that the ranking of leading causes of death was substantially changed due to changes in classification system from ICD-9 to ICD-10. For example, chronic liver disease and cirrhosis, the 10th cause of death under ICD-9, was dropped out from the top 10 list under ICD-10, and Alzheimer's disease became one of the top 10 causes of death in ICD-10. Janssen and Kunst (2004) analyzed long-term cause-specific mortality in six European countries and noticed discontinuities in trends in cause-specific mortality due to changes in the coding system. Kokotailo and Hill (2005) reviewed charts from ICD-9-CM and ICD-10 admission records to determine whether the ICD-10 coding system had potential improvements over ICD-9-CM for stroke and stroke risk factors. They found that stroke and stroke risk factors were coded equally well with ICD-9-CM and ICD-10. Further, the factors of atrial fibrillation, coronary artery disease/ischemic heart disease, diabetes mellitus, and hypertension were recorded significantly better than the factors of history of cerebrovascular disease, hyperlipidemia, renal failure, and tobacco use in both ICD-9-CM and ICD-10 databases. Henderson, Shephard, and Sundararajan (2006) compared routinely coded ICD-10 data with audit data from public hospitals in Australia and demonstrated that the transition of the coding from ICD-9-CM to ICD-10 did not noticeably affect the quality of administrative data. Our study of dually coded data thus adds to this growing body of literature on ICD-10 validity, and like previous studies suggests that ICD-10 data have generally comparable validity, but that they do not (at least yet) have better validity than do ICD-9-CM data.

A number of conditions had poor validity in both ICD-9-CM and ICD-10 administrative data. The poor coding of certain conditions such as weight lost, obesity, and certain anemia may relate to the fact that coders do not code

these conditions even if they are documented in charts, because they may not be explicitly mentioned by nurses or physicians in clinical notes, and also because they may not affect length of stay, health care, or therapeutic treatment. Additionally, coders may intentionally not code these conditions due to the limited amount of time given to code each chart.

This study has limitations. A first limitation is that we reviewed charts only in teaching hospitals. We acknowledge that a study of nonteaching hospitals is also needed. Iezzoni et al. (1988, 1990) reported that the validity of administrative data varies between teaching and nonteaching hospitals. At nonteaching hospitals, acute clinical conditions tend to be more accurately documented but chronic coexisting diseases are less completely recorded than at teaching hospitals. A second limitation is that we employed chart data extracted by reviewers as a "reference standard" to assess the validity of ICD-9-CM and ICD-10 data. Such a criterion standard depends on the quality of charts and could only reflect part of the validity of administrative data. Ideally, a validity study should assess whether a condition that is truly present in a patient, and this depends on whether a condition is recorded correctly in the chart, and then subsequently coded precisely in the administrative data. Therefore, this study does not capture errors that could occur when clinicians take histories, make diagnoses, or record clinical information on charts (O'Malley et al. 2005). A third limitation is that the validity of administrative data may vary across hospitals, across regions, and across countries. Therefore, our findings may not be applicable to other regions.

Weighing against these limitations are some notable study strengths. Our study is perhaps the first to undertake a direct comparison of ICD-9-CM versus ICD-10 in dually coded administrative data. We studied a large number of hospital discharge records and thus achieved good precision of our validity measures for many of the conditions studied. We also used new ICD-9-CM and ICD-10 coding algorithms (Quan et al. 2005) to define conditions that are likely to optimize administrative data validity for capturing the clinical conditions.

In conclusion, our analysis of a unique dually coded database demonstrated that ICD-9-CM and ICD-10 administrative data were coded reasonably well and had similar validity in recording clinical condition information. The implementation of ICD-10 coding did not lead to an improvement in the coding of clinical conditions. However, we assessed hospital discharge data quality relatively early after implementation of ICD-10. The longer term impact of ICD-10 on data quality will need to be assessed in future studies.

ACKNOWLEDGMENTS

This study was supported by an operating grant from the Canadian Institutes of Health Research, Canada. Dr. Quan is supported by a Population Health Investigator Award from the Alberta Heritage Foundation for Medical Research, Edmonton, Alberta, Canada and by a New Investigator Award from the Canadian Institutes of Health Research. Dr. Ghali is supported by a Senior Health Scholar Award from the Alberta Heritage Foundation for Medical Research, Alberta, Canada, and by a Government of Canada Research Chair in Health Services Research. The authors thank 3M for providing 3M™ *Codefinder*™ ICD-9-CM code searching software.

IMECCHI (International Methodology Consortium for Coded Health Information) investigators include Bernard Burnand, University of Lausanne, Switzerland; Cyrille Colin, University of Lyon, France; Chantal Couris, University of Lyon, France; Carolyn De Coster, University of Manitoba, Canada; Saskia Drössler, Niederrhein University of Applied Sciences, Germany; Alan Finlayson, the National Health Service in Scotland, U.K.; Kiyohide Fushimi, Tokyo Medical and Dental University Graduate School, Japan; Min Gao, British Columbia Provincial Public Health Services Authority, Canada; William Ghali, University of Calgary, Canada; Patricia Halfon, University of Lausanne, Switzerland; Brenda Hemmelgarn, University of Calgary, Canada; Karin Humphries, University of British Columbia, Canada; Jean-Marie Januel, University of Lausanne, Switzerland; Helen Johansen, Statistics Canada; Lisa Lix, University of Manitoba, Canada; Jean-Christophe Luthi, University of Lausanne, Switzerland; Jin Ma, Jiaotong University, China; Hude Quan, University of Calgary, Canada; Patrick Romano, University of California at Davis, U.S.A.; Leslie Roos, University of Manitoba, Canada; Fiona Shrive, University of Calgary, Canada; Vijaya Sundararajan, Victorian Department of Human Services, Australia; Jack Tu, University of Toronto, Canada; Sandrine Touzet, University of Lyon, France; and Greg Webster, Canadian Institute of Health Information, Canada.

Disclosures: No any conflicts of interest.

Disclaimers: None.

REFERENCES

- Anderson, R.N., and H.M. Rosenberg. 2003. "Disease Classification: Measuring the Effect of the Tenth Revision of the International Classification of Diseases on Cause-of-Death Data in the United States." *Statistics in Medicine* 22: 1551–70.

- Best, W.R., S.F. Khuri, M. Phelan, K. Hur, W.G. Henderson, J.G. Demakis, and J. Daley. 2002. "Identifying Patient Preoperative Risk Factors and Postoperative Adverse Events in Administrative Databases: Results from the Department of Veterans Affairs National Surgical Quality Improvement Program." *Journal of American College of Surgeons* 194: 257–66.
- Canadian Institute for Health Information. 2003. *International Statistical Classification of Diseases and Related Health Problems Tenth Revision, Canada [ICD-10-CA]*. Ottawa, ON: Canadian Institute for Health Information.
- . 2007. *Canadian Coding Standards for ICD-10-CA and CCI for 2007*. Ottawa: Canadian Institute of Health Information.
- Charlson, M.E., P. Pompei, K.L. Ales, and C.R. MacKenzie. 1987. "A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation." *Journal of Chronic Diseases* 40: 373–83.
- Commission on Professional and Hospital Activities. 1986. *Annotated ICD-9-CM International Classification of Diseases, 9th revision, Clinical Modification*. Ann Arbor, MI: Edwards Brothers.
- Elixhauser, A., C. Steiner, D.R. Harris, and R.M. Coffey. 1998. "Comorbidity Measures for Use with Administrative Data." *Medical Care* 36: 8–27.
- Geraci, J.M., C.M. Ashton, D.H. Kuykendall, M.L. Johnson, and L. Wu. 1997. "International Classification of Diseases, 9th Revision, Clinical Modification Codes in Discharge Abstracts are Poor Measures of Complication Occurrence in Medical Inpatients." *Medical Care* 35: 589–602.
- Henderson, T., J. Shephard, and V. Sundararajan. 2006. "Quality of Diagnosis and Procedure Coding in ICD-10 Administrative Data." *Medical Care* 44: 1011–9.
- Hsia, D.C., C.A. Ahern, B.P. Ritchie, L.M. Moscoe, and W.M. Krushat. 1992. "Medicare Reimbursement Accuracy under the Prospective Payment System, 1985 to 1988." *Journal of the American Medical Association* 268: 896–9.
- Iezzoni, L.I., S. Burnside, L. Sickles, M.A. Moskowitz, E. Sawitz, and P.A. Levine. 1988. "Coding of Acute Myocardial Infarction. Clinical and Policy Implications." *Annals of Internal Medicine* 109: 745–51.
- Iezzoni, L.I., M. Shwartz, M.A. Moskowitz, A.S. Ash, E. Sawitz, and S. Burnside. 1990. "Illness Severity and Costs of Admissions at Teaching and Nonteaching Hospitals." *Journal of the American Medical Association* 264: 1426–31.
- Janssen, F., and A.E. Kunst. 2004. "ICD Coding Changes and Discontinuities in Trends in Cause-Specific Mortality in Six European Countries, 1950–99." *Bulletin of World Health Organization* 82: 904–13.
- Jollis, J.G., M. Ancukiewicz, E.R. DeLong, D.B. Pryor, L.H. Muhlbaier, and D.B. Mark. 1993. "Discordance of Databases Designed for Claims Payment versus Clinical Information Systems. Implications for Outcomes Research." *Annals of Internal Medicine* 119: 844–50.
- Kokotailo, R.A., and M.D. Hill. 2005. "Coding of Stroke Risk Factors Using ICD-10 and ICD-9." *Stroke* 36: 1776–81.
- Landis, J.R., and G.G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33: 159–74.

- Lee, D.S., L. Donovan, P.C. Austin, Y. Gong, P.P. Liu, J.L. Rouleau, and J.V. Tu. 2005. "Comparison of Coding of Heart Failure and Comorbidities in Administrative and Clinical Data for Use in Outcomes Research." *Medical Care* 43: 182-8.
- Muhajarine, N., C. Mustard, L.L. Roos, T.K. Young, and D.E. Gelskey. 1997. "Comparison of Survey and Physician Claims Data for Detecting Hypertension." *Journal of Clinical Epidemiology* 50: 711-8.
- Needham, D.M., DC. Scales, A. Laupacis, and P.J. Pronovost. 2005. "A Systematic Review of the Charlson Comorbidity Index Using Canadian Administrative Databases: A Perspective on Risk Adjustment in Critical Care Research." *Journal of Critical Care* 20: 12-9.
- O'Malley, K.J., K.F. Cook, M.D. Price, K.R. Wildes, J.F. Hurdle, and C.M. Ashton. 2005. "Measuring Diagnoses: ICD Code Accuracy." *Health Service Research* 40: 1620-39.
- Quan, H., G. Parson, and W. Ghali. 2002. "Validity of Information on Comorbidity Derived from ICD-9-CM Administrative Data." *Medical Care* 40: 675-85.
- Quan, H., V. Sundararajan, P. Halfon, A. Fong, B. Burnand, J.C. Luthi, L.D. Saunders, C.A. Beck, T.E. Feasby, and W.A. Ghali. 2005. "Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data." *Medical Care* 43: 1130-9.
- Romano, P.S., B.K. Chan, M.E. Schembri, and J.A. Rainwater. 2002. "Can Administrative Data Be Used to Compare Postoperative Complication Rates across Hospitals?" *Medical Care* 40: 856-67.
- Romano, P.S., and D.H. Mark. 1994. "Bias in the Coding of Hospital Discharge Data and Its Implications for Quality Assessment." *Medical Care* 32: 81-90.
- Southern, D.A., H. Quan, and W.A. Ghali. 2004. "Comparison of the Elixhauser and Charlson/Deyo Methods of Comorbidity Measurement in Administrative Data." *Medical Care* 42: 355-60.
- Sundararajan, V., T. Henderson, C. Perry, A. Muggivan, H. Quan, and W. Ghali. 2004. "New ICD-10 Version of the Charlson Comorbidity Index Predicts in Hospital Mortality." *Journal of Clinical Epidemiology* 57: 1288-94.
- Weingart, S.N., L.I. Iezzoni, R.B. Davis, R.H. Palmer, M. Cahalane, M.B. Hamel, K. Mukamal, R.S. Phillips, D.T. Davies Jr., and N.J. Banks. 2000. "Use of Administrative Data to Find Substandard Care: Validation of the Complications Screening Program." *Medical Care* 38: 796-806.
- World Health Organization. 1992. *International Statistical Classification of Disease and Related Health Problems, Tenth Revision (ICD-10)*. Geneva: World Health Organization.
- Yasmeen, S., P.S. Romano, M.E. Schembri, J.M. Keyzer, and W.M. Gilbert. 2006. "Accuracy of Obstetric Diagnoses and Procedures in Hospital Discharge Data." *American Journal of Obstetrics and Gynecology* 194: 992-1001.