Accuracy of ICD-10 Coding System for Identifying Comorbidities and Infectious Conditions Using Data from a Thai University Hospital Administrative Database

Pinyo Rattanaumpawan MD, MSCE, PhD*,**,
Thanyarak Wongkamhla MD**, Visanu Thamlikitkul MD*,**

** Division of Infectious Diseases and Tropical Medicine, Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand*
*** Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand*

Objective: *To determine the accuracy of International Statistical Classification of Disease and Related Health Problems, 10th Revision (ICD-10) coding system in identifying comorbidities and infectious conditions using data from a Thai university hospital administrative database.*

Material and Method: *A retrospective cross-sectional study was conducted among patients hospitalized in six general medicine wards at Siriraj Hospital. ICD-10 code data was identified and retrieved directly from the hospital administrative database. Patient comorbidities were captured using the ICD-10 coding algorithm for the Charlson comorbidity index. Infectious conditions were captured using the groups of ICD-10 diagnostic codes that were carefully prepared by two independent infectious disease specialists. Accuracy of ICD-10 codes combined with microbiological data for diagnosis of urinary tract infection (UTI) and bloodstream infection (BSI) was evaluated. Clinical data gathered from chart review was considered the gold standard in this study.*

Results: *Between February 1 and May 31, 2013, a chart review of 546 hospitalization records was conducted. The mean age of hospitalized patients was 62.8±17.8 years and 65.9% of patients were female. Median length of stay [range] was 10.0 [1.0-353.0] days and hospital mortality was 21.8%. Conditions with ICD-10 codes that had good sensitivity (90% or higher) were diabetes mellitus and HIV infection. Conditions with ICD-10 codes that had good specificity (90% or higher) were cerebrovascular disease, chronic lung disease, diabetes mellitus, cancer, HIV infection, and all infectious conditions. By combining ICD-10 codes with microbiological results, sensitivity increased from 49.5 to 66% for UTI and from 78.3 to 92.8% for BSI.*

Conclusion: *The ICD-10 coding algorithm is reliable only in some selected conditions, including underlying diabetes mellitus and HIV infection. Combining microbiological results with ICD-10 codes increased sensitivity of ICD-10 codes for identifying BSI. Future research is needed to improve the accuracy of hospital administrative coding system in Thailand.*

Keywords: *Comorbidities, ICD-10, Infectious conditions*

J Med Assoc Thai 2016; 99 (4): 368-73
Full text. e-Journal: *http://www.jmatonline.com*

The International Statistical Classification of Disease and Related Health Problems, tenth Revision (ICD-10) coding system has been widely used for recording morbidity and clinical data in electronic medical databases. While the ICD-10 coding system has been widely used by the Thailand healthcare system for reimbursement and resource allocation purposes, it is not commonly used for research.

A number of international studies have confirmed that ICD coding is a reliable and acceptable tool for identifying comorbidities and infectious conditions[1]. Based on results from a Canadian study, ICD-10 coding provided a positive predictive value (PPV) of 78.8% for identifying patients with ischemic heart disease when the system was used to evaluate patient data on the Electronic Medical Record Administrative data Linked Database (EMRALD)[2]. A study from the United Kingdom revealed that use of ICD-10 coding in the primary care electronic medical record for identifying patients with pneumonia provided positive predictive value of 86%, sensitivity of 87.5%, and specificity of 99.4%[3]. Moreover, a study conducted at a rural hospital in Phitsanulok, Thailand found that ICD-10 coding used for identifying patients with atrial fibrillation provided positive predictive value of 87.6%[4].

Correspondence to:
Rattanaumpawan P, Division of Infectious Diseases and Tropical Medicine, Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, 2 Wanglang Road, Bangkoknoi, Bangkok 10700, Thailand.
Phone: +66-2-419774, Fax: +66-2-4197783
E-mail: pinyo.rat@mahidol.ac.th

Until now, the accuracy of ICD-10 coding for identifying comorbidities and infectious conditions for research and surveillance purposes in Thailand remains uncertain. Accordingly, this study aimed to evaluate the accuracy of ICD-10 coding for identifying comorbidities and infectious conditions among hospitalized patients at Siriraj Hospital in Thailand.

Material and Method

Study design and population

A retrospective cross-sectional study was conducted at Siriraj Hospital between February 1 and May 31, 2013. Siriraj Hospital is a 2,200-bed university hospital and the largest referral center in Thailand. Inclusion criteria were all hospitalized patients aged 15 years or more who were admitted at any of six general medicine wards during the study period. A given patient could be enrolled in the study more than once if she/he had multiple hospitalizations during the study period. The study protocol was approved by the Siriraj Institutional Review Board (SIRB) with waiver of informed consent.

Data collection

ICD-10 code data was identified and retrieved directly from the Siriraj Hospital database. We obtained the ICD-10 codes for comorbidities and infectious conditions for all diagnosis, principle diagnosis, comorbidities, complication, other diagnosis, and procedure fields. Baseline characteristics, clinical data, and microbiological data were obtained by performing chart review. All necessary data is consisted of demographics, previous hospitalization, type of referral, underlying diseases, and diagnosis upon admission and during hospitalization. Among patients with infection, we also retrieved data regarding details of infection, microbiological data, antibiotic treatment, and treatment outcomes.

Identification of ICD-10 codes for comorbidities and infectious conditions

Our gold standard in this study was diagnosis based on data from chart review. Patient comorbidities were captured using the ICD-10 coding algorithm for the Charlson comorbidity index, which was described in a previously published report[5]. Infectious conditions were captured using the groups of ICD-10 diagnostic codes that were carefully prepared by two independent infectious disease specialists. Detail regarding ICD-10 codes and the study gold standard are shown in the supplementary section. Table 1 shows the groups of ICD-10 codes used for identifying Charlson comorbidities, while Table 2 shows the groups of ICD-10 codes used for identifying infectious conditions.

Microbiological results were also combined with the ICD-10 code for diagnosis of urinary tract infection (UTI) and bloodstream infection (BSI). UTI was documented if urine culture grew any uropathogen(s) of at least 10^5 CFU/ml. BSI was documented if at least one blood culture grew any causative pathogen(s).

Statistical analysis and sample size calculation

Based on findings from our pilot study, it was estimated that the sensitivity and specificity of ICD-10 coding for identifying comorbidities and organ-specific infections would be higher than 80% and 90%, respectively. Using a sample size calculation with an acceptable error of 10%, a sample size of 311 subjects was needed.

Categorical variables were reported as frequency and percentage. Continuous variables were reported as mean ± standard deviation or median and range, according to the distribution. Accuracy of ICD-10 coding for identification of comorbidities and infectious conditions was reported as positive predictive value, sensitivity, and specificity and 95% confidence interval [95% CI]. All analyses were performed using the Stata/IC version 14.0 (StataCorp LP, College Station, TX, USA).

Results

During the 4-month study period (February 1 to May 31, 2013), a chart review of 546 medical records (441 unique patients) was successfully performed. Of these 546 hospitalizations, the mean age of patients was 62.8±17.8 years and 65.9% were female. Median length of stay [range] was 10.0 [1.0-353.0] days and hospital mortality was 21.8%. Baseline characteristics of 546 hospitalizations are described in Table 3.

Sensitivity, specificity, and positive predictive value of the ICD-10 coding algorithm for identifying comorbidities and infectious conditions are shown in Table 4. The sensitivity of some selected conditions, including diabetes mellitus and HIV infection are higher than 90%. The specificity of some selected conditions, including cerebrovascular disease, chronic lung disease, diabetes mellitus, cancer, HIV infection, and all other infections are higher than 90%. Only ICD-10 codes for underlying diabetes mellitus and HIV infection provided both sensitivity and specificity of

90% or more. None of the ICD-10 codes for infectious conditions provided both sensitivity and specificity of 90% or more. When we focused on high positive predictive value, only underlying diabetes mellitus, HIV infection, and overall infections provided a positive predictive value of 90% or more.

To increase sensitivity for diagnosis of infectious conditions, we combined ICD-10 codes and microbiological results for purposes of capturing UTI and BSI. Sensitivity, specificity, and positive predictive value of the ICD-10 codes and/or microbiological results for identifying UTI and BSI are shown in

Table 1. ICD-10 codes for Charlson comorbidity index

Comorbidities	ICD-10 codes
Myocardial infarction	I21.x, I22.x, I25.2
Congestive heart failure	I09.9, I11.0, I13.0, I13.2, I25.5, I42.0, I42.5-I42.9, I43.x, I50.x, P29.0
Peripheral vascular disease	I70.x, I71.x, I73.1, I73.8, I73.9, I77.1, I79.0, I79.2, K55.1, K55.8, K55.9, Z95.8, Z95.9
Cerebrovascular disease	G45.x, G46.x, H34.0, I60.x-I69.x
Dementia	F00.x-F03.x, F05.1, G30.x, G31.1
Chronic pulmonary disease	I27.8, I27.9, J40.x–J47.x, J60.x-J67.x, J68.4, J70.1, J70.3
Rheumatic disease	M05.x, M06.x, M31.5, M32.x-M34.x, M35.1, M35.3, M36.0
Peptic ulcer disease	K25.x-K28.x
Mild liver disease	B18.x, K70.0-K70.3, K70.9, K71.3-K71.5, K71.7, K73.x, K74.x, K76.0, K76.2-K76.4, K76.8, K76.9, Z94.4
Diabetes without chronic complication	E10.0, E10.1, E10.6, E10.8, E10.9, E11.0, E11.1, E11.6, E11.8, E11.9, E12.0, E12.1, E12.6, E12.8, E12.9, E13.0, E13.1, E13.6, E13.8, E13.9, E14.0, E14.1, E14.6, E14.8, E14.9
Diabetes with chronic complication	E10.2-E10.5, E10.7, E11.2-E11.5, E11.7, E12.2-E12.5, E12.7, E13.2-E13.5, E13.7, E14.2-E14.5, E14.7
Hemiplegia or paraplegia	G04.1, G11.4, G80.1, G80.2, G81.x, G82.x, G83.0-G83.4, G83.9
Renal disease	I12.0, I13.1, N03.2-N03.7, N05.2-N05.7, N18.x, N19.x, N25.0, Z49.0-Z49.2, Z94.0, Z99.2
Any malignancy, including lymphoma, and leukemia (except malignant neoplasm of skin)	C00.x-C26.x, C30.x-C34.x, C37.x-C41.x, C43.x, C45.x-C58.x, C60.x-C76.x, C81.x-C85.x, C88.x, C90.x-C97.x
Moderate or severe liver disease	I85.0, I85.9, I86.4, I98.2, K70.4, K71.1, K72.1, K72.9, K76.5, K76.6, K76.7
Metastatic solid tumor	C77.x-C80.x
AIDS/HIV	B20.x-B22.x, B24.x

Table 2. ICD-10 codes for infectious conditions

Infectious conditions	ICD-10
All infections	A00-A99, B00-B99
CNS infection	G0, I22
CVS infection	I00-I02, I30-I33, I38-I41
Lower respiratory tract infection	A15-A16, A19, A150-153, A157-162, A167-169, A202, A221, A221, A310, A420, A430, A481 B012, B052, B206, B250, B334, B371, B380-382, B390-392, B400-402, B410, B420, B440-441, B450, B460, B583, B59, B671 J100, J110, J160, J12-18, J120-123, J128-129, J150-159, J168, J180-182, J188-189 P23, P230-239
Urinary tract	A181, A590 B374 N080, N160, N290-291, N298, N30, N300-303, N308-309, N330, N34, N340-342, N390 O23, O230-O234, O239, O862-863 P393
Skin infections	L00-L08
Gastrointestinal tract infection	A00-A09
Blood stream infection	A40-A41, A021, A207, A227, A267, A327, A392-394, A400-403, A408-415, A418-419, A427, A440, A483, B50-B54, B500, B508-509, B518, B520, B528-531, B538, O85, P36, P360-365, P368-369

Table 5. By combining ICD-10 codes and microbiological results, sensitivity increased from 49.5 to 66% for UTI and from 78.3 to 92.8% for BSI. Unfortunately, the positive predictive value decreased from 77.3 to 63.5% for UTI and from 81.7 to 62.5% for BSI.

Table 3. Baseline characteristics of 546 hospitalizations from 441 unique patients

Characteristics	Number (n = 546)	%
Mean age (year), (± SD)	62.8±17.8	
Median length of stay (day), [range]	10.0 [1.0-353.0]	
Male gender	186	34.1
Transfer status		
Home	499	91.4
Another hospital	40	7.3
Long-term care facility	6	1.1
Unknown	1	0.2
Previous hospitalization in the past 3 months	205	37.6
Clinically documented infection	416	76.2
Type of infection*		
Community-acquired	258	62.0
Hospital-acquired	188	45.2
Discharge status		
Alive	413	75.6
Dead	119	21.8
Transfer	12	2.2
Against advice	2	0.4

* A given patient may have more than one infection episode

Discussion

Our study revealed that sensitivity of ICD-10 codes for identifying comorbidities and infectious conditions was quite low. Only ICD-10 codes for diabetes mellitus and HIV-infection provided both acceptable sensitivity and specificity (90.0% or higher). None of the ICD-10 codes for any infectious conditions provided acceptable level of both sensitivity and specificity. By combining ICD-10 codes with microbiological results, sensitivity of ICD-10 codes for diagnosis of UTI and BSI were moderately improved. Given our intended use of the ICD-10 coding algorithm as a screening tool for research and surveillance purposes, high sensitivity is of greater importance than high specificity.

The relative low sensitivity of the ICD-10 coding algorithm in our study may be explained by several reasons. Given that ICD-10 codes are primarily used for reimbursement and resource allocation in Thailand, it is plausible that the ICD coder may tend to document only diagnoses that result in reimbursement. Put another way, diagnoses lacking monetary benefit may be overlooked and not recorded in the system. Second, the ICD system was developed by the World Health Organization (WHO) health system and was not specifically designed to account for diseases and conditions in Thailand.

Regarding potential strengths of this study, this was the first study conducted in Thailand to

Table 4. Sensitivity, specificity, and positive predictive value of the Charlson algorithm in identifying comorbidities relative to ICD-10 codes for infectious conditions

Conditions	% sensitivity [95% CI]	% specificity [95% CI]	% positive predictive value [95% CI]
Comorbid conditions			
Cerebrovascular disease	65.0 [55.8-73.5]	95.3 [92.8-97.1]	79.6 [70.3-87.1]
Chronic lung disease	58.4 [46.7-69.6]	98.5 [96.9-99.4]	86.5 [74.2-94.4]
Chronic renal disease	15.0 [8.0-24.7]	79.2 [75.2-82.8]	11.0 [5.8-18.4]
Diabetes mellitus	91.0 [85.9-94.8]	95.9 [93.3-97.7]	91.5 [86.5-95.2]
Cancer	83.5 [75.4-89.7]	96.9 [94.9-98.3]	88.0 [80.5-93.5]
Chronic liver disease	17.3 [10.6-26.0]	83.9 [80.2-87.2]	20.2 [12.4-30.1]
HIV infection	95.0 [75.1-99.9]	99.8 [98.9-99.9]	95.0 [75.1-99.9]
Infectious conditions			
Any infection	63.9 [59.1-68.6]	84.6 [77.2-90.3]	93.0 [89.4-95.7]
CNS infection	75.0 [47.6-92.7]	99.4 [98.3-99.9]	80.0 [51.9-95.7]
CVS infection	75.0 [19.4-99.4]	98.3 [96.9-99.2]	25.0 [5.4-57.2]
Lower respiratory tract infection	57.2 [50.2-64.0]	93.5 [90.3-95.9]	84.4 [77.3-90.0]
Urinary tract infection	49.5 [39.5-59.5]	96.6 [94.5-98.1]	77.3 [65.3-86.7]
Skin infection	28.0 [16.2-42.4]	98.8 [97.4-99.5]	70.0 [45.7-88.1]
Gastrointestinal infection	16.4 [8.8-26.9]	96.6 [94.6-98.1]	42.8 [24.5-62.8]
Bloodstream infection	78.3 [68.8-86.1]	96.2 [94.0-97.8]	81.7 [72.4-89.0]

ICD-10 = International Statistical Classification of Disease and Related Health Problems, 10th Revision; CNS = central nervous system; CVS = cardiovascular system

Table 5. Sensitivity, specificity, and positive predictive value of ICD-10 codes and/or microbiological results in identifying urinary tract infection and bloodstream infection

Infectious condition	% sensitivity [95% CI]	% specificity [95% CI]	% positive predictive value [95% CI]
Urinary tract infection			
Diagnosis by ICD-10 codes only	49.5 [39.5-59.5]	96.6 [94.5-98.1]	77.3 [65.3-86.7]
Diagnosis by microbiological results only	62.1 [52.0-71.5]	98.0 [96.2-99.1]	88.9 [79.2-95.1]
Diagnosis by ICD-10 codes and microbiological results	66.0 [56.0-75.1]	91.2 [88.2-93.7]	63.5 [53.7-72.6]
Bloodstream infection			
Diagnosis by ICD-10 codes only	78.3 [68.8-86.1]	96.2 [94.0-97.8]	81.7 [72.4-89.0]
Diagnosis by microbiological results only	86.6 [78.2-92.7]	97.6 [95.6-98.8]	89.4 [81.3-94.8]
Diagnosis by ICD-10 codes and microbiological results	92.8 [85.7-97.0]	79.2 [75.3-82.6]	62.5 [54.1-70.4]

evaluate accuracy of ICD-10 coding for identifying comorbidities and infectious conditions. Second, the groups of diagnosis codes for diagnosis of infectious conditions were carefully prepared by two infectious disease specialists, a factor that would increase the accuracy of our coding algorithm. The ICD-coding algorithms for diagnosis of diabetes mellitus and HIV infection, as well as the combination of ICD-coding and microbiological results for diagnosis of BSI, are promising for research purposes.

Our study also had some potential limitations. First, only one investigator reviewed a given patient's medical record. Therefore, the reliability of our gold standard for this study relied solely on the accuracy and thoroughness of each individual investigator. Given that all investigators who performed chart review were physicians, the probability of over diagnosis would have been minimal. Second, given the retrospective design of this study, missing or incomplete data would have been encountered in some cases. Third, our study was conducted in general medicine wards at a university tertiary care hospital; as such, results of our study may not be applicable to other settings.

Conclusion

Accuracy of ICD-10 coding for identifying comorbidities and infectious conditions was quite low in our study. The ICD-10 coding algorithm can be used only in selected conditions, including underlying diabetes mellitus and HIV infection. Combining the ICD-10 coding algorithm and microbiological results for diagnosis of BSI showed promising sensitivity for research purposes. Further study is needed to identify strategies for improving the accuracy and overall efficacy of the hospital administrative database.

What is already known on this topic?

ICD-10 coding system is a tool for recording morbidity and clinical data in medical databases.

Although the ICD-10 coding system has been widely used for research purpose worldwide, it has been mainly used for reimbursement in Thailand.

What this study adds?

The ICD-10 coding algorithm is reliable only in some selected conditions, including underlying diabetes mellitus (91.0% sensitivity and 95.9% specificity) and HIV infection (95.0% sensitivity and 99.8% specificity).

Acknowledgement

The authors wish to thank all ward nurses of Department of Medicine, Siriraj Hospital and Miss Yadawadee Wongthanasuporn for their cooperation and support of this study.

Financial support

This study was primarily supported by grants from the Health Systems Research and Development Project (Faculty of Medicine Siriraj Hospital) and the Health Systems Research Institute (Thailand).

Potential conflicts of interest

None.

References

1. Jhung MA, Banerjee SN. Administrative coding data and health care-associated infections. Clin Infect Dis 2009; 49: 949-55.
2. Tu K, Mitiku T, Lee DS, Guo H, Tu JV. Validation of physician billing and hospitalization data to identify patients with ischemic heart disease using data from the Electronic Medical Record Administrative data Linked Database (EMRALD).

Can J Cardiol 2010; 26: e225-8.
3. Meropol SB, Metlay JP. Accuracy of pneumonia hospital admissions in a primary care electronic medical record database. Pharmacoepidemiol Drug Saf 2012; 21: 659-65.
4. Chotchaisuwatana S, Jedsadayanmata A, Chaiyakunapruk N, Jampachaisri K. Validation of electronic medical database in patients with atrial fibrillation in community hospitals. J Med Assoc Thai 2011; 94: 686-92.
5. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care 2005; 43: 1130-9.

ความแม่นยำในการใช้ ICD-10 ในการสืบค้นหาโรคร่วมและภาวะโรคติดเชื้อ จากฐานข้อมูลของโรงพยาบาลศิริราช

ภิญโญ รัตนาอัมพวัลย์, ธัญญารักษ์ วงค์คำหล้า, วิษณุ ธรรมลิขิตกุล

วัตถุประสงค์: *เพื่อประเมินความแม่นยำของการใช้ICD-10 เพื่อสืบค้นหาโรคร่วม และภาวะโรคติดเชื้อจากฐานข้อมูลของโรงพยาบาลระดับมหาวิทยาลัย*

วัสดุและวิธีการ: *ทำการศึกษาแบบ cross-sectional study ในผู้ป่วยในหอผู้ป่วยอายุรกรรมสามัญทั้ง 6 แห่ง ในโรงพยาบาลศิริราช โดยรวบรวมข้อมูลรหัสโรค ICD-10 จากฐานข้อมูลอิเล็กทรอนิกส์ของโรงพยาบาล ทำการทบทวนเวชระเบียนผู้ป่วยเพื่อนำมาใช้เป็นเกณฑ์มาตรฐาน (gold standard) ในการวินิจฉัย*

ผลการศึกษา: *ตั้งแต่วันที่ 1 กุมภาพันธ์ ถึง 31 พฤษภาคม พ.ศ. 2556 ได้ทำการทบทวนเวชระเบียนของผู้ป่วยจำนวน 546 เวชระเบียน พบว่าเป็นผู้ป่วยหญิงร้อยละ 65.9 อายุเฉลี่ย 62.8±17.8 ปี มัธยฐานของระยะเวลาการนอนโรงพยาบาล 10.0 [1.0-353.0] วัน และมีอัตราตายร้อยละ 21.8 พบว่าความไว (sensitivity) ของรหัส ICD-10 ในการวินิจฉัยโรคเบาหวาน และการติดเชื้อ HIV นั้นอยู่ในระดับที่มากกว่าร้อยละ 90 ในขณะที่ความจำเพาะ (specificity) ของรหัส ICD-10 ในการวินิจฉัยโรคหลอดเลือดสมอง โรคปอดเรื้อรัง โรคเบาหวาน โรคมะเร็ง โรคติดเชื้อ HIV และโรคติดเชื้ออื่น ๆ นั้นสูงกว่าร้อยละ 90 อย่างไรก็ตาม มีเฉพาะกลุ่มรหัส ICD-10 ของโรคเบาหวาน และโรคติดเชื้อ HIV เท่านั้นที่มีทั้งความไวและความจำเพาะมากกว่าร้อยละ 90 เมื่อนำข้อมูลทางด้านจุลชีววิทยามาประกอบการสืบค้นหาผู้ป่วยที่มีภาวะการติดเชื้อในระบบทางเดินปัสสาวะ และการติดเชื้อในกระแสเลือด พบว่าความไวเพิ่มขึ้นเป็นร้อยละ 66.0 และ 92.8 ตามลำดับ*

สรุป: *จากการศึกษานี้พบว่าความแม่นยำในการใช้กลุ่มรหัสโรค ICD-10 ในการสืบค้นหาโรคร่วม และภาวะโรคติดเชื้อนั้นมีความแม่นยำค่อนข้างต่ำ ยกเว้นโรคเบาหวาน และการติดเชื้อ HIV อย่างไรก็เมื่อนำข้อมูลทางด้านจุลชีววิทยามาประกอบการสืบค้นหาภาวะการติดเชื้อในกระแสเลือดก็พบว่ามีความไวสูงขึ้น ในอนาคตควรมีการศึกษาถึงวิธีการที่จะช่วยเพิ่มความแม่นยำของฐานข้อมูลของโรงพยาบาลในประเทศไทย*
