RESEARCH ARTICLE

Disagreement of ICD-10 Codes Between a Local Hospital Information System and a Cancer Registry

Hutcha Sriplung[1], Tirada Kantipundee[1], Cheamjit Tassanapitak[2]

Abstract

Background: In the field of cancer, the ICD-10 coding convention is based on the site of a neoplasm in the body and usually ignores the morphology, thus the same code may be assigned to tumors of different morphologic types in an organ. Nowadays, all general (provincial) and center hospitals in Thailand are equipped with the hospital information system (HIS) database. Objective: This study aimed to find the characteristics and magnitude of agreement represented by the positive predictive value (PPV) of provisional cancer diagnoses in the HIS database in Pattani Hospital in Thailand in comparison with the final cancer diagnosis of the ICD-10 codes generated from a well established cancer registry in Songklanagarind Hospital, the medical school hospital of Prince of Songkla University. Materials and Methods: Data on cancer patients residing in Pattani province who visited Pattani Hospital from January 2007 to May 2011 were obtained from the HIS database. The ICD-10 codes of the HIS computer database of Pattani Hospital were compared against the ICD-10 codes of the same person recorded in the hospital-based cancer registry of Songklanagarind Hospital. The degree of agreement or positive predictive value (PPV) was calculated for each sex and for both sexes combined. Results: A total of 313 cases (15.9%) could be matched in the two databases. Some 222 cases, 109 males and 113 females, fulfilled the criteria of referral from Pattani to Songklanagarind Hospitals. Of 109 male cancer cases, 76 had the same ICD-10 codes in both hospitals, thus, the PPV was 69.7% (95%CI: 60.2-78.2%). Agreement in 76 out of 113 females gave a PPV of 67.3% (95%CI: 57.8-75.8%). The two percentages were found non-significant with Fisher's exact p-value of 0.773. The PPV for combined cases of both sexes was 68.5% (95%CI: 61.9-74.5%). Conclusions: Changes in final diagnosis in the referral system are common, thus the summary statistics of a hospital without full investigation facilities must be used with care, as the statistics are biased towards simple diseases able to be investigated by available facilities. A systematic feedback of patient information from a tertiary to a referring hospital should be considered to increase the accuracy of statistics and to improve the comprehensive care of cancer patients.

Keywords: ICD-10 codes - cancer registry - local hospital information system - discrepancy

Asian Pac J Cancer Prev, **16 (1)**, 259-263

Introduction

The purposes and uses of the ICD-10 and ICD-O coding systems are different. Sponsored by the World Health Organization, the ICD-10 (WHO, 1992) aims to promote international comparability in the collection, classification, processing, and presentation of mortality statistics. The ICD-10 coding convention for neoplasms usually ignores the morphology (pathological nomenclature) of the tumor, as the primary aim of this code is to differentiate the behavioral nature of tumors into malignant, metastatic, in situ, benign, uncertain and unknown cancers by assigning different codes for these characteristics. For tumors, 'C' code is assigned for malignant tumors and 'D' code is for benign, borderline, in situ and unknown behavior. Although, the 'D' code should be used for a tumor of unknown behavioral nature, however, when a patient is diagnosed with a mass, a 'C' code is usually used, implying a malignant tumor in an organ without consideration of whether the mass still has a chance to be non-malignant when investigation is completed. With the ICD-10, the same code is assigned to tumors of different morphologic types as long as they appear in the same organ. For this reason, it is not possible, by simply knowing the code, to distinguish between an adenocarcinoma and a squamous cell carcinoma of the lung since both are coded to C34.9. Further complicating the situation is hematologic malignancies, in which the histology is coded rather than the site where the tumor is found. For example, C83.1 is for mantle cell lymphoma, regardless of its place of origin. Simply referring to such codes, it is not possible to differentiate extra-nodal lymphoma from the same type of lymphoma occurring in a lymph node.

To overcome this ambiguity of the ICD-10 among

*[1]Epidemiology Unit, [2]Nursing Division, Faculty of Medicine, Prince of Songkla University, Hat Yai, Songkhla, Thailand *For correspondence: hutcha.s@psu.ac.th*

medical and health professions and also researchers, the ICD-O was devised, with the word 'O' indicating oncology. The current version is the ICD-O3 (Fritz et al., 2000). To comprehensively describe a tumor, two main code sections, topography and morphology, are used. In ICD-O system, topography is coded according to the primary site of cancer while the morphology portion of the ICD-O code consists of histology, behavior, and grade. When a tumor is coded with both topography and morphology, the code can be translated back to the pathological diagnosis of primary cancer occurring in the body. Though a metastatic cancer has the behavior code of 6 within the component of morphology, behavior code 6 is not allowed in the cancer registration system. Instead, it is recommended that ill-defined topography codes should be used when a metastatic tumor is suspected. The ICD-O codes and TNM stages are widely used in cancer registries worldwide to describe cancer cases in detail while the ICD-10 which describes cancer cases in brief is used as an overview of cancer problems.

Currently in Thailand, cancer registries start cancer case finding procedures by searching for cases with ICD-10 'C' and some 'D' codes which are considered malignant in the ICD-O version 3 in the HIS database. After abstraction of essential variables in the HIS and medical records, variables are coded and entered into the cancer registry database. An ICD-O3 code is used in data entry and an ICD-10 code is generated back to the case in two ways, by the database application and/or the IARCcrgTools (IARC, 2013b), a stand-alone software created by the IARC. After the ICD-O code is verified and checked against a few general characteristics of the patient, CanReg (IARC, 2013a) software created by the International Agency for Research on Cancer (IARC) automatically generates an ICD-10 code and stores it in the database. Before analysis of cancer statistics and creation of a report, cancer registries use IARCcrgTools, to do a similar process but in a batch mode.

Currently, all general (provincial) and center hospitals in Thailand are equipped with an HIS database. The HOSxP (Bangkok Medical Software, Bangkok, Thailand) is the most commonly used HIS system by hospitals in Thailand. In the current system of medical care for cancer patients, general hospitals are not equipped with full investigation and treatment modalities. Suspected cancer patients are referred to a higher-level hospital for specialized treatment. The National Health Security Office (NHSO) and the other two governmental health insurance schemes cover the diagnosis and treatment fee for cancer patients when the procedures occur in a hospital designated as a cancer center. This referral system serves cancer patients and medical care providers well.

Diagnoses coded with the ICD-10 system in the HIS database in a general hospital are usually provisional rather than definite. However, using this process, the same patient can get a different code from the original and referral hospitals. In addition, there are cases where a patient with one disease is coded with multiple ICD-10 cancer codes, which can occur when a series of provisional codes for the same patient are kept in the HIS database of a general hospital and are used for case tracking, even though the final diagnosis was different from the provisional ones. Another instance of different diagnoses in different databases can occur when a patient is referred to another institution, and the diagnosis at the end is not the same as that recorded in the HIS system of the first hospital in the referral system. Thus, in cancer registration, provisional diagnoses are mainly used for case finding and all the information including the diagnosis in the referred hospital must be retrieved and recorded as the final diagnosis.

Summary statistics of cancer cases in a general hospital using the ICD-10 from an HIS are usually different from those reported in a cancer registry using the IARC standard protocols (Jensen et al., 1991). However, there has been no study on the discrepancy of cancer statistics derived from both sources in the referral system established in any country. The objective of this study was to examine the characteristics and magnitude of agreement or disagreement represented by the positive predictive value (PPV) of provisional cancer diagnoses in the HIS database in a general hospital in comparison with the final diagnosis of cancer by the ICD-10 code generated from a well-established cancer registry in a medical school hospital with higher level of investigation and treatment facilities than available in the smaller local hospitals.

Materials and Methods

Study setting

Pattani hospital is a general hospital and is the largest hospital in the province, with a total of 230 beds. It provides surgical resection for breast cancer patients but has limited facilities for chemotherapy, and no radiotherapy or other sophisticated treatment modalities. Cancer patients needing further treatment are transferred to Yala hospital in Yala province or to Songkhlanagarind Hospital, the teaching hospital associated with the Faculty of Medicine of Prince of Songkla University in Songkhla province north of Pattani, which has its own hospital-based cancer registry in which cancer cases of Pattani residents are also collected.

Selection of cases

The records of cancer patients who were residents of Pattani province were obtained from the HIS database of Pattani Hospital in two separate database sections, out-patient and in-patient visits, from January 2007 to May 2011. In these records, the ICD-10 codes of primary diagnosis and 2 co-morbidities for each patient visit were assigned by medical statisticians and keyed into the HIS computer database of Pattani Hospital.

At Songklanagarind Hospital, staff of the Cancer Registry unit identified cancer cases diagnosed and/or treated in the hospital by finding cases with ICD-10 codes 'C' and some 'D' codes which were also determined as malignant according to the ICD-O3 rating. The related information of these cases, including the place of residence, was abstracted from the electronic medical records. The morphological diagnosis (M code) and organ site of origin of cancer (topography or T code)

were coded according to the ICD-O3 protocols. When codes were keyed into the cancer registry database, the appropriate ICD-10 code according to the primary site of the malignancy was then automatically generated by the standard algorithm. Only cases present in both databases were analyzed.

Statistical analysis

R software (R Core Team, 2013) was used for all statistical analysis. The provisional ICD-10 code from Pattani Hospital was plotted against the code generated from the Songklanagarind Hospital Cancer Registry by sex. Since the direction of diagnosis was from provisional diagnosis in Pattani hospital to final diagnosis in Songklanagarind Hospital, PPVs were calculated for each sex and for both sexes combined.

Results

The combined in-patient and out-patient files with diagnosis of cancer obtained from Pattani Hospital, there were 7,268 hospital visits. Of these 7,268 visits, there were 1,964 patients, 859 males and 1,105 females, recorded as meeting the 'C' category of the ICD-10 classification in at least one of the three indicants (primary and 2 co-morbidity diagnoses) in the two groups of files obtained from the Pattani Hospital HIS database. When those 1,964 cases were compared with cancer cases in the cancer registry database of Songklanagarind Hospital through 13-digit personal identification numbers, a total of 313 cases (15.9%) were found to be present in both databases. Among these 313 matched cases, 222 cases (11.3%) had visited Pattani Hospital before Songklanagarind Hospital, and 91 cases had been diagnosed before January 2007, or in other words had by-passed the official referral system to Songklanagarind Hospital where they had originally been diagnosed before going back for follow-up care near their residence or had visited Pattani Hospital for treatment of other diseases. Only the final 222 cases, 109 males and 113 females, who had visited Pattani Hospital first and then been referred to Songklanagarind Hospital were analyzed in this study.

Of the 109 male cancer cases first diagnosed in Pattani Hospital and referred to Songklanagarind Hospital for further diagnosis and/or treatment, 76 had been assigned the same ICD-10 codes in both the Pattani Hospital HIS database and the Cancer Registry database of Songklanagarind Hospital. Thus, the agreement or PPV for primary topographic site of cancer cases first diagnosed in Pattani Hospital was 69.7% (95%CI: 60.2-78.2%) in males. Agreement in 76 out of 113 females gave a PPV of 67.3% (95%CI: 57.8-75.8%). There were no differences in the two percentages of agreement by sex with Fisher's exact test p-value of 0.773. The number of those who had the same ICD-10 code was 152 out of 222 cases of both sexes combined, producing an overall percentage of agreement or PPV of 68.5% (95%CI: 61.9-74.5%).

Of all these 222 cancer cases, there were 152 Muslims (68.5%) and 70 Buddhists (31.5%). The agreement rates between Muslims and Buddhists were not significantly different, 69.7% (95%CI: 61.8-76,9%) and 65.7% (95%CI: 53.4-76.7%) respectively (the Fisher's exact test

Table 1. Distribution of Definite Diagnosis by Primary Organ/Type of Cancer in Songklanagarind Hospital of Patients with Provisional Diagnosis as Other and Unspecified (O&U) Categories in Pattani Hospital

Primary organ/ type of malignant neoplasm	ICD-10 Code	Total (21 cases)	
		Males (9 cases)	Females (12 cases)
Tongue	C01-02	1	-
Nasopharynx	C11	-	1
Stomach	C16	1	-
Colon, rectum	C19-20	-	2
Larynx	C32	1	-
Trachea, bronchus, lung	C33-34	3	-
Skin, non-melanoma	C44	1	-
Breast	C50	-	3
Cervix uteri	C53	-	2
Ovary	C56	-	1
Thyroid	C73	1	-
Non-Hodgkin lymphoma	C82-85, C96	1	1
Multiple myeloma	C90	-	1
Other and unspecified	O&U	-	1

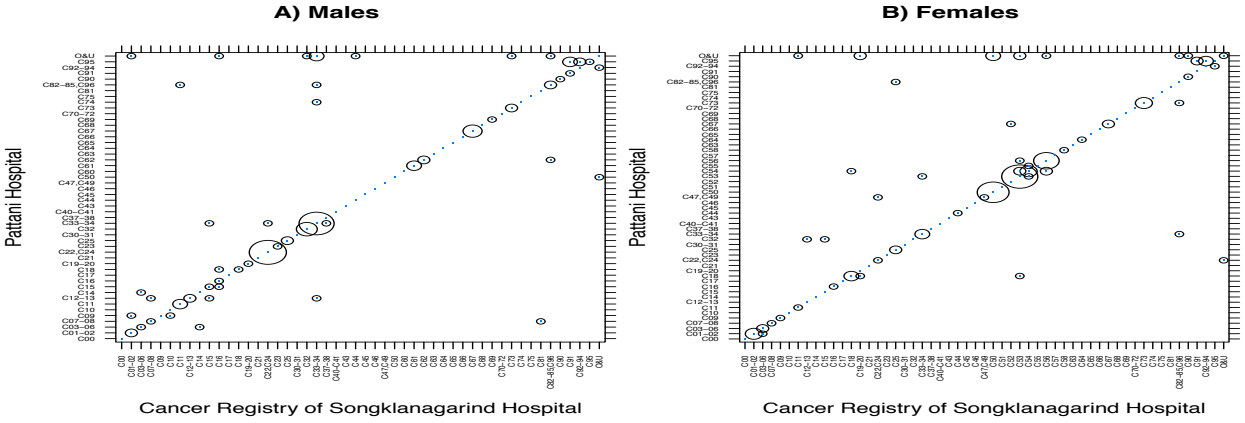

Figure 1. Comparison of ICD-10 Codes of Cancer Cases in Pattani Hospital with Codes Assigned by the Cancer Registry of Songklanagarind Hospital. The area of the circles corresponds to the frequency of cases. A) In males, cancers of the liver and bile duct (C22 and C24) and bronchus and lung (C33-34) had high agreement between the two centers. B) The areas of highest agreement in females were cervix uteri (C53), breast (C50) and ovarian (C56) cancers

p-value for the difference is 0.663).

Figures 1A and 1B demonstrate the agreement of ICD-10 codes assigned to cancer cases in Pattani Hospital with the codes assigned by the Cancer Registry unit of Songklanagarind Hospital. Liver and bile duct (C22 and C24) and bronchus and lung (C33-34) were the most common organs of cancers with the highest rate of agreement between the two databases among males, while cervix uteri (C53), breast (C50 and ovary (C56) had the highest concurrence among females. Some of cases who had provisional diagnosis from Pattani hospital falling into other and unspecified (O&U) category, the top row in Figure 1, primary site of cancer could be documented at Songklanagarind hospital in both males and females. Table 1 shows the definite diagnoses in Songklanagarind Hospital in 21 patients referred from Pattani Hospital as O&U. The most frequent definite diagnoses were malignancies of the lung in males, and of breast, cervix uteri, and colon-rectum in females. The primary site was never identified in one case. Hematologic malignancies were the diseases with the highest rates of disagreement between the two hospitals (right upper corner of Figure 1).

Discussion

Using the final diagnosis in Songklanagarind Hospital as the gold standard, the agreement or PPV for primary topographic site of male cancer cases first diagnosed in Pattani Hospital was 69.7% (95%CI: 60.2-78.2%) while it was 67.3% (95%CI: 57.8-75.8%) among females. The proportions were not statistically different in either sex. The overall PPV for both sexes combined was 68.5% (95%CI: 61.9-74.5%). Viewed from the other hand, the overall false negative rate was 31.5% (95%CI: 26.5%-38.1%). Thus we can say that the ICD-10 diagnosis codes for cancer in provincial hospitals must be used with care in reporting cancer statistics as the misclassification rate can be as high as 30% or more. The PPV for diagnosis of cancer in other general hospitals in Thailand is not known since this is the first study on this issue. The agreement rates between the hospital diagnoses in terms of religion, Buddhist or Muslim, were not statistically different, indicating it is the disease itself, not ethnic group, that influences a doctor's opinion on cancer diagnosis, and systematic misclassification of diagnosis is related solely to incorrect conclusion of malignancy.

The results from this study can be applied to most general hospitals in which investigation facilities and cancer specialists are limited and a large proportion of cases are referred to specialized cancer centers elsewhere for definite diagnosis and treatment. In Thailand, pathologists and oncologists are often unavailable in general hospitals. The referral system was established and is routinely utilized as standard medical care for cancer patients under all public health insurance schemes provided by the government. Thus, we can say that this situation is built in to the medical care system and such misclassifications in cancer diagnoses at the level of the general hospital is unavoidable and the degree of misclassifications is big enough to distort cancer statistics in any provinces in the country.

To make cancer statistics valid within this situation, cancer registries, both hospital-based and population-based ones, must be done as an intensive information network where demographic and clinical data of cancer patients must be transferred through the referral system. However, the information transfer protocol has not yet been well-established in many medical centers in Thailand. Lack of trained personnel, limited budgets, and lack of clear policies are among the barriers to establishing standardized cancer registries in Thailand. General (provincial) hospitals are required by the service plan policy of the Ministry of Public Health to report statistics of common cancers in the province. However, there is no strategic plan requiring these hospitals to provide essential knowledge of standard procedures of case ascertainment and data abstraction as required by the International Agency for Research on Cancer (Jensen et al., 1991). To deal with the resulting data limitations, one of the easiest ways to get the statistics is to use existing HIS data bases to summarize cancer cases by ICD-10 and distortion of cancer statistics is the major problem to be solved by some means of adjustment.

A certain degree of inaccuracy of ICD-9 and 10 coding within hospitals has been reported (George and Maddocks 1997; Gibson and Bridgman 1998; O'Malley et al., 2005; Gologorsky et al., 2014). The main error sources include information in the medical records coders received in coding, experience of coders in tackle with ICD code ambiguity, and implementation of regular quality control protocols (O'Malley et al., 2005; Gologorsky et al., 2014). Though a routine audit of coding processes would improve the accuracy of routine diagnostic codes (Gibson and Bridgman 1998), such audits are notably lacking in routine practice in Thailand. Researchers, policymakers, national health insurance payers and physicians should recognize the systemic coding limitations and inaccuracies and use the data and reports generated from this imperfect system in making clinical and policy decision with caution (Gologorsky et al., 2014).

In summary, this study demonstrates the phenomenon of misclassification in cancer statistics generated from general hospitals in Thailand. The summary statistics in a hospital without full investigation facilities must be used with care, as such statistics do not represent the true distribution of cancer in a hospital with limited investigation facilities and only simple diseases can be diagnosed with some high degree of accuracy. Such untrustworthy data can misinform policy decisions and thus lead to suboptimal cancer control plans. A system where feedback of patient information can be transferred back from a tertiary to the referring hospital should be considered to increase the accuracy of cancer statistics and to facilitate comprehensive care of cancer patients when final diagnosis and treatment plan is open to the referring hospital.

This study was limited by the small number of cancer cases and the fact it was done in only one pair of hospitals in the countrywide referring system of Thailand. Analysis of other pairs of hospitals in other regions of the country may give a better insight of PPV, which could be different with different degrees of basic and tertiary medical care.

Acknowledgements

We would like to thank the cancer registry of Songklanagarind Hospital for their cooperation in providing the cancer data and Pattani General Hospital for the details on the cancer cases in Pattani province. Dr. Sriplung acknowledges the support of the National Science and Technology Development Agency (P-10-10307).

References

Bangkok Medical Software. HOSxP. [accessed on 2014 October 1]. Available from: http://hosxp.net/joomla25/

Fritz A, Percy C, Jack A, et al (2000). International Classification of Diseases for Oncology. 3rd ed., Geneva, Switzerland: World Health Organisation.

George AM, Maddocks GB. (1997) Accuracy of diagnostic content of hospital activity analysis in infectious diseases. *Br Med J*, **1**, 1332-4.

Gibson N, Bridgman SA (1998). A novel method for the assessment of the accuracy of diagnostic codes in general surgery. *Ann R Coll Surg Engl*, **80**, 293-6.

Gologorsky Y, Knightly JJ, Lu Y, Chi JH, Groff MW (2014) Improving discharge data fidelity for use in large administrative databases. *Neurosurg Focus*, **36**, 2.

IARC (2013a). [Internet] CanReg5. [accessed on 2013 April 15]. Available from: http://www.iacr.com.fr/canreg5.htm.

IARC (2013b). [Internet] IARCcrgTools. [accessed on 2013 April 15]. Available from: http://www.iacr.com.fr/iarccrgtools.htm.

Jensen OM, Parkin DM, McLennan R, Muir CS, Skeet RG (1991). Cancer registration: principles and methods. IARC scientific publications No. 95. Lyon, France: International agency for research on cancer.

O'Malley KJ, Cook KF, Price MD, et al (2005). Measuring Diagnoses: ICD code accuracy. *Health Serv Res*, **40**, 1620-39.

R Core Team. (2013). A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

World Health Organisation (1992). ICD-10 Classifications of mental and behavioural disorder: clinical descriptions and diagnostic guidelines. geneva, Switzerland: World Health Organisation.