



Assessment of Text Documentation Accompanying Uncoded Diagnoses in Computerized Health Insurance Claims in Japan

Shinichi Tanihara

Department of Public Health and Preventive Medicine, School of Medicine, Fukuoka University, Fukuoka, Japan

Received May 27, 2014; accepted September 30, 2014; released online February 7, 2015

Copyright © 2015 Shinichi Tanihara. This is an open access article distributed under the terms of Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Background: Uncoded diagnoses in health insurance claims (HICs) may introduce bias into Japanese health statistics dependent on computerized HICs. This study's aim was to identify the causes and characteristics of uncoded diagnoses.

Methods: Uncoded diagnoses from computerized HICs (outpatient, inpatient, and the diagnosis procedure-combination per-diem payment system [DPC/PDPS]) submitted to the National Health Insurance Organization of Kumamoto Prefecture in May 2010 were analyzed. The text documentation accompanying the uncoded diagnoses was used to classify diagnoses in accordance with the International Classification of Diseases-10 (ICD-10). The text documentation was also classified into four categories using the standard descriptions of diagnoses defined in the master files of the computerized HIC system: 1) standard descriptions of diagnoses, 2) standard descriptions with a modifier, 3) non-standard descriptions of diagnoses, and 4) unclassifiable text documentation. Using these classifications, the proportions of uncoded diagnoses by ICD-10 disease category were calculated.

Results: Of the uncoded diagnoses analyzed ($n = 363\,753$), non-standard descriptions of diagnoses for outpatient, inpatient, and DPC/PDPS HICs comprised 12.1%, 14.6%, and 1.0% of uncoded diagnoses, respectively. The proportion of uncoded diagnoses with standard descriptions with a modifier for *Diseases of the eye and adnexa* was significantly higher than the overall proportion of uncoded diagnoses among every HIC type.

Conclusions: The pattern of uncoded diagnoses differed by HIC type and disease category. Evaluating the proportion of uncoded diagnoses in all medical facilities and developing effective coding methods for diagnoses with modifiers, prefixes, and suffixes should reduce number of uncoded diagnoses in computerized HICs and improve the quality of HIC databases.

Key words: coding; health insurance claim; ICD-10; uncoded diagnoses

INTRODUCTION

A precise evaluation of the burden of disease is required to set priorities in health policy. Health insurance claims (HICs) are prepared by healthcare providers for reimbursement of their services. In Japan, HIC records contain information about health insurance qualification status, healthcare costs, clinical procedures, and diagnoses. Statistics on medical expenditures in Japan, such as those reported in the Estimation of National Medical Expenditures, Social Insurance Claims Survey, and the National Health Insurance Medical Benefit Surveys, are based on HICs. The following list outlines healthcare indicators that recently have been calculated using HIC information: the quality of care of patients with diabetes,¹ the

completeness of the infectious disease surveillance system,² the relationship between disease type and patients' health-seeking behaviors,³ the relationships between health guidance for metabolic syndrome and outpatient charges and drug costs and metabolic syndrome,⁴ the association between hospital case volume and mortality in non-elderly pneumonia patients,⁵ and regional differences in the performance of bone marrow transplants.⁶ There are limitations in the information recorded in HICs in Japan due to regulations on medical cost reimbursement. First, most HICs contain more than one diagnosis⁷⁻⁹ because healthcare providers submit only one HIC describing all of the healthcare services provided to an individual during a calendar month. Previously, HICs were submitted on paper, and it was

Address for correspondence. Shinichi Tanihara, Department of Public Health and Preventive Medicine, School of Medicine, Fukuoka University, 7-45-1 Nanakuma, Jonan-ku, Fukuoka 814-0180, Japan (e-mail: taniyan@cis.fukuoka-u.ac.jp).

common to select only one principal diagnosis from a HIC for the database, even if there were multiple diagnoses. This practice resulted from the labor costs of inputting the information from the HIC into a database. It has been reported that hypertension tended to be selected as the principal diagnosis among the elderly insured by the National Health Insurance for Medical Services for the Aged.⁷ Thus, the estimation of disease-specific medical expenditures using HICs may be biased.^{8,10}

In Japan, confirmed, unconfirmed, and disproved diagnoses are all included in HICs.^{10,11} Furthermore, health insurance coverage is based primarily on fee-for-service reimbursement, with regulations dictating that each clinical procedure be described and justified by a corresponding diagnosis. For certain diagnostic categories (eg, neoplasm), various examinations are usually conducted, including consultations, imaging, and investigations using tumor markers. When test results indicate that a suspected disease is not present, rule-out diagnoses are included in the HIC to ensure reimbursement for the clinical procedures conducted. Technical limitations prevent the assessment of all the information mentioned in a HIC; therefore, information about multiple and rule-out diagnoses have been ignored in estimates of disease-specific medical expenditures, which may introduce bias. For example, more than one-third of the medical expenditures for outpatient care of neoplasms are spent on rule-out diagnoses,¹⁰ and there are differences in procedures and medical expenditures for inpatients with and without a rule-out diagnosis of sepsis.⁹

Until recently, most of the technical limitations of using information contained in HICs were caused by HICs that were submitted on paper. The labor cost of maintaining a database containing all of the information on the claim forms prevented assessments of complete HICs. Therefore, the information about multiple and rule-out diagnoses in HICs were often ignored in estimates of disease-specific medical expenditures.⁸⁻¹² After August 2010, however, all hospitals and medical clinics were mandated to submit computerized HICs to claim reimbursement for the cost of healthcare services. Thus, 96.6% of HICs were computerized by March 2014.¹²

Uncoded diagnoses in computerized HICs have been recognized as a new problem in the management of these data.^{13,14} Medical facilities are required to code all of the diagnoses according to the International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10) before submitting the computerized HIC.¹⁵ However, if the medical facilities are unable to code a diagnosis, it is listed as “uncoded”; in this case, the HICs are submitted with text documentation related to the uncoded diagnoses. The proportion of uncoded diagnoses electronically submitted to National Health Insurance Organization (NHIO) of Kumamoto Prefecture in 2010 was reported as approximately 10%.^{13,14}

Uncoded diagnoses are not used in the summaries of key health statistics, such as the Social Insurance Claims Survey and the National Health Insurance Medical Benefit Surveys. The proportion of uncoded diagnoses varies by the type of HIC and disease category, and uncoded diagnoses may have introduced bias into Japanese health statistics that are based on HICs.¹⁴ The proportion of uncoded diagnoses was relatively high for some diagnoses, especially *Diseases of the ear and mastoid process* and *Injury, poisoning, and certain other consequences of external causes*, which are supposed to be accompanied by information about the disease sites (eg, the right or left side of the body). The causes underlying uncoded diagnoses in computerized HICs must be identified to improve the accuracy of Japanese health statistics that are based on computerized HIC data, such as the estimation of disease-specific medical expenditures. The aim of this study is to investigate the causes of uncoded diagnoses in computerized HICs by analyzing the text documentation accompanying the uncoded diagnoses.

METHODS

Health insurance claims

This study reviewed all 363 753 uncoded diagnoses of the 3 804 246 diagnoses in HICs that were electronically submitted to the NHIO of Kumamoto Prefecture in May 2010. The details of HICs in Japan have been described in our previous report.¹⁴

Medical facilities are required to code all diagnoses according to the standard descriptions of diagnoses defined in the “Manual about Specifications of the Computerized Health Insurance Claim System Master File”¹⁵ before submitting HICs. The coding of the standard descriptions of diagnoses, which is defined in the master file, is based on the ICD-10. In addition, every diagnosis is accompanied by supporting text documentation. We first classified text documentation accompanying the uncoded diagnoses according to the disease categories in the ICD-10, which are subdivided into chapters. Next, we classified text documentation as follows: 1) a standard description of diagnoses, 2) a standard description with a modifier, 3) a non-standard description of diagnoses, and 4) unclassifiable text documentation. The “standard description of diagnoses” classification refers to text documentation that is identical to one of the standard descriptions. The “standard description with a modifier” classification is a combination of the standard description of diagnoses plus information, such as disease sites (eg, right or left side of the body). A “non-standard description of diagnoses” classification occurs when the description of the diagnosis in the text documentation is different from the “standard description” but can still be classified according to the ICD-10 categories. The “unclassifiable text documentation” classification refers to text documentation that cannot be classified in one of the other

Table 1. The number of diagnoses and uncoded diagnoses according to type of claim and type of text documentation accompanying uncoded diagnoses

Type of diagnoses	Type of HIC								P
	Outpatient		Inpatient		DPC/PDPS		Total		
Total diagnoses	3 393 106		325 968		85 172		3 804 246		
Uncoded diagnoses	316 151	9.3%	35 493	10.9%	12 109	14.2%	363 753	9.6%	<0.001
Types of text documentation accompanying uncoded diagnoses									
Standard descriptions of diagnoses	107 451	34.0%	12 663	35.7%	6 515	53.8%	126 629	34.8%	
Standard descriptions with a modifier	146 529	46.3%	13 931	39.2%	5 408	44.7%	165 868	45.6%	<0.001
Non-standard descriptions of diagnoses	38 217	12.1%	5 196	14.6%	118	1.0%	43 531	12.0%	
Unclassifiable text documentation	23 954	7.6%	3 703	10.4%	68	0.6%	27 725	7.6%	

DPC/PDPS, Diagnosis Procedure Combination/Per-Diem Payment System; HIC, health insurance claim.

three categories. The latter two categories are classified according to the presence or absence of a modifier.

Statistical analysis

The proportions of the types of text documentation accompanying the uncoded diagnoses from outpatient care, inpatient care, and DPC/PDPS were compared, and the diagnoses were classified according to the disease categories in the ICD-10. The proportions of uncoded diagnoses according to the type of text documentation and HIC were compared. Descriptive summary statistics are presented as frequencies and proportions for the categorical data. The χ^2 test was used to compare the proportions of uncoded diagnoses among the disease categories and the overall proportion of uncoded diagnoses according to the type of HIC. A two-tailed *P* value of <0.05 was considered statistically significant. All analyses were performed using IBM SPSS Statistics, Version 19 (International Business Machines Corporation, Armonk, NY, USA).

Ethical concerns

All personal information from the HIC data was deleted by the NHIO before the data were delivered to the researchers. This study was approved by the Institutional Review Committee of Fukuoka University.

RESULTS

Table 1 shows the number of uncoded diagnoses analyzed according to the type of claim and type of accompanying text documentation. Among the 363 753 uncoded diagnoses included in the analyses, 316 151 (86.9%) were from outpatient medical HICs, 35 493 (9.8%) were from inpatient medical HICs, and 12 109 (3.3%) were from DPC/PDPS HICs. Standard descriptions of diagnoses were included with approximately one-third of the text documentations submitted with the uncoded diagnoses from outpatient (34.0%) and inpatient (35.7%) HICs; however, standard descriptions of diagnoses comprised more than half of the DPC/PDPS HICs (53.8%). The proportions of uncoded diagnoses with standard descriptions with a modifier were 46.3%, 39.2%, and

44.7% for the outpatient, inpatient, and DPC/PDPS HICs, respectively. The proportions of uncoded diagnoses with non-standard descriptions of diagnoses were 12.1% for outpatient HICs, 14.6% for inpatient HICs, and only 1.0% for DPC/PDPS HICs. The proportions of uncoded diagnoses with unclassifiable text documentation from outpatient and inpatient HICs were 7.6% and 10.4%, respectively; only 0.6% of uncoded diagnoses in DPC/PDPS HICs included unclassifiable text documentation. The proportion of the types of text documentation accompanying uncoded diagnoses differed significantly according to type of HIC.

Table 2 shows the proportion of uncoded diagnoses according to the type of text documentation accompanying the diagnoses in the outpatient HICs after the documentation data were classified into major disease categories. For every disease category, the proportions of uncoded diagnoses were significantly different from the overall proportion of the uncoded diagnoses. The proportion of uncoded diagnoses with the standard description for diagnoses was lowest in the category of *Pregnancy, childbirth, and the puerperium* (4.5%). However, only seven diagnoses in total fell into this category. The proportion and number of uncoded diagnoses related to *Diseases of the eye and adnexa* were 5.4% and 2082 respectively. The proportion of uncoded diagnoses with standard descriptions of diagnoses was the highest in the *Mental and behavioral disorders* category (72.2%), followed by the *Diseases of the respiratory system* category (69.3%).

The proportion of uncoded diagnoses with the standard description with a modifier was the largest in the *Diseases of the eye and adnexa* category (91.1%); *Neoplasms* had the second largest proportion (68.7%), and *Pregnancy, childbirth, and the puerperium* (15.4%) had the lowest proportion; however, only 24 diagnoses were classified in the latter category. The proportion and number of uncoded diagnoses in the *Diseases of the respiratory system* category, which was the second least frequently represented disease in this text documentation category, were 19.4% and 1712, respectively.

The proportion of uncoded diagnoses with non-standard descriptions of diagnoses was highest for *Pregnancy, childbirth, and the puerperium* (80.1%), followed by

Table 2. The proportion of uncoded diagnoses according to type of text documentation accompanying diagnoses in outpatient HICs

ICD-10 major disease categories	Types of text documentation accompanying uncoded diagnoses						P
	Standard descriptions of diagnoses		Standard descriptions with a modifier		Non-standard descriptions of diagnoses		
	Uncoded	(%)	Uncoded	(%)	Uncoded	(%)	
1 Certain infectious and parasitic diseases	2234	32.6%	3729	54.4%	895	13.1%	<0.001
2 Neoplasms	1498	15.2%	6774	68.7%	1583	16.1%	<0.001
3 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	732	42.1%	880	50.6%	128	7.4%	<0.001
4 Endocrine, nutritional, and metabolic diseases	9289	60.5%	5100	33.2%	969	6.3%	<0.001
5 Mental and behavioral disorders	10030	72.2%	2995	21.6%	865	6.2%	<0.001
6 Diseases of the nervous system	7369	58.4%	4187	33.2%	1070	8.5%	<0.001
7 Diseases of the eye and adnexa	2082	5.4%	35325	91.1%	1364	3.5%	<0.001
8 Diseases of the ear and mastoid process	1984	41.8%	2280	48.0%	487	10.3%	<0.001
9 Diseases of the circulatory system	17848	49.5%	16387	45.4%	1847	5.1%	<0.001
10 Diseases of the respiratory system	6104	69.3%	1712	19.4%	995	11.3%	<0.001
11 Diseases of the digestive system	14346	42.6%	14502	43.1%	4811	14.3%	<0.001
12 Diseases of the skin and subcutaneous tissue	4022	21.3%	10002	53.0%	4830	25.6%	<0.001
13 Diseases of the musculoskeletal system and connective tissue	14114	33.6%	21467	51.1%	6427	15.3%	<0.001
14 Diseases of the genitourinary system	6227	47.3%	5169	39.3%	1760	13.4%	<0.001
15 Pregnancy, childbirth, and the puerperium	7	4.5%	24	15.4%	125	80.1%	<0.001
16 Certain conditions originating in the perinatal period	15	44.1%	8	23.5%	11	32.4%	<0.001
17 Congenital malformations, deformations, and chromosomal abnormalities	708	22.4%	1344	42.6%	1103	35.0%	<0.001
18 Symptoms, signs, and abnormal clinical and laboratory findings not elsewhere classified	6712	47.7%	4709	33.5%	2654	18.9%	<0.001
19 Injury, poisoning, and certain other consequences of external causes	2130	11.6%	9935	54.1%	6293	34.3%	<0.001
Total	107451	36.8%	146529	50.2%	38217	13.1%	

HIC, health insurance claim; ICD-10, the International Statistical Classification of Diseases and Related Health Problems, 10th Revision.

The number of uncoded diagnoses with unclassifiable text documentation was 23 954. Major disease category 20 (External causes of morbidity and mortality) was not found in outpatient HICs.

Congenital malformations, deformations, and chromosomal abnormalities (35.0%). The lowest proportion of uncoded diagnoses with non-standard descriptions of diagnoses was in the *Diseases of the eye and adnexa* category (3.5%), followed by *Diseases of the circulatory system* (5.1%).

Table 3 shows the proportion of uncoded diagnoses according to the type of text documentation accompanying the diagnoses for inpatient HICs. Except for *Certain conditions originating in the perinatal period*, the proportions of uncoded diagnoses were significantly different from the overall proportion of uncoded diagnoses among all disease categories. The proportion of uncoded diagnoses with standard descriptions of diagnoses was highest for *Certain conditions originating in the perinatal period* (71.4%). However, there were only seven diagnoses in this disease category and the difference was not statistically significant. For *Mental and behavioral disorders*, the category with the second highest proportion of uncoded diagnoses with standard descriptions, the proportion was 68.9%; the number of diagnoses classified in this category was 2077. The proportion of uncoded diagnoses with the standard descriptions of diagnoses was the lowest for *Pregnancy, childbirth, and the puerperium* (9.5%). However, there were only 21 diagnoses in this disease category. For *Injury, poisoning, and certain other consequences of external causes*, which was the category with the second lowest proportion (10.8%), there were 311 diagnoses.

The proportion of uncoded diagnoses with the standard description with a modifier was highest for *Diseases of the eye and adnexa* (65.1%) and second highest proportion for *Neoplasms* (60.1%), which was the same as observed in the outpatient HICs. However, the proportions in both of these disease categories were lower than those in the outpatient HICs. The proportion of uncoded diagnoses with the standard description with a modifier was lowest for *Pregnancy, childbirth, and the puerperium* (4.8%). However, there were only 21 diagnoses in this category. The second lowest proportion was observed in the *Mental and behavioral disorders* category (24.4%), which consisted of 735 diagnoses.

The proportion of uncoded diagnoses with non-standard descriptions of diagnoses was highest for *Pregnancy, childbirth, and the puerperium* (85.7%), as was observed among outpatient HICs. Once again, however, the total number of diagnoses in this disease category was small (18). The category with the second highest proportion was *Injury, poisoning, and certain other consequences of external causes* (35.4%). No uncoded diagnoses with the non-standard descriptions of diagnoses were observed in *Certain conditions originating in the perinatal period*. There were seven diagnoses classified in this disease category and the difference was not statistically significant. The second lowest proportion was in the *Diseases of the circulatory system* category (6.5%), with 260 diagnoses.

Table 3. The proportion of uncoded diagnoses according to type of text documentation accompanying diagnoses in inpatient HICs

ICD-10 major disease categories	Types of text documentation accompanying uncoded diagnoses						P
	Standard descriptions of diagnoses		Standard descriptions with a modifier		Non-standard descriptions of diagnoses		
	Uncoded	(%)	Uncoded	(%)	Uncoded	(%)	
1 Certain infectious and parasitic diseases	340	31.5%	562	52.1%	176	16.3%	<0.001
2 Neoplasms	173	18.0%	579	60.1%	211	21.9%	<0.001
3 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	173	40.0%	229	53.0%	30	6.9%	<0.001
4 Endocrine, nutritional, and metabolic diseases	1098	53.4%	757	36.8%	202	9.8%	<0.001
5 Mental and behavioral disorders	2077	68.9%	735	24.4%	204	6.8%	<0.001
6 Diseases of the nervous system	990	44.8%	1015	45.9%	205	9.3%	<0.001
7 Diseases of the eye and adnexa	225	28.4%	516	65.1%	52	6.6%	<0.001
8 Diseases of the ear and mastoid process	81	60.0%	38	28.1%	16	11.9%	<0.001
9 Diseases of the circulatory system	1749	43.5%	2010	50.0%	260	6.5%	<0.001
10 Diseases of the respiratory system	749	55.0%	412	30.2%	202	14.8%	<0.001
11 Diseases of the digestive system	1932	46.3%	1738	41.6%	504	12.1%	<0.001
12 Diseases of the skin and subcutaneous tissue	581	25.5%	1028	45.2%	666	29.3%	<0.001
13 Diseases of the musculoskeletal system and connective tissue	814	29.3%	1482	53.3%	484	17.4%	<0.001
14 Diseases of the genitourinary system	559	40.8%	534	39.0%	277	20.2%	<0.001
15 Pregnancy, childbirth, and the puerperium	2	9.5%	1	4.8%	18	85.7%	<0.001
16 Certain conditions originating in the perinatal period	5	71.4%	2	28.6%	0	0.0%	0.194
17 Congenital malformations, deformations, and chromosomal abnormalities	39	25.3%	70	45.5%	45	29.2%	<0.001
18 Symptoms, signs, and abnormal clinical and laboratory findings not elsewhere classified	765	37.2%	669	32.5%	623	30.3%	<0.001
19 Injury, poisoning, and certain other consequences of external causes	311	10.8%	1554	53.8%	1021	35.4%	<0.001
Total	12663	39.8%	13931	43.8%	5196	16.3%	

HIC, health insurance claim; ICD-10, the International Statistical Classification of Diseases and Related Health Problems, 10th Revision.

The number of uncoded diagnoses with unclassifiable text documentation was 3703. Major disease category 20 (External causes of morbidity and mortality) was not found in inpatient HICs.

Table 4 shows the proportion of uncoded diagnoses according to the type of text documentation accompanying the diagnoses in DPC/PDPS HICs. Except for *Diseases of the ear and mastoid process*, the proportions of uncoded diagnoses were significantly different from the overall proportion of uncoded diagnoses among all of the disease categories. The proportion of uncoded diagnoses with the standard descriptions of diagnoses was highest for *Pregnancy, childbirth, and the puerperium* (93.5%), followed by *Mental and behavioral disorders* (93.0%). There were fewer than 100 diagnoses in each of these two disease categories. However, in the disease category with the fourth highest proportion of uncoded diagnoses with standard descriptions of diagnoses, *Endocrine, nutritional, and metabolic diseases*, 545 (81.3%) diagnoses were accompanied by standard descriptions of diagnoses. The proportion of uncoded diagnoses with standard descriptions of diagnoses was the second lowest for *Diseases of the eye and adnexa* (27.2%) and lowest for *Injury, poisoning, and certain other consequences of external causes* (13.3%).

The proportion of uncoded diagnoses with the standard description with a modifier was highest for *Injury, poisoning, and certain other consequences of external causes* (84.5%), followed by *Diseases of the eye and adnexa* (72.8%). The proportions of uncoded diagnoses with the standard description with a modifier for *Diseases of the eye and adnexa* were high in all three types of HICs. The proportion of uncoded diagnoses with the standard description with a

modifier was lowest for *Pregnancy, childbirth, and the puerperium* (4.3%), followed by the *Mental and behavioral disorders* category (5.0%). However, few diagnoses were categorized in the three least frequently represented disease categories.

For almost every disease category, the proportions of uncoded diagnoses with non-standard descriptions of diagnoses were lower in outpatient than inpatient HICs. The highest proportion of uncoded diagnoses with non-standard descriptions of diagnoses was found for *Diseases of the skin and subcutaneous tissue* (4.2%). No diagnosis was found for *Diseases of the eye and adnexa*, *Diseases of the ear and mastoid process*, or *Certain conditions originating in the perinatal period*.

For all three types of HICs, the proportions of uncoded diagnoses with standard descriptions of diagnoses were highest for *Endocrine, nutritional, and metabolic diseases* and *Mental and behavioral disorders* and were lowest for *Diseases of the skin and subcutaneous tissue* and *Injury, poisoning, and certain other consequences of external causes*. The proportions of uncoded diagnoses with standard descriptions with a modifier were highest for *Neoplasms*, *Diseases of the eye and adnexa*, and *Injury, poisoning, and certain other consequences of external causes* and lowest for *Mental and behavioral disorders*; *Pregnancy, childbirth, and the puerperium*; and *Certain conditions originating in the perinatal period*. The proportions of uncoded diagnoses with non-standard descriptions of diagnoses were highest for

Table 4. The proportion of uncoded diagnoses according to type of text documentation accompanying diagnoses in DPC/PDPS HICs

ICD-10 major disease categories	Types of text documentation accompanying uncoded diagnoses						P
	Standard descriptions of diagnoses		Standard descriptions with a modifier		Non-standard descriptions of diagnoses		
	Uncoded	(%)	Uncoded	(%)	Uncoded	(%)	
1 Certain infectious and parasitic diseases	230	72.8%	83	26.3%	3	0.9%	<0.001
2 Neoplasms	1619	58.5%	1137	41.1%	11	0.4%	<0.001
3 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	201	77.6%	57	22.0%	1	0.4%	<0.001
4 Endocrine, nutritional, and metabolic diseases	545	81.3%	119	17.8%	6	0.9%	<0.001
5 Mental and behavioral disorders	93	93.0%	5	5.0%	2	2.0%	<0.001
6 Diseases of the nervous system	226	64.2%	123	34.9%	3	0.9%	<0.001
7 Diseases of the eye and adnexa	137	27.2%	367	72.8%	0	0.0%	<0.001
8 Diseases of the ear and mastoid process	21	42.0%	29	58.0%	0	0.0%	0.152
9 Diseases of the circulatory system	1067	60.4%	697	39.5%	2	0.1%	<0.001
10 Diseases of the respiratory system	471	65.5%	245	34.1%	3	0.4%	<0.001
11 Diseases of the digestive system	742	66.6%	352	31.6%	20	1.8%	<0.001
12 Diseases of the skin and subcutaneous tissue	84	39.6%	119	56.1%	9	4.2%	<0.001
13 Diseases of the musculoskeletal system and connective tissue	258	28.7%	631	70.2%	10	1.1%	<0.001
14 Diseases of the genitourinary system	251	70.1%	103	28.8%	4	1.1%	<0.001
15 Pregnancy, childbirth, and the puerperium	43	93.5%	2	4.3%	1	2.2%	<0.001
16 Certain conditions originating in the perinatal period	21	87.5%	3	12.5%	0	0.0%	0.005
17 Congenital malformations, deformations, and chromosomal abnormalities	27	77.1%	7	20.0%	1	2.9%	0.008
18 Symptoms, signs, and abnormal clinical and laboratory findings not elsewhere classified	281	76.8%	75	20.5%	10	2.7%	<0.001
19 Injury, poisoning, and certain other consequences of external causes	198	13.3%	1254	84.5%	32	2.2%	<0.001
Total	6515	54.1%	5408	44.9%	118	1.0%	

DPC/PDPS, Diagnosis Procedure Combination/Per-Diem Payment System; HIC, health insurance claim; ICD-10, the International Statistical Classification of Diseases and Related Health Problems, 10th Revision.

The number of uncoded diagnoses with unclassifiable text documentation was 68. There were 58 125 diagnoses classified as Major disease category 20 (External causes of morbidity and mortality) in the DPC/PDPS HICs, but no diagnoses were uncoded diagnoses in this disease category.

Diseases of the skin and subcutaneous tissue; Pregnancy, childbirth, and the puerperium; Congenital malformations, deformations, and chromosomal abnormalities; and Injury, poisoning, and certain other consequences of external causes. The proportions of uncoded diagnoses with non-standard descriptions of diagnoses were lowest for *Diseases of the blood and blood-forming organs, Certain disorders involving the immune mechanism, Diseases of the eye and adnexa, and Diseases of the circulatory system.*

DISCUSSION

The present study was the first in Japan to investigate the reasons for uncoded diagnoses in computerized HICs by analyzing the text documentation that accompanied the uncoded diagnoses. The three main findings were as follows: 1) the pattern of text documentation that accompanied uncoded diagnoses varied by type of HIC, 2) the proportions of uncoded diagnoses with standard descriptions of diagnoses with a modifier comprised approximately 40%–45% of all types of HICs, and 3) the proportions of uncoded diagnoses with standard descriptions of diagnoses with a modifier varied by disease category.

The proportions of uncoded diagnoses with non-standard descriptions of diagnoses and unclassifiable text documentation comprised approximately 10% of both

outpatient and inpatient HICs and approximately 1% of DPC/PDPS HICs. Certification to provide DPC/PDPS care is granted only to hospitals that submit the required data using the format specified by the regulations on medical cost reimbursement. Thus, medical facilities certified for DPC/PDPS are motivated to use the standard descriptions of diagnoses. In comparison, there is no penalty for using non-standard descriptions of diagnoses for outpatient and inpatient HICs.

Approximately one-half of the uncoded diagnoses included standard description of diagnoses with a modifier. The proportions of uncoded diagnoses with standard descriptions of diagnoses with a modifier were relatively high in disease categories in which it is important to distinguish the affected sites on the body, such as *Diseases of the eye and adnexa*. Since the diagnostic codes for the Japanese computerized HICs do not distinguish such sites because they are based on the ICD-10, diagnoses with additional information, such as the right or left side of the body, might appear difficult to code. Development of methods with acceptable labor costs for medical facilities to code diagnoses accompanied by additional information is required to reduce the number of uncoded diagnoses. In addition, evaluating the proportion of uncoded diagnoses from each medical facility and using incentives and penalties might be an effective method of reducing numbers of uncoded diagnoses.

The proportions of uncoded diagnoses with standard descriptions of diagnoses with a modifier were relatively high in the categories of *Diseases of the eye and adnexa* and *Injury, poisoning, and certain other consequences of external causes*, and were relatively low for *Certain conditions originating in the perinatal period*; *Mental and behavioral disorders*; and *Pregnancy, childbirth, and the puerperium*. The proportion of uncoded diagnoses varied by the type of HIC and disease category.¹⁴ If there was a tendency for certain types of diagnoses requiring additional information regarding body sites to be uncoded, it may have created a bias in estimating disease-specific medical expenditures.¹⁴

Approximately one-third of the accompanying text documentation with uncoded diagnoses were classified as standard descriptions of diagnoses. As there is no penalty for submitting uncoded diagnoses, some medical facilities may lack the motivation to complete the coding. Hospitalization charges per day are determined according to the principal diagnosis, which may account for the high proportion of uncoded diagnoses with standard descriptions of diagnoses in the DPC/PDPS HICs. This reimbursement scheme may be an additional reason why medical facility staff may not be motivated to complete the coding. There are various types of diagnoses in DPC/PDPS claims, including 1) principal diagnosis, 2) most resource-intensive diagnoses, 3) diagnoses prompting hospitalization, 4) comorbidity present at the time of admission, and 5) complications developed in the course of hospitalization. Further research on the relationship between the types of diagnoses in DPC/PDPS claims and uncoded diagnoses is required.

There are some limitations in this study. First, I investigated information only from the computerized HICs. Therefore, the characteristics of the medical facilities and the validity of the diagnoses were not assessed directly. An evaluation of the characteristics of the medical facilities, the validity of the diagnoses, and the accuracy of coding of the diagnoses in the HICs requires additional information using medical charts^{16–19} or telephone interviews.²⁰

Second, this study investigated only whether the standard descriptions of the diagnoses and modifier were included in the text documentation that accompanied the uncoded diagnoses. If a description currently used for a specific disease is not endorsed by a society of specialists, the description may not be used in the future and, therefore, will be classified as an abolished description. Such abolished description of the diagnoses will be deleted from the standard description of diagnoses after one year. From the results of this study, it is presumed that most of the modifiers are related to disease sites (eg, right or left side of the body). Modifiers other than disease sites, such as “severe,” are defined in the manual on specifications of the computerized HIC system master file.¹⁵ Further investigation of the effect of such changes in definitions on the standard descriptions of diagnoses and the type of modifier, prefix, and suffix is necessary.

Third, this study assessed HICs for reimbursement services provided only in May 2010. The current providers of health statistics in Japan, including The Social Insurance Claims Survey and National Health Insurance Medical Benefit Surveys, also used HIC data from May only. In the past, most of the technical limitations associated with data from HICs were because they were derived from paper submissions of HICs.^{7,8,10,11} This study reviewed HICs that were electronically submitted; it has been reported that 96.6% of HICs were computerized by March 2014.¹² Investigation of electronic HICs over a longer period is necessary.

In conclusion, the pattern of uncoded diagnoses varied by the type of HIC and disease category. Most of the text documentation that accompanied uncoded diagnoses was classified using the standard descriptions of diagnoses defined by the Japanese reimbursement rules and the standard descriptions of diagnoses with a modifier. An evaluation of the proportion of uncoded diagnoses in all medical facilities and the implementation of effective methods for coding diagnoses using modifiers, prefixes, and suffixes should reduce the proportion of uncoded diagnoses in the computerized HICs in Japan and improve the quality of HIC databases.

ONLINE ONLY MATERIAL

Abstract in Japanese.

ACKNOWLEDGMENTS

This work was partly funded by Grants-in-Aid for Scientific Research (No. 21119006 and 25460646) from the Ministry of Science and Education and the Research Project for Patients Aged 20 and Over with Intractable Chronic Diseases approved by Medical Aid for Children (No. H23-Jisedai-shitei-007) from the Ministry of Health, Labour and Welfare of Japan.

Conflicts of interest: None declared.

REFERENCES

1. Tomio J, Toyokawa S, Tanihara S, Inoue K, Kobayashi Y. Quality of care for diabetes patients using National Health Insurance claims data in Japan. *J Eval Clin Pract*. 2010;16:1164–9.
2. Tanihara S, Okamoto E, Imatoh T, Momose Y, Kaetsu A, Miyazaki M, et al. Evaluating measles surveillance: comparison of sentinel surveillance, mandatory notification, and data from health insurance claims. *Epidemiol Infect*. 2011;139:516–23.
3. Kisa K, Kawabata H, Terashita T, Nakamura T, Maezawa M. Medical reimbursement receipt analysis to determine the relationship between disease type and patients' healthcare-seeking behavior. *Gend Med*. 2012;13:77–84.
4. Okamoto E. Effects of health guidance on outpatient and pharmacy expenditures: a disease- and drug-specific 3-year

- observational study using propensity-score matching. *J Epidemiol.* 2013;23:262–9.
5. Kumamaru H, Tsugawa Y, Horiguchi H, Kumamaru KK, Hashimoto H, Yasunaga H. Association between hospital case volume and mortality in non-elderly pneumonia patients stratified by severity: a retrospective cohort study. *BMC Health Serv Res.* 2014;14:302.
 6. Maeda T, Babazono A, Nishi T, Matsuda S, Fushimi K, Fujimori K. Regional differences in performance of bone marrow transplantation, care-resource use and outcome for adult T-cell leukaemia in Japan. *BMC Health Serv Res.* 2014;14:337.
 7. Tanihara S, Yamagata Z, Une H. [Reliability of health insurance claim statistical data based on the principal diagnosis method]. *Nihon Eiseigaku Zasshi.* 2008;63:29–35 (in Japanese).
 8. Tanihara S, Okamoto E, Une H. A comparison of disease-specific medical expenditures in Japan using the principal diagnosis method and the proportional distribution method. *J Eval Clin Pract.* 2012;18:616–22.
 9. Tanihara S, Imatoh T, Momose Y. Assessment of medical expenditures for sepsis: Differentiating between cases with and without ruled-out diagnoses. *Acta Med Okayama.* 2014;68:1–6.
 10. Tanihara S, Okamoto E, Une H. Estimating medical expenditures spent on rule-out diagnoses in Japan. *J Eval Clin Pract.* 2012;18:426–32.
 11. Tanihara S, Okamoto E, Une H. A statistical analysis of ‘rule-out’ diagnoses in outpatient health insurance claims in Japan. *J Eval Clin Pract.* 2011;17:1070–4.
 12. Health Insurance Claims Review and Reimbursement Services. Press Release No. 499. 2014 [cited 2014 May 7]. Available from: http://www.ssk.or.jp/pressrelease/pdf/pressrelease_499.pdf.
 13. Tanihara S, Sairenchi T, Noda T, Tokumoto S, Uehara R, Yamagata Z, et al. [The distribution of uncoded diagnoses in health insurance claims electrically submitted]. *Kousei no Shihyo.* 2013;60(4):20–5 (in Japanese).
 14. Tanihara S. The proportion of uncoded diagnoses in computerized health insurance claims in Japan in May 2010 according to ICD-10 disease categories. *J Epidemiol.* 2014;24:392–6.
 15. Ministry of Health, Labour and Welfare of Japan. The manual about specifications of the computerized health insurance claim system master file [cited 2014 August 21]. Available from: <http://www.iryohoken.go.jp/shinryohoshu/standardMenu/doStandardMasterBz>.
 16. Wilchesky M, Tamblyn RM, Huang A. Validation of diagnostic codes within medical services claims. *J Clin Epidemiol.* 2004;57:131–41.
 17. Jetté N, Reid AY, Quan H, Hill MD, Wiebe S. How accurate is ICD coding for epilepsy? *Epilepsia.* 2010;51:62–9.
 18. Funch D, Holick C, Velentgas P, Clifford R, Wahl PM, McMahill-Walraven C, et al. Algorithms for identification of Guillain-Barré Syndrome among adolescents in claims databases. *Vaccine.* 2013;31:2075–9.
 19. Allen LA, Yood MU, Wagner EH, Aiello Bowles EJ, Pardee R, Wellman R, et al. Performance of claims-based algorithms for identifying heart failure and cardiomyopathy among patients diagnosed with breast cancer. *Med Care.* 2014;52:e30–8.
 20. Kolodner K, Lipton RB, Lafata JE, Leotta C, Liberman JN, Chee E, et al. Pharmacy and medical claims data identified migraine sufferers with high specificity but modest sensitivity. *J Clin Epidemiol.* 2004;57:962–72.