Comparison of Mayo Clinic Coding Systems

Jennifer St. Sauver
James Buntrock, Diana Rademacher,
Debra Albrecht, Melissa Gregg,
Donna Ihrke, Vinod Kaggal, Amy Weaver

Technical Report #83
Date: 12/14/2010

Department of Health Sciences Research
Mayo Clinic
Rochester, Minnesota

Abstract

This goal of this project was to determine how three different disease coding systems (HICDA, ICD-9, and SNOMED CT) compared at retrieving lists of individuals with specific medical conditions from among patients seen at Mayo Medical Center in 2004. Fifteen medical conditions were chosen to assess the concordance of the systems in retrieving lists of patients with those conditions. Additionally, we reviewed a sample of the medical records of patients with at least one Parkinsonism code from at least one coding system. For Parkinsonism, the sensitivity and specificity of each coding system was assessed compared to hand review of the medical records. The percent positive agreement for each of the three coding systems varied widely depending on the medical condition being studied, while percent negative agreement was consistently very high. Comparison of the sensitivity and specificity of the coding systems for Parkinsonism indicated that the SNOMED CT codes had the highest sensitivity, while the HICDA codes had the highest specificity. These data indicate that the systems are not identical in their ability to retrieve groups of people with specific diseases or conditions. However, the sensitivity and specificity of both SNOMED CT and ICD-9 are relatively high, and both may be useful replacements for the HICDA coding system.

Introduction

To conduct research studies on a disease or condition, it is necessary to first identify individuals with the disease or condition of interest. Clinical research on a huge array of diseases and conditions occurring in patients of Mayo Medical Center (MMC) is possible because MMC has developed and maintained a medical indexing and retrieval system since the early 1900s. Through this system, all diagnoses received by MMC patients have historically been given a numeric diagnostic code. Lists of people with diagnostic codes of interest may then be retrieved, and these individuals may be contacted, or their medical records may be reviewed, to answer the specific research questions.

Over the years, MMC diagnosis data have been coded using different coding systems, depending on the year of the diagnosis. Beginning in 1929, Mayo physicians were required to enter patient diagnoses following each visit onto a summary "master sheet" of the unit medical record, which was then forwarded to the Department of Health Sciences Research to be indexed by trained nosologists. The original Mayo diagnostic classification system was enlarged by Dr. Joseph Berkson in 1935 (“Berkson Coding System”) to provide rapid identification of MMC records for any of 20,000 diagnostic categories. In 1975, MMC nosologists began coding medical diagnoses using the Hospital Adaptation of the International Classification of Diseases, Second Edition (HICDA; a modification of the International Classification of Diseases, version 8; ICD-8).[1] HICDA codes consist of 8-digit numbers: the first five define the general disease or condition category, but three extra digits have been added as an MMC modification in order to increase specificity. For example, “acute anterior myocardial infarction” has the general code 04100, while “acute anteroseptal myocardial infarction” has the more specific 04100112 code.

The process of assigning HICDA codes has changed over time. First, the “master sheet” of the MMC paper medical record was eventually replaced by a “final diagnosis section” in the clinical note portion of the MMC electronic medical record, and the

conditions in this section received HICDA codes beginning in 2000 (Table 1). Additionally, beginning in 2002, MMC implemented an ”autocoder” computer tool. This tool uses Natural Language Processing techniques with historic coding data to assign the extended HICDA codes to all word-based diagnoses in the clinical notes of the MMC medical record.[2] A portion of these codes are considered provisional, and a team of nosologists review codes accepting or correcting HICDA codes as appropriate.

In 2009, it became clear that the human review necessary to ensure accuracy in the HICDA coding was no longer feasible. The sheer volume of medical records that needed to be checked could only be accomplished by a very large group of trained nosologists. Declining resources made maintaining such a large group impossible. Additionally, the HICDA coding system was based on the ICD-8 coding system, while most institutions throughout the U.S. relied on the ICD-9 coding system. Two alternate coding systems were available to MMC in 2009 and 2010. The first was the ICD-9 coding system and the second was the SNOMED CT coding system. ICD-9 codes were assigned to diseases or conditions summarized in the “clinical problem list” of the MMC medical record as part of the MMC billing process (Table 1). SNOMED CT codes were assigned to the full clinical note text through Natural Language Processing of the electronic medical record (Table 1). As the coding systems differ in specificity, and as each coding system is used to code a slightly different portion of the MMC medical record, it is unclear which system should be best optimized and used for future research studies. Therefore, this study was undertaken to determine how the ICD-9 and SNOMED CT coding systems compared to the HICDA coding system in retrieving lists of patients with various diseases. Maintenance of continuity between these systems ensures that long-term studies of disease trends and outcomes across coding system changes remain feasible.

Methods

Identification of Diseases and Codes.

A list of all medical index retrievals conducted in 2002 was obtained from the MMC Medical Index Retrieval Specialists. A list of 15 single-disease queries was chosen in

an attempt to represent common and rare conditions as well as chronic and acute conditions. Two retrieval specialists (DA and DI) identified the HICDA, ICD-9, and SNOMED CT codes that they felt would most accurately identify individuals with the disease or condition of interest. The retrieval specialists then met together to ensure that the codes chosen for each disease identified the same condition. Codes chosen for each condition are shown in Appendix 1.

Retrievals.

The year 2004 was chosen as the target year for the study, as all three coding systems were available, and because hand-verification of HICDA codes was complete through this year. Using the codes chosen by the retrieval specialists, three separate lists of patients with at least one disease code were retrieved from the three systems. Lists of individuals with HICDA codes were retrieved from Medical Index; lists of individuals with ICD-9 codes were retrieved from Decision Support Services (DSS; the Mayo Clinic billing system); lists of individuals with SNOMED CT codes were retrieved from Mayo Clinic Life Science System (MCLSS) using the process described below.

Clinical notes that are generated as part of clinical care are included in MCLSS, and the Natural Language Processing (NLP) pipeline reads these clinical notes in a clinical document architecture format.[3] The text information is tagged with appropriate codes for disorders, signs/symptoms, drug prescriptions, etc. SNOMED CT codes[4] are used for disorders and signs/symptoms, while RxNorm codes[5] are used for prescriptions or current medications. These coded clinical notes are also stored in MCLSS. For the 15 medical conditions being studied, a list of Unified Medical Language System concept unique identifiers (CUIs) was created using the SNOMED CT relationships in the Metathesaurus. Even though both SNOMED CT and UMLS CUI codes are recorded, we used the UMLS CUI for this evaluation. Once the CUIs listed were created, all "children" CUIs using the Lexgrid model for terminologies were added.[6] Patient records with the final CUIs of interest were retrieved from the coded clinical notes stored in MCLSS.

Parkinsonism.

A single condition was chosen to examine the sensitivity and specificity of each of the coding systems. A 3% random sample of individuals was chosen from the pooled list of all individuals identified by any of the three Parkinsonism retrievals for hand review. The electronic notes from these medical records were reviewed and each individual was identified as either having Parkinsonism or not having Parkinsonism by a single reviewer (MG). Instances where it was difficult to determine whether the individual had the condition or not were re-reviewed by a second reviewer (JS). Reasons for retrieval errors from each system were also recorded. We note that it was possible for some cases of Parkinsonism to have been missed by all three coding systems, but assumed that error rate was very small.

Analysis.

The total numbers of codes and people identified by each system were summarized, and percent positive and percent negative agreements were calculated among the three systems.[7] An example of how positive and negative agreements were calculated, together with the formulas that were used, is shown in Figure 2. Finally, sensitivity and specificity of each coding system was assessed for Parkinsonism, as compared to hand-review of the medical record notes. Reasons for discrepancies between the coding systems and hand-review were summarized.

Results

Overall, 294,623 individuals visited the Mayo Clinic at least once between January 1 and December 31, 2004. The three retrieval systems (HICDA, ICD-9, and SNOMED CT) identified different numbers of diagnostic codes and different numbers of individuals as having at least one of the 15 conditions of interest (Table 2). Though not universal for each of the 15 conditions studied, the SNOMED CT coding system identified the most individuals (39,136), while the H-ICDA coding system identified the fewest individuals (20,425) as having at least one visit for at least one of the 15 conditions.

Percent positive agreement among the systems varied dramatically depending on the coding system and on the condition being studied (Table 3). For example, the percent positive agreement varied from a low of 0.54% between the SnoMed and ICD-9 systems for non-functioning kidney to a high of 81.91% between HICDA and ICD-9 for primary malignancy/neoplasm of the lung. Percent negative agreement between the various coding systems was generally very high, ranging from 98.8% to 100%.

Overall, 8,333 people were identified as having at least one Parkinsonism code from at least one coding system during 2004. Therefore, 286,290 individuals did not have a Parkinsonism code during the same time period. A 3% random sample (n=250) of the people identified with at least one code was hand-reviewed. The overlap among the coding systems for identifying people with Parkinsonism from the three coding systems is shown in Figure 1. Of the 250 records that were hand-reviewed, 100 (40%) indicated that the patient had Parkinsonism. Sensitivity and specificity of each coding system are shown in Table 4. The sensitivity of the SNOMED CT coding system for identifying people with Parkinsonism was the highest (100%), while the sensitivity of the ICD-9 coding system was the lowest (75.0%). Specificity was highest for the H-ICDA coding system (99.9%), and equivalent for the ICD-9 and SNOMED CT coding systems (99.1%). Reasons for the errors in the coding systems are summarized in Table 5.

Discussion and Conclusions

Overall, the percent positive agreement among the three coding systems varied widely depending on the disease being studied, while percent negative agreement was very high. Comparison of the sensitivity and specificity of the coding systems for a single condition (Parkinsonism) indicated that the SNOMED CT codes had the highest sensitivity, while the HICDA codes had the highest specificity. These data indicate that the systems are not identical in their ability to retrieve groups of people with specific diseases or conditions. However, the sensitivity and specificity of both SNOMED CT and ICD-9 for Parkinsonism are relatively high, and both may be useful replacements for the HICDA coding system.

Overall comparisons of the three coding systems indicated that overlap among the systems varied widely for the same condition. For example, the HICDA coding system identified 97 people with “nonfunctioning kidney”, SNOMED CT identified 35 people, and ICD-9 identified 5,118 people with the same condition. This wide variation likely reflects both the coding system itself as well as the portion of the medical record that is coded by each system. For instance, in this example, the ICD-9 coding system does not contain a specific code for “nonfunctioning kidney”. Instead, the closest code available to retrieve this condition was code 593.9 for “unspecified disorder of kidney and ureter”. This code is clearly much less specific than either the HICDA code of 05869120 “nonfunctioning, kidney” or the SNOMED CT code C0232808 “absent renal function”, and is therefore likely to return many more people. This variation in the coding systems themselves is also reflected in the wide range of positive agreements among the systems – ranging from 0.54% between ICD-9 and SNOMED CT for nonfunctioning kidney to 81.91% between HICDA and ICD-9 for primary malignancy/neoplasm of the lung. These data also suggest that the best system for retrieving patients for a given study may depend on the disease or condition of interest.

It has long been recognized that no coding system will perfectly identify all individuals with a disease or condition of interest. Therefore, the goal has typically been to cast a wide net to capture all individuals who might have a disease. The medical records of patients with a particular code or group of codes are then hand-reviewed by trained nurse abstractors to determine whether a patient truly does or does not have the disease. To assess how the coding systems compared in retrieving people with a condition compared to hand-review of medical records, we reviewed a random sample of people with at least one Parkinsonism code. We found that for this condition, SNOMED CT was the most sensitive at detecting true cases (100% sensitivity), followed by HICDA (81% sensitivity), and ICD-9 (75% sensitivity). This may be because the SNOMED CT coding system is not limited to just the diagnosis section of the clinical note, but instead codes information found in the entire clinical note, including notes related to phone conversations or written communications with the patient. It is not entirely clear why the HICDA and ICD-9 coding systems missed some of the cases

detected by the SNOMED CT coding system; however, if the visit was not strictly for Parkinsonism, HICDA and ICD-9 may have been less likely to detect the case. As Parkinsonism is a chronic, long-term condition, extending the years searched beyond just 2004 may have improved detection of the patients that were missed by these systems.

Specificity of each of the coding systems for Parkinsonism was very high, with the HICDA system being the most specific (99.9%) and the SNOMED CT and ICD-9 systems having 99.1% specificity. It should be noted, however, that these numbers are somewhat artificial, as Parkinsonism is a relatively rare condition, and nearly 300,000 people visited Mayo Clinic for some reason in 2004. Instead, it is probably best to focus on the raw numbers of individuals who were identified by each system, but, on hand review, were found to not have Parkinsonism. When these numbers are considered, investigators who used the HICDA codes to detect Parkinsonism would have had to review 85 charts to obtain 81 true cases (95% yield). Investigators who used ICD-9 and SNOMED CT codes would have had to review 151 and 180 charts, respectively, to obtain 75 and 100 true cases (50% and 56% yield rates). When the reasons for the low specificity in the ICD-9 coding system were examined, it was clear that “restless leg syndrome” was frequently coded using ICD-9 code 333.99 (Other extrapyramidal diseases and abnormal movement disorders) rather than the more specific 333.94 (Restless leg syndrome) code. The more general 333.99 code was therefore pulled with the Parkinsonism ICD-9 code retrieval. Exclusion of this more general code would have significantly improved the specificity of the ICD-9 retrieval. The reasons for the low specificity of the SNOMED CT coding system were more complex, but an explicit search to avoid retrieving people with negation terms would have eliminated 15 of those who were retrieved, bringing the yield rate up to 61% (data not shown). Additionally, if the SNOMED CT retrieval was limited only to the current visit clinical information, and did not encompass the family or social history portions of the medical record, an additional 41 people may also have been eliminated, bringing the yield rate up to 81%.

In summary, the HICDA, ICD-9, and SNOMED CT coding systems are not completely comparable. Differences are inherent to the coding systems themselves, and the coding systems are used to code different portions of the medical record for different purposes. However, when we examined the sensitivity and specificity of each system for retrieving patients with Parkinsonism, the sensitivity of each system was quite good, ranging from 75-100%. Additionally, although the specificity of the ICD-9 and SNOMED CT systems were not ideal, additional experience with each system could likely improve the specificity of these retrievals. In conclusion, the ICD-9 and SNOMED CT coding systems may represent viable alternatives to the current HICDA coding system.

Funding

This project was funded by the Department of Health Sciences Research Executive Committee.

References

1. Commission on Professional and Hospital Activities: H-ICDA, hospital adaptation of ICDA. Ann Arbor, National Center for Health Statistics, 1973.
2. Pakhomov S, Buntrock JD and Chute CG: Using compound codes for automatic classification of clinical diagnoses. *Stud Health Technol Inform.* 107: 411-415, 2004.
3. Wikipedia: Clinical Document Architecture, 2010, vol. 2010. http://en.wikipedia.org/wiki/Clinical_Document_Architecture
4. SNOMED CT Clinical Terms User Guide, International Health Terminology Standards Development Organisation, 2009. http://www.ihtsdo.org/fileadmin/user_upload/Docs_01/SNOMED CT_CT/About_SNOMED CT_CT/Use_of_SNOMED CT_CT/SNOMED CT_CT_User_Guide_20090731.pdf
5. Liu S, Ma W, Moore R, Ganesan V and Nelson S: RxNorm: Prescription for electronic drug information exchange: *IT Pro*, 2005, pp 17-23. http://www.nlm.nih.gov/research/umls/rxnorm/RxNorm.pdf
6. National Cancer Institute: LexBig and LexEVS, 2010. https://cabig-kc.nci.nih.gov/Vocab/KC/index.php/LexBig_and_LexEVS
7. Cicchetti DV and Feinstein AR: High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol*. 43: 551-8, 1990.

Table 1. Portion of the Mayo Medical Record coded by each coding system

Coding System	What is coded?	Coding Process
HICDA	Final diagnosis section of clinical note	Auto-coded and human-verification of low confidence records
SNOMED CT	Full clinical note text	Natural language processing
ICD-9	Clinical problem list	Provided selected problems sent to finance with human review and assembly by finance group

Table 2. Numbers of codes and people retrieved by each coding system for 15 diseases and conditions (2004)

	HICDA		SNOMED CT		ICD-9	
Medical condition or disease	**No. of codes**	**No. of individuals**	**No. of codes**	**No. of individuals**	**No. of Codes**	**No. of individuals**
Nonfunctioning kidney	181	97	43	35	9040	5118
Nasal polyps	1301	740	15879	3574	779	473
Acute pancreatitis	1809	812	5958	930	1239	628
Parkinsonism	7640	2525	30764	5922	8899	4607
Carcinoma of the eyelid	358	161	3913	476	408	197
Pulmonary fibrosis	1846	910	4181	1888	3796	2440
Nephritic syndrome	4402	2188	29428	4956	8337	3918
Pneumocystis	66	28	582	165	42	23
Preterm labor	256	172	1664	716	1832	923
Primary malignancy/ neoplasm lung	11107	2581	4958	1023	9078	2270
Vena cava thrombosis	181	90	2	1	178	91
Primary sclerosing cholangitis	2874	962	3138	1098	4582	1485
Progressive supranuclear palsy	2070	988	6618	3221	4676	2290
Atrial fibrillation	20131	8153	59429	15012	32415	10723
C1 enterase deficiency	23	18	193	119	191	118

Table 3. Percent positive (%POS) and percent negative (%NEG) agreement between number of people retrieved by each coding system for 15 diseases or conditions

Disease/Condition	HICDA/ICD-9		HICDA/SNOMED CT		ICD-9/SNOMED CT	
	%POS	%NEG	%POS	%NEG	%POS	%NEG
Nonfunctioning kidney	1.69	99.12	20.31	99.98	0.54	99.12
Nasal polyps	63.48	99.92	31.94	99.50	20.16	99.45
Acute pancreatitis	64.31	99.91	48.89	99.85	51.63	99.87
Parkinsonism	59.45	99.50	58.87	99.41	46.27	99.04
Carcinoma of the eyelid	64.58	99.98	1.22	99.89	1.50	99.89
Pulmonary fibrosis	29.96	99.60	51.71	99.77	32.65	99.51
Nephritic syndrome	52.99	99.51	49.62	99.39	48.70	99.22
Pneumocystis	65.38	100.00	29.95	99.98	20.99	99.98
Preterm labor	26.98	99.86	27.29	99.89	21.59	99.78
Primary malignancy neoplasm lung	81.91	99.85	35.51	99.61	36.29	99.64
Vena cava thrombosis	53.04	99.99	2.20	99.98	0.00	99.98
Primary sclerosing cholangitis	62.91	99.85	65.43	99.88	51.67	99.79
Progressive supranuclear palsy	49.94	99.72	46.76	99.63	60.91	99.64
Atrial fibrillation	74.99	99.17	69.83	98.78	72.96	98.78
C1 enterase deficiency	8.82	99.98	15.04	99.98	11.16	99.96
Median Agreement	59.45	99.85	35.51	99.85	32.65	99.64

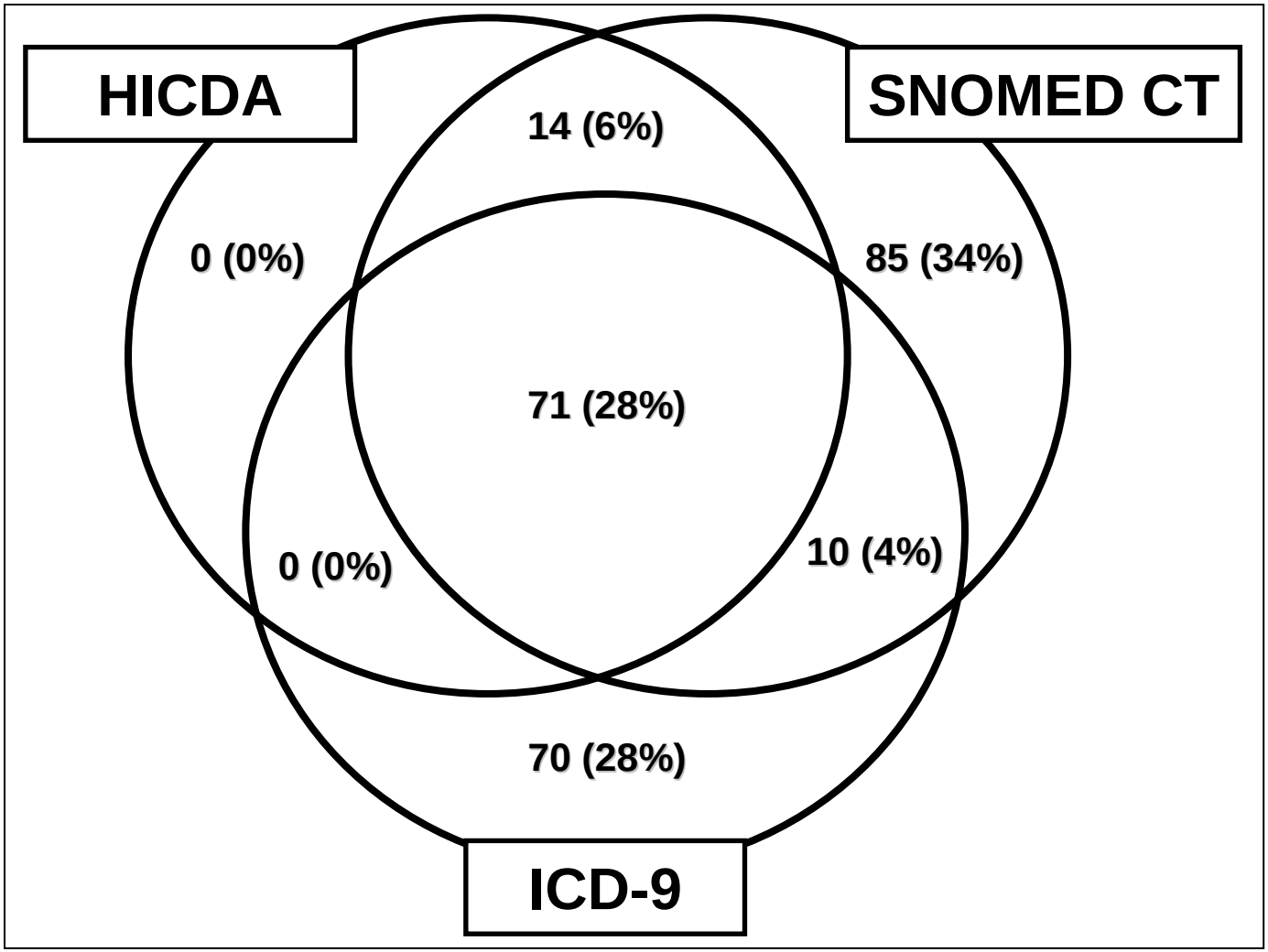
Table 4. Sensitivity and specificity of each coding system for identifying cases of Parkinsonism

Coding System	Coding System + /Hand Review +	Coding System - /Hand Review +	Coding System + /Hand Review -	Coding System - /Hand Review -	Sensitivity	Specificity
HICDA	81	19	4	8,735	81.0	99.9
ICD-9	75	25	76	8,663	75.0	99.1
SNOMED CT	100	0	80	8,659	100.0	99.1

Table 5. Reasons coding systems were incorrect

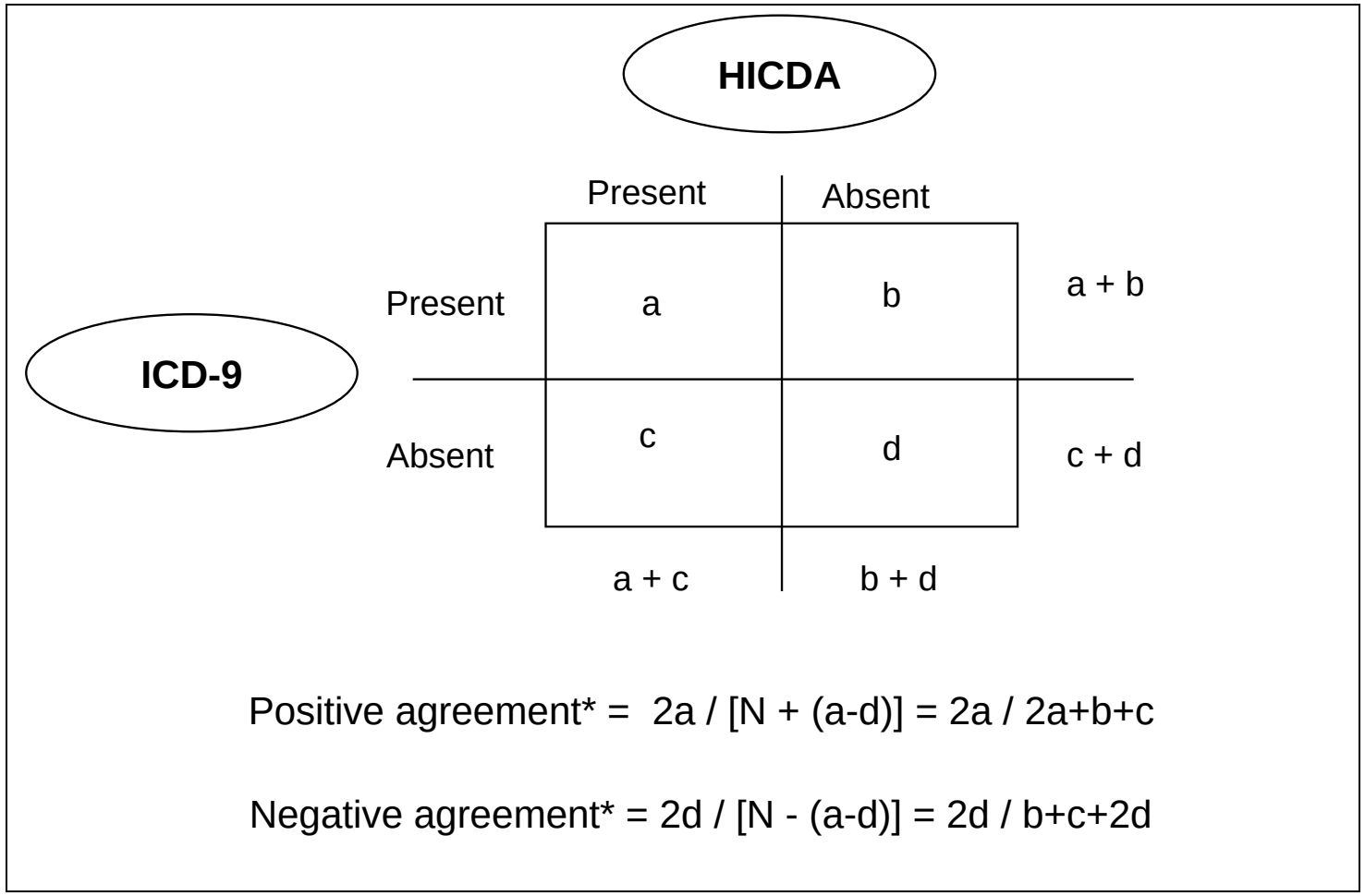
Coding System	Coding System -/Hand Review +		Coding System + /Hand Review -	
	N (%)	Reason	N (%)	Reason
HICDA	12 (63.2)	Parkinsonism was mentioned in note, but did not receive a HICDA code	4 (100)	Unclear; no mention of Parkinsonism in clinical notes
	5 (26.3)	Visit was not for Parkinsonism		
	2 (10.5)	Parkinsonism mentioned in a phone call		
ICD-9	25 (100)	Parkinsonism was mentioned in clinical note, but did not receive an ICD-9 code	68 (89.5)	Patient had restless leg syndrome (similar ICD-9 code to Parkinsonism)
			8 (9.3)	Unclear; no mention of Parkinsonism in clinical notes
SNOMED CT			41 (48.8)	Patient had a family history of Parkinsonism, or was caring for someone with condition
			33 (39.3)	Negation term in text (i.e. “rule out Parkinsonism”)
			5 (6.0)	Patient part of a Parkinson Disease research study
			1 (1.2)	Unclear; no mention of Parkinsonism in clinical notes

Figure 1. Overlap among the coding systems for identifying people with a Parkinsonism disease code*



*Some cases of Parkinsonism may have been missed by all three coding systems.
For this report, the number of missed cases was assumed to be negligible.

Figure 2. Example of how positive and negative agreement were calculated[7]



Appendix 1a. Final HICDA diagnostic codes chosen for the study

Disease/Condition	HICDA Code	Code Description
Nonfunctioning kidney	05869120	Nonfunctioning, kidney
Nasal polyps	05050-31-34	Nasal polyps
Acute pancreatitis	05770	Acute pancreatitis
Parkinsonism	03420 07732-13-3 03479-72-0-3 03479-82-1 03449-71-1-7	Paralysis agitans Actinic rigid syndrome Extrapyramidal movements Multiple systems Atrophy Supranuclear palsy
Carcinoma of the eyelid	01721 01731 34095-11-2	Melanoma eyelid Malignant neoplasm of eyelid
Pulmonary fibrosis	05170-22-0-3 05170-33-0-9 05193-21	Pneumonitis with fibrosis Chronic fibrosis of lung
Nephritic Syndrome	05800-12 thru 05800-25 05810-11 -12 05820-19-20-21-23-25-26-27 05820-31-32 05830-11 0-2-3 05830-12-14-15 05830-13 0-1 05840-22 1 05850-11-12 05851-11 05860-11-1 05861-22-0 05862-20-21-22-23-24 05901-31 05903-22 05935-31-8 02740-61-1 07598-40 07161-11-6 07531-42 02506-15 05810-24-0 05810-22-0-2 05810-14-23-25-26-31-32	Acute , Chronic and unspecified Glomerulonephritis Nephrotic syndrome Acute Renal Failure
Pneumocystis carinii pneumonia	04830-43	Pneumocystis Carinii Pneumonia
Premature labor	06345 06348 06643	Premature labor threatened or with delivery False labor Braxton Hicks
Primary malignancy/neoplasm lung	01621 34090 01792-87-88	Neoplasm malignant bronchus and lung Lung cancer morphology codes

Vena cava thrombosis	04530-33-0	Vena cava thrombosis
Primary sclerosing cholangitis	05760-31 05761-11-21-31 05762-11-12-21-22 05762-20-0	Sclerosing cholangitis Common bile duct stenosis obstruction Bile duct obstruction, stenosis
Progressive supranuclear palsy	03449-71-1-5-6-7 03420-11-4 03420-12-0 03420-60-4	Supranuclear palsy Parkinsonism Parkinson's plus
Atrial fibrillation	04163	Atrial fibrillation
C1 enterase deficiency	02751-11-2 07080-11	C1 enterase deficiency Angioneurotic edema

Appendix 1b. Final ICD-9 diagnostic codes chosen for the study

Disease/Condition	ICD-9CM Code	Code Description
Nonfunctioning kidney	593.9	Unspecified disorder of kidney and ureter
Nasal polyps	471.0 471.9	Polyp of nasal cavity (includes choanal) Unspecified nasal polyp
Acute pancreatitis	577.0	Acute pancreatitis
Parkinsonism	332.0 333.0 332.1 331.82 333.90 333.99	Paralysis agitans Other degenerative diseases of the basal ganglia Secondary Parkinsonism Dementia with Lewy bodies Unspecified extrapyramidal disease and abnormal movement disorder Other extrapyramidal diseases and abnormal movement disorders
Carcinoma of the eyelid	173.1 172.1	Malignant neoplasm skin of eyelid, including canthus Malignant melanoma of skin of Eyelid, including canthus
Pulmonary fibrosis	515	Postinflammatory pulmonary fibrosis
Nephritic syndrome	580- 581- 583- 584-	Acute glomerulonephritis Nephrotic syndrome Nephritis and nephropathy, not specified as acute or chronic Acute renal failure
Pneumocystis carinii pneumonia without HIV infect	136.3	Pneumocystosis
Preterm labor	644	Early onset of delivery
Primary malignancy/neoplasm lung	162.2 162.3 162.4 162.5 162.8 162.9	Malignant neoplasm of main bronchus Malignant neoplasm of upper lobe, bronchus or lung Malignant neoplasm of middle lobe, bronchus or lung 55 Malignant neoplasm of lower lobe, bronchus or lung Other parts of bronchus or lung Malignant neoplasm of contiguous or overlapping sites of bronchus or lung whose point of origin cannot be determined Bronchus and lung, unspecified
Vena cava thrombosis	453.2	Other venous embolism and thrombosis of inferior vena cava
Primary sclerosing cholangitis	576.1 576.2 576.8	Biliary cirrhosis Obstruction of bile duct Other specified disorders of biliary tract

Progressive supranuclear palsy	333.0 332.0	Other degenerative diseases of the basal ganglia Paralysis agitans
Atrial fibrillation	427.31	Atrial fibrillation
C1 enterase deficiency	277.6	Other deficiencies of circulating enzymes Hereditary angioedema

Appendix 1c. Final SNOMED CT diagnostic codes chosen for the study

Disease/Condition	SNOMED CT Concept ID	Description
Nonfunctioning kidney	C0232808	Absent renal function
Nasal polyps	C0027430	Nasal Polyps
Acute pancreatitis	C0001339	Acute pancreatitis unspecified
Parkinsonism	C0242422 C0030567 C0393571 C0038868 C0270729 C0752347 C0015371	Parkinsonian Disorders Parkinson Disease Multiple System Atrophy Progressive supranuclear palsy Parkinsonism due to drug Lewy Body Disease Extrapyramidal Disorders
Carcinoma of the eyelid	C0339111 C0346725 C1828015 C0339114 C0339113 C0339115 C0339116 C0007129 C0149722	Malignant tumor of eyelid Malignant neoplasm of eyelid including canthus Malignant neoplasm of skin of eyelid Basal cell carcinoma of eyelid Squamous cell carcinoma of eyelid Sebaceous adenocarcinoma of eyelid Malignant melanoma of eyelid Merkel cell carcinoma Hutchinson`s Melanotic Freckle
Pulmonary fibrosis	C0034069	Pulmonary Fibrosis
Nephritic Syndrome	C0017658 C0156221 C1263744 C0027697 C0022660	Glomerulonephritis Acute glomerulonephritis NOS Subacute glomerulonephritis Nephritis Kidney Failure, Acute
Pneumocystis carinii pneumonia (without hiv infect)	C0032305	Pneumonia, Pneumocystis carinii

Preterm labor	C0022876 C0156718 C0015944 C0085598 C0233187 C0565404	Premature Obstetric Labor Early onset of delivery, unspecified as to episode of care Fetal Membranes, Premature Rupture False labor False uterine contraction Premature/false labor NOS
Primary malignancy/neoplasm lung	C1306460 C0007121 C0149925 C0153491 C0346602 C0153492 C0346604 C0024624 C0346600	Primary malignant neoplasm of lung Bronchogenic Carcinoma Small cell carcinoma of lung Malignant neoplasm of middle lobe, bronchus or lung NOS Malignant neoplasm of middle lobe of lung Malignant neoplasm of lower lobe, bronchus or lung NOS Malignant neoplasm of lower lobe of lung Malignant neoplasm of upper lobe, bronchus or lung NOS Malignant neoplasm of upper lobe of lung
Vena cava thrombosis	C0265050 C0549289 C0235513	Thrombosis of vena cava Thrombosis of inferior vena cava Thrombophlebitis of vena cava
Primary sclerosing cholangitis	C0566602 C0597984 C0520571	Primary sclerosing cholangitis Biliary stricture Fibrosis of bile duct
Progressive supranuclear palsy	C0038868 C0242422	Progressive supranuclear palsy Parkinsonian Disorders
Atrial fibrillation	C0004238 C0038454	Atrial fibrillation Cerebrovascular Disease accident
C1 enterase deficiency	C0019243	Hereditary angioneurotic edema