A Randomized Controlled Trial of the Accuracy of Clinical Record Retrieval using SNOMED-RT as Compared with ICD9-CM

Peter L. Elkin, MD, Alexander P. Ruggieri, MD, Steven H. Brown, MD, James Buntrock, Brent A. Bauer, MD, Dietlind Wahner-Roedler, MD, Scott C. Litin, MD, Julie Beinborn, Kent R. Bailey, PhD, Larry Bergstrom, MD

Abstract:

Background: *Concept-based Indexing is purported to provide more granular data representation for clinical records[1,2] This implies that a detailed clinical terminology should be able to provide improved access to clinical records. To date there is no data to show that a clinical reference terminology is superior to a precoordinated terminology in its ability to provide access to the clinical record. Today, ICD9-CM is the most commonly used method of retrieving clinical records.*

Objective: *In this study, we compare the sensitivity, specificity, positive likelihood ratio, positive predictive value and accuracy of SNOMED-RT vs. ICD9-CM in retrieving ten diagnoses from a random sample of 2,022 episodes of care.*

Method: *We randomly selected 1,014 episodes of care from the inpatient setting and 1,008 episodes of care from the outpatient setting. Each record had associated with it, the free text final diagnoses from the Master Sheet Index at the Mayo Clinic and the ICD9-CM codes used to bill for the encounters within the episode of care. The free text diagnoses were coded by two expert indexers (disagreements were addressed by a Staff Clinician) as to whether queries regarding one of 5 common or 5 uncommon diagnoses should return this encounter. The free text entries were automatically coded using the Mayo Vocabulary Processor. Each of the ten diagnoses was exploded in both SNOMED-RT and ICD9-CM and using these entry points, a retrieval set was generated from the underlying corpus of records. Each retrieval set was compared with the Gold Standard created by the expert indexers.*

Results: *SNOMED-RT produced significantly greater specificity in its retrieval sets (99.8% vs. 98.3%, p<0.001 McNemar Test). The positive likelihood ratios were significantly better for SNOMED-RT retrieval sets (264.9 vs. 33.8, p<0.001 McNemar Test). The positive predictive value of a SNOMED-RT retrieval was also significantly better than ICD9-CM (92.9% vs. 62.4%, p<0.001 McNemar Test). The accuracy defined as 1 – (the total error rate (FP+FN) / Total # episodes queried (20,220)) was significantly greater for SNOMED-RT (98.2% vs. 96.8%, p=0.002 McNemar Test). Interestingly, the sensitivity of the SNOMED-RT generated retrieval set was not significantly different from ICD9-CM, but there was a trend toward significance (60.4% vs. 57.6%, p=0.067 McNemar Test). However, if we examine only the outpatient practice SNOMED-RT produced a more sensitive retrieval set than ICD9-CM (54.8% vs. 46.4%, p=0.002 McNemar Test).*

Conclusions: *Our data clearly shows that information regarding both common and rare disorders is more accurately identified with automated SNOMED-RT indexing using the Mayo Vocabulary Processor than it is with traditional hand picked constellations of codes using ICD9-CM. SNOMED-RT provided more sensitive retrievals of outpatient episodes of care than ICD9-CM.*

Introduction:

ICD9-CM is the current standard for billing and morbidity indexing in the United States.[3] This reporting responsibility makes the assignment of ICD9-CM codes a necessity for most or all healthcare organizations. The availability of this coded data has lead to the use of ICD9-CM codes for research involving the clinical record. The availability of clinical reference terminologies such as SNOMED-RT, offers the potential to provide much more granular coding of the data found within the clinical record.[4,5] This advantage is achieved by compositional encoding of health concepts.[6] This means that a clinician or system can put together multiple concepts in order to form a more detailed clinical expression.[8,8] An example of this is, "Cellulitis of the Left foot with Osteomyelitis of the third Metatarsal without Lymphangitis." This notion is represented by an expression containing six concepts within SNOMED-RT. The ability to record arbitrary complex structure within a clinical reference terminology holds the promise to significantly address the problem of content completeness.[9]

1067—5027/01/$5.00 © 2001 AMIA, Inc.

Vocabularies such as SNOMED-RT, the UMLS or one of several other vocabulary efforts have common characteristics that are helpful to understand within the scope of Mayo Vocabulary Processor (MVP).[2,9,10,11] A vocabulary is essentially a set of concepts that are identified by a unique identifier and described by terms and relationships. For each concept, there exists one or more terms that belong to that concept. For example, in SNOMED-RT the concept "Myocardial Infarction" is identified by a code and contains the terms "heart attack," "infarction of heart," and "cardiac infarction." The concept is also described by its relationships to other concepts. These relationships commonly include hierarchies built on parent/child relationships but often extend to many other types of relationships such as morphology, topology and etiology.

Theoretically, this should lead to improved information retrieval of relevant records when presented with a clinical query. We have designed this experiment to test the hypothesis that the use of SNOMED-RT will lead to more specific information retrieval of clinical records when compared with ICD9-CM.

Method

The Mayo Clinic has a long history of exacting manual record keeping. The Master Sheet Index relates to Mayo's view of a clinical episode. At Mayo, an episode of care is defined at the divisional level and is related to the notion that there is a point in the time course of caring for a particular patient when the case is stable and is therefore no longer in flux. Often this is after the diagnoses have been established, or after the major interventions have been accomplished. At this point in time, the Primary Physician caring for the patient is required to list the "Final" Diagnoses for this episode of care. This requires the physician to perform the actions of Filtering, Subsumption, and Prioritization. Filtering would include eliminating diagnoses, originally contained in one's differential diagnosis, that were ruled out during this episode. Subsumption allows the clinician to state that the patient's presenting problem of Chest Pain was in fact due to Atherosclerotic Coronary Artery Disease (CAD). Prioritization relates to the notion that although the patient's problem list may in fact be very long, the clinician may have only dealt with a subset of the patient's problems during this episode. For example, our patient with Chest Pain may also have Seborrheic Keratosis, which we may not have had occasion to discuss during the work up and treatment of this patient's CAD.

Mayo Vocabulary Processor

The Mayo Vocabulary Processor (MVP) is a set of tools that facilitate medical vocabulary indexing. This allows the mapping of free text into coded data that is both expressive and comparable for clinical, educational and research purposes. There are two main divisions within the set of tools corresponding to the server side and the client side. On the server side, there are vocabulary services that contain the preprocessing of the underlying reference vocabulary. This application is designed to be platform and programming language neutral. On the client side, there are GUI building blocks for searching and navigating the underlying vocabulary, building complex coded expressions (Using our Automated Term Composition[6] and Dissection[12] Methods), and maintaining a personal list of commonly used coded expressions.

Decision Support System

The Decision Support System (DSS) is a software and data base management system that aggregates coded clinical and financial data from various sources within Mayo Foundation for both inpatient and outpatient care. DSS captures ICD9-CM codes for each episode of care, which occurs within the Mayo Foundation. DSS allows an analyst to directly access and model the information to support clinical and financial decision-making.

Study Design

We randomly selected 1,050 episodes of care from the inpatient setting and 1,050 episodes of care from the Outpatient setting (note: all episodes were from year 2000). This yielded 1,014 inpatient and 1,008 outpatient unique episodes of care, which contained DSS assigned ICD9-CM codes. Each record had associated with it the free text final diagnoses from the Master Sheet Index at the Mayo Clinic and the ICD9 codes used to bill for the encounters within the episode of care. The free text diagnoses were coded by two expert indexers (disagreements were addressed by a Staff Clinician, 607 out of 2022 (30%) episodes were reviewed in this fashion) as to whether queries regarding one of 5 common or 5 uncommon diagnoses should return this encounter. The five most common diagnoses at Saint Mary's Hospital (Mayo's largest teaching Hospital) were Congestive Heart Failure, Coronary Artery Disease, Chronic Obstructive Pulmonary Disease, Pneumonia and Atrial Fibrillation. The five uncommon (rare) diagnoses were number 50 through 53 on the same frequency list of Master Sheet entries (Ulcerative

Colitis, Restrictive Lung Disease, Vasculitis, Nephrotic Syndrome, Histoplasmosis). The free text entries were automatically coded using the Mayo Vocabulary Processor. Each of the ten diagnoses was exploded in both SNOMED-RT and ICD9-CM and using these entry points, a retrieval set was generated from the underlying corpus of records. For SNOMED-RT, all concepts associated with the primary mapping by either an "isa" relationship or a co-occurrences of the "Assoc-Morph" and "Assoc-Topo" concept relations were included in the retrieval set. The ICD9-CM explosions were performed by hand by an expert nosologist with over 20 years of experience with formulating ICD9-CM queries. Each retrieval set was compared with the Gold Standard created by the expert indexers (See Figure #1). Formal statistical comparison of sensitivity, specificity, and accuracy was performed by use of McNemar's test applied to the 2-way tables SNOMED by ICD9, within the sets of cases with or without each diagnosis (by the gold standard).

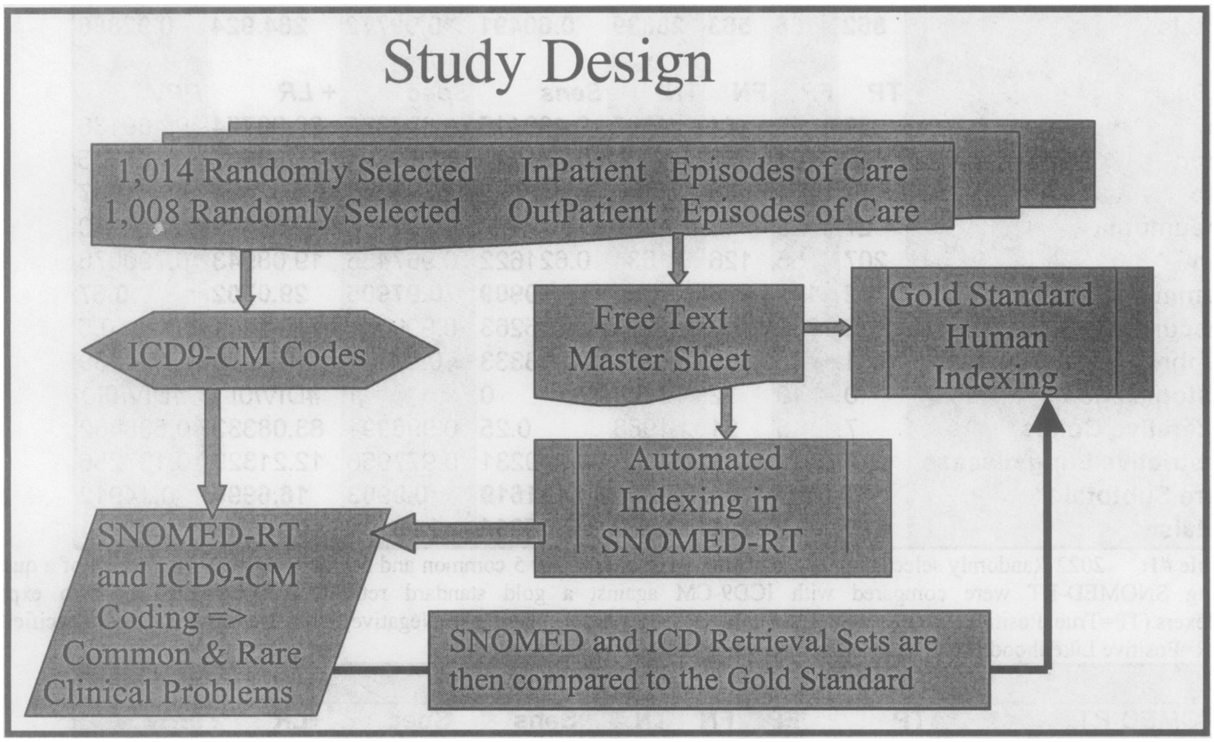

Figure #1 The free text master sheet entries were hand indexed by professional medical indexers and this was compared to the retrieval sets generated using SNOMED-RT and ICD9-CM respectively.

Results

SNOMED-RT provided significantly greater specificity in its retrieval sets (99.8% vs. 98.3%, $p<0.001$ McNemar Test), (See Table #1). The positive likelihood ratios were significantly better for SNOMED-RT retrieval sets (264.9 vs. 33.8, $p<0.001$ McNemar Test). The positive predictive value of a SNOMED-RT retrieval was also significantly better than ICD9-CM (92.9% vs. 62.4%, $p<0.001$ McNemar Test). The accuracy defined as 1 – (the total error rate (FP+FN) / Total # episodes queried (20,220)) was significantly greater for SNOMED-RT (98.2% vs. 96.8%, $p=0.002$ McNemar Test). Interestingly, the sensitivity of the SNOMED-RT generated retrieval set was not significantly different from ICD9-CM, but there was a trend toward significance (60.4% vs. 57.6%, $p=0.067$ McNemar Test). However, if we examine only the outpatient practice SNOMED-RT produced a more sensitive retrieval set than ICD9-CM (54.8% vs. 46.4%, $p=0.002$ McNemar Test) (See Table #2).

Retrievals of common diagnoses were more sensitive but not more specific than rare diagnoses and overall had higher positive likelihood ratios and positive predictive values.

SNOMED-RT	TP	FP	FN	TN	Sens	Spec	+ LR	PPV
chf	39	3	43	1937	0.47561	0.998454	307.561	0.928571
copd	60	5	36	1921	0.625	0.997404	240.75	0.923077
afib	75	3	29	1915	0.721154	0.998436	461.0577	0.961538
pneumonia	26	7	19	1970	0.577778	0.996459	163.181	0.787879
cad	220	10	113	1679	0.660661	0.994079	111.5856	0.956522
Common Subtotal	**420**	**28**	**240**	**9422**	**0.63636**	**0.99704**	**214.773**	**0.9375**
vasculitis	9	7	10	1996	0.473684	0.996505	135.5414	0.5625
Nephrotic Syndrome	1	0	29	1992	0.033333	1	#DIV/0!	1
Histoplasmosis	0	0	2	2020	0	1	#DIV/0!	#DIV/0!
Ulcerative Colitis	11	3	17	1991	0.392857	0.998495	261.119	0.785714
Restrictive Lung disease	1	0	25	1996	0.038462	1	#DIV/0!	1
Rare Subtotal	**22**	**10**	**83**	**9995**	**0.20952**	**0.999**	**209.629**	**0.6875**
Totals	**862**	**66**	**563**	**28839**	**0.60491**	**0.99772**	**264.924**	**0.92888**
ICD9	TP	FP	FN	TN	Sens	Spec	+ LR	PPV
chf	38	43	44	1897	0.463415	0.977835	20.90754	0.469136
copd	59	21	37	1905	0.614583	0.989097	56.36607	0.7375
afib	71	51	33	1867	0.682692	0.97341	25.67459	0.581967
pneumonia	27	28	18	1949	0.6	0.985837	42.36429	0.490909
cad	207	55	126	1634	0.621622	0.967436	19.08943	0.790076
Common Subtotal	**402**	**198**	**258**	**9252**	**0.60909**	**0.97905**	**29.0702**	**0.67**
vasculitis	2	2	17	2001	0.105263	0.999001	105.4211	0.5
Nephrotic Syndrome	1	45	29	1947	0.033333	0.97741	1.475556	0.021739
Histoplasmosis	0	0	2	2020	0	1	#DIV/0!	#DIV/0!
Ulcerative Colitis	7	6	21	1988	0.25	0.996991	83.08333	0.538462
Restrictive Lung disease	7	44	19	1952	0.269231	0.977956	12.21329	0.137255
Rare Subtotal	**17**	**97**	**88**	**9908**	**0.1619**	**0.9903**	**16.6996**	**0.14912**
Totals	**821**	**493**	**604**	**28412**	**0.57614**	**0.98294**	**33.7796**	**0.62481**

Table #1: 2022 Randomly selected episodes of care were queried for 5 common and 5 rare disorders. The results of a query using SNOMED-RT were compared with ICD9-CM against a gold standard retrieval set compiled by two expert indexers.(TP=True Positive, TN=True Negative, FP=False Positive, FN=False Negative, Sens=Sensitivity, Spec=Specificity, +LR=Positive Likelihood Ratio, and PPV=Positive Predictive Value).

	Inpatient							
SNOMED-RT	TP	FP	FN	TN	Sens	Spec	+LR	PPV
Common Subtotal	284	22	144	4620	0.66355	0.99526	140.009	0.9281
Rare Subtotal	16	9	46	4999	0.25806	0.9982	143.599	0.64
Totals	**300**	**40**	**190**	**9619**	**0.61224**	**0.99586**	**147.842**	**0.88235**
ICD9-CM	TP	FP	FN	TN	Sens	Spec	+LR	PPV
Common Subtotal	286	159	142	4483	0.66822	0.96575	19.5088	0.6427
Rare Subtotal	14	61	48	4947	0.22581	0.98782	18.5383	0.18667
Totals	**300**	**281**	**190**	**9430**	**0.61224**	**0.97106**	**21.1584**	**0.51635**
	Out-patient							
SNOMED-RT	TP	FP	FN	TN	Sens	Spec	+LR	PPV
Common Subtotal	136	6	96	4802	0.58621	0.99875	469.747	0.95775
Rare Subtotal	6	1	37	4996	0.13953	0.9998	697.256	0.85714
Totals	**142**	**8**	**133**	**9798**	**0.51636**	**0.99918**	**632.933**	**0.94667**
ICD9-CM	TP	FP	FN	TN	Sens	Spec	+LR	PPV
Common Subtotal	116	39	116	4769	0.5	0.99189	61.641	0.74839
Rare Subtotal	3	36	40	4961	0.06977	0.9928	9.68411	0.07692
Totals	**119**	**111**	**156**	**9730**	**0.43273**	**0.98872**	**38.3646**	**0.51739**

Table #2: Summary Data for both the inpatient and the outpatient subsets are presented.

Discussion

Clinical terminologies that provide clinicians and health systems with the ability to code their clinical data using a robust reference terminology, which support compositional expressions, can significantly improve the accuracy and specificity of information retrieval as compared with the current indexing provided by ICD9-CM.[1] This improvement in the specificity of clinical retrieval sets represents a distinct advantage to the organization adopting this technology. Aggregation and retrieval of data for research, education, improving the practice and administrating the practice of medicine will become easier and more accurate. For outpatient episodes, SNOMED-RT provided more sensitive retrievals than ICD9-CM. The authors suspect that more rigorous assignment of ICD9-CM coding in the inpatient setting may account for the lack of a significant difference in the sensitivities of the retrieval sets, also up to 9 diagnoses are codeable from a hospitalization while only 3 are allowed for an outpatient episode. SNOMED would obviously show a greater advantage for concepts whose accurate representation requires a compositional approach. The greater availability of high quality information will in some, perhaps many, cases be the defining competitive advantage that one healthcare organization has over its competitors in the healthcare marketplace.

The overall low sensitivity of SNOMED-RT for retrieving clinically relevant episodes of care for both common and rare disorders highlights the need for the continued development of more robust relationships, both hierarchical and non-hierarchical (e.g. "Diastolic Dysfunction" in SNOMED-RT is not listed as a type or etiology of "CHF" whereas clinically CHF secondary to diastolic dysfunction is one of the important distinctions in the diagnosis of chf). Of note, the description logic provided no benefit in terms of increased retrievals from this dataset. Some of the false negative retrievals stemmed from the cases where the diagnosis was discussed but not mentioned. Other failures were due to the co-occurrence of concepts, which to the indexers implied a high probability that the episode was relevant.

Our data clearly shows that information regarding both common and rare disorders is more accurately identified with automated SNOMED-RT indexing using the Mayo Vocabulary Processor than it is with traditional hand picked constellations of codes using ICD9-CM. SNOMED-RT provided more sensitive retrievals of outpatient episodes of care than ICD9-CM.

References:

1. Campbell KE, Musen MA. Representation of Clinical Data Using SNOMED III and Conceptual Graphs. Symposium on Computer Applications in Medical Care 1992;16:354-358.
2. Elkin PL, et al; "Standard Problem List Generation, Utilizing the Mayo Canonical Vocabulary Embedded within the Unified Medical Language System"; JAMIA Suppl 1997, p. 500-504.
3. http://www.cdc.gov/nchs/about/otheract/icd9/abt icd9.htm
4. Spackman KA, Campbell KE. Compositional Concept Representation Using SNOMED: Towards Further Convergence of Clinical Terminologies. JAMIA 1998;Symp. Suppl(1998):740-744.
5. Cote R, Rothwell D, Palotay J, Beckett R, Brochu L, eds. The Systemized Nomenclature of Human and Veterinary Medicine: SNOMED International. Northfield: College of American Pathologists; 1993.
6. Elkin PL, et al.. A Randomized Controlled Trial of Automated Term Composition. JAMIA 1998;SympSuppl:765-769.
7. Elkin P, et al. Medical Knowledge Representation: From a Clinically Derived Terminology to Understanding. In: Thirteenth Symposium: TEPR 1997.
8. Elkin PL, Harris M, Brown SH, et al; "Semantic Augmentation of Description Logic Based Terminologies", IMIA WG6 Symp. 2000.
9. Bousquet c, et al. Using Semantic Distance for the Efficient Coding of Medical Concepts; JAMIA Suppl. 2000.
10. Chute C, Elkin P. A Clinically Derived Lexicon: Qualifications to Reduction. JAMIA 1997:570-4.
11. Elkin PL, Tuttle MS, et al. The Role of Compositionality in Standardized Problem List Generation. In: Cesnik B, McCray AT, Scherrer J-R, ed. Ninth World Congress on Medical Informatics; IOS Press; 1998. p. 660-664.
12. PL Elkin, BA Bauer, et al; "A Randomized Double-Blind Controlled Trial of Automated Term Dissection", JAMIA Suppl. 1999

Acknowledgements

This work has been supported, in part, by a Mayo: Clinician Engaged in Education Grant and the Mayo Clinic's Department of Internal Medicine. The authors wish to thank Dawn Bergen and Philip Ogren for their technical support of this project.