MEDINFO 2010
C. Safran et al. (Eds.)
IOS Press, 2010

doi:10.3233/978-1-60750-588-4-1100

Visualization of disease distribution with SNOMED CT and ICD-10

Mikael Nyström[a], Anna Vikström[b], Gunnar Nilsson[b], Håkan Örman[a], Hans Åhlfeldt[a]

[a] *Department of Biomedical Engineering, Linköping University, Sweden*
[b] *Center for Family and Community Medicine, Karolinska institutet, Sweden*

Abstract

Methods for presentation of disease and health problem distribution in a health care environment rely among other things on the inherent structure of the controlled terminology used for coding. In the present study, this aspect is explored with a focus on ICD-10 and SNOMED CT. The distribution of 2,5 million diagnostic codes from primary health care in the Stockholm region is presented and analyzed through the "lenses" of ICD-10 and SNOMED CT. The patient encounters, originally coded with a reduced set of ICD-10 codes used in primary health care in Sweden, were mapped to SNOMED CT concepts through a mapping table. The method used for utilizing the richer structure of SNOMED CT as compared to ICD-10 is presented, together with examples of produced disease distributions. Implications of the proposed method for enriching a traditional classification such as ICD-10 through mappings to SNOMED CT are discussed.

Keywords:

Visualization, Disease distribution, Health problems, ICD-10, SNOMED CT, Terminology models

Introduction

Methods for visualization of disease distribution in a health care environment rely on the structure of the controlled terminology used for coding. In this study this is explored with a focus on the use of ICD-10 and SNOMED CT.

ICD-10 [1] is primarily intended for statistical reporting and administrative tasks such as disease monitoring and quality assurance. Although neither based on nor intended as a model of health problems, the ICD classifications are by far the most used terminology systems in electronic health records [2]. The grouping of diseases in the ICD classifications, and still reflected in ICD-10, is done based on categorization of health problems into epidemic diseases, constitutional or general diseases, local diseases arranged by site, developmental diseases and injuries [3].

While ICD-10 is mono-hierarchical, SNOMED CT is poly-hierarchical, allowing one concept to have multiple parents [4]. SNOMED CT is also more fine-granular as compared to ICD-10. The SNOMED CT hierarchy *Clinical findings* contain approximately 110,000 concepts as compared to 12,000 disease categories in ICD-10. SNOMED CT is a clinical terminology intended for clinical documentation and reporting [4], and could be used as the basis for coding and aggregation in Electronic Health Records (EHRs). However, there is still a need to explore the potential of SNOMED CT in EHRs and other clinical information systems.

Use of Electronic Health Records (EHRs) by general practitioners is almost universal in Sweden [5]. EHRs also support diagnostic coding, which is mandatory, and has made it possible to systematically collect information on health problems [6]. In Swedish primary health care practice there is a widespread tradition of using a primary care version of ICD-10 [7] with the Swedish abbreviation KSH97-P. Sweden, being a member of the IHTSDO-organization [8], is in the process of introducing SNOMED CT as a terminological resource in the health care sector.

The primary objective of this study was to explore the use of SNOMED CT as a mean to enrich visualization of disease distribution through mapping from KSH97-P to SNOMED CT. A second objective was to analyze the distribution of health problems from a large primary health care database and to describe and compare the results from using the KSH97-P/ICD-10 versus the SNOMED CT structure.

Materials and Methods

Diagnostic data

The diagnostic data used in this study were coded by primary care physicians and collected from Electronic Health Records (EHRs) in Stockholm County during 2006, in all approximately 2,5 million encounters. Diagnostic codes were reported in an average of 78% of the encounters. Up to 15 diagnostic codes were allowed for each care contact. 82% of all care contacts had one (1) code, 13% of all care contacts had two (2) codes, and 2% of the contacts had three or more diagnostic codes.

Coding systems

The Swedish National Board of Health and Welfare has worked out a primary health care version of ICD-10 with the Swedish abbreviation KSH97-P [7]. KSH97-P contains 972 categories which relate to diseases and health related problems

that are common in primary health care. KSH97-P has the same chapter division as ICD-10. The exceptions are that chapter XX, External causes of morbidity and mortality is left out from KSH97-P and chapter XXII, Codes for special purposes, is left out in both the Swedish version of ICD-10 and KSH97-P.

SNOMED CT is intended both for clinical documentation and reporting [4]. SNOMED CT consists of concepts, descriptions and relationships. Relationships link concepts to each other and relationships are of different relationship types. The generic relationship type "Is a" relates from subtypes to supertypes and is always defining relationships. All concepts, except for the root concept, have at least one "Is a" relation to a supertype concept. The other relationship types that are defining relationships are the defining attribute relationships. The defining relationships logically represent a concept by establishing relationships between the concepts.

Mapping between KSH97-P and SNOMED CT

We used a category mapping from KSH97-P to SNOMED CT that was based on a previous mapping reliability study [9]. Of the 972 categories in KSH97-P, 14 (1%) did not have a matched concept in SNOMED CT and 67 (7%) were mapped to more than one concept. We applied an additional mapping on an ICD-10 chapter level, described in another study [10]. Two ICD-10 chapters (Symptoms, signs, abnormal clinical and laboratory findings, not elsewhere classified (XVIII) and Factors influencing health status and contact with health services (XXI)) had no direct mapping match to SNOMED CT concepts.

Computational methods

We used the mapping results of categories and chapters to aggregate the diagnostic data through SNOMED CT "Is a" relationships to describe the data and make comparisons between ICD-10 and SNOMED CT on the chapter level. For each chapter we extracted the mapped chapter concept(s) together with the mapped concepts' "Is a" descendants. All diagnoses belonging to a category that were mapped to any of the extracted concepts were assumed to belong to the specific chapter. This implies that a specific category could belong to zero, one or more chapters.

To explore complementary ways of aggregating diagnoses, we used the category mapping. We carried out aggregations through the defining "Is a" relationships and defining attribute relationships. For each category the mapped concepts were extracted together with their ancestors (all supertypes) to a mapped set. All defining attribute relationships from concepts in the mapped set were then followed, and the target concepts were included in a specific attribute value set for each relationship type. In each attribute value set the concepts that were ancestors of another concept in the same attribute value set were removed. The remaining concepts in each attribute value set were assumed to be attribute values of the respective attribute types in the category.

The computational methods are performed in a relational database management system (PostgreSQL).

Results

A chapter level comparison of the diagnostic data between KSH97-P/ICD-10 and SNOMED CT is presented in Figure 1. The frequency distribution differs somewhat between ICD-10 and SNOMED CT due to the poly-hierarchical structure of SNOMED CT, allowing one ICD-10 category to be mapped to more than one SNOMED CT concept. This means that each category may belong to zero, one or more new chapter(s), which is shown in Table 1.

The use of "Is a" relationships in SNOMED CT aggregated the diagnostic data to 2861 concepts, showing a new, multidimensional view of different medical aspects, where every view can be further explored. A subset of such concept views is shown in Table 2, with a diagnosis percentage share cut-off at 10 %. Examples of new concept views below the 10 % cut-off are "Neurological finding" (8.0%), "Acute disease" (9.5%), "Pain/sensation finding" (6.6%), and Acute inflammatory disease (5.8 %). Aggregations into high-level concepts such as Clinical finding or Disorder by body site are not presented.

The attribute relationships in SNOMED CT can also be used for creating new views of the diagnostic data. Unlike the "Is a" relationship showing disorders concepts in Table 2, Table 3 could be used for further analysis of specific perspectives of e.g. causative agents, finding sites or associated morphologies for disorders and clinical findings.

Discussion

The results of this article show hidden information about health problems and diagnoses, coded with KSH97-P/ICD-10. The multiple views that are explored in this study illustrate the advantage of a poly-hierarchy. The multiple views of diagnostic data offer new possibilities for follow-up of specific aspects of the disease panorama. Clinically relevant views could also be used for navigation in order to support classification, thereby possibly improving coding validity and reliability.

The results presented rely on the soundness of the chapter and category mapping between KSH97-P and SNOMED CT which have been developed by the research group [9, 10]. Although rigorously developed, these mapping tables need to be further validated in an international research context. Another limitation is that the impact of possible quality errors in the diagnostic coding process that have been shown in previous studies is unknown [5, 6, 11, 12].

The results are also dependent on the soundness of the SNOMED CT concept model. Analysis and further refinement of the SNOMED CT model are beyond the scope of this study, but is obviously a topic for further research.

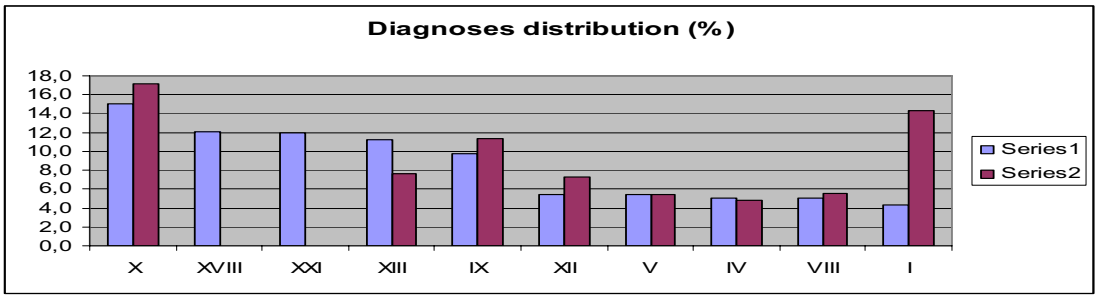

Figure 1- Distribution of diagnoses in top-ten ICD-10 chapters (left bars) and corresponding SNOMED CT concepts (right bars). (The table is truncated at a cut-off level of 5 %).

Table 1- Multiple chapter division of ICD-10/KSH97-P chapters. (The table is truncated due to space limitation – only chapters I-XIII are presented).

		Original chapter													
		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	All
Multiple chapter division chapter	I	98 528									184 971	19	31 953	69	316 391
	II		29 184										5 641		34 878
	III		262	12 480			78			176	49 908		2 332		67 378
	IV		135		101 187			76						4 108	105 988
	V					115 424	1 535						18		118 342
	VI	380	502		212	25 862	16 509	590	111	7 622				2 157	55 061
	VII	375	6					38 628						591	40 818
	VIII		92						120 150						120 980
	IX	3				498	7 096	85		240 451		42	138	1 064	251 984
	X	56	723							2 536	371 927				377 837
	XI	10 614	1 578		49 889					6 507	76 197	47 470	531	591	194 904
	XII	24 013	2 630					3 027					116 700		160 319
	XIII	17	72				348		35		280	2 494	189	130 869	168 815

Table 2- Aggregation of diagnostic data in new groups through the SNOMED CT "Is a" relationship.

Snomed CT concept	Percentage share (%)
Finding of head and neck region	21.9
Finding of trunk structure	18.9
Inflammatory disorder	17.9
Inflammation of specific body structures	17.7
Inflammation of specific body systems	17.5
Respiratory finding	17.3
Inflammation of specific body organs	15.6
Ear, nose and throat finding	15.6
Disorder of body cavity	15.6
Ear, nose and throat disorder	15.4
Disorder of respiratory system	15.1
Disorder of trunk	14.9
Head finding	14.3
Disorder of head	13.5
Infectious disease	12.6
Viscus structure finding	11.0
Upper respiratory tract finding	10.6
Cardiovascular finding	10.6
Disorder of upper respiratory system	10.6
General finding of abdomen	10.3

Table 3- Number of disease categories with specific attributes.

Attribute	Number
Finding site	722
Associated morphology	457
Causative agent	126
Interprets	61
Has definitional manifestation	59
Occurrence	54
Clinical course	31
Has interpretation	20
Due to	16
Associated with	14
Method	14
After	10
Finding method	10
Subject relationship context	6
Temporal context	6
Finding context	5
Has focus	4
Finding informer	3
Pathological process	2
Procedure site	2

Acknowledgement

This study was supported by grants from The Swedish Research Council and The National Board of Health and Welfare.

References

[1] Gersenovic M. The ICD family of classifications. Methods Inform Med. 1995 Mar;34(1-2):172-5.

[2] Hayrinen K, Saranto K, Nykanen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. Int J Med Inform. 2008 May;77(5):291-304.

[3] World Health Organization. International statistical classification of diseases and related health problems, tenth revision, vol. 2 [Instruction manual]. Geneva: World Health Organization; 1992

[4] International Health Terminology Standards Development Organisation. SNOMED Clinical Terms User Guide - July 2008 International Release.

[5] Nilsson GH, Strender LE, Månsson J, Åhlfeldt H, Gunnarsson G. Patients, general practitioners, diseases and health problems in urban general practice: a cross-sectional study on electronic patient records. Primary Health Care Research & Development 2008; 9:119-125

[6] Nilsson G, Ahlfeldt H, Strender LE. Textual content, health problems and diagnostic codes in electronic patient records in general practice. Scand J Prim Health Care. 2003 Mar;21(1):33-6.

[7] Socialstyrelsen. Klassifikation av sjukdomar och hälsoproblem 1997. Primärvård. Stockholm: Socialstyrelsen; 1997

[8] International Health Terminology Standards Development Organisation, www.ihtsdo.org

[9] Vikström A, Skånér Y, Strender LE, Nilsson GH. Mapping the categories of the Swedish primary health care version of ICD-10 to SNOMED CT concepts: rule development and intercoder reliability in a mapping trial. BMC Med Inform Decision Making. 2007 May 2;7:9

[10] Nyström M, Vikström A, Nilsson GH, Åhlfeldt H, Petersson H. Enriching a primary health care version of ICD-10 using SNOMED CT mapping. Submitted to J Biomed Semantics

[11] Nilsson G, Petersson H, Åhlfeldt H, Strender L-E. Evaluation of Three Swedish ICD-10 Primary Care Versions: Reliability and Ease of Use in Diagnostic Coding. Methods Inform Med. 2000;39(4-5):325-31

[12] Stausberg J, Lehmann N, Kaczmarek D, Stein M. Reliability of diagnoses coding with ICD-10. Int J Med Inform 2008 Jan;77(1):50-7

Address for correspondence

Mikael Nyström, Dept Biomedical Engineering, Linköping University, S-581 83 Linköping, Sweden.

E-mail: mikael.nystrom@liu.se