

Using SNOMED CT in combination with MedDRA for reporting signal detection and adverse drug reactions reporting

Olivier Bodenreider, MD, PhD

National Library of Medicine, Bethesda, Maryland, USA

olivier@nlm.nih.gov

Abstract

Objective: To investigate the feasibility of using SNOMED CT as an entry point for coding adverse drug reactions and map them automatically to MedDRA for reporting purposes and interoperability with legacy repositories. **Methods:** On the one hand, we attempt to map SNOMED CT concepts to MedDRA concepts through the UMLS, using synonymy and explicit mapping relations. On the other, we compute the set of all fine-grained concepts that can be reached from concepts having a mapping to MedDRA. **Results:** 58% of the Preferred Terms in MedDRA have a mapping to SNOMED CT. Through the descendants in SNOMED CT, 108,305 additional SNOMED CT concepts can be linked to MedDRA. **Conclusions:** Fine-grained SNOMED CT concepts can be mapped automatically to MedDRA. This approach has the potential to enable the collection of adverse events related to drugs directly from clinical repositories. The quality of the mapping needs to be evaluated.

Introduction

Adverse events related to drugs have traditionally been reported to regulatory agencies using controlled terminologies such as MedDRA. These reports can be used for signal detection, i.e., for identifying clusters of similar reactions related to a given drug. Controlled vocabularies such as MedDRA are crafted in such a way as to support the aggregation of cases.

However, in addition to case reporting, self-reporting and signal detection from clinical databases are important elements for pharmacovigilance. With the promise of rapid deployment of electronic health records in the US over the next few years, signal detection from clinical repositories is likely to become more important.

The terminologies used in electronic health records are clinical terminologies such as SNOMED CT. Therefore the integration of adverse events collected from clinical repositories with adverse events reported through the traditional channels will require some level of interoperability between the terminologies to which clinical repositories and legacy report-

ing databases are coded. In particular, the extent to which adverse events coded with SNOMED CT can automatically be “translated” into MedDRA for reporting and analysis purposes remains to be determined.

The objective of this study is to investigate the feasibility of using SNOMED CT as an entry point for coding adverse drug reactions and mapping them automatically to MedDRA for reporting purposes and interoperability with legacy repositories.

Background

MedDRA. The Medical Dictionary for Regulatory Activities (MedDRA) is a controlled terminology developed for reporting adverse events related to drugs to regulatory agencies [1]. MedDRA has a shallow hierarchical structure with five levels: System Organ Class (SOC), High-Level Group Term (HGLT), High-Level Term (HLT), Preferred Term (PT) and Lowest-Level Term (LLT). MedDRA is organized in 26 classes (SOCs). PTs are the main descriptors in MedDRA. Each PT is linked to at least one SOC. LLTs correspond to synonyms, lexical variants, or subtypes of the PT. In addition to hierarchical relations between terms, MedDRA also records mapping relation to other adverse reaction vocabularies (e.g., WHO-ART), but not to clinical vocabularies. All MedDRA terms are integrated in the UMLS Metathesaurus. The version of MedDRA used in this study is version 11.0 dated March 2008.

SNOMED CT is a comprehensive concept system for healthcare developed by the International Health Terminology Standard Development Organization (IHTSDO). SNOMED CT provides broad coverage of clinical medicine, including findings, diseases, and procedures, and is used in electronic medical records [2]. SNOMED CT uses description logics for its representation. Unlike MedDRA, SNOMED CT is not limited to a few levels for its hierarchies, which can span more than 10 levels. In general, SNOMED CT is finer-grained than MedDRA. All SNOMED CT concepts are integrated in the UMLS Metathesaurus. The version of SNOMED CT used in this study is dated July 31, 2008 and comprises 315,550 active concepts.

UMLS. The Unified Medical Language System[®] (UMLS[®]) is a terminology integration system developed at the National Library of Medicine. The UMLS Metathesaurus[®] integrates almost 150 biomedical vocabularies, including SNOMED CT and MedDRA. Synonymous terms from the various source vocabularies are grouped into one concept. Additionally, the Metathesaurus records the relations asserted among terms in the source vocabularies, including hierarchical, associative and mapping relations. These features make the Metathesaurus a popular resource for mapping across vocabularies. Version 2008AB of the UMLS is used in this study. This version contains approximately 1.8M concepts and 40M relations.

Related work. Interoperability issues have been investigated among terminologies for adverse events, including MedDRA and SNOMED CT, but essentially from the perspective of their structural characteristics [3]. In a series of investigations, Jaulent's group in Paris has shown the influence of a rich set of relations on the ability of a terminological system to completely and appropriately classify adverse drug reactions. In particular, they used SNOMED CT as a source of relations to enrich terminologies such as WHO-ART and MedDRA [4-7]. To our knowledge, however, the interoperability between MedDRA and SNOMED CT has not been studied from the perspective of using SNOMED CT as an entry point into MedDRA. The contribution of this study is to investigate the interoperability between these two terminologies for reporting purposes.

Methods

Since our goal is to associate SNOMED CT concepts with MedDRA concepts, this investigation can be thought of as evaluating the proportion of SNOMED CT concepts for which a path can be found to MedDRA concepts. Toward this end, we explore two major approaches. On the one hand, we attempt to map SNOMED CT concepts to MedDRA concepts through the UMLS. On the other, as SNOMED CT is finer-grained than MedDRA, we exploit the rich hierarchical structure of SNOMED CT to aggregate SNOMED CT concepts to the granularity of the corresponding MedDRA concepts.

Mapping through UMLS. SNOMED CT concepts can be mapped to MedDRA concepts, either directly (i.e., through synonymy) or through explicit mapping relations.

The *direct mapping through synonymy* leverages synonymy in the UMLS. In the Metathesaurus, synonymous terms are grouped into the same concept. Therefore, SNOMED CT terms synonymous with

MedDRA terms will share the same UMLS concept identifier. For example, the MedDRA PT term *Congenital hip deformity* [10061066] and the SNOMED CT term *Congenital deformity of hip joint* [2749000] are synonymous names for the UMLS concept C0265615.

Mapping through explicit mapping relations exploits one of the features of the UMLS, namely the fact that mapping relations asserted by some source vocabularies are recorded as relations in the Metathesaurus. We focus on those relations using the *mapped_from* and *mapped_to* relationships. It is worth noting that the mapping relations need not be specifically asserted between MedDRA and SNOMED CT terms, but can be asserted between terms from other vocabularies, with which the MedDRA and SNOMED CT terms happen to be synonymous. For example, the MedDRA PT term *Pseudomonas mallei infection* [10037136] has an explicit mapping relation to the SNOMED CT term *Glanders* [4639008] contributed by the source ICPC2ICD10ENG (mapping between the International Classification of Primary Care and the International Classification of Diseases).

Exploring descendants in SNOMED CT. SNOMED CT is finer-grained than MedDRA. Therefore, when a mapping to MedDRA is found for a given SNOMED CT term, the MedDRA term mapped to is likely to be the closest mapping for all the descendants of this SNOMED CT term. We exploit the rich hierarchical structure of SNOMED CT to compute the set of all descendants, direct or not, for each SNOMED CT term for which a mapping to MedDRA was identified. For example, the SNOMED CT term *Glaucoma associated with ocular trauma* [68241007] is mapped to the MedDRA term *Glaucoma traumatic* [10018330] through synonymy. Its three descendants are *Glaucoma due to perforating injury* [66725002], *Angle recession glaucoma* [392352004] and *Traumatic glaucoma due to birth trauma* [206248004]. None of them is mapped to any term in MedDRA. All three can be associated with the MedDRA term *Glaucoma traumatic* to which their ancestor *Glaucoma associated with ocular trauma* is associated.

Results

Mapping through UMLS. Overall, 10,852 (55.5%) of the 19,570 MedDRA terms had mappings to SNOMED terms through the UMLS. As shown in **Table 1**, a mapping is found for a higher proportion of PT terms compared to other types of terms. Inter-

mediary categories such as HLT and HGLT terms have the lowest mapping rate (below 30%).

Type	Yes	%	No	%	Total
SOC	14	53.8	12	46.2	26
HGLT	82	29.8	193	70.2	275
HLT	409	27.2	1,096	72.8	1,505
PT	10,351	58.3	7,417	41.7	17,768
Total	10,856	55.5	8,718	44.5	19,574

Table 1. Number of MedDRA terms with mapping to SNOMED CT through the UMLS (for each type of terms in MedDRA)

As illustrated in **Table 2**, the vast majority of mappings are found through synonymy in the UMLS. The mapping rate for PT terms increases slightly when LLT terms are used in addition to PT terms for identifying mappings to SNOMED CT. For example, while no direct mapping is found for the PT *Bladder squamous cell carcinoma stage unspecified* [10005081], its LLT *Bladder squamous cell carcinoma* [10005074] is mapped to *Squamous cell carcinoma of bladder* [255111004] in SNOMED CT through the UMLS concept C0279681.

Type	Syn.	%	Rel.	%	Total
SOC	14	100.0	0	0.0	14
HGLT	80	97.6	2	2.4	82
HLT	392	95.8	17	4.2	409
PT alone	9,168	96.7	316	3.3	9,484
PT / LLT	799	92.2	68	7.8	867
Total	10,453	96.3	403	3.7	10,856

Table 2. Direct mapping through synonymy in the UMLS and through explicit mapping relations in the UMLS (for each type of terms in MedDRA)

The overall mapping performance of PT terms by MedDRA system organ class (SOC) is presented in **Table 3**. The mapping rate ranges from 30.1% for *Investigations* to 83.3% for *Congenital, familial and genetic disorders*. Half of the SOCs have a mapping rate of 70% or more and only 6 SOCs have a mapping rate below 60%.

Specifically for PT terms, a total of 14,071 mappings were identified between a PT term from MedDRA and a SNOMED CT concept. One typical example is the mapping of the PT *Vagus nerve disorder* [10061403] to *Disorder of vagus nerve* [73765005]

in SNOMED CT through the UMLS concept C0152179.

From the perspective of MedDRA PT terms, a total of 9,484 PT terms are mapped to at least one SNOMED CT concept. The number of SNOMED CT concepts mapped to ranges from 1 to 21. A vast majority of PT terms map to 1 SNOMED CT concept (78%) or 2 (17%). For example, the PT *Acrophobia* [10000605] is mapped to both *Acrophobia* [58963008] and *Fear of heights* [276241001] in SNOMED CT through the UMLS concept C0233701.

From the perspective of SNOMED CT, a total of 12,843 unique SNOMED CT concepts mapped to at least one PT terms from MedDRA. 11,736 mapped to exactly one PT term, 999 to two, 96 to three, 11 to four and 1 to five PT terms. For example, the two PT terms *Gardnerella infection* [10017728] and *Vaginitis gardnerella* [10046957] are mapped to *Gardnerella vaginitis* [419468003] in SNOMED CT (of which *Gardnerella infection* is a synonym) through the UMLS concept C1622505.

The mapping of one SNOMED CT to several PT terms (or the other way around) through one or several UMLS concepts is possible as the UMLS Meta-thesaurus, MedDRA and SNOMED CT might have a slightly different notion of what a concept is. For example, the UMLS groups into one single concept (C1704214) the two PT terms *Lipogranuloma* [10049940] and *Xanthogranuloma* [10051251], as well as the following three concepts from SNOMED CT, *Lipogranuloma (disorder)* [416439000], *Lipogranuloma (morphologic abnormality)* [36279001] and *Xanthogranuloma (disorder)* [189099001].

Exploring descendants in SNOMED CT. For each SNOMED CT concept identified as a mapping for a MedDRA term, we computed the list of all its descendants in SNOMED CT by traversing the *isa* relations recursively. Among the 12,843 unique SNOMED CT concepts mapped to PT terms in MedDRA, 7,384 (57%) have at least one descendant. The number of descendants (direct or not) of these SNOMED CT concepts ranges from 1 to 17,648 (median = 6). A total of 114,709 unique SNOMED CT concepts are found in the descendants of the 7,384 concepts with mapping to MedDRA that have at least one descendant. Some of the SNOMED CT concepts mapped to directly from MedDRA PT terms are also found in the descendants of other SNOMED CT concepts. In fact, 6,404 SNOMED CT concepts are both mapped to directly and found among the descendants. Overall, a total of 108,305 additional SNOMED CT concepts can be linked to MedDRA

PT terms through the descendants of the SNOMED CT concepts to which they are mapped directly.

For example, through the mapping of the PT *Uterine cyst* [10048931] to *Cyst of uterus* [758002] in SNOMED CT through the UMLS concept C0269188, the five descendants of this SNOMED CT can also be linked to this PT. These are *Embryonic cyst of cervix* [253833001], *Nabothian follicles on cervix* [24565001], *Endocervicitis with Nabothian cyst* [198206001], *Cyst of cervix* [81956008] and *Cervicitis with Nabothian cyst* [198203009]. Of note, one of the descendants (*Cyst of cervix*) already has a direct mapping to the PT *Cervical cyst* [10008254].

Discussion

Practical implications. Overall, the mapping rate of MedDRA PT terms to SNOMED CT is limited (58.3%). That is, only 9,484 PT terms from MedDRA have a direct mapping to SNOMED CT, and only 12,843 concepts from SNOMED CT have a direct mapping to MedDRA through synonymy and explicit mapping relations in the UMLS Metathesaurus. On the other hand, due to the difference in granularity between MedDRA and SNOMED CT, while most PT terms are leaf nodes in the MedDRA hierarchy, many of the SNOMED CT concepts having a mapping to MedDRA have descendants. Through these 7,384 SNOMED CT concepts, 108,305 additional SNOMED CT concepts automatically acquire a link to some coarser term in MedDRA. The practical implication of this finding is that this approach could be used to sift through clinical databases coded with SNOMED CT and automatically aggregate fine-grained clinical findings not only to the appropriate level of granularity for reporting, but also to the terminology used for reporting. In other words, this approach leverages the structure of SNOMED CT for aggregation purposes, while the mapping between MedDRA and SNOMED is used for “translating” SNOMED CT concepts into MedDRA terms.

Limitations. Evaluating the quality of the mapping is beyond the scope of this study. As suggested by the existence of one-to-many mappings between SNOMED CT and MedDRA through the UMLS, it might be impossible to derive a high-quality mapping completely automatically for all concepts. Further research is needed involving the manual review of some mappings by domain experts to assess their quality.

Another limitation of this study is that it is disconnected from actual clinical repositories and case reporting databases. Not knowing the prevalence of the phenomena coded with the two terminologies under

investigation, it is impossible to fully evaluate the practical consequences of relatively low mapping rates (58% for PT). In fact, if the MedDRA codes for which there is no mapping in SNOMED CT are never used in practice, the absence of mapping might not be detrimental to pharmacovigilance. On the other hand, missing mappings for frequent or important concepts would preclude the use of this approach. Of note, such frequency analyses in MedDRA would also help SNOMED CT developers identify those rare manifestations that might have been overlooked in the terminology.

Conclusion

We investigated the feasibility of using SNOMED CT as an entry point for coding adverse drug reactions and mapping them automatically to MedDRA for reporting purposes and interoperability with legacy repositories. This mapping exploits features from the UMLS. From this purely quantitative study, it appears that large numbers of fine-grained SNOMED CT concepts can be mapped automatically to MedDRA. This approach has the potential to enable the collection of adverse events related to drugs directly from clinical repositories. Further research is needed to evaluate the quality of the mapping. The mapping is available upon request to the author.

Acknowledgements

This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM). We wish to thank Raffael Jovine, Stephen Evans, Julie James and Hugh Glover for stimulating discussions at the beginning of this project. A MedDRA license was provided to the author for research purposes by MedDRA MSSO / Northrop Grumman Corporation.

References

1. Giannangelo K. Healthcare code sets, clinical terminologies, and classification systems. Chicago, Ill.: American Health Information Management Association; 2006
2. Donnelly K. SNOMED-CT: The advanced terminology and coding system for eHealth. *Stud Health Technol Inform* 2006;121:279-90
3. Richesson RL, Fung KW, Krischer JP. Heterogeneous but "standard" coding systems for adverse events: Issues in achieving interoperability between apples and oranges. *Contemp Clin Trials* 2008;29(5):635-45
4. Alecu I, Bousquet C, Jaulent MC. A case report: using SNOMED CT for grouping Adverse Drug

Reactions Terms. BMC Med Inform Decis Mak 2008;8 Suppl 1:S4

5. Alecu I, Bousquet C, Mougin F, Jaulent MC. Mapping of the WHO-ART terminology on Snomed CT to improve grouping of related adverse drug reactions. Stud Health Technol Inform 2006;124:833-8
6. Bousquet C, Lagier G, Lillo-Le Louet A, Le Beller C, Venot A, Jaulent MC. Appraisal of the

MedDRA conceptual structure for describing and grouping adverse drug reactions. Drug Saf 2005;28(1):19-34

7. Henegar C, Bousquet C, Lillo-Le Louet A, Degoulet P, Jaulent MC. Building an ontology of adverse drug reactions for automated signal generation in pharmacovigilance. Comput Biol Med 2006;36(7-8):748-67

System Organ Class (SOC)	Yes	%	No	%	Total
Blood and lymphatic system disorders	450	56.7%	343	43.3%	793
Cardiac disorders	352	73.0%	130	27.0%	482
Congenital, familial and genetic disorders	858	83.3%	172	16.7%	1,030
Ear and labyrinth disorders	132	82.0%	29	18.0%	161
Endocrine disorders	302	76.8%	91	23.2%	393
Eye disorders	628	79.4%	163	20.6%	791
Gastrointestinal disorders	923	69.5%	405	30.5%	1,328
General disorders and administration site conditions	250	43.1%	330	56.9%	580
Hepatobiliary disorders	227	67.6%	109	32.4%	336
Immune system disorders	303	71.8%	119	28.2%	422
Infections and infestations	1,139	68.9%	515	31.1%	1,654
Injury, poisoning and procedural complications	677	52.8%	606	47.2%	1,283
Investigations	1,371	30.1%	3,183	69.9%	4,554
Metabolism and nutrition disorders	465	80.0%	116	20.0%	581
Musculoskeletal and connective tissue disorders	685	75.8%	219	24.2%	904
Neoplasms benign, malignant and unspecified (incl cysts and polyps)	901	50.7%	876	49.3%	1,777
Nervous system disorders	1,110	78.6%	303	21.4%	1,413
Pregnancy, puerperium and perinatal conditions	341	74.9%	114	25.1%	455
Psychiatric disorders	500	80.5%	121	19.5%	621
Renal and urinary disorders	399	68.6%	183	31.4%	582
Reproductive system and breast disorders	573	64.8%	311	35.2%	884
Respiratory, thoracic and mediastinal disorders	594	67.4%	287	32.6%	881
Skin and subcutaneous tissue disorders	688	76.1%	216	23.9%	904
Social circumstances	157	63.3%	91	36.7%	248
Surgical and medical procedures	1,016	58.8%	712	41.2%	1,728
Vascular disorders	806	72.0%	314	28.0%	1,120
Total	15,847	61.2%	10,058	38.8%	25,905

Table 3. Overall mapping performance of PT terms by MedDRA system organ class (NB: Since PT terms can be associated with more than one category, the total number of mappings does not reflect the total number of PT terms)