To what entities does an ICD-9-CM code refer? A realist approach.

William R. Hogan*

Division of Biomedical Informatics,
University of Arkansas for Medical Sciences, Little Rock, AR

Abstract. I take a view of ICD-9-CM codes as diagnostic statements, where these statements are about entities that exist in reality. I then represent these entities according to a realist view of disease, disorder, and diagnosis as defined by the Ontology for General Medical Science and using Referent Tracking templates. I illustrate the approach using ICD-9-CM codes that refer to systemic arterial hypertension. The approach is general and could help enable interoperability of data encoded with realist ontologies and data encoded with terminologies and administrative classifications, such as data from electronic health records and insurance claims.

1. Introduction

As several researchers have noted, the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) is not truly a classification of disease, but rather a classification of patients or statements about patients [1, 2]. For example, Bodenreider et al. note that the term *Tuberculosis of adrenal glands, tubercle bacilli not found (in sputum) by microscopy, but found by bacterial culture* is a sentence schema that contains information about how the disease was diagnosed, which has no bearing on the disease itself or the type of which it is an instance [1].

The *ICD-9-CM Official Guidelines for Coding and Reporting* document lends credence to this view. First, it uniformly refers to ICD-9-CM codes as *diagnosis* codes, not disease codes [3]. Second, it also describes *combination codes*, noting that they refer to multiple diagnoses (and by extension, denote multiple diseases) [3]. Many ICD-9-CM codes therefore classify multiple disease particulars in a single patient. They do not assert that a single disease instantiates multiple types. For example, all codes beginning with 404 combine heart disease and chronic kidney disease resulting from hypertension. Thus these diagnosis codes, when applied to a particular patient, refer to at least three disease particulars that are instances of hypertension, heart disease, and chronic kidney disease.

Besides combination codes, the existence of epistemological criteria in ICD-9-CM also supports the diagnosis perspective on ICD-9-CM. For example, all codes starting with 010 through 018 are differentiated by the manner in which tuberculosis was diagnosed

(e.g., *tubercle bacilli found (in sputum) by microscopy*). A realist-based ontology should not have an exact match for these codes because they contain an epistemological criterion. Thus, conversion of these codes to a realist representation is more complex than a simple assertion of equivalence among codes.

If we view each ICD-9-CM code as a diagnostic statement, then what are these statements about? As the name of the classification suggests, they are about diseases, at least. We take the realist view that distinguishes carefully between a diagnosis, a disorder, and a disease. In this view, diagnostic statements refer to diseases, disorders, and possibly other entities that exist in reality (which in the case of ICD-9-CM, include symptoms, findings, and epistemological methods).

The purpose of this work, therefore, is to generate a framework for creating representations of the entities to which ICD-9-CM codes as diagnostic statements—or perhaps more generally *clinical statements*—refer. That is, when a clinician assigns an ICD-9-CM code to a particular patient, this framework enables a software application to generate automatically representations of the disease, disorder, and other entities that are thereby implied.

Another motivation for this approach is to facilitate interoperability of data annotated using the Referent Tracking (RT) paradigm [4]. This paradigm accommodates data annotated using ontologies constructed according to the realist approach, as well as data annotated using "concept"-based terminologies and even administrative classifications such as ICD-9-CM. If both types of such data are extant within a single RT system (RTS), however, the data may not be fully interoperable. For example, a query for type 2 diabetes mellitus using DOID:9352 from the Disease Ontology (DO) will not return patients whose type 2 diabetes mellitus is recorded using only ICD-9-CM and/or SNOMED-CT.

In this paper, I outline a general approach to representing the disease, disorder, etc. particulars in a given patient that are referred to by ICD-9-CM codes relating to systemic arterial hypertension. These data are the most common form of diagnostic data, occurring in electronic medical records (EMRs) and insurance claims data. I also describe an implementation of the approach in the RT paradigm. The approach is general and could be applied to ICD-10-CM and terminologies like SNOMED-CT. SNOMED-CT has built-in constructs that might enable a more automated approach. However, the vast majority of clinical diagnoses to be integrated with basic science data are encoded with ICD-9-CM.

2. Ontological representation

2.1. Selecting a subset of ICD-9-CM

I searched the tabular list of the current version of ICD-9-CM on the word 'hypertension', selecting only codes that represent statements about systemic arterial hypertension. Hypertension is the most common chronic condition that results in visits to physician offices [5]. It causes substantial morbidity and mortality [6]. Furthermore, hypertension is a significant subject of translational research [7, 8], and thus linking clinical to genomics and proteomics data about patients with hypertension will increasingly be important.

I excluded ICD-9-CM codes for pulmonary hypertension, hypertension complicating pregnancy and childbirth (for which there are another 55 ICD-9-CM codes), and hypertension of other body substances in liquid phase. For example, I excluded 365.04 *Ocular hypertension*, 416.0 *Primary pulmonary hypertension*, 459.3 *Chronic venous hypertension (idiopathic)*, and all codes beginning with 642 *Hypertension complicating pregnancy, childbirth, and the puerperium*.

2.2. ICD-9-CM diagnoses of hypertension and related conditions

I found a total of 51 ICD-9-CM codes representing statements that implied either directly (e.g., essential hypertension) or indirectly (e.g., hypertensive heart disease) that the patient suffers from systemic arterial hypertension.[1] This set excludes 55 ICD-9-CM codes beginning with 642, which refer to hypertension complicating pregnancy and childbirth.

A key issue was how to treat ICD-9-CM codes with the phrases 'unspecified' or 'not elsewhere classified' in their title. Although one could handle 'unspecified' and 'not elsewhere classified' in the RT paradigm, I did not. Typically, ICD-9-CM codes are assigned within the context of a single chart, and thus the 'unspecified' reference refers to statements made within only that chart. Indeed, the ICD-9-CM Guidelines specifically state codes … *titled "unspecified" are for use when the information in the medical record is insufficient to assign a more specific code* [3].

Given that the vision for the RT paradigm includes semantic interoperability across organizations within a community, it would be inappropriate to include templates that deny the presence of other templates within the larger community as a whole. However, anyone wishing to include templates to handle 'unspecified' and 'not elsewhere classified' could adapt this work.

2.3. Definitions of disease, disorder, and diagnosis

In this work, I use the realist definitions of disease, disorder, and diagnosis of Scheuermann et al. [9] and the Ontology of General Medical Science (OGMS) [10]. Specifically, the definitions from version 0.5 of OGMS[2] are as follows:

- *Disorder: A material entity which is clinically abnormal and part of an extended organism. Disorders are the physical basis of disease.*
- *Disease: A disposition (i) to undergo pathological processes that (ii) exists in an organism because of one or more disorders in that organism.*
- *Diagnosis: The representation of a conclusion of an interpretive process that has as input a clinical picture of a given patient and as output an assertion … to the effect that the patient has a disease of such and such a type.*

The notion is that a disorder is a physical aberration in the body that confers a disposition to undergo a particular pathological process. Note that absence of realization of the disposition (e.g., in the presence of medications that maintain blood

---

[1] This list of 51 codes is publicly available at: http://spreadsheets.google.com/ccc?key=0AnlWz1vGIs2hdF9acjBjRWpheG9DWXdwMTFhaGc4NGc&hl=en

[2] The author is one of the developers of OGMS.

pressure in the normal range) does not imply absence of the disease (disposition). Patients with hypertension have a disposition to high blood pressure regardless of treatment: cessation of treatment results in a return to high blood pressure because the disposition never ceased to exist.

2.4. Representing Hypertension

In the case of hypertension (both the so-called essential and secondary varieties), significant evidence exists that the underlying disorder is an abnormally elevated quantity of salt in the blood, which itself is the consequence of a wide variety of disorders that confer a disposition to enhanced reabsorption of salt by the kidneys [11]. The disease is then the resulting disposition of the blood to exert higher-than-normal pressure on the walls of blood vessels.

In what follows I use *italics* for particulars and universals[3] and **bold** for relations.[4] Where necessary, I number particulars as *dz_1*, *dz_2*, … to distinguish among them.

Now, for a particular patient, to whom one of the 51 ICD-9-CM codes has been assigned, I say the following:

hs **instance_of** *Human*

dz **instance_of** *Hypertension*

do **instance_of** *Disorder*

do **part_of** *hs*

dz **inheres_in** *hs*

dz **disposition_of** *do*

Now, I treat the portion of salt in the blood as the disorder:

do **instance_of** *Scattered molecular aggregate*

The portion of salt has at least one molecule of salt as its grain (else the system will not know it is a *salt* aggregate):

mo **instance_of** *NaCl*

do **has_grain** *mo*

There is a portion of blood that is part of the patient, and the portion of salt is part of it:

pb **instance_of** *Portion of blood*

do **part_of** *pb*

pb **part_of** *hs*

2.5. Essential vs. Secondary vs. Complication of Treatment

Many of the 51 ICD-9-CM codes distinguish ‘essential’ vs. ‘secondary’ hypertension. The term ‘essential hypertension’ refers to hypertension where the underlying disorder that disposes to renal salt retention is unknown, whereas secondary hypertension refers

---

[3] I set particulars in lower-case and universals with an upper-case first letter

[4] All relations in this work are instance-level relations except **instance_of** and **lacks**, which are instance–universal relations.

to cases whose underlying disorders/diseases are known. As an example of the latter, reduced blood flow to the kidney due to stricture or stenosis of a renal artery causes the kidney to retain salt to increase blood pressure (to in turn increase blood flow to itself).

These distinctions have an epistemological basis, and except for ICD-9-CM codes that explicitly state the disorder that predisposes to renal retention of salt, we add nothing further. Since the final common pathway of hypertension regardless of these distinctions is elevated salt content of the blood, which in turn results from a different disposition (disease) which in turn has a different physical basis (disorder), these distinctions do not affect our expansions.

There is also an ICD-9-CM code for secondary hypertension resulting from iatrogenesis (i.e., caused by treatment). Here, a disorder was caused by an instance of treatment, giving rise to a disposition to renal salt resorption, whose realization caused the disorder underlying hypertension:

tr **instance_of** *Treatment*
tr **results_in** *do_2*
do_2 **part_of** *hs*
dz_2 **disposition_of** *do_2*
dz_2 **realized_by** *sr*
sr **instance_of** *Exaggerated renal salt resorption*
sr **results_in** *do*

2.6. Benign vs. Malignant Hypertension

The words 'benign' and 'malignant' in ICD-9-CM diagnoses refer to hypertension, not to other diseases mentioned such as heart disease, chronic kidney disease, etc. Whether hypertension is benign or malignant refers to the *course* of the disease. Malignant hypertension … *is a sudden and rapid development of extremely high blood pressure* [12].

Thus, for malignant hypertension, an acute disease course:

dc **instance_of** *Acute disease course*

And for benign hypertension, a chronic disease course:

dc **instance_of** *Chronic disease course*

And in both cases, associate the course with the disease:

dz **realized_by** *dc*

2.7. Hypertensive Heart and Kidney Disease

When 'hypertensive heart disease' appears in the title of an ICD-9-CM code, there also exists a disease of the heart:

dz_2 **instance_of** *Heart disease*
do_2 **instance_of** *Disorder*
dz_2 **inheres_in** *hs*
dz_2 **disposition_of** *do_2*

We assert that realization of the hypertension (pathological process) resulted in the disorder of the heart:

pp **instance_of** *Pathological process*

dz **realized_by** *pp*

pp **results_in** *do_2*

And we need to say that the disorder is part of the heart, and the heart part of the patient:

ht **instance_of** *Heart*

do_2 **part_of** *ht*

ht **part_of** *hs*

We represent hypertensive chronic kidney disease in the same manner, but additionally assert a chronic course, and that the disease is realized by the course.

2.8. Presence vs. Absence of Heart Failure

The title of several ICD-9-CM codes refers to the presence or absence of heart failure. 'Heart failure' is an ambiguous term that may refer to either a pathological process or the disposition to it. We use 'heart failure' in the latter sense, which is consistent with medical usage: we treat heart failure, and when the heart is not malfunctioning, the disposition to undergo malfunctioning remains.

dz_3 **instance_of** *Heart failure*

do_3 **instance_of** *Disorder*

dz_3 **disposition_of** *do_3*

dz_3 **inheres_in** *pt*

do_3 **part_of** *ht*

Conversely, when the title says 'without heart failure', the patient is not the bearer of any instance of heart failure:

pt **lacks** *Heart failure* with respect to **bearer_of**

3. Implementation in RT templates

RT is a realist approach to managing data by asserting the existence of particulars and their instantiations and relationships over time [4]. RT has several types of templates that capture different information about particulars [13]. An A-template asserts the existence of a particular:

$$A_i = < IUI_p, IUI_a, t_{ap} >$$

Specifically, it …*captures the assignment of an* IUI_p *to a particular at time* t_{ap} *by the … author* IUI_a [13]. Each particular to which the new RT templates refer—including diseases, disorders, pathological processes, disease courses, and so on—requires an A-template. For example,

$$A_1 = <IUI_{dz}, IUI_{hogan}, 5/14/10>$$

A PtoU template states that a particular IUI_p instantiates a universal *u* at time t_r, where t_a is time of authorship, *o* is the ontology from which the universal *u* derives, and *inst* is the **instance_of** relation from the Relation Ontology (RO):

$$U_i = < IUI_a, t_a, inst, o, IUI_p, u, t_r >$$

$$Ex: U_1 = < IUI_h, 5/14/10, inst, DO, IUI_{dz}, Hypertension, 5/7/10 >$$

A PtoP template states that a set of particulars *P* stand in relation *r* from ontology *o* to each other:

$$R_i = < IUI_a, t_a, r, o, P, t_r >$$

$$Ex: R_1 = < IUI_h, t_a, part\_of, RO, <IUI_{do}, IUI_{pt}>, t_r >$$

A PtoLackU template states that a particular IUI_p lacks a relation *r* from ontology *o* to any instance of a universal *u*:

$$U_{\neg i} = < IUI_a, t_a, r, o, IUI_p, u, t_r >$$

$$Ex: U_1 =< IUI_h, t_a, bearer\_of, RO, IUI_{pt}, Heart\ failure, t_r >$$

A PtoCO template states that the author has annotated a particular with a 'concept' code *co* from a terminology *cbs*:

$$Co_i = < IUI_a, t_a, cbs, IUI_p, co, t_r >$$

$$Ex: Co_1 =< IUI_h, t_a, ICD\text{-}9\text{-}CM, IUI_{pt}, 401.1, t_r >$$

I refer to the process of converting the PtoCO template to templates that represent the entities implied by a diagnosis—as outlined above—as an *expansion*. I use the term 'expansion template' to refer to an RT template created to represent these entities and their relationships one another. I propose a new subsystem of the RTS called the 'expansion subsystem (ESS)' that performs the expansion. The templates I described above are sufficient to represent all the statements about particulars I have listed here.[5]

Each expansion template requires the IUI of an author (IUI_a). One possibility is to use the IUI_a from the original PtoCO template that contained the ICD-9-CM code in question. However, this approach would incorrectly assign any errors made in the expansion of the code to that author. Other alternatives include having a single IUI_a for the ESS or having a different IUI_a for each expansion of a single ICD-9-CM. I adopt a single IUI_a for the ESS.

Except for A-templates and Meta-templates (which we do not discuss further here), each template has two timestamps: one to capture the time at which the assertion holds in reality, or t_r and one to capture the time at which the assertion was made, or t_a. The latter timestamp is straightforward: I simply use the time at which the ESS completed the expansion. For the former timestamp, I use the t_r of the PtoCO template, because it is the time at which the diagnostic statement of the ICD-9-CM code holds, and thus the time at which the assertions implied by the ICD-9-CM code hold.

Finally, the expansion templates could refer to a particular already assigned an IUI in the RTS. The problem of ensuring unique reference of IUIs and potentially merging duplicates is not unique to this work: it is a problem for referent tracking in general. I assume the existence of a solution in the RTS to this problem that the ESS may employ.

---

[5] The complete set of expansions is in the same document with the codes.

4. Discussion

I have defended a view of ICD-9-CM codes as representing not diseases, but diagnostic or even clinical statements. Using realist definitions of disease and its related entities, and the referent-tracking paradigm, I have enabled representation of the entities in reality referred to by 51 ICD-9-CM codes that imply the existence of hypertension. This approach may help unlock EMR data for translational science.

Acknowledgements

This work was supported by award number 1UL1RR029884 from the National Center for Research Resources.

References

1. Bodenreider O, Smith B, Burgun A. The ontology-epistemology divide: A case study in medical terminology. Proceedings of FOIS 2004; 2004; Turin.
2. Ceusters W, Smith B, DeMoor G. Ontology-based integration of medical coding systems and electronic patient records 2006 [cited 2010 Apr 16]. Available from: http://ontology.buffalo.edu/medo/CodingAndEHCR.pdf
3. ICD-9-CM Official Guidelines for Coding and Reporting2009 [cited 2010 Apr 16]: Available from: http://www.cdc.gov/nchs/data/icd9/icdguide09.pdf.
4. Ceusters W, Smith B. Strategies for referent tracking in electronic health records. J Biomed Inform. 2006 Jun;39(3):362-78.
5. Cherry D, Woodwell D, Rechtsteiner E. National Ambulatory Medical Care Survey: 2005 Summary. Hyattsville, MD: National Center for Health Statistics,2007.
6. Chobanian AV, et al. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. Jama. 2003 May 21;289(19):2560-72.
7. Cowley AW. The genetic dissection of essential hypertension. Nat Rev Genet. 2006 Nov;7(11):829-40.
8. Tomaszewski M, Zimmerli L, Charchar FJ, Dominiczak AF. Genetic information in the diagnosis and treatment of hypertension. Curr Hypertens Rep. 2006 Aug;8(4):309-16.
9. Scheuermann RH, Ceusters W, Smith B. Toward an ontological treatment of disease and diagnosis. AMIA Summit on Translational Bioinformatics; 2009.
10. Ontology for General Medical Science. 2009 [cited 2009 11/24]; Available from: http://code.google.com/p/ogms/.
11. Lifton RP, Gharavi AG, Geller DS. Molecular mechanisms of human hypertension. Cell. 2001 Feb 23;104(4):545-56.
12. Malignant hypertension. 2009 [cited 2010 Apr 16]; Available from: http://www.nlm.nih.gov/medlineplus/ency/article/000491.htm.
13. Manzoor S, Ceusters W, Rudnicki R. Implementation of a referent tracking system. Int J of Healthcare Information Systems and Informatics. 2007 Oct-Dec;2(4):41-58.