University of Groningen

A New Coding System for Metabolic Disorders Demonstrates Gaps in the International Disease Classifications ICD-10 and SNOMED-CT, Which Can Be Barriers to Genotype-Phenotype Data Sharing

Sollie, Annet; Sijmons, Roelof; Lindhout, Dick; van der Ploeg, Ans T.; Gozalbo, M. Estela Rubio; Smit, Gerrit; Verheijen, Frans; Waterham, Hans R.; van Weely, Sonja; Wijburg, Frits A.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Link to publication in University of Groningen/UMCG research database

Human Mutation

A New Coding System for Metabolic Disorders Demonstrates Gaps in the International Disease Classifications ICD-10 and SNOMED-CT, Which Can Be Barriers to Genotype–Phenotype Data Sharing

Annet Sollie,[1]* Rolf H. Sijmons,[1] Dick Lindhout,[2] Ans T. van der Ploeg,[3] M. Estela Rubio Gozalbo,[4] G. Peter A. Smit,[5] Frans Verheijen,[6] Hans R. Waterham,[7] Sonja van Weely,[8] Frits A. Wijburg,[9] Rudolph Wijburg,[10] and Gepke Visser[10]

[1]Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; [2]Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; [3]Department of Pediatrics, Erasmus Medical Center, Rotterdam, The Netherlands; [4]Department of Pediatrics and Laboratory Genetic Metabolic Diseases, Maastricht University Medical Center, Maastricht, The Netherlands; [5]Department of Pediatrics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; [6]Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands; [7]Department of Clinical Chemistry, Academic Medical Centre, Amsterdam, The Netherlands; [8]Dutch Steering Committee on Orphan Drugs, The Hague, The Netherlands; [9]Department of Pediatrics, Academic Medical Centre, Amsterdam, The Netherlands; [10]Department of Metabolic Diseases, University Medical Center Utrecht, Utrecht, The Netherlands

ABSTRACT: Data sharing is essential for a better understanding of genetic disorders. Good phenotype coding plays a key role in this process. Unfortunately, the two most widely used coding systems in medicine, ICD-10 and SNOMED-CT, lack information necessary for the detailed classification and annotation of rare and genetic disorders. This prevents the optimal registration of such patients in databases and thus data-sharing efforts. To improve care and to facilitate research for patients with metabolic disorders, we developed a new coding system for metabolic diseases with a dedicated group of clinical specialists. Next, we compared the resulting codes with those in ICD and SNOMED-CT. No matches were found in 76% of cases in ICD-10 and in 54% in SNOMED-CT. We conclude that there are sizable gaps in the SNOMED-CT and ICD coding systems for metabolic disorders. There may be similar gaps for other classes of rare and genetic disorders. We have demonstrated that expert groups can help in addressing such coding issues. Our coding system has been made available to the ICD and SNOMED-CT organizations as well as to the Orphanet and HPO organizations for further public application and updates will be published online (www.ddrmd.nl and www.cineas.org).
Hum Mutat 34:967–973, 2013. © 2013 Wiley Periodicals, Inc.

KEY WORDS: data sharing; rare diseases; ICD; SNOMED; ontology; phenotyping

*Correspondence to: Annet Sollie, Department of Genetics, University Medical Center Groningen, PO Box 30001, 9700 RB Groningen, The Netherlands. E-mail: a.sollie@umcg.nl

Introduction

Data sharing is essential for a better understanding of rare genetic disorders and the underlying genetic defects. Good phenotype coding plays a key role in this process and also in general in processes where phenotype data need to be entered into clinical registries, genotype–phenotype databases, and biobanks, and shared between them. Such initiatives to register, combine, and exchange clinical and research data are pivotal in supporting research and improving healthcare [Jones et al., 2011; Richesson and Vehik, 2010].

Rare diseases are life threatening or chronically debilitating diseases with a prevalence of up to five per 10,000 inhabitants in the European Union (EU). It is estimated that there are at least 5,000 rare diseases, many of them genetic, affecting 6%–8% of the total population in the EU, which implies a minimum 27 million people in the EU are affected [European Medicines Agency (EMA), 2007]. In the United States, the Rare Disease Act of 2002 also defined rare disease according to prevalence, specifically "any disease or condition that affects less than 200,000 persons in the United States," or about one in 1,500 people. Although there are no disease-modifying therapies for most rare diseases, the passing of the 1983 U.S. Orphan Drug Act [Food and Drug Administration] and European legislation in 2000 [European Parliament. Regulation (EC) No. 141/2000] stimulated new research lines by creating financial incentives and other supportive measures for developers of new drugs to treat people with rare diseases [Talele et al., 2010]. It is expected that many more rare diseases will become amenable to treatment within the next few decades.

The need to improve research and care in the field of rare disorders, which can be strongly supported by the sharing and combining of data on these rare patients, has also been recognized by the Council of the European Union. Through their European Action in the Field of Rare Diseases [Official Journal of the European Union, Council recommendation of 8 June 2009 on an action in the field of rare disease], signed in 2009, the EU member states committed themselves to establishing and implementing a national rare disease action plan and to cooperating at a European level on this health issue. The European Action stated that member states should "aim

to ensure that rare diseases are adequately coded and traceable in all health information systems, encouraging an adequate recognition of the disease in the national healthcare and reimbursement systems based on the ICD."

Unfortunately, the two most widely used coding systems in medicine—ICD (the WHO's International Classification of Diseases, www.who.int/classifications/icd) and SNOMED-CT (Systematized Nomenclature of Medicine Clinical Terminology, www.ihtsdo.org)—lack essential details for classifying and annotating rare and hereditary disorders. This is a barrier to the optimal registration of patients with these disorders in databases, and to much needed data-sharing efforts, such as those in the Human Variome Project (http://www.humanvariomeproject.org).

Our study addresses this problem for metabolic diseases, a particular hereditary subgroup of rare disorders. Metabolic diseases, also referred to as inborn errors of metabolism, are generally monogenic defects resulting in a deficient activity in an enzyme or a transporter in a pathway of cellular metabolism [Scriver et al., 2001]. The number of recognized metabolic diseases is continually increasing because of the advances in knowledge and diagnostic laboratory techniques. Most metabolic diseases are extremely rare (less than one per 50,000 inhabitants), although all metabolic diseases combined have an estimated, relatively high birth prevalence of up to one per 800 newborns [Sanderson et al., 2006]. In the Netherlands, we decided to build a registry for patients with metabolic disorders and also to optimize the codes for national use in medical and clinical genetics. With these purposes in mind, we developed, with a dedicated group of clinical specialists, a clinically oriented annotation system for metabolic disorders based on two existing national coding systems. To assess the potential value of adding our annotation system to ICD and SNOMED-CT, we compared the three systems and identified large gaps in both ICD and SNOMED-CT. To the best of our knowledge, we are the first to actually quantify these gaps for a specific field of rare diseases.

Materials and Methods

Study Overview

We combined and expanded two existing coding systems for metabolic diseases, the DDRMD (Dutch Diagnosis and Registration of Metabolic Diseases, www.ddrmd.nl) and a subset of CINEAS (Dutch center for disease code development and distribution to the clinical genetics community, www.cineas.org) to develop a more detailed and strongly clinically oriented coding system. The DDRMD was set up by specialists in metabolic disorders, whereas CINEAS was initiated by clinical geneticists. Both systems were originally developed independent of each other and born out of the need to have more extensive coding system available than the ones offered by SMOMED and ICD. The primary purpose of each of our original coding systems was improving patient classification and retrieval. We used the DDRMD as a starting point for extending the coding system of metabolic diseases because this system had already been used for more than 10 years by metabolic specialists in clinical practice. We matched and enriched these systems in a three-step process, exemplified in a flowchart (Figure 1). A list of criteria for including codes in the coding system was drawn up for the matching process (Table 1). To facilitate cross-linking, but also to investigate the extent to which codes were lacking in ICD and SNOMED-CT, we checked and updated existing mappings to these two international systems.

DDRMD (Background, Origin, and Objective)

The DDRMD is a collaborative project of all the clinical metabolic centers in the Netherlands. It was started in 2001 and over 5,000 patients have been registered so far, with almost 300 different metabolic diseases. The main reason for initiating the DDRMD was that despite the various diagnosis registration systems used in hospitals, it was proving difficult to retrieve patients with metabolic diseases from these registers. Because there was no disease-specific registration for metabolic diseases, it was impossible to analyze relevant patient data, either for research or for care purposes.

In the DDRMD, patient data are registered by one metabolic specialist per metabolic center (see Figure 2) via a secure Web server. In addition, relevant data on newborns referred because of an abnormal neonatal screening result indicative for metabolic disease are also included. The data are used to facilitate research on metabolic diseases and to provide information on the outcome of the Dutch newborn screening procedure for metabolic diseases.

CINEAS (Background, Origin, and Objective)

CINEAS is the Dutch center for disease code development and distribution to the clinical genetics community. It was initiated by the eight clinical genetics centers responsible for genetic counseling and diagnostics in the Netherlands in 1992 [Zwamborn-Hanssen et al., 1997]. It is used in daily practice by the Dutch clinical geneticists and genetic counselors to assign diseases to patients. Presently, the 55th edition of CINEAS lists more than 5,500 diseases, most of them rare, and the metabolic diseases form a distinct subset (Figure 3). A number of Dutch diagnostic DNA laboratories use the CINEAS system as well, and recently the Danish genetics centers have decided to adopt CINEAS. Each new edition of the list contains new disease entries submitted by users, after they have been discussed and approved by a group of experts. The entire process of submitting and adding new entries to the database is supported by a Website (www.cineas.nl or www.cineas.org), a paid professional curator and a quickly responding national expert panel, which has reduced throughput time to an average of 2 weeks. Local system administrators upload new editions to their own patient information systems, and the Website facilitates searching of the CINEAS database and is used to publish the new editions. Codes are never removed from the system, but can be made obsolete and thus no longer assigned to patients. Entry, modifications, and obsoletion of codes including dates are saved in the Diagnosis History. Cross-links are provided to OMIM, Online Mendelian Inheritance in Man, a catalogue of hereditary disorders and their genes (www.omim.org), and to SNOMED and ICD. Although the codes in CINEAS, including the metabolic codes, are nonhierarchical (see *Discussion*), individual codes can easily be found in the system using the onboard search engine.

Existing Coding Systems

The most widely used system in medical practice is the ICD, published by the World Health Organization (version 9 published in 1977, or version 10 in 1999). It is categorized by the affected organ system, which makes it difficult to use for diseases in which more than one organ is affected, as is the case for many rare genetic diseases. The WHO is working on the revision of ICD-10, with the aim of publishing ICD-11 in 2015 (www.who.int/classifications/icd/revision).
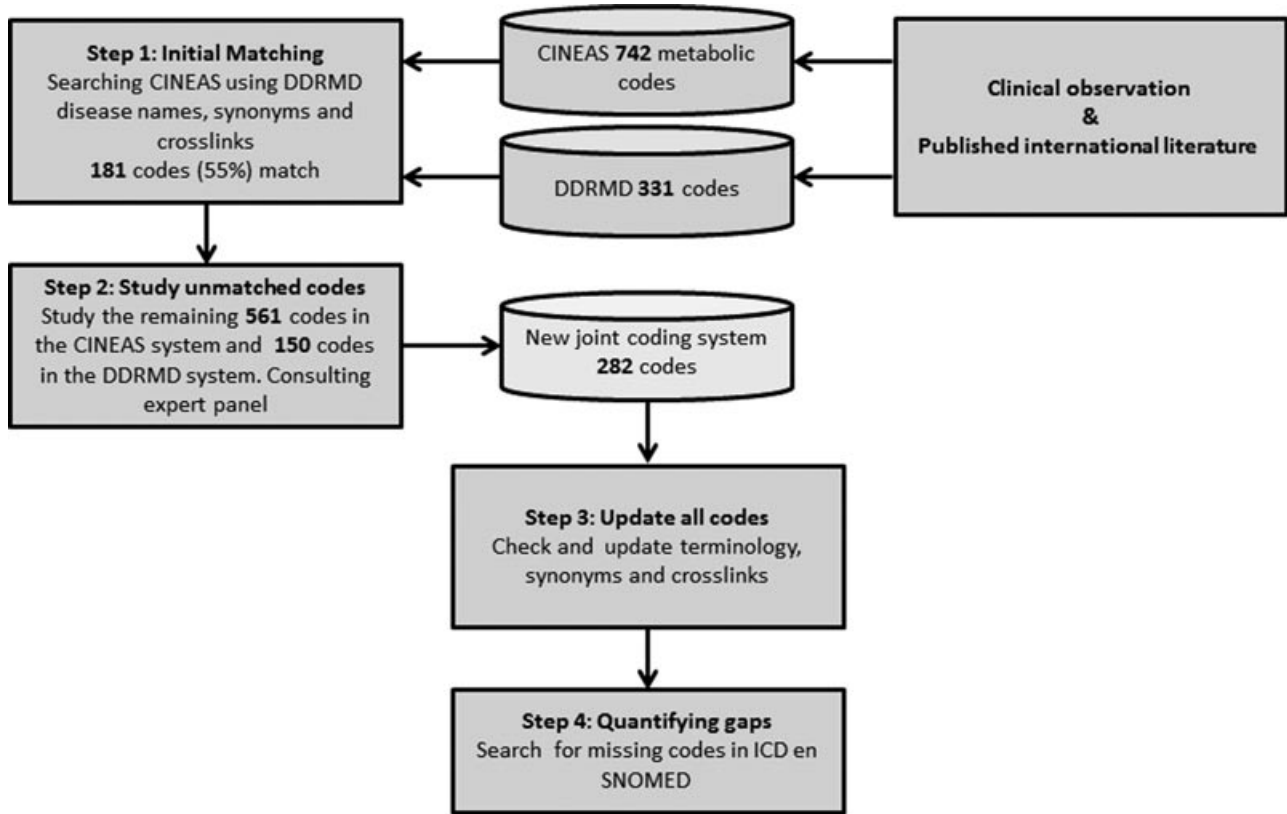
Figure 1. Flowchart of building the coding system and quantifying the gaps in ICD-10 and SNOMED-CT.

SNOMED-CT is a coding system, which has been adopted by many hospital information systems and standards organizations worldwide as a key coding system. Since 1974, SNOMED-CT has evolved from a pathology-specific nomenclature into a healthcare terminology system. There are many studies that have shown the value of SNOMED-CT in theory, but studies on its use in clinical practice are relatively rare [Cornet and de Keizer, 2008].

Assessment of Gaps for Metabolic Disease in ICD and SNOMED-CT

During the final steps of our matching process (Figure 1), for each code in our system, we chose the most appropriate ICD-10 code as a cross-link, using the online WHO browser (http://apps.who.int/classifications/icd10/browse/2010/en) and searching with disease names and synonyms. When no specific disease code was available, we chose a group name or nonspecific code based on the etiology, for example, "E79.8 Other disorders of purine and pyrimidine metabolism" or "E88.8 other specified metabolic disorder" (Table 2).

In addition, we mapped as many entries as possible to the SNOMED-CT International Edition of January 2011, using CliniClue Explore (http://www.cliniclue.com/) and by searching SNOMED-CT using disease names and synonyms. We recorded all the unambiguous mappings and all the possible mappings if more than one SNOMED-CT code was available. Finally, we calculated the gaps for metabolic diseases in ICD-10 and SNOMED-CT as percentages of codes with matches in our coding system.

Table 1. List of Criteria for Including Codes in our Coding System

Number	Criteria
1	The disease has to be a separate clinical entity.
2	Is must be likely that the disease is a separate clinical entity; just one case report in the literature is not enough, unless an enzyme deficiency or transport defect was demonstrated.
3	No separate entries for gene defects; a gene can, however, be connected to a disease (no specific mutation is mentioned).
4	No specific entries for groups of diseases.
5	One enzyme defect leads to only one separate code.

Results

We have developed a specific coding system for metabolic diseases, currently containing almost 300 different disorders. Every item in our system has a unique identifier and includes a disease name, existing synonyms, and mappings to the OMIM catalogue, ICD-10 and SNOMED-CT (example in Table 2). For the unique identifiers, we used the existing CINEAS codes. The mappings to the other coding systems can be used for data exchange with other databases and provide extra search possibilities.

Note that we deviated from our inclusion criteria (Table 1) for the group of mitochondrial diseases. Apart from the separate respiratory chain disorders, we added two general codes for diseases caused by mitochondrial DNA variations. This is because this particular area is evolving rapidly and clear classification is not yet possible in all cases. We expect to be able to create more specific entries for these diseases in the coming years.

*Dutch Diagnosis & Registration of Metabolic Diseases

Abbreviations:

bsn = Dutch national identification number

nbs = newborn screening

Figure 2. Data model for DDRMD.

For 214 (76%) of the diseases in our coding system, there was no specific matching ICD-10 code and only an ICD-10 group name that was too general for our clinical classification purposes was available (e.g., ICD-10 code E88.8 "other specified metabolic disorders").

For 155 (54%) of our codes, it was not possible to map unambiguously to SNOMED-CT because for 81 codes (29%), there was no SNOMED-CT code available and for 72 codes (25%) SNOMED-CT contained double codes. These duplicates were counted mostly because the disease and enzyme deficiency were given separate codes in SNOMED-CT, but also because too much detail in SNOMED-CT made it difficult to distinguish between group codes and subcodes. An example is the disease "alpha-N-acetylgalactosaminidase deficiency" for which SNOMED provided mistakingly two codes, one with three subcodes (Table 2). This shows that despite the size of the SNOMED system, the unambiguous detail needed in clinical practice for metabolic diseases is often not available.

We aim to publish incidences and prevalence of individual metabolic diseases in our coding system online in the spring of 2013. Our coding system is being continuously updated and is published on www.ddrmd.nl and www.cineas.org in pdf and xml formats. CINEAS and DDRMD keep existing as two different organizations, each with a different purpose, now sharing the code system for metabolic disorders. Requests for additions to and alterations of DDRMD and/or CINEAS users are submitted by email to info@cineas.nl or ddrmd@umcutrecht.nl or by using an online Webform on the CINEAS Website for registered organizations. These requests are subsequently discussed in the national CINEAS online expert panel for approval. The national coordinator of DDRMD is now a member of the CINEAS expert panel. The coding system has

already been updated using these procedures and now contains 285 diseases.

Continued funding for the classification efforts is provided by CINEAS (Dutch national disease code development and distribution center for the clinical genetics community). Our novel coding system has recently been donated to ICD, SNOMED-CT, and also to the Human Phenotype Ontology (HPO) (www.human-phenotype-ontology.org), a promising emerging ontology for phenotypic abnormalities, and to Orphanet (www.orpha.net), an important reference portal for information on rare diseases and orphan drugs, for further public application. Continuous updates of our system will be published online.

Discussion

The most widely used classification and coding systems in medical databases are ICD and SNOMED-CT. Historically, the focus in the development of these systems has been directed toward classifying common disorders. The development and updating process for international broad medical coding systems is a highly demanding task and we acknowledge the important contribution of ICD and SNOMED-CT to the annotation of common disorders. However, annotation for rare disorders has been left behind. Collectively, this group is large and growing steadily because of the identification of new diseases and improved clinician awareness. Our study demonstrates large gaps in both ICD (76%) and SNOMED-CT (54%) for metabolic disorders. On the basis of our clinical experience, we

*Dutch national disease code development and distribution center for the clinical genetics community
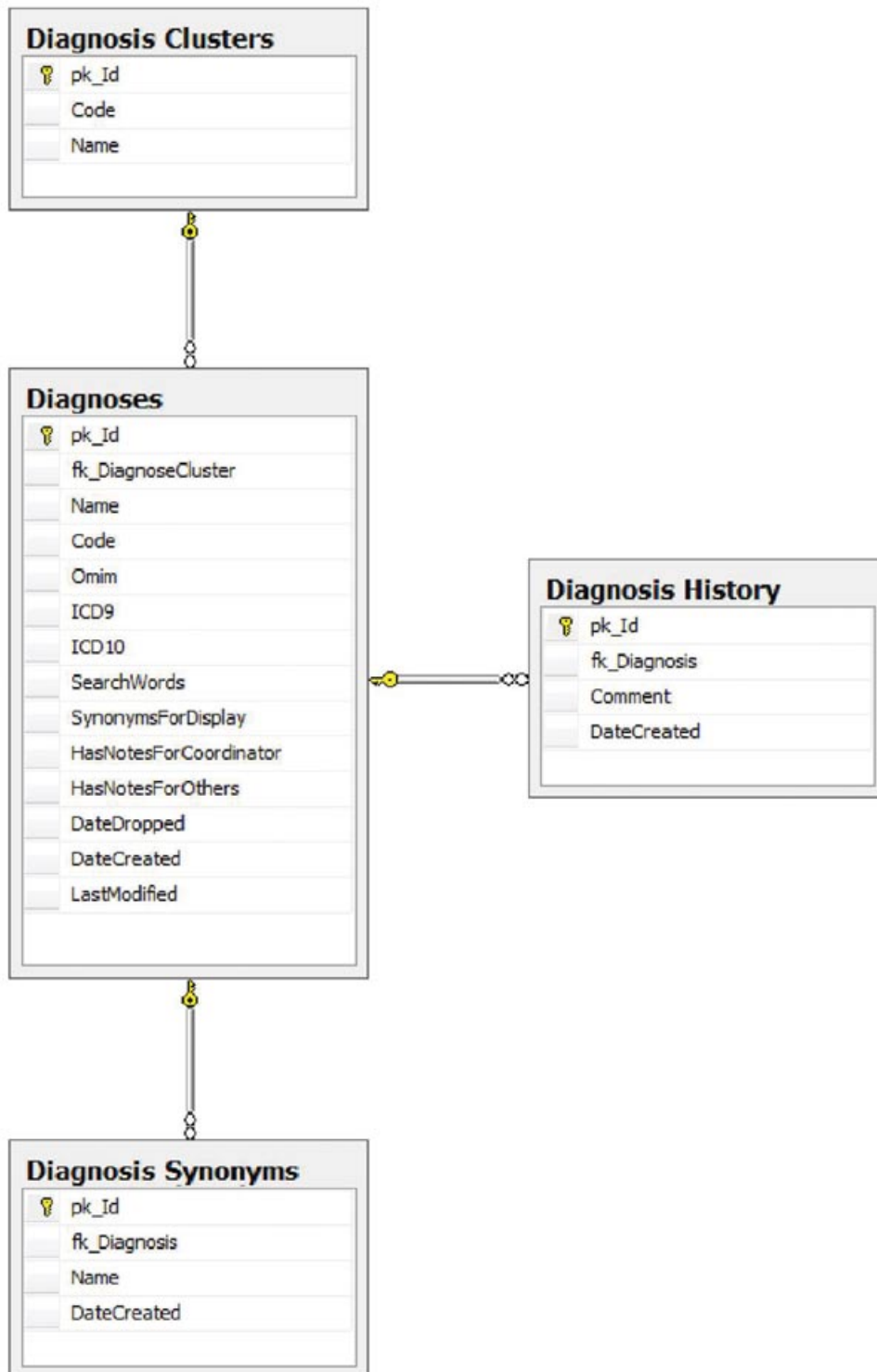
Figure 3. Data model for CINEAS—only core tables.

Table 2. Example from Our Coding System

Disease	Synonyms	Identifier	OMIM	ICD-10	SNOMED
Adenylosuccinase deficiency	Adenylosuccinate lyase	1573	103050	E79.8	73843004
Aldolase-B deficiency	Hereditary fructose intolerance	1318	229600	E74.1	20052008
Alpha-aminoadipic aciduria	2-Amino-/2-oxoadipic aciduria 2-Aminoadipic 2-Oxoadipic	1599	204750	E88.8	Not available
Alpha-aminoadipic semialdehyde dehydrogenase deficiency	Pyridoxine dependent epilepsy AASA Folinic acid responsive convulsions Antiquitine gene	2286	266100	E88.8	Not available
Alpha-N-acetylgalactosaminidase deficiency	NAGA Neuroaxonal dystrophia Schindler disease	1539	609241	E88.8	Double codes: 238048001 and 230365004 with three subcodes

suspect that there may be similar gaps for other types of rare disorders. Such gaps are a barrier to database- and data-sharing efforts. We have shown that with the help of dedicated clinicians and code development agencies, the problem of coding gaps for rare disorders can be successfully addressed.

Developing codes for a rare field of medicine has special challenges. We observed during the development of our system that existing hierarchical, "tree," classification structures, such as those used in SNOMED-CT and in ICD-10, were proving inconvenient for our purpose. Such structures, when well developed for the particular branches, allow for the selection of patients from groups of disorders rather than those with particular individual disorders. However, in the rare field of metabolic disorders, these existing tree structures turned out to be problematic and we dropped our initial hierarchical approach for several reasons. Firstly, several diseases did not fit into any specific group or category leading to a risk of misclassification. Secondly, several diseases fitted into more than one group or category leading to significant risk of double entries for the same disorder. Furthermore, given the explosion of knowledge in this field of rare genetic diseases, extensive and continuous expertise is needed to update the accuracy of a specialist tree structure. Given the aim of our coding system to assign diagnostic end codes to patients and to obtain incidence and prevalence data from our registry on individual metabolic diseases, a nonhierarchical design turned out to be functional. With growing knowledge on underlying molecular pathways, well-fitting metabolic branches of the coding trees are likely to be developed in the future in the international community and this will support better data handling on the level of groups of metabolic disorders.

The World Health Organization has signaled the need to improve ICD-10 for use in the field of rare diseases. A special Topic Advisory Group (http://www.who.int/classifications/icd/TAGs/en/index.html) has been assigned to the subject of rare diseases to advise the WHO on the current updating and revision process from ICD-10 to ICD-11 (anticipated publication in 2015). We recently donated our work to both the ICD and SNOMED-CT communities to support further code development, and to the Orphanet and HPO organizations as well. These organizations are also contributing to solving annotation problems. The Orphanet organization (www.orpha.net) has stressed the need to provide well-designed codes for rare diseases, especially for the purposes of data sharing and it puts much effort into this field [Rath et al., 2012]. The HPO (http://www.human-phenotype-ontology.org/) is another important international initiative to support the annotation of genetic disorders and we are presently collaborating with HPO to further enrich both coding systems.

We are convinced that the approach we adopted—of code development driven by particular clinical and epidemiological needs, and support for that development from experts working in the clinical and medical fields of interest—can contribute to the quality of annotation for rare diseases, and thus to healthcare for patients with these diseases.

Contributors

A.S. and G.V. had full access to all the data in the coding systems used in this study and had final responsibility for the decision to submit for publication. A.S., R.H.S., and G.V. conceptualized the study. Data acquisition was done by A.S. and G.V. A.S., R.H.S., D.L., A.T.P., M.E.R.G., G.P.A.S., F.V., H.W., S.W., F.A.W., and G.V. analyzed and interpreted the data. A.S., G.V., and R.H.S. wrote the report. Critical revision of the manuscript was done by F.A.W., R.W., F.V., H.W., A.T.P., G.P.A.S., M.E.R.G., D.L., and S.W. R.W. provided technical support and J.S.R. revised the manuscript.

References

Official Journal of the European Union. Council recommendation of 8 June 2009 on an action in the field of rare disease.

Cornet R, de Keizer N. 2008. Forty years of SNOMED: a literature review. BMC Med Inform Decis Mak 8, Suppl 1:S2.

European Medicines Agency (EMA). Orphan drugs and rare diseases at a glance. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Other/2010/01/WC500069805.pdf. Accessed October 20, 2011. Document ref. EMEA/290072/2007.

European Parliament. Regulation (EC) No. 141/2000 of the European Parliament and the Council of 16 December 1999 on Orphan Medicinal Products. Off J Eur Union. Available at: http://eur-lex.europa.eu.proxy-ub.rug.nl/LexUriServ/LexUriServ.do?uri = OJ:L:2000:018:0001:0005:En:PDF. Accessed January 22, 2000.

Food and Drug Administration (USA). Orphan drug act. Available at: http://www.fda.gov/RegulatoryInformation/Legislation/FederalFoodDrugandCosmeticActFDC

Act/SignificantAmendmentstotheFDCAct/OrphanDrugAct/default.htm. Accessed October 20, 2011.

Jones S, James E, Prasad S. 2011. Disease registries and outcomes research in children: focus on lysosomal storage disorders. Paediatr Drugs 13:33–47.

Rath A, Olry A, Dhombres F, Brandt MM, Urbero B, Ayme S. 2012. Representation of rare diseases in health information systems: the Orphanet approach to serve a wide range of end users. Hum Mutat 33:803–808.

Richesson R, Vehik K. 2010. Patient registries: utility, validity and inference. Adv Exp Med Biol 686:87–104.

Sanderson S, Green A, Preece MA, Burton H. 2006. The incidence of inherited metabolic disorders in the West Midlands, UK. Arch Dis Child 91:896–899.

Scriver CR, Beaudet AL, Sly WS, Valle MD. 2001. *The metabolic and molecular bases of inherited disease.* New York: McGraw-Hill, Medical Publishing Division. p xliiii.

Talele SS, Xu K, Pariser AR, Braun MM, Farag-El-Massah S, Phillips MI, Thompson BH, Cote TR. 2010. Therapies for inborn errors of metabolism: what has the orphan drug act delivered? Pediatrics 126:101–106.

Zwamborn-Hanssen AM, Bijlsma JB, Hennekam EF, Lindhout D, Beemer FA, Bakker E, Kleijer WJ, de France HF, de Die-Smulders CE, Duran M, van Gennip AH, van Mens JT, et al. 1997. The Dutch uniform multicenter registration system for genetic disorders and malformation syndromes. Am J Med Genet 70:444–447.