An infrastructure for quality assessment in intensive care. Prognostics models and terminological systems

de Keizer, N.F.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 9

Analysis and design of a terminological system for Intensive Care diagnoses

N.F. de Keizer, A. Abu-Hanna, R. Cornet, J.H.M Zwetsloot-Schonk, C.P. Stoutenbeek

Based on
Analysis and Design of an ontology for intensive care diagnoses.
Abstract
Information about the patient's health status, and about medical problems in general, play an important role in stratifying a patient population for quality assessment of intensive care. A terminological system which supports both the description of health problems for daily care practice and the aggregation of diagnostic information for evaluative research, is desirable for the description of the patient population. This chapter describes the engineering of an ontology that is at the heart of a terminological system for intensive care diagnoses. We analyse the criteria for such an ontology and evaluate existing terminological systems according to these criteria. The analysis shows that none of the existing terminological systems completely satisfies all our criteria. We describe choices regarding design, content and representation of a new ontology on which an adequate terminological system is based. The proposed ontology is characterised by the meta-specification of its concepts, terms and relations, the explicit and formal representation of the domain model, the vocabulary to define the concepts and the nomenclature to support the composition of new concepts.

9.1 Introduction
As a consequence of the ageing population and the continuous development of new medical technologies, there is an increasing volume and intensity of medical services. Due to limited resources, issues such as budgeting, quality assessment and benchmarking [1] are introduced to control costs and quality of care [2]. In intensive care medicine, where costs are high due to expensive technologies and highly specialised personnel, assessment and accountability are particularly important, and are carried out world wide [3-5]. In the Netherlands, a foundation for National Intensive Care Evaluation (NICE) was established in 1996 [6]. The objective of NICE is quality assessment and quality improvement of intensive care units (ICUs) by comparing case mix adjusted outcome in various ICUs. Case mix adjusted outcome means the correction of expected outcome for variation in population characteristics (correction for age, diagnoses, severity of illness etc.). Initially, only the outcome measures mortality and length of stay (the latter used as a crude indication of cost) are considered in this quality assessment, because survival of critically ill patients is the primary concern in intensive care and the chosen outcome measures are objective to measure. At a later stage other outcome measures will be considered as well including quality of life and costs of care. The use of mortality as an outcome measure in quality assessment of intensive care is also found in the prognostic models, such as APACHE II, III [7, 8], SAPS II [9], MPMII/24II [10], to estimate case mix adjusted hospital mortality for intensive care patients. These models are mainly based on the measure of disturbance of physiological parameters in the first 24 hours of admission. The APACHE II and III systems are the only systems that also include the primary reason for intensive care admission to adjust for case mix differences. However, the diagnostic categories included in these models have not been empirically derived and are of varying levels of detail. We believe that the use of more structured diagnostic information will improve case mix adjustment and allow for more effective stratification. To date, classification and registration of diagnoses in intensive care is still unsatisfactory. In the Netherlands, as in many other countries, intensive care is not a
Chapter 9

recognised speciality and it is not mandatory to record intensive care diagnoses for
national registers. However, intensivists do record diagnoses in free text in the patient
record during the care process. With the introduction of Patient Data Management
Systems [11], a structured and systematic registration of diagnoses became essential.
This allows for the evaluation of the patient’s health problems over time and its relation
with therapeutic actions and outcome. Diagnostic information in intensive care will only
have sufficient quality and is therefore only usable for research studies when the
registration of diagnoses is part of daily care [12, 13], for example when the recorded
diagnostic information in the PDMS is used during shifts.
This chapter describes the development of an ontology that facilitates a terminological
system for intensive care diagnoses. A terminological system is a system that, based on
a specification of concepts and their interrelationship, provides terms that denote these
concepts. An ontology (formally) describes this specification of the concepts,
relationships and functions in a domain of interest [14]. Hence the notion of ontology
plays an important role in this study. Although in the literature “classification” is often
commonly used for all types of terminological systems, we will be specific in
terminology about terminological systems themselves. In this paper we use the
terminology conform the description in chapter 7.
The first step in the engineering process of an ontology that facilitates a terminological
system for intensive care diagnoses consists of an analysis of the objectives and
requirements for such an ontology, and the evaluation of existing terminological
systems on the basis of these requirements (section 9.2 and 9.3). None of the existing
terminological systems proved to completely satisfy the requirements. We therefore
develop a new ontology according to the steps mentioned by Rossi Mori [15] and
describe our choices concerning the design, the content and the representation of this
ontology (section 9.4). The result is a formal specification existing of a meta-model,
describing concepts, relationships between them and between concepts and terms in
general, such as constraints on the number of terms a concept may have; a domain
model describing concepts and their relationships in the IC-domain such as “Health
problem” and “Aetiology”; the IC-domain, which is the filling in of the domain model
formed by concepts such as “Viral hepatitis”; their definition in a vocabulary, e.g. viral
hepatitis is an viral infection of the liver; and a nomenclature to support the composition
of new concepts, e.g. viral hepatitis can be caused by the Epstein Barr virus, the
cyto megal o virus or hepatitis viruses.

9.2 Objectives and criteria for a terminological system to record ICU diagnoses

9.2.1 Objectives
To support the registration of diagnoses in intensive care in a structured and non-
ambiguous way, a terminological system for intensive care diagnoses is necessary. For
our purposes it has to serve two main objectives:
1. Provide a terminology to adequately describe the health status of a patient to support
daily care practice;
2. Provide a structure by which health problems of the patient population can be
described for the purpose of analysis and evaluation of medical and nursing care.
To achieve the first objective, comprehensive terms are necessary to designate medical problems requiring monitoring and treatment and to improve the communication during shifts between doctors and nurses in and outside the ICU.

To enable the use of diagnostic information for analysis and evaluation of care, which is the second objective, the structure of the terminological system should support aggregation on different levels of granularity. A terminological system with a compositional character (a so-called post-coordinated system) provides the possibility of composing new complex concepts by joining more basic concepts, e.g. infection + meninges + meningococcus. In a purely pre-coordinated classification each concept, no matter how complex, corresponds to a comprehensive term defined beforehand and has a single representational code, e.g. meningococcal meningitis. Pre-coordinated concepts are more comprehensible by readily available terms chosen by experts, compared to post-coordinated concepts in which syntax rules are necessary to guarantee sensible composites [16]. In a post-coordinated system, however, processing the information is better facilitated, for example the selection of terms may proceed faster because the structure provides different search entries by using the basic concepts, e.g. a meningitis caused by meningococcus can be reached by infection, meninges and meningococcus. Pre-coordinated terminological systems are usually harder to use for aggregation, only the hierarchy or structure in which the concepts are arranged can be used to aggregate diagnostic information. In case a pre-coordinated terminological system does have semantic relationships between concepts, aggregation and search are in principle made possible, but most pre-coordinated terminological systems do not represent these relationships, and, moreover, these cannot be generated dynamically.

9.2.2 Description of criteria

In many ICUs, the PDMS has been introduced to continuously collect and integrate various data from bedside devices, a hospital information system and manually entered data, and to present these in a clear format. To integrate the registration of diagnostic information into these systems, a terminological system for IC diagnoses should be incorporated in the information architecture of which the PDMS is a part (Figure 1.2). Most terminological systems include a structure which is based on hierarchical relations. Formally these structures respect the features of a semi-lattice in the sense that any two different concepts have a most specific ancestor that belongs to the hierarchy. This feature imposes constraints on the connectivity of the graph implied by the hierarchy. Next to this general feature many authors have attempted to identify the essential conceptual features of an ideal terminological system. The criteria for an ideal terminological system mentioned by Cimino et al [17] and Campbell et al [18] have been formalised in chapter 8. Below, we describe these criteria, related to the two objectives (daily care practice and research) of the terminological system for intensive care diagnoses.
The terminological system must support the intensivists in describing the diagnoses of patients in a structured way. For daily clinical use, the demand for domain completeness of the terminological system is probably the most important; the user must be able to classify a health problem in its full detail, such as “acute anterior myocardial infarction”. Attributes can be used to add extra detail, like the attribute value “anterior” could qualify the anatomical concept “myocard”. In this study we use domain completeness as the theoretical completeness of the domain, i.e. whether the domain can be extended in breadth (number of direct subordinates or children of a concept) as well as depth (number of levels in the hierarchy). We hence do not examine whether a term is really present in the terminology but rather if it could be in principle incorporated in it.

Synonyms (in more than one language) are highly desirable, as individual users will have their own preferred terms to search for concepts in the domain. Non-vagueness and non-ambiguity guarantee respectively that a term refers to a specific object in the domain and that the meaning of a term is unique. For browsing the concept hierarchy, multiple classification is needed, in order to allow the user to choose their own path to find the desired concept. Non-redundancy should prevent the user from finding different concepts for the same clinical object when following different paths.

As described in section 9.2.1, post-coordinated classifications support the aggregation of concepts in a natural way. Syntax rules should guarantee clinical sensibility of composed concepts. Multiple classification and explicit relationships are also important structure features to make aggregation for research available in a natural way. The next criterion is the use of context free codes. Context free codes have to be used so that the code itself does not hint at the concept it represents. The use of significant (context related) codes restricts the extendibility of the terminological system. When clinical insights evolve, the structure should not restrict the extension or changes in the ontology.

Since admission to the ICU is not an isolated event, information such as reason for ICU admission and complications during ICU admission, has to be available during the rest of the hospital admission. Because diagnostic information of the total hospital admission is usually recorded in a hospital information system using ICD-codes, mapping of ICU diagnostic information to ICD-9-CM or ICD-10 is desirable.

Chapter 8 extensively describes the evaluation of existing terminological systems with respect to the above described criteria. Since none of the evaluated terminological systems completely satisfied all criteria (see chapter 8), we developed a new terminological system that incorporates the strengths of the existing terminological systems. We formally describe this new terminological system in an attempt to overcome the limitations of the informal descriptions in the literature of existing terminological systems.
9.3 Design choices

In this section we clarify the representation formalism used and the design choices made for an ontology which is part of a new terminological system for intensive care diagnoses. Choices are based on the criteria, and the analysis of the features of the existing terminological systems, described in chapter 8.

9.3.1 Representation formalism

A good conceptualisation of a domain is a description which captures with fidelity the underlying concepts in that domain and the relationships between them. We require a representation formalism of our model which (1) is intuitive and conceptual - i.e. it lends itself for human comprehension and communication, (2) has adequate expressive power, and (3) has a formal basis to avoid ambiguity. The Entity-Relationship (ER) formalism and its extended form (EER), which are widely used instruments in the database community, satisfy most of these criteria [19]. ER is an attractive conceptual model that is capable of expressing concepts (entity types), attributes, relationships and constraints (e.g. cardinality constraints on relationships). However, the ER formalism may not always be suitable for expressing complex constraints. Hence, a more expressive formal instrument is needed to provide the semantics of the ER model and to complement it in order to capture complex constraints. This formal specification formalism is chosen to be many-sorted first order predicate logic (FOL hereafter for brevity) due to its universality, expressive power and its proven usefulness in representing clinical data [20]. There are, of course, many other formalisms which could have been selected. For example, one could have used the popular object oriented (OO) modelling notations such as OMT [21] or UML [22] instead of ER. Note, however, that the OO approaches provide more notational machinery, such as for the functional and dynamic aspects, than is strictly necessary for our purpose here. Similarly, one can consider conceptual graphs [23] or KIF structures in Ontolingua [24] as good alternatives to pure FOL. FOL is however more familiar by various research communities.

9.3.2 Concepts and terms

Synonymy, as well as non-redundancy and multi-lingual support can be achieved by means of separating concepts from their terms, that provide their names. This is shown in the ER diagram in Figure 9.1. The diagram indicates that a term may denote more than one concept and that a concept is described by terms. We would also like to express the fact that a term can be preferred for only one concept in a language and that each concept has one preferred term and zero or more synonymous terms per language. This sharpened description of the situation is shown in the lower part of Figure 9.1 by using FOL. Note that this is a meta-model of the domain where a 'model concept' in Figure 9.1 represents any concept in the domain model, such as "Health problems" and "Aetiology", which are described in section 9.4.1.
9.3.3 Pre-coordination versus post-coordination

As described in section 9.2.1 two fundamentally different ways are distinguished for the structuring of a terminological system: pre-coordination and post-coordination.

Pre-coordination is characterised by (a) sensible, non-ambiguous and non-vague terms with single representational codes; (b) terms which are useful in daily practice; (c) limited power of expression; (d) combinatorial explosion of terms in attempting to cover a domain. Post-coordination is characterised by (a) composing complex concepts by combining more basic ones; (b) syntax rules to control compositions; (c) great power of expression; (d) compactness of representation.

Terms used in pre-coordinated classifications can be more natural and useful compared to composed terms, e.g. “Meningococcal meningitis” is more comprehensible than the composition of the health problem “Infection” located in “meningen” caused by “Meningococcus”. The power of expression is closely related to domain completeness.

As trying to define all possible concepts (pre-coordination) will lead to combinatorial explosion with respect to the number of concepts, the classification should enable users to compose their own concepts (post-coordination) which makes compact classification possible. A set of syntax rules is necessary to prevent the composition of clinically nonsensical concepts. To strike a balance in expressiveness and compactness, a combination of pre-coordination and post-coordination can be sought in which a collection of selected pre-coordinated terms can be extended, turning it into a post-coordinated terminological system. A mechanism to check whether “new” composed concepts already exist as a pre-coordinated concept is essential to control consistency within the system e.g. during the maintenance of the system. In post-coordinated systems a concept can be extended with detail by adding associated concepts or
attributes to make a new composition. Whether a concept should be defined as a basic concept or as a composed concept is somewhat arbitrary. We have chosen to define health problem concepts at least at a level of clinical relevance for intensivists and IC-nurses or we mark a selected set of health problems, that themselves are not sensible, for compulsory extension to result in clinically relevant concepts. For example "meningitis" is a clinically sensible health problem concept although in most cases extra information about the causing micro-organism is available. However, the concept "fracture" without an anatomical location is not sensible as a health problem (it is a sensible dysfunction concept) because one does not know how to clinically interpret this health problem. This concept must be composed with at least an anatomical component such as "femur fracture", which is a clinically sensible health problem. The fact whether it is an "open" or "closed" fracture can then be further defined by making a new composite concept with an attribute value "open" or "closed".

9.3.4 Concept codes

In a terminological system in which new concepts can be composed, a unique code, which will be recorded in the PDMS, has to be generated automatically. This code, though context-free, cannot be chosen at random: if the terminological system is used at a number of independent sites, the same composed concept should generate the same code in order to facilitate interchangeability of information between the sites. To solve this problem, composed concepts get composed codes based on the unique codes of the basic concepts which are context-free. Although we are aware that this code construction may conflict with the context-free criterion in the sense that changes in the model can lead to invalid codes, the code itself will never be used to interpret the meaning of the concept. Instead the nomenclature, which will be described in section 9.4.2, will always be used to provide the underlying codes. In case all sites are connected to a shared terminology server each composed concept can get a random code which will be stored by the server.

9.3.5 Choices related to the content of the domain model and IC-model

In section 9.3.1 we described the formalism of our choice for representing the ontology. The next step is to decide on the content of the ontology, i.e. which concepts, relationships and restrictions are to be represented in the domain model which can be used to describe intensive care diagnoses, which we call health problems in the rest of this paper.

The divisions and semantic categories (concepts at higher level of aggregation) used in the existing terminological systems are mostly based on "anatomy or topography", "pathophysiology" and "aetiology or morphology". The domain model of the new ontology will be conformed as much as possible to those concepts generally accepted. Because of the similar domain, especially the divisions used in the ICNARC coding method are used as a basis for our domain model.

The choices of the concepts in the domain model, e.g. "anatomical components", and their instances in the IC-model, e.g. "heart", are refined and appended by discussing these with the domain experts. This discussion was meant to ascertain the concepts, relations and features the domain experts use to select and retrieve diagnoses or diagnostic categories. The acquisition of the IC-model follows the specifications and
restrictions in the domain model. Besides the concept “health problem”, we have chosen four other concepts at high level of aggregation in the domain model: “Systems” contains the body systems involved in the health problems. To support the translation of clinical concepts into ICD-10 concepts, we use those chapters of ICD-10 which are concerned with body systems, such as chapter 10 “Diseases of the respiratory system”. This division is also comparable with the second hierarchical level used in the ICNARC Coding Method. In Table 9.1 the relationship between ICD-10 chapters and our chosen concepts is shown.

“Anatomical components” contains all anatomical components associated with health problems. We used the Anatomical Localisation Classification from the Handbook Standardisation of Classification and Definitions in Health Care of the Dutch Classification and Terminology Committee for Health [25] as basis for this category. “Dysfunctions or abnormalities” contains organ failures, traumas and processes such as infections and haemorrhages. We mainly used the third hierarchical level, “process”, of the ICNARC Coding Method and a small number of concepts from the UMLS. “Aetiology” contains causes of health problems. It contains causes like micro-organisms causing infections, toxical substances causing intoxication and causes of trauma.

Health problems caused by other health problems, like “respiratory failure caused by meningitis” have a different kind of causality relationship than aetiology. This type of relationship is situated in a grey area because of the question whether it is part of the terminological system proper (as a concept) or whether it is part of the registration where two different concepts are recorded along with the relationship between them. The chosen concept at high level of aggregation in the domain model and the concepts in the IC-domain model serve three purposes:

1. support the user to select the precise diagnosis for individual patients by providing different search-entries to reach a diagnosis. For example “hepatitis” can be reached by the anatomical component “liver” or the dysfunction/abnormality “infection”;
2. facilitate evaluation research by supporting aggregation of diagnostic categories on the basis of (combinations of) associated concepts;
3. support the maintenance of the classification when new health problems are to be added, changed or declared obsolete. For example the terminological system creates a warning and asks for confirmation when a concept is added while there already exists a concept with the same characteristics.
### Table 9.1 Relationship between ICD-10 chapters and concepts in our IC-domain model.

<table>
<thead>
<tr>
<th>Chapters ICD-10</th>
<th>Systems</th>
<th>Anatomical components</th>
<th>Dysfunctions / abnormalities</th>
<th>Aetiology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infectious and Parasitic Diseases</td>
<td>N/A</td>
<td>N/A</td>
<td>Infection</td>
<td>Micro-organisms</td>
</tr>
<tr>
<td>Neoplasms</td>
<td>N/A</td>
<td>N/A</td>
<td>Tumor/malignancy</td>
<td>Hyperplasia</td>
</tr>
<tr>
<td>Endocrine, Nutritional and Metabolic Diseases</td>
<td>Endocrine system</td>
<td>Metabolic system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diseases of the Blood and Blood-forming Organs</td>
<td>Blood and blood-forming organs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mental Disorders</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diseases of the Nervous System</td>
<td>Nervous system</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diseases of the eye and adnexa</td>
<td>Eye</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diseases of the ear and mastoid process</td>
<td>Ear</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diseases of Circulatory System</td>
<td>Circulatory system</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diseases of Respiratory System</td>
<td>Respiratory system</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diseases of Digestive System</td>
<td>Digestive system</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diseases of Genitourinary System</td>
<td>Urogenital system</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conditions of Pregnancy, Childbirth and the Puerperium</td>
<td>Pregnancy related</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diseases of Skin and Subcutaneous Tissue</td>
<td>Skin and subcutaneous system</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diseases of Musculoskeletal System and Connective Tissue</td>
<td>Musculoskeletal system</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congenital anomalies</td>
<td>Congenital deformity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Certain Conditions Originating in the Perinatal Period</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Symptoms, Signs, and III-Defined Conditions</td>
<td>Dysfunctions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Injury and Poisoning</td>
<td>Trauma, Intoxication, Envenomation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classification of Factors Influencing Health Status and Contact With Health Services</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Classification of External Causes of Injury and Poisoning</td>
<td>Accident, Toxical substance</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

121
9.3.6 Composition of terms and concepts

The vocabulary and nomenclature, part of our terminological system, are (partially) specified by formal statements and constraints that describe the health problems and that support the composition of new concepts in a way which guarantees that the concepts are clinically sensible. It is possible to specify attribute values and compose new concepts on the basis of the relationships and their specifications defined in the vocabulary and nomenclature, giving the terminological system a post-coordinated character. The formal statements and constraints in the vocabulary and nomenclature support the maintenance of the terminological system by providing a basis for checking whether concepts with the same characteristics can potentially be equivalent. The (formal) statements can also support (semi)-automated classification, because new composed concepts are always subordinate concepts generically related to (at least) one parent which is already present in the nomenclature.

The relationships between the concepts belonging to the class “Health problems” and the associated concepts classes “Systems”, “Anatomical components”, “Dysfunctions/abnormalities” and “Aetiology” can be described by two kinds of specifications. Many of the relationships have the specification “define” which means that the health problem concept is (at least partly) defined by an associated concept e.g. the anatomical component “liver” is part of the definition of hepatitis. These relationships form the vocabulary. The second kind of specification of a relationship is the “refine” specification which means that extra detail in some given category can or should be added to further specialise a clinical concept, e.g. possible micro-organisms can be specified for the category “Aetiology” as the cause of a health problem. These relationships form the nomenclature. The way “define” and “refine” specifications in the vocabulary and nomenclature are used, is described in the following section.

9.4 Description of the ontology

The architecture of our ontology is inspired by [26] and consists of six components (Figure 9.2): (1) the meta-model and (2) the domain model represented by an ER diagram, (3) their formal specification including complementary information described in first order logic, (4) the IC-domain which is formed by the specialisation of the domain model, (5) the vocabulary for a (partial) definition of health problems and (6) the nomenclature for the specification of the allowed terminology compositions.

<table>
<thead>
<tr>
<th>Nomenclature</th>
<th>Vocabulary</th>
</tr>
</thead>
</table>
| Meta-model (Figure 9.1)  
e.g. “Concept” | Domain model (Figure 9.3)  
e.g. “Health problem” |
| IC domain model  
e.g. “meningococcal meningitis” | Formal specification in FOL |

Figure 9.2 Six components of the ontology for intensive care diagnoses.
9.4.1 The domain model and its formal description

At the heart of the domain model (see Figure 9.3) is the semantic category “Health problems”, representing diagnoses in the broad sense of the word. Each health problem may be related to specialised kinds of the health problem, like “Hepatitis B” is a specialisation of “Viral hepatitis” or to subpart health problems e.g. “Tetralogy of Fallot” consists of the health problems “Ventricular septal defect”, “Pulmonic valve stenosis”, “Infundibular stenosis”, and “Dextroposition of the aorta”. Two different subclasses of health problems can be distinguished: “Direct Health problems” and “SP_Health problems” (monitoring and care after a Surgical Procedure). The semantic category “Direct Health problems” is associated with the four semantic categories discussed earlier: “Systems”, “Anatomical components”, “Dysfunctions or abnormalities”, “Aetiology”. With the exception of “Systems”, all semantic categories are top nodes of sub-classifications, in which different levels of detail can be found (this is not shown in Figure 9.3 for brevity). This supports the user in selecting and retrieving diagnoses from different views and different levels of detail.

In ICUs, a large group of patients are to be monitored after an extensive operative procedure. For this reason the semantic categories “SP_Health problem” and “Surgical procedures” are parts of the conceptual model. This is an example where the domain (intensive care) affects the choice of the concepts. The semantic category “Surgical procedures” is linked to the semantic categories “Systems”, “Anatomical components”, “Dysfunctions or abnormalities”. If a patient arrives for care after a surgical procedure, it is the procedure and not the health problem which is related to the associated semantic categories. For example, a patient arriving at intensive care after a Coronary Artery Bypass Graft (CABG), the health problem “monitoring after CABG” is only linked to the surgical procedure “CABG” which is related to the “cardiovascular system”, the anatomical component “coronary arteries” and to the dysfunction “obstruction”.

Currently we restrict the number of concepts in the domain model to the five ones mentioned above, but in the future an extension of concepts, which is one of the principles for the design of an ontology [14], is possible. This is contingent on the use of the ontology in daily practice.

Recall that there is a relationship between every concept and its associated terms. Each concept is described by one or more terms, preferred or synonymous, originating from different languages. This relationship is not shown in Figure 9.3 since it is already specified as a shortcut in the meta-model appearing in Figure 9.1. The formal specification in FOL includes additional information that does not appear in the ER model. For example constraints concerning the description of specialised health problems are formally described in FOL. Asymmetry for example is described in FOL by $(\forall h1.h2. \text{Health\ problem\ Specialised\ problem}(h1,h2) \rightarrow \neg\text{Specialised\ problem}(h2,h1))$ which means that when a health problem “h1” is a specialised health problem of the health problem “h2”, this health problem “h2” cannot be a specialised health problem of health problem “h1”. This constraint, among others, is to be enforced by the acquisition tool of the concepts in the first place but can not be easily described in the ER model.
Figure 9.3 The (simplified) conceptual model of the domain model. Note that, for brevity, attributes are not showed. We used the notations in [27] where numbers indicate cardinality constraints. Note the convention of placing the numbers at the edge of the diamonds denoting relationships, instead of at the edge of the arrows.
9.4.2 The vocabulary and nomenclature

The last components of our model are the vocabulary and nomenclature which define the concepts and support the composition of new concepts. In Table 9.2 a simplified part of the vocabulary and nomenclature is presented. In this example “Direct Health problems” and their related concepts of the four associated semantic categories and the relationship with “Specialised_problem” are presented. “SP_Health problems” are not presented for brevity. Although we here represent the definition of “Viral hepatitis” in a table, it can be formalised in FOL by:

\[ VH(x) \land \text{Anatomy}(x, \text{liver}) \land \text{Dysfunction}(x, \text{infection}) \land \exists_{\text{ve} \text{virus}} \text{Aetiology}(x,v). \]

Table 9.2 (Simplified) Part of the vocabulary and nomenclature:

<table>
<thead>
<tr>
<th>Direct Health problem</th>
<th>Systems</th>
<th>Anatomical components</th>
<th>Dysfunctions/Abnormalities</th>
<th>Aetiology</th>
<th>Specialised_problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viral hepatitis</td>
<td>Def(Digestive system)</td>
<td>Def(Liver)</td>
<td>Def(Infection)</td>
<td>OR_specialisation (Hepatitis viruses, Epstein-Barr virus, Cytomegalo virus)</td>
<td>Hepatitis A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hepatitis B</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&quot;......&quot;</td>
</tr>
<tr>
<td>Hepatitis B</td>
<td>Def(Digestive system)</td>
<td>Def(Liver)</td>
<td>Def(Infection)</td>
<td>Def(Hepatitis B virus)</td>
<td>&quot;......&quot;</td>
</tr>
</tbody>
</table>

- Def (X): X is an intrinsic characteristic of the direct health problem
- OR_specialisation(X): a new concept, based on the concept mentioned between brackets, can be composed using X or one or more of X’s subordinates
- XOR_specialisation(X): a new concept, based on the concept mentioned between brackets, can be composed using X or exactly one of X’s subordinates
- DOR_specialisation(X): X is an intrinsic characteristic of a concept, a new concept can be composed using one or more of X’s subordinates
- DXOR_specialisation(X): X is an intrinsic characteristic of a concept, a new concept can be composed using exactly one of X’s subordinates

The first row of Table 9.2 shows that the concept “Viral hepatitis” of the concept “Direct Health problem” is defined by “Digestive system”, “Liver” and “Infection”. This implies that it is not useful to combine the term “Viral hepatitis” with “Liver”, since “Liver” is already part of the definition. Furthermore a refine-specification, “OR_specialisation”, exists between “Viral hepatitis” and several viruses mentioned in the semantic category “Aetiology”. As indicated in the legend below Table 9.2, “OR_specialisation” prescribes that the concepts mentioned under “Aetiology” or one or more of their subordinated concepts can be chosen to further refine the concept “Viral hepatitis”, e.g. theoretically a health problem can be composed with two causing micro-organisms “Viral hepatitis caused by Cytomegalo virus and Epstein-Barr virus”. The concept “Hepatitis viruses” under “Aetiology”, is related to subordinated concepts, e.g. hepatitis A virus, hepatitis B virus etc. (Note: although the specification is called “OR_specialisation”, because it allows one or more concepts to be considered, the result is a conjunction of the concepts chosen and not a disjunction).
“XOR_specialisation” is used to indicate that only one of the concepts or one of their subordinated concepts may be used to refine the health problem, e.g. “Viral hepatitis caused by Epstein-Barr virus”. If we would indicate DOR_specialisation(Hepatitis viruses), an OR_specialisation with a defining component, we restrict the refinement of “Viral hepatitis” to one or more of the subordinated concepts of “Hepatitis viruses” (the concept “Hepatitis viruses” by itself may not be used as it is a defining property). In accordance to the semantics of “XOR_specialisation”, DXOR(Hepatitis viruses) restricts the refinement to only one of the subordinated concepts of “Hepatitis viruses”. The concepts “Hepatitis A”, “Hepatitis B” under “Specialised_problem” in the first row indicates that these are specialised concepts of “Viral hepatitis”. The second row in Table 9.2 is the representation of “Hepatitis B”. All associated concepts of “Hepatitis B” define this health problem.

In some cases we decided to include complex composed health problems as distinct concepts in the vocabulary and nomenclature although they could also be composed from a health problem on a higher aggregation level (like “Hepatitis B” can be composed from “Viral hepatitis” in Table 9.2). The decision to do this is based on the prevalence of the health problem in the intensive care population or on the fact whether intensivists use different terms for the complex composed health problem than can be derived from the composition of the terms describing the composed concepts. Rules will be defined to guarantee similar codes for the same concept, no matter if it is composed or not (section 9.3.4). The knowledge engineer is responsible for this process.

Changes in the vocabulary and nomenclature are not expected to occur frequently but they may occur due to either new medical insights or to modelling decisions of the designer especially during the phase of setting up the vocabulary and nomenclature. Changes in the vocabulary and nomenclature are local in nature. A typical change in a concept usually implies looking at its immediate parents and immediate children and performing simple checks on them resulting in the modification of their entries in the vocabulary and nomenclature. In this way a change in a concept (such as its move within the hierarchy) does not have to propagate beyond its direct parents and children. This is because of the way concepts are defined in the vocabulary and the modularity of the model which helps to control the effects of changes.

9.4.3 Implementation

The ontology is implemented in a Java application. We have chosen to develop the software in Java because of its platform independence. This is an important feature because ICUs use various Patient Data Management Systems running on different platforms. In the future the terminological system has to be incorporated into these PDMSs. At this moment approximately 1000 health problems are defined. All health problems and their relationships are defined with the help of domain experts. In a pilot study, described in chapter 10, two intensivists have evaluated the health problems and their relationships to the associated concepts defined in the vocabulary and nomenclature for the use of selecting and aggregating diagnoses.
9.5 Summary and Discussion

In this section we summarise the objectives of our study, the analysis of the requirements of a terminological system for intensive care diagnoses as described in section 9.2.2 and chapter 8 and the design of a new ontology that facilitates such a terminological system. Finally, we provide some critical observations about existing terminological systems and about our design choices.

Information about the patient’s health status and the actual medical problems play an important role in stratifying the patient population for quality assurance of intensive care. To date no systematic registration of diagnoses, health status and medical problems has been used in daily practice in intensive care, mainly due to the lack of appropriate terminological systems. A terminological system which supports the description of the patient’s health problems as part of daily care practice and which supports the aggregation of diagnostic information, is essential. These two objectives resulted in the prioritisation of criteria for terminological systems which were defined by Cimino et al. and Campbell et al. Because none of the evaluated terminological systems completely satisfied all our objectives we decided to design a new terminological system using the strengths of the evaluated systems.

We have engineered an ontology for intensive care diagnoses, but the engineering approach and the general ideas are applicable to a broader spectrum of medical domains. We used ER representation techniques to design the meta-model and domain model of this ontology. Because the ER modelling technique is not appropriate to describe all information about the domain we also described the requirements formally in FOL, which is more expressive. However, other equally conceptual and expressive formalisms, such as UML and OMT (for conceptual modelling) and conceptual graphs and Ontolingua (for the formal aspects), could have been used as well. Whatever specific notations are used, the combination of conceptual and formal specifications provides important advantages. The great majority of current terminological systems miss formal specifications. A challenging research issue is the automatic reasoning with (part of) these specification for example as attempted in ontology based knowledge acquisition approaches such as in GAMES [28] and PROTEGE [29].

In the domain model, concepts at high level of aggregation, which we believe are useful in describing health problems in the intensive care population, were derived from existing terminological systems. Although the UMLS contains numerous concepts (called semantic types in UMLS), only a very limited number of these concepts were useful for our purposes.

There are some similarities between our ontology and the designs of existing terminological systems. The evaluation of terminological systems according to formalised criteria as described in chapter 8 supported us in our design choices.

We used the ICM levels in our domain model since ICM is the only terminological system intended for IC diagnoses.

There is a similarity between our model and SNOMED. Since we believe not all health problems can be defined intensionally and since clinical relevance and non-redundancy are important requirements, we did not follow the post-coordination and compositional terms used in SNOMED because they lack syntax rules. However, there is now some
research to incorporate semantic information to SNOMED, called SNOMED RT, using
conceptual graphs in order to constrain the post-coordination [30, 31].
The NHS Clinical Terms and our model also show some similarities. The main
difference lies in the structure: NHS Clinical Terms uses a strictly hierarchical list of
concepts which can be modified by attributes in a controlled way, whereas we use a
conceptual model which includes explicit semantic relationships with explicit
constraints. The attributes in NHS Clinical Terms are not consequently used for all
concepts in the domain, which restricts the use of these attributes for aggregation, e.g.
although the attribute "site" exists, not every concept related to a site is actually linked
with a site. Furthermore, formal specifications, which could control consistency
between combinations of attributes of a concept, are lacking.
We are conscious of similarity with the GALEN project [32-34] which is an ambitious
project aiming at a formal description of the total medical domain. The fact that
GALEN is very expressive, implies that a great effort should go into the syntax rules to
guarantee sensible composed concepts. We chose for another balance in expressiveness
and compactness also because we are tackling a specific medical domain i.e. intensive
care. In our ontology, domain knowledge is presented in a domain model and
composition and definition rules are included in a nomenclature and vocabulary. The
CEN prestandard ENV 1828 [16] also includes composition rules but these are informal
and concentrates only on surgical procedures. The existence of our "refine"-
specification on relationships, and the explicit constraints on the relationships between
concepts give more expressiveness than in pre-coordinated terminological systems such
as ICD-10 but take far less effort in sensibility checks than used in GALEN. We think
there are also some practical disadvantages to the use of GALEN in our set up like the
extensive training course necessary before the GALEN software can be used and before
a part of the CORE model can be implemented and extended. Furthermore we require
total control over the software by which the ontology and terminological system are
developed and this could not be realised with the GALEN software at the time of
development. However, because of the opportunities described by the GALEN project
we intend to co-operate with the GALEN organisation and perform an assessment of
the CORE model and the GALEN software to get better insight into the strengths and
weaknesses of both GALEN and our methods.
Currently the ontology which is implemented in a Java application is filled with
approximately 1000 health problems. We will incorporate it in the Patient Data
Management Systems. The next chapter describes a test-phase in which the
terminological system has been evaluated. After sufficient testing and adaptation of the
IC-domain it could be implemented into daily care practice. Once this terminological
system has been integrated into the daily care practice and into the existing PDMSs, a
better insight into the patient population will be possible. This information is essential
in the quality assurance and improvement of intensive care.

Acknowledgements
We would like to thank the intensivists Evert de Jonge and Rob Bosman for their co-
operation and valuable assistance in developing the conceptual model of the intensive
care diagnostic domain and the preliminary evaluation of the prototype. Thanks are also
due to Peter Lucas for providing helpful remarks on an earlier draft of this manuscript.
References


