Terminological systems and prognostic models as instruments for quality assessment in intensive care

Raiez, F.

Citation for published version (APA):
Introduction
1.1 Introduction

In recent years, measuring the quality of health care has received growing attention by health care processonals as well as policy makers. In their quest to provide efficient and high quality care and to limit health care expenditures, health care providers and health care managers increasingly rely on performance indicators, i.e. outcome summaries such as mortality, which are believed to reflect the quality of health care. For instance, by comparing performance indicators across different institutions, it becomes possible to identify high and low performance institutions and thereby to identify the processes of care that result in better or worse outcomes and to learn from the better performing institutions (1-3). Lately, there is also a trend towards publishing performance indicators to stimulate public dialogue on health care performance (4;5). Such publications might lead to changes in the behavior of patients and health care providers (6;7).

This evolving accreditation of quality performance necessitates the need for standardized performance indicators (i.e. output of the quality assessment process such as aggregated outcomes) and performance data to calculate these performance indicators (i.e. input for the quality assessment process such as patient outcomes).

It is impossible to reliably assess the quality of health care by solely interpreting aggregated outcomes without accounting for case-mix differences. Differences in mortality rates, for example, could reflect a complex patient case-mix rather than poor quality of health care. Consequently, when comparing quality of care across institutions we must ensure that the groups of patients are sufficiently similar to make the comparison fair and meaningful.

Therefore, case-mix adjustment is the heart of quality assessment and often a prognostic model is used to adjust the raw performance data for patient case-mix (e.g. diagnoses, age, and sex) differences (8). To date, several prognostic models have been developed for different patient groups such as the Intensive Care (IC) population.

Without high quality performance data, quality assessment using prognostic models has little meaning and ongoing evaluation and improvement of performance data is essential (8). The completeness and quality of performance data used for case-mix adjustment can vary between institutions, resulting in erroneous performance measurements. Hence, the quality of performance data should be optimized by standardizing data collection through standard registration procedures, unambiguous data definitions, and standardization of documentation by means of coded data (9). The increased use of enhanced Information Technologies (IT) to electronically store patient data in Electronic Health Records has contributed to the standardization of a large number of the numeric data items needed for case-mix correction. Nevertheless, little attention has been paid to the use of terminological systems to standardize health care–related phrases in medical files such as diagnoses and clinical encounters, which also forms an important element of case-mix correction for quality assessment. Terminological systems are systems that compile the concepts from a
medical domain and aid consistent and computer-readable coding and storage of clinical data (10).
This thesis addresses the use and application of terminological systems for consistent data entry and the use and application of prognostic models as instruments to facilitate quality assessment in the intensive care. This chapter first introduces the domain of intensive care and describes the significance of quality assessment in the intensive care. Next, an introduction to the use of terminological systems for structured data capture is provided. Finally, the use of prognostic models in the intensive care for quality assessment is described. The chapter concludes with an outline of the thesis.

1.2 Intensive Care

Intensive care has been defined as “a service for patients with potentially recoverable conditions who can benefit from more detailed observation and more invasive treatment than can safely be provided in general wards or high dependency areas” (11). The idea of intensive care originates from the second half of the 20th century and was first established as a response to a polio epidemic where many patients required constant ventilation and surveillance. In the next decades, the concept of intensive care expanded enormously. The advances in supportive care and monitoring resulted in significant improvements in the care of severely ill patients, such as patients with life-threatening infections. Consequently, in most countries the demand for intensive care services in terms of number and size of units has been growing (12). Nowadays, Intensive Care Units (ICUs) are complex organizations involving various medical domains to treat patients with a substantial risk of mortality (13). Also the costs of these services are increasing, with highly educated personnel and new technologies consuming constantly larger portions of the health care budget (14).

The prominent role of intensive care in healthcare systems makes it one of the major targets of the quality assessment process in hospitals. For quality assessment insight is required in the clinical and organizational process of the ICUs. Yet, defining and measuring intensive care performance are complex tasks. Detailed observation of severely ill patients involves specially trained personnel with a broad range of clinical experience in a highly technical environment. Various (monitoring) devices are available at the patient’s bedside, generating a mass of data. And so, ICUs are not only labor-intensive environments, but also data intensive. The large amount of data can overwhelm care providers when interpreting and acting upon it. Also, different involved parties require different levels and amounts of information for several purposes, including quality assessment, publication of performance indicators to inform the public, for benchmarking against peers for feedback and improvement, and for research (8). Standardization of the performance data and presenting it by means of quality indicators can support the involved parties in the decision process.
To this end, several initiatives have been undertaken by ICUs to collect the data in medical registries for quality monitoring and quality improving purposes (15-18). In the Netherlands, the National Intensive Care Evaluation (NICE) foundation was set up by the professional group of intensivists to focus on quality assessment of the Dutch ICUs.

1.2.1 The NICE registry

The NICE foundation was established in 1996 (15). In that year, on the initiative of intensive care physicians from the NICE foundation, six institutions started collecting data on IC admissions in the NICE registry. The aim of the NICE registry was to facilitate quality monitoring and quality improvement initiatives in Dutch ICUs. The registry contains all data on demographics, severity of illness in the first 24-hours of admission, diagnoses, and outcomes, necessary to calculate among others the Acute Physiology and Chronic Health Evaluation (APACHE), the Simplified Acute Physiological Score (SAPS), and the Mortality Probability Model (MPM) prognostic models (19-21). By 2008 also some structure and process indicators as defined by the Dutch Society of Intensive Care (NVIC) were added to the registry. Standardized Mortality Ratios (SMR) from the NICE registry are used by intensive care physicians and intensive care managers as the case-mix-adjusted outcome indicator to measure quality of intensive care, to compare the outcomes of their own ICU to outcomes of other ICUs or to the national average, or to perform longitudinal analyses.

At the moment (June 2010), 80 ICUs are voluntarily participating in the NICE registry, which amounts to about 80% of all ICUs in the Netherlands. The participating ICUs are a mixture of tertiary, teaching and non-teaching settings in urban and non-urban hospitals. The registry contains information on more than 400,000 admissions providing a rich basis for quality assessment. Furthermore, given its wide variety of data items, the high quality of this data and its representativeness for the Dutch ICU population, the NICE registry forms also a valuable instrument to perform scientific research in the field of intensive care (15;22).

1.2.2 Diagnosis coding in the Dutch intensive care

To be useful for quality assessment and research purposes, data from medical registries such as the NICE registry must be consistent and of good quality. In the last decade, the increased use of Patient Data Management Systems (PDMSs) in the Dutch ICUs has enhanced the collection, processing and retrieval of large parts of the data items produced by monitoring devices and laboratories. Nevertheless, little attention is paid to the standardisation of diagnoses, i.e. clinical terms used to describe a clinical finding. Diagnoses and in particular the “reasons for IC admission” are essential components of
case-mix correction in the prognostic models for quality assessment in the intensive care and thus in the NICE registry. Currently, diagnostic information is often recorded more than once in different systems in free text (unstructured data) or using special-purpose classifications. For calculation of mortality risks, for instance, this data is captured using the APACHE II and APACHE IV reasons for IC admission classifications (19;23), while the same information is often also captured separately in free text in medical files for day-to-day care processes and once again in discharge letters. Systematic recording of diagnoses is central to the use, exchange, analysis, and interpretation of this data.

1.3 Standardized diagnosis coding by means of terminological systems

Terminological systems are meant to support the process of capturing patient data in a structured and standardized way by organizing and interrelating concepts of a particular domain and providing concepts with related terms and possibly definitions and codes (10). During the last decades, many terminological systems have been developed for different purposes, serving different functional requirements with regard to their intended use and domain (24). Classifications, also known as aggregate terminologies, such as the International Classification of Diseases (ICD-9), are used to categorize disease encounters into disjoint classes (24;25). Classification systems allow granular clinical phrases to be aggregated into manageable categories for quality assessment, monitoring resource utilization, or reimbursement. The lack of formal semantic definitions for classes and the lack of detail in most aggregate terminologies results in shortcomings when aiming for reuse and logical reasoning. Therefore, classification systems are considered “output” rather than “input” systems and are not intended for the primary documentation of clinical processes (26). Reference terminologies such as Systematized Nomenclature of Medicine-Clinical Terms (SNOMED CT) are meant to address these issues by providing detail and precise meaning to data by formal concept definitions, required for complete and consistent coding of clinical data and logical reasoning (27). Nevertheless, reference terminologies are inadequate for serving the purposes for which classification systems are used because of their immense size, considerable granularity, complex hierarchies, and lack of reporting rules. Therefore, neither a reference terminology nor a classification system can, by itself, serve all of the purposes for which health information is currently used or will be used in the future. Finally, interface terminologies are used for actual data entry into electronic medical records, facilitating display and collection of clinical data in a simple way while simultaneously linking users’ own descriptions to structured data elements in a reference terminology or aggregate terminology (27). Interface terminologies provide easier adoption to the user requirements, i.e. the non-relevant content of the underlying reference
terminology or classification system is excluded and relevant concepts and terms which are not present in the underlying reference terminology or classification are added to the interface terminology.

1.3.1 SNOMED CT
SNOMED CT is the world's largest concept-based terminological system containing 310,000 medical concepts, associated with 990,000 description terms for these concepts, and related to each other by about 1.38 million relationships (January 2010 release). SNOMED CT provides the clinical content and expressivity required for precise clinical documentation and aims for capturing this information in a way that the information can be retrieved for reuse. To this end, each concept is uniquely identified and is placed in a hierarchy by which it is related to other SNOMED CT concepts. Individual concepts may appear in more than one place in the hierarchy. The relationships are used to define a concept where it can be expressed in terms of other concepts in the terminological system (e.g. Bacterial pneumonia is defined as: Is A Bacterial lower respiratory infection; Is A Infective pneumonia, with Causative Agent: Bacteria; Finding Site: Lung structure). SNOMED CT enables also post-coordination, i.e. composing new concepts by qualifying existing concepts with more detail (e.g. qualifying the concept “bacterial pneumonia” with additional information on the microbiological agent). Relationships are also used as a mechanism which allows logical reasoning about the information in SNOMED CT.

SNOMED CT is regarded as by far the most comprehensive reference terminology for coding clinical data. SNOMED CT is comprehensive on its own, but also maps to (aggregate) terminologies already in use. Mapping from SNOMED CT to other medical terminologies and classification systems avoids duplicate data capture, while facilitating granular coding of data for day-to-day care processes, enhanced quality assessment, and statistical analysis.

In the intensive care, SNOMED CT can be used to capture diagnostic information which will not only aid the calculation of mortality risks in prognostic models, but will also facilitate sharing and aggregation of data for other purposes.

Yet, although SNOMED CT does include interface terminology capabilities, its use as an interface terminology in a specific clinical setting is doubtful (28;29).

Instead of providing all of SNOMED CT to the users, an interface terminology based on SNOMED CT is proposed to provide easier adoption to the user requirements, i.e. the non-relevant content of SNOMED CT is excluded and the relevant concepts and terms which are not present in SNOMED CT are added to the interface terminology (27).
1.3.2 Terminological System Maintenance

The dynamic character of medical knowledge makes the maintenance of medical terminological systems such as SNOMED CT necessary. With the regular changes in the terminology content, incorrect representation of the knowledge might occur in the terminological system. Furthermore, frequent changes in the terminology content can lead to difficulties in the processing and tracking down of historical data. Therefore, to sustain their utility, it is necessary that the maintenance of terminological systems is performed in an organized and structured way. Without a well-structured and standardized maintenance process, terminological systems cannot provide the quality required by today’s medical applications (30;31).

Research questions

It has been argued that the terminology-based data entry in clinical settings does not only depend on the terminological system’s content but also on performance and the design of the graphical user interface of the terminology client (32). Regarding the use of terminological systems for structured data entry in the intensive care, this thesis focused on the methodologies underlying the process of designing, implementing, and maintaining terminological systems, in particular SNOMED CT, in the intensive care by addressing the following research question.

*Can a terminological system be used to standardize diagnosis coding in the intensive care and what characterizes its development, presentation and maintenance?*

1.4 Standardized quality indicators by means of prognostic models

Next to consistent performance data, appropriate performance indicators are also important for reliable quality measurement. Risk adjustment is the term commonly applied to those methods that account for case-mix differences, making performance indicators comparable across providers or institutions. This procedure provides the opportunity to uncover and institute “best practices” and to improve the quality of health care (33).

The existing intensive care prognostic models such as APACHE II and IV, SAPS II and MPM0/24II are logistic regression models mainly concerning mortality by providing an estimate of the mortality in specific patients based on their characteristics (e.g. severity of illness, demographics, reasons for IC admission etc.) (19;34-36). Case-mix adjusted outcome is typically assessed by calculating the SMR, which is the ratio of observed and expected mortality (based on the prognostic models), in a population. The SMR gives an overall estimation of performance over a certain period (cross-sectional) and can be used
for benchmark purposes, i.e. the process of comparing a quality indicator of one ICU to for example the national average or to other ICUs (37).

The use of performance indicators as a basis for performance assessment and the use of prognostic models to obtain case-mix adjusted performance indicators has been the subject of discussion in many studies (1;3;38-43). Research has shown that different prognostic models frequently produce different impressions about health care performance (44-46). These differences may be caused by choices that were made during the model development process; prognostic models are only an approach to reality and different prognostic models emphasize different aspects of the case mix; the difference may also be caused by the fact that the models were built at different times and for different patient populations. It has been argued that in the latter case, customization of the model is necessary to avoid poor prognostic reliability (23;47;48).

Research question:
When the case-mix adjustments by a prognostic model are unreliable, then the same may hold for the predictions and thus for the results of the quality assessment based on that model. Regarding the use and evaluation of prognostic models for quality assessment, NICE data was used to address the following research question:

How do prognostic models perform in the Dutch ICUs and does the choice of a prognostic model and its customization affect the results of quality assessment?

1.5 Effect of standardized data on quality assessment
As mentioned before, the family of APACHE prognostic models is widely applied in the intensive care to adjust observed raw mortalities for case mix differences in order to assess the quality of health care (49). Although the detailed registration of the related reasons for admission in the recent APACHE IV model enables a more accurate prediction of mortality risks, to preserve the continuity of the health care assessment, also the older APACHE II prognostic model is still applied. The APACHE IV prognostic model is backward compatible with its forerunner, the APACHE II prognostic model, except for the related reasons for IC admission classification. Consequently, the primary reasons for IC admission in the NICE registry are collected based on both APACHE classifications, rendering registration inefficiency.

To reduce registration inefficiency in the intensive care, clinical information in the ICUs should preferably be captured on a detailed level in the PDMS using a reference terminology such as SNOMED CT which holds the promise to aggregate clinical encounters in terms of different classification systems. Since this is not yet the case, to reduce the registration burden in the NICE registry, and to preserve the continuity and
interoperability of health care documentation, a mapping between these successor classification systems is necessary, i.e. a mapping from APACHE IV to APACHE II. It has been argued that one way to create a mapping between such classification systems is to use a granular reference terminology as an intermediate (50). An advantage of using a reference terminology to create a mapping is that once a mapping has been created between the reference terminology and the target classification system, this mapping can be re-used to identify crossmaps to all other (versions of this) classification systems which are mapped to the reference terminology. In the future, it is not inconceivable that an even more detailed successor of the APACHE model will be available. In that case, the mappings between the reference terminology and the APACHE II classification and the APACHE IV classification can be re-used to identify crossmaps between the existing and the newly available APACHE classifications.

For instance, reference terminologies such as SNOMED CT are aimed to facilitate data entry in the PDMS during the course of patient. Reference terminologies are inadequate for serving the purposes for which classification systems, e.g. the ICD, are used. The benefits of using SNOMED CT increase if the reference terminology is mapped to standard classification systems. Such mapping enables data re-use for the purpose of generating health information necessary for statistical analysis and reimbursement. Furthermore, the 10th revision of the International Classification of Diseases, ICD-10, for instance, was developed in 1993 to replace ICD-9, which stems from the 1970s. Many changes were made to coding in ICD-10-CM (51). These changes have a direct impact on the health care systems in that the changes produce inconsistencies and discontinuities in the health care statistics (52-54). A crossmap between such classification systems, would assist the continuity and interoperability of health care documentation and avoid double registration. It has been argued that a reference terminology can serve as an intermediary to support the creation of a mapping between two (versions of a) classification systems (50).

**Research question**

In the intensive care, SNOMED CT could be used as an intermediary to create a crossmap between the APACHE IV and APACHE II classifications to ensure continuity of quality assessment in the NICE registry. To evaluate this, the following research question is addressed:

*What is the effect of using SNOMED CT to crossmap two intensive care specific classification systems on quality assessment?*

**1.6 Outline of this thesis**

To address the research questions mentioned above, different studies were performed.
Chapter 1

Chapter 2 describes a study that evaluated the agreement between free-text reasons for IC admission captured in a PDMS and reasons for IC admission that were recorded using a locally developed post-coordinated terminological system embedded in the PDMS. This study was conducted to evaluate whether clinicians adequately capture reasons for admission by using a terminological system.

Chapter 3 and Chapter 4 focus on the development, application and evaluation of a domain-specific interface terminology based on SNOMED CT. Chapter 3 provides a generic approach for developing a domain-specific interface terminology based on SNOMED CT and describes an application of this methodology to the domain of intensive care. Chapter 4 describes the usability evaluation of this interface terminology to register the reasons for IC admission in a PDMS using the terminology service from Chapter 2.

Chapter 5 provides a framework that summarizes the features that the maintenance process of a terminological system should cover to sustain its content accuracy. This framework is applied to gain insight into the current maintenance practice of existing terminological system.

Chapter 6 and 7 focus on the use and evaluation of prognostic models for quality assessment in the intensive care. Chapter 6 illustrates the influence of choice of a prognostic model and the effect of customization of these models on quality assessment and quality comparison between ICUs. Chapter 7 focuses on the performance of a newly available prognostic model to predict hospital mortality in comparison to the performance of the widely applied older prognostic models.

Chapter 8 combines the different objectives of this thesis by evaluating the effect of the use of a terminological system to crossmap two intensive care specific classification systems on the calculation of case-mix adjusted mortality risks for quality assessment.

This thesis concludes with Chapter 9, which provides an overall discussion of the work in this thesis. The merits and limitations of the different studies are addressed and recommendations for further research are presented.

Bibliography


