Terminological systems and prognostic models as instruments for quality assessment in intensive care
Raiez, F.

Citation for published version (APA):
Raiez, F. (2010). Terminological systems and prognostic models as instruments for quality assessment in intensive care

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.
Summary
Over the last two decades there has been an increasing interest in measuring the quality of health care and, in particular, in quantitative comparisons of performance between health care institutions. In recent years there is also an increased trend towards publishing these performance outcomes. As such publications can lead to disciplinary measures against hospital organizations and to changes in the behaviour of patients and providers, it is important that the instruments used for performance assessment are reliable.

Quality assessments are usually based on a quality indicator, e.g. an outcome summary such as mortality, which is believed to reflect the quality of care. These outcomes summaries might be influenced by case-mix, i.e. specific characteristics of a population such as diagnosis, age and sex. A prognostic model is often used to adjust these crude outcomes for case-mix differences, making quality indicators comparable across providers or institutions. When the case-mix adjustments by a prognostic model are unreliable, then the same may hold for the outcome predictions and thus for the results of the quality assessment based on that model. At the same time, when the quality of the performance data, i.e. data used to predict the case-mix adjusted outcomes, is poor, quality assessment has little meaning. The completeness and quality of performance data used for case-mix adjustment may vary between institutions, resulting in erroneous performance comparisons.

The use of terminological systems might help overcome (parts of) these variations between institutions.

This thesis elaborates on these topics by addressing the use and application of terminological systems for consistent data entry (Chapters 2 to 5) and the use and application of prognostic models (Chapters 6 to 7) as instruments to facilitate quality assessment in the intensive care (IC). Chapter 8 then combines the two subjects 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.

Chapter 2 describes a study that evaluated the agreement between reasons for IC admission that were recorded in a patient data management systems (PDMS) as free-text and reasons for IC admission that were recorded using a locally developed post-coordinated terminological system embedded in the PDMS. Both reasons for IC admission were captured in the PDMS by clinicians during regular care practice. Each pair was judged as exact match, partial match or mismatch by two independent raters. Partial matches were further analyzed to investigate whether free-text or terminological system-based reasons for IC admissions included more detail and whether these differences could be explained by the content, the interface of the terminological system or by user or usability characteristics. It was shown that the correctness and specificity of both ways of registration were comparable. The use of post-coordination led to some difference in detail but the level of
Summary

Detail was comparable to free-text registration of reasons for IC admission. Details missing in the terminological system-based registration of reasons for IC admission were most often available in the terminological system, indicating that user interaction with the system is more of an impediment than the contents of the terminological system. These results are encouraging to replace free text recording of reasons for IC admission by terminological system-based recording, thereby contributing to patient data re-use for e.g. clinical research or quality assessment.

Chapter 3 defines a generic approach for developing a domain-specific interface terminology based on SNOMED CT for IC. This process was regarded as consisting of six sequential phases: domain analysis, mapping from the domain concepts to SNOMED CT concepts, creating the SNOMED CT subset guided by the mapping, extending the subset with non-covered concepts, constraining the subset by removing irrelevant content, and deploying the subset in a terminology server. This approach was applied to develop an IC-specific interface terminology based on SNOMED CT. The APACHE IV classification, a standard in the IC with 445 diagnostic categories, served as the starting point for designing the interface terminology. The majority (89.2%) of the diagnostic categories from the APACHE IV could be mapped to SNOMED CT concepts and for the remaining concepts a partial match was identified. The resulting set of mapped concepts consisted of 404 SNOMED CT concepts. This set could be extended to 83,125 concepts if all taxonomic children (descendants) of these concepts were included in the interface terminology. In the final phase, the interface terminology was deployed in a locally developed terminology server to collect the reasons for IC admission. The study provides a structure for the process of identifying a domain-specific interface terminology based on SNOMED CT and is of value for other researchers who intend to build a domain-specific interface terminology based on SNOMED CT.

Chapter 4 describes the usability evaluation of the interface terminology from chapter 3 to register the reasons for IC admission in a PDMS. Usability was defined by five aspects: effectiveness, efficiency, learnability, overall user satisfaction, and experienced usability problems. Qualitative (the Think-Aloud method) and quantitative (Modified version of System Usability Scale (SUS) questionnaire and Time-on-Task method) methods were used to examine the usability aspects. The results of the evaluation study revealed that the usability of the interface terminology fell short (SUS score 37.5 from 100). The qualitative measurements revealed a high number of distinct usability problems (n=35), resulting in ineffective and inefficient registration of reasons for IC admissions. This remained steady over time. One third of the encountered usability problems could be related to the interface terminology based on SNOMED CT, while the remaining problems concerned the
terminology client. The problems related to the interface terminology were more severe than the problems related to the terminology client. This study provided a detailed look at how clinicians interact with a controlled terminological system. The study concludes that adjusted terminology content and well designed terminology clients can facilitate the use of a complex interface terminology based on SNOMED CT.

Chapter 5 provides a framework that summarizes the criteria for the management of the maintenance processes of medical terminological systems. By means of a literature study, criteria for the maintenance process of medical terminological systems were obtained and categorized into a framework. The primary component of the maintenance process is "Execution" which concerns the core activities of the maintenance process. The other three components "Process management", "Change specifications" and "Editing tools" support the core activities of the component "Execution". Next, the current practice in the maintenance of medical terminological systems was explored by a survey among organizations that maintain a medical terminological system. The survey had a response rate of 40% (37 of 93). The answers reflected the large variation in the number of criteria that are satisfied for the participating organizations. Overall, for larger terminological systems most of the criteria from the framework are fulfilled whereas for smaller terminological systems more criteria failed, probably as this incurs costs and overhead that are too high and/or unnecessary. The survey showed that there is ample room to improve the maintenance processes of medical terminological systems, especially for the smaller terminological system. The framework is an important step towards standardization of the maintenance process of medical terminological systems.

Chapter 6 illustrates the influence of the choice of a prognostic model and the effect of customization of these models on quality assessment in league tables in which intensive care units (ICUs) are ranked by standardized mortality ratios (SMR) using APACHE II, SAPS II and MPM24 II prognostic models. The study used a dataset from the NICE registry containing data on 86,427 ICU admissions from 40 Dutch ICUs over the period of January 2002 to October 2006. The league tables associated with the different models were compared to evaluate their agreement. Bootstrapping was used to quantify the uncertainty in the difference for ICUs. It was shown that assessments of institutional performance in case-mix adjusted league tables are sensitive to the model that is used for case-mix adjustment, irrespective of the customization of the models. Therefore, performance data in league tables should only be used as a rough indication of the quality of care in an individual ICU. Additional information on the practices of care is needed before firm conclusions can be drawn.
Summary

Chapter 7 focuses on the performance of the newly available APACHE IV prognostic model to predict hospital mortality in comparison to the performance of the widely applied older prognostic models, i.e. APACHE II and SAPS II models. The study used data from the NICE registry between 2006 and 2009 from 59 Dutch ICUs. Measures of discrimination, accuracy, and calibration (Area Under the receiver operating characteristic Curve, Brier score, R-squared, and Hosmer-Lemeshow $\hat{C}$-statistic) were calculated using bootstrapping. Additionally, the SMRs were calculated. The original APACHE IV showed good discrimination and accuracy but poor calibration. The overall discrimination and accuracy of the customized APACHE IV model were statistically better and the overall $\hat{C}$-statistics was inferior to those of the customized APACHE II and SAPS II models, but these differences were small in perspective of clinical use. The study showed that the three models have comparable capabilities for benchmarking purposes after customization. The main advantage of the APACHE IV model is the large number of diagnoses which enable subgroup analysis. In addition, the APACHE IV coronary artery bypass grafting (CABG) model has a good performance in the Dutch ICU population and can be used to complement any of the three models.

Chapter 8 evaluates the effect of the use of SNOMED CT to crossmap two IC specific classification systems, i.e. from APACHE IV to APACHE II, and to evaluate the effect of this crossmap on the calculation of case-mix adjusted mortality risks for quality assessment. Firstly, the SNOMED CT crossmap was compared with an expert-based and a data-driven crossmap. Next, the influence of these crossmap strategies on the health care outcome was evaluated. For 50% of the analyzed cases, the three mapping strategies resulted in the same crossmaps. In other cases, there was an overlap between the SNOMED CT crossmaps and the crossmaps provided by one of the two other strategies. Differences in the crossmap results however had no significant influence on the case-mix adjusted mortality risks. The study showed that SNOMED CT can be used as an intermediary to solve the problem of crossmapping between versions of diagnostic classification systems.

Chapter 9 provides an overall discussion of the work presented in this thesis. The merits and limitations of the different studies are addressed and opportunities for further research are presented. This study showed that physicians can use a compositional terminological system to record clinical data. However, further research is needed to fully explore the possibilities for implementing SNOMED CT or other reference terminologies in the PDMS in Dutch ICUs. The usefulness and applicability of direct data entry and natural language processing techniques to ease the data entry based on a compositional terminological system in clinical practice should be investigated in the future. In the mean time, the
registration burden for the NICE participants can be reduced by applying the SNOMED CT crossmap between the APACHE IV and APACHE II classifications. By using this mapping, the continuity of health care assessment based on APACHE II or APACHE IV prognostic models is preserved and inefficient capture of the related reasons for IC admission is prevented.

The second part of the thesis focused on the use and evaluation of prognostic models for quality assessment in the Dutch ICUs. It was shown that assessments of institutional performance in case-mix adjusted league tables are sensitive to the model that is used for case-mix adjustment, irrespective of the customization of the models. The results of this thesis even more underline the importance of realizing that quality indicators based on prognostic models in the NICE registry should not be used for judgment (e.g. accountability) but for quality improvement (e.g. change initiatives).