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

de Keizer, N.F.

Publication date
2000

Chapter 11

Summary and general discussion

11.1 Summary

The topic of this thesis is quality assessment of intensive care. We explored, designed and implemented an infrastructure for empirical "evidence based" quality assessment of intensive care from two perspectives: a clinical epidemiological perspective concerning data analysis, and a medical informatics perspective concerning structured diagnoses and data collection.

From the clinical epidemiological perspective we conclude that a national database intensive care is an important tool to assess and assure quality of intensive care (chapter 3). NICE has successfully implemented a national database and uses it to compare outcome, in this case mortality and length of stay, between Dutch ICUs and the national average. Furthermore we conclude that existing prognostic models used in this quality assessment program have to be evaluated and recalibrated before they can be used to compare case mix adjusted mortality between Dutch ICUs (chapter 4). Although diagnosis has a secondary role as explanatory (adjustment) variable in most prognostic models, we hypothesized that this role could be increased. From the collaborative research with the UK audit center for intensive care, ICNARC [1] we conclude that the added value of increasing the level of detail of diagnostic information to a well-known prognostic model mainly based on physiological disturbance, was less important than we expected. The model could even be simplified by replacing the 53 APACHE II diagnostic categories by 9 body system based categories (chapter 5).
In this thesis special attention is paid to the role of diagnostic information in prognostic models, in communication and in the context of research. From the medical informatics perspective we conclude that registration of intensive care diagnoses is not adequate in current PDMSs due to the lack of a terminological system for intensive care diagnoses (chapter 6). We proposed a framework for understanding terminological systems, existing of a standard terminology and typology to support communication about terminological systems themselves on the one hand, and a conceptual and formal representation formalism to describe the structure of terminological systems on the other hand. This framework was used to explore five well-known terminological systems with a general medical domain and one terminological system for intensive care. From our study we conclude that this framework has been useful to explore the merits and limitations of existing terminological systems (chapter 7 and 8). Furthermore, the framework formed an adequate basis for developing a new terminological system for intensive care diagnoses (chapter 9 and 10).

We will now summarise the results and their implications. In section 11.2 we discuss the merits and limitations of a national quality assessment and assurance program. Then, the use of prognostic models to measure quality of intensive care and the added value of increasing the level of diagnostic information to improve the performance of the APACHE II [2] prognostic model to predict in-hospital mortality of adult intensive care patients is discussed in 11.2.2. In section 11.3 we discuss the use of PDMSs for daily care practice and research. The framework for understanding terminological systems and the specific terminological system designed and implemented in order to facilitate structured registration of diagnoses in intensive care are discussed in 11.4. Directions for further research are proposed throughout the chapter. The chapter concludes with remarks in section 11.5, concerning the role of terminological systems in enabling quality measurement of intensive care.

11.2 Quality assessment and quality assurance in Intensive Care

In this section we discuss the results from chapters 3, 4 and 5. First we discuss the NICE initiative and the Dutch national database intensive care. Next we discuss the use of prognostic models to measure the quality of intensive care.

11.2.1 NICE and the national database intensive care

For the purpose of quality assessment and quality assurance of intensive care, large scale multicenter databases have been developed [3-9]. Information from such a database enables benchmarking, comparing the effectiveness and efficiency of individual units with a standard, the benchmark. As part of the study described in this thesis, the NICE (National Intensive Care Evaluation) foundation with its Dutch national database intensive care has been established. NICE uses the database to compare outcome, at this moment restricted to mortality and length of stay, of individual ICUs with the national average (see chapter 3).

From the definitions of quality assessment and quality assurance described in chapter 1 we can conclude that the NICE initiative is at this moment only a quality *assessment* program, using the national average as a benchmark. Although quality *assurance* is the ultimate goal of the NICE initiative, it is not yet implemented. With data from the national database it is possible to compare outcome of an individual ICU with the national average or to compare outcome of a particular patient category with outcome of another patient category. Differences in outcome among ICUs or between an ICU and the national average have already been discussed in meetings with all participating intensivists. However, a thorough search for the causes of these differences and implementation of activities in order to reduce these discrepancies in as far as these are judged avoidable, have not been performed yet. The development of the national database intensive care and the implementation of the data collection process in the ICU makes participating intensivists aware of policy and issues related to quality. Furthermore, the national database is an important tool to get insight into trends in the demographical and physiological characteristics, and the health status and outcome of the intensive care population. It can also support clinical researchers to derive hypotheses and test them or to assist them in study design issues (e.g. sample size).

The value of benchmarking depends on the standard used, in this case the national average. We are aware of the limitations of this benchmark, viz. the national average does not automatically imply a good standard, e.g. it is imaginable that a small number of ICUs perform very badly which will reduce the standard to an inadequately benchmark. At this moment NICE includes ten participating hospitals, but at the moment of the study described in chapter 4, data of only three hospitals could be used which can not be regarded representative for the Netherlands. Hence, the average is not a real *national* average. In view of the willingness to participate in the NICE initiative now shown by many Dutch ICUs, coverage will be broader soon. But, even when NICE will include a large proportion of Dutch ICUs, as long as not every ICU is included, one has to be careful with overall conclusions about the quality of Dutch intensive care. Participation in NICE is voluntary and selection bias may limit the validity of benchmarks based on the current participants.

With more ICUs participating in NICE, additional benchmarks will become available, for example to accommodate specialized ICUs (neurological, cardiac surgical etc.). Next to general benchmarks, NICE also has the intention to pay attention to trends in time.

A higher participation level of ICUs implies new challenges. The success of an initiative like NICE strongly depends on the commitment of the participants. The current participants are the founders of NICE, who also developed the minimal data set and are strongly committed to the initiative, which is essential for the quality of data collected and for the quality judgements derived from the data. This level of commitment can not be expected and to some extend is not necessary from new participants. However, the financial contribution and the extra work this initiative requires, show adequate commitment of each voluntary participating ICU. To some extend the high quality of data collection can be achieved by technical means, such as an extensive data dictionary describing definitions, examples and technical information for each variable; obligatory training to reduce variability in data collection before one may start data collection; and obligatory consistency checking on domain values and inter-variable consistency during local (electronically) data collection in the ICU and before importing the data in the national database. An audit-program, including site visits and sample recollection [10], to investigate how these data quality improvement activities are implemented in the participating ICUs, is currently under development. We hope, this audit activity will further add to the social process of acceptation of the professional belief of the founders.

As described in chapter 3 the outcome measures currently considered in NICE are restricted to (ICU and in-hospital) mortality and length of stay. In chapter 4 we mentioned that differences in ICU mortality (adjusted for case-mix) can be caused by differences in the availability of medium care and ICU discharge policy. The same applies to hospital discharge policy. Therefore, it would be more objective to use mortality figures at a fixed time point, for example 6 or 12 months after ICU discharge. Information about the location or survival after hospital discharge is not included in the NICE database intensive care. Due to privacy regulation it is not easy to collect these follow-up data. NICE will perform at least a pilot in which the differences in using hospital mortality and 12-months mortality will be investigated. Efficiency measures are not under consideration yet. However, length of stay can be an approximation of efficiency because costs are closely related to length of stay. NICE has the intention to explore other outcome measures such as costs and quality of life in the future.

11.2.2 Scoring systems and prognostic models

Scoring systems and prognostic models can be applied at three aggregation levels [11, 12]. The first level is population-based, such as guiding policy and management or guiding research and education. The second level is sub-population based, such as guiding the care of groups of patients (quality assessment and assurance) or grading patient outcomes in clinical trials. The third level is individual patient-based, such as guiding choice of treatment in individual patients. The demands in terms of accuracy, discrimination and calibration on a prognostic model increase considerably as one progresses from the population-based level (focussing on multigroup means of prognosis) to individual patient-based level (focussing on

an individual prognostic point estimate with confidence interval). For example in a population based application the overall accuracy (overall correct classification) is the important issue regardless of the fact whether the model fits well to particular subgroups or individuals. In a sub-population based application the model should have sufficient accuracy for each distinguished group of patients and should also provide sufficient discrimination between different groups. The accuracy in individual cases again is less important. For the individual patient based applications calibration per case is essential because the calculated probability is the input for decisions about the patient's treatment. The scoring and prognostic models APACHE II, SAPS II, MPM_0 II, MPM_{24} II and LODS, used in intensive care, were developed for population and sub-population based applications. The score is often used for stratification of patients within a study, for example a clinical trial. The probability of mortality, which can be derived from the score, is mainly used for quality assessment of intensive care. One of the reasons for the development of new prognostic systems or to improve on the older ones, e.g. APACHE III, is to rise the accuracy to a level suitable for individual application. Sofar, progress is slow, as witnessed by the divergence in individual hospital mortality risks estimated by the two most commonly used prognostic models, APACHE II and SAPS II (see figure 11.1). This divergence still is poorly understood.

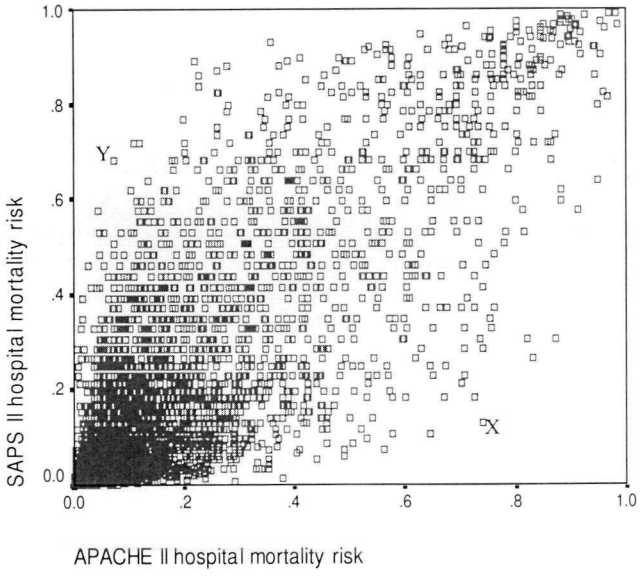

Figure 11.1 APACHE II versus SAPS II hospital mortality risk. Large differences exist for individual patients, e.g. X shows low SAPS II and large APACHE II mortality risk, the opposite holds for Y (data extracted from Dutch national database intensive care 1/1/1999 to 1/10/1999, n=8414).

In chapter 2 we mentioned the disadvantages of accuracy and discrimination in evaluating the performance of a model: (1) dependence on the cut-off point chosen and (2) prevalence of the outcome. Prevalence dependency implies that when a population mainly exists of e.g. low risk patients, the specificity will be large for many cut-off points, because in a large range of middle and high value cut-off points many survivors will be correctly predicted to survive.

For the same reasons the sensitivity will be large in a high risk population. Calibration, another aspect considered in the evaluation of a model's performance, partially deals with the problem of prevalence dependency because it includes different risk strata. The Hosmer-Lemeshow statistics, commonly used to quantify calibration, are based on stratification of patients into groups with either equal expected probability ranges (H-statistics) or equal percentiles of patients per probability range ($\hat{C}$-statistics). Although every characteristic could be the basis for the division of groups, probability of hospital mortality is always reported in articles to define the groups. It is arguable whether the probability is the most appropriate characteristic because now probability of hospital mortality is both used as characteristic of partition and as the outcome of interest. Depending on the interest of the study in which the models are used, it can at least be interesting to additionally use for example diagnostic categories or score categories to explore the agreement between observed and expected survival and non-survival in these groups.

The study described in chapter 4 demonstrated that the discrimination of all prognostic models used in the study was good, especially of APACHE II [2] and SAPS II [13]. Because calibration showed lack of fit in all models, we recalibrated one model, viz. we recalculated the coefficients of the variables in the model, to improve discrimination and calibration. SAPS II was chosen for recalibration because this model is most commonly used and has best discrimination next to APACHE II. APACHE II could not be used in our case because the number of patients per APACHE II diagnostic category was too small at the time of this study to allow for categorical analysis.
In the study described in chapter 4 we used data of only 3 hospitals. Application of the original prognostic models suggests that one of the three hospitals had significantly lower mortality than expected. However, this hypothesis was at first instance rejected after recalibration of the SAPS II model. The small number of patients may have been partially responsible for lack of statistical significant differences. Recalibration and reanalysis will be repeated if more data per hospital will be available, and if more hospitals participate.
In the process of model development we can choose to recalibrate existing scoring models and prognostic models by (1) reweighting the coefficients per variable (e.g the score) in the prognostic model, (2) reweighting the number of points of variables which result in the score or (3) reconsidering the independent variables. Recalibration or reweighting seems appropriate when there exists an overall over- or underprediction. Bad fit in particular patient or severity categories indicates reweighting coefficients or even reconsideration of independent variables. As shown in table 2.2 the variables used vary between the prognostic models. Because the NICE national database intensive care includes all these variables, new selections could be undertaken. A related hypothesis to be investigated is whether the change in physiological condition instead of worst state of each physiological variable or diagnostic information is a better predictor.

The study described in chapter 5 showed that the additional value of increasing the level of detail of diagnostic information to the APACHE II model was low. Based on the results of this study the APACHE II model recalibrated for the UK population [14] could even be simplified by reducing the amount of diagnostic information in the APACHE II without losing performance. However, further research has to explore if specific diagnostic categories instead of fixed sets of diagnostic information ("body system" or "anatomical localization"

and/or "process or dysfunction") add to the performance of the prognostic model. To enable this research structured registration of diagnostic information is needed. Although the ICNARC Coding Method which is used in this study is an important step into this direction, the lack of a conceptual and formal representation of the diagnostic information limits it use for this purpose (see chapter 7, 8 and 9). For example the lack of explicit definitions of diagnoses harms the non-ambiguous and complete selection of patients with particular diagnostic characteristics such as tumors in the gastrointestinal system.
The next section deals with an important part of the information architecture in intensive care, the Patient Data Management System, which could facilitate the collection of routinely prognostic data.

11.3 Patient Data Management Systems (PDMS)

During the last years ICUs become more and more interested in PDMSs to manage the large amount of data available in intensive care. PDMSs were not only introduced with the promise to support nurses and physicians in (automatically) collecting a large amount of data, they also promised to improve the quality of care by integrating and interpreting the data and thereby support decisions during treatment and care. Another promise concerned the continuous availability of data from patients admitted to the ICU and the use of these data for research.
As described in chapter 6 some Dutch ICUs implemented a PDMS and used it in daily care practice. We investigated whether the functionality of Dutch PDMS configuration agreed with predefined requirement concerning "automated charting", "care planning" and "management information". Most PDMSs support nursing activities by automated registration of data from bedside devices or the Hospital Information System. However, there is much functionality in "care planning" not yet optimally available in the current generation of PDMSs and maybe therefore the success of the PDMS stays away. An integrated and problem oriented care plan for all involved disciplines is often lacking or inadequate. The support for structured data entry for e.g. physical examination or medical history is lacking and the retrieval of data of a patient's earlier ICU admission is hard or impossible.
Because much relevant data about the patients and their treatment and outcome is available in the PDMS, it seems a valuable tool for "management information" and thereby useful for data collection for a national database intensive care. From chapter 6 we can conclude that this promise has not been fulfilled yet. The most important reasons why the expected value of the PDMS for management information and research have not been met yet seems the lack of structured data entry which is essential for data retrieval. Another important limitation for success is the complex database structure combined with the lack of a user-friendly facility for data extraction for research and management purposes.
When data are not structured, it is hard to use this information for management information, research or to integrate and reason with it in daily care practice. For example integration of protocols, knowledge based decision support and critiquing systems are the functionalities which make a PDMS valuable. An important reason why structured registration of *diagnoses* is lacking in PDMSs is mainly because of the lack of a terminological system which is tailored to the intensive care and which is integrated in the PDMS (see figure 1.2).

11.4 Terminological systems

Large scale classification of medical data started at the end of the 18th century. Historically, medical data was coded mainly for epidemiological and administrative purposes. The introduction of computerized information systems and the tendency to an information-driven and "evidence-based" medicine increased the importance of these data in daily care practice. These data are especially important in interdisciplinary communication, integration with protocols and critiquing systems and for clinical research. This shift in use of medical data implies new requirements on terminological systems concerning for example the level of detail and the structure of the terminological systems.

In intensive care no terminological system was available to support systematic registration of diagnoses, health status and medical problems. Such a terminological system should support the description of the patient's health problems as part of daily care practice and should support the aggregation of diagnostic information for research. These two objectives resulted in the prioritisation of criteria for terminological systems which were defined by Cimino et al. [15] and Campbell et al [16].
In this study we explored whether existing terminological systems (for at least medical diagnoses) could meet these criteria. A good understanding of terminological systems is essential before one can assess whether an existing terminological system is appropriate for use in certain circumstances or when one has to develop a new system. Therefore we suggested a referential framework for understanding terminological systems. It includes two components. First, terminology and typology of terminological systems to facilitate the communication about notions used in this field. Existing standards such as ISO and CEN [17, 18] intend to describe the first part of such a framework. They are rather dry enumerations of definitions about notions in the field. Therefore we enriched the framework with interrelations between these notions, including a typology of terminological systems.
The second part of the framework is a uniform (conceptual and formal) representation formalism to describe the criteria for terminological systems and the structure of terminological systems. The representation of the structure of a terminological system in a conceptual and formal way has more advantages next to merely *understanding* the terminological system [19-23]. A conceptual and formal representation of a structure of a system supports the communication about the meaning of a system. Furthermore, it supports the development of new systems by using the "desired patterns", the essential criteria for terminological systems, in the design. By making knowledge of the underlying domain explicit with formal specifications, these specifications can be used in a knowledge acquisition tool, such as GAMES [21] and PROTÉGÉ [24], to support consistency checks within the terminological system. Formalised knowledge is also a prerequisite for inference of new knowledge. Knowledge acquisition and inference are important for the management of knowledge in the terminological system.

We applied both parts of the framework to five widely disseminated terminological systems with a general medical domain, ICD-9-CM /ICD-10 [25, 26], NHS Clinical Terms [27], SNOMED [28], UMLS [29], and GALEN [30] and to one terminological system for ICU reasons for admission, the ICNARC Coding Method (ICM) [31].

In our experience, formalization supports understanding a terminological system and facilitates the comparison between the criteria on terminological systems and the structure of existing terminological systems. Formalization resulted in a reference design that helped us to observe anomalities in some terminological systems. From the comparison between formalized criteria and structure of existing terminological systems we conclude that none of the existing terminological systems adequately satisfy our objectives for intensive care. Therefore, the formalization of criteria formed the basis for the development of a new terminological system for intensive care diagnoses.

We have engineered an ontology, a description of concepts and relations, in the domain of intensive care diagnoses. Our engineering approach and the general ideas behind it are not restricted to IC but rather applicable to a broader spectrum of medical domains. We used Entity-relationship (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, such as detailed constraints, we described some specifications formally in First Order Logic (FOL), which is more expressive. Of course, other conceptual and expressive formalisms, such as UML [32] and OMT [33] could also be used instead of ER. Our choice for ER with FOL means that the descriptions in this formalism could easily be translated to and from other logic-based formalisms such as Ontolingua [34], conceptual graphs [23] and description logics [35] when their expressivity allows this.

Our ontology was implemented in an application called DICE (Diagnoses for Intensive Care Evaluation) which consists of three parts (see Figure 9.2): the meta-model (Figure 9.1), the domain model (Figure 9.3) and the IC-domain (Table 9.3). The *meta-model* describes concepts, relationships between them and between concepts and terms in general, such as constraints on the number of terms a concept may have. The *domain model* describes concepts and their relationships in the IC-domain such as "health problem" has localization "Anatomical component". The *IC-domain* is formed by concepts such as "Meningitis", their definition in a *vocabulary*, e.g. "Viral hepatitis" *has location* "Liver" and *has abnormality* "Infection"; and a *nomenclature* with rules to support the composition of new concepts, e.g. "viral hepatitis" *Is caused by* {"hepatitis virus" or "Cytomegalo virus" or "Epstein Barr virus"}.
Although the domain model is focused on intensive care, the chosen concept and relations may be extended in the future to other domains or may be extended because one might need additional features to aggregate diagnostic information for research purposes. From our experiences with the application of this model to the domain of internal medicine (not described in this thesis) we can conclude that the ideas behind DICE are general enough to extend the model without major problems to another domain.

DICE is a JAVA application, which implies that it can run on the different platforms the PDMSs are running on. A preliminary evaluation showed that both the knowledge modeller, who is responsible for the implementation and maintenance of the knowledge and terms in intensive care, and the intensivists, who want to find and assign diagnoses for patients, judged the application as a promising and acceptable tool.
During the pilot, aimed at a preliminary evaluation of the implementation of our ideas, the knowledge modeller could model all relationships and all possible terms for the concepts. Formalisation of knowledge facilitates automated reasoning such as automated classification when a new concept is added to the classification [35]. However, this functionality is not yet implemented, but is planned for the future. Since in formalisation there is always a trade-off between expressiveness and the level of automation of reasoning we have to explore in which way the application can facilitate and implement automated reasoning. The main weakness observed in the pilot by the modeller was the impossibility to define cardinality constraints and dependencies *between* attributes due to lack of expressiveness. Cardinality constraints are conceptually and formally modelled, but it is not yet sufficiently implemented in DICE. Although this weakness requires attention in the near feature, this problem is infrequent (in this pilot only for one diagnosis) and some responsibility of the user while recording this diagnosis can be assumed. From the modeller's perspective, it has to be mentioned that implementing the ontology by describing hundreds of ICU diagnoses with their relations and characteristics is a very labour-intensive and specialized job.
Our evaluation showed that the interface for the intensivists needs some changes. Although it was possible to select plural operative procedure, the interface was not very intuitive to easily compose e.g. "CABG *and* aortic valve resection". When more than one procedure was selected, it was unclear which attributes belong to which procedure. Furthermore, the knowledge in DICE is also not completed yet (some concepts and their relationships are missing or need adaptation). The knowledge has to be under continuous development anyhow, to accommodate the emergence of new diseases or the redefinition of existing ones. Maintenance of knowledge is an important issue for further research. Although formalisation can support the consistency of knowledge during maintenance, especially when more than one knowledge modeller is responsible for the maintenance, it is hard to protect the consistency of knowledge [36].

11.5 The role of terminological systems in measuring quality of intensive care

In this section we integrate the clinical epidemiological and medical informatics perspectives by discussing the role of terminological systems in measuring quality of intensive care.
At the beginning of this study we hypothesized that diagnoses play an important role in case mix description of intensive care patients and that diagnostic information should therefore be included in the data collection part of the quality assessment and assurance program for intensive care. The study described in chapter 5 does not support our hypothesis that increasing the level of diagnostic information in the prognostic model APACHE II does improve the predictive power of the model to estimate risk of in-hospital mortality for adult intensive care patients. However, we still believe diagnoses are important in clinical studies, among other quality assessment and assurance, in intensive care. Diagnostic information is at least important to stratify the patient population and to select patient categories for specific research questions. Therefore, structured and formally defined diagnostic information is

essential to unambiguously define patient categories on diagnostic characteristics. Further research should then prove whether specific diagnostic categories instead of fixed sets of diagnostic information ("body system" or "anatomical localization" and/or "process or dysfunction") add to the performance of the prognostic model. Anyhow, to enable stratification and selection of patients from the intensive care population and to further explore the role of diagnostic information in prognoses we need structured and formally defined diagnoses. The research described in this thesis is aimed at the development of a terminological system to facilitate this.

We believe that DICE satisfies our ideas about a terminological system for intensive care and that it is useful to enable structured and formally defined diagnoses. Although there are some point for improvement, it seems to be a promising application for both the knowledge modeller and the intensivists. Once the mentioned shortcomings concerning the user interface are solved and once the knowledge base includes most of the intensive care diagnoses, the terminological system has to be integrated with the PDMSs. The possibility to record reasons for admission and complications during ICU stay in a structured and formally defined way will bridge an important gap in current information architecture in intensive care in which the PDMS plays a central role. It will improve communication between different physicians and between different disciplines during daily patient care by standardisation of language. Furthermore, the data set of the Dutch national database intensive care of NICE will be extended with data about the reason for admission and comorbidity occurring during ICU stay so that stratification and patient selection on diagnostic information will become feasible for research purposes.

Quality assessment and assurance of intensive care is difficult because of the abundance and diversity of reasons for admission, the severity of illness and the many parallel therapies given to treat the patients. Validated prognostic models to adjust mortality estimations in the Dutch intensive care population for severity of illness and a terminological system to describe reasons for admission and complications during ICU stay will be important improvements in the infrastructure for quality assessment and quality assurance in Dutch intensive care.

References

1. ICNARC. ICNARC Case Mix Programme Dataset Specification. London: Intensive Care National Audit & Research Centre, 1995.
2. Knaus W, Draper E, Wagner D, Zimmerman J. APACHE II: a severity of disease classification system. Crit Care Med 1985;13:818-829.
3. Anonymous. ANZICS/CACCN annual scientific meeting on intensive care. Anaesth & Intens Care 1995;23:379-401.
4. Nelson L. Data, data everywere. Editorial comment. Crit Care Med 1997;25:1265.
5. Rowan K, Kerr J, Major E, McPherson K, Short A, Vessey M. Intensive Care Society's APACHE II study in Britain and Ireland-I: Variation in case mix of adult admissions to general intensive care units and impac on outcome. BMJ 1993;307:972-976.
6. Rowan K, Kerr J, Major E, McPherson K, Short A, Vessey M. Intensive Care Society's APACHE II study in Britain and Ireland-II: Outcome comparisons of intensive care units after adjustment for case mix by the American APACHE II method. BMJ 1993;307:977-981.
7. Apolone G, Bertolini G, D'Amico R, et al. The performance of SAPS II in a cohort of patients admitted to 99 Italian ICUs: results from GiViTI. Gruppo Italiano per la Valutazione degli interventi in Terapia Intensiva. Intens Care Med 1996;22:1368-1378.
8. Metnitz P, Valentin A, Vesely H, et al. Prognostic performance and customization of the SAPS II: results of a multicenter Austrian study. Intens Care med 1999;25:192-197.
9. Moreno R, Morais P. Outcome prediction in intensive care: results of a prospective, multicentre, Portuguese study. Intens Care Med 1997;23:177-186.
10. Knatterud G, Rockhold F, George S, Barton F, Davis C, Fairweather W. Guidelines for Quality Assurance in Multicenter trials: A position paper. Control Clin Trials 1998;20(1):477-493.
11. Wyatt J. Nervous about artifical neural networks? Lancet 1995;346:1175-1177.
12. Wyatt J, Altman D. Prognostic models: clinically useful or quickly forgotten? BMJ 1995;311:1539-1541.
13. Le Gall J, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. JAMA 1993;270:2957-2963.
14. Rowan K, Kerr J, Major E, McPherson K, Short A, Vessey M. Intensive Care Society's Acute Physiology and Chronic Health Evaluation (APACHE II) study in Britain and Ireland: a prospective, multicenter, cohort study comparing two methods for predicting outcome for adult intensive care patients. Crit Care Med 1994;22:1392-1401.
15. Cimino J, Clayton P, Hripcsak G, Johnson S. Knowledge-based approaches to the maintenance of a large controlled medical terminology. J Am Med Inform Assoc 1994;1:35-50.
16. Campbell J, Carpenter P, Sneiderman C, et al. Phase II evaluation of clinical coding schemes: completeness, definitions and clarity. J Am Med Inform Assoc 1997;4:238-251.
17. Coding Methods and principles. ISO, 1989.
18. European prestandard. Medical informatics - Categorical structures of systems of concepts - Model for representation of semantics. Brussel: CEN, 1997.
19. Abu-Hanna A, Jansweijer W. Modeling application domain knowledge using explicit conceptualization. IEEE-Expert 1994.
20. Gruber T. Towards principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies 1995;43:907-928.
21. Heijst G, Falasconi S, Abu-Hanna A, Schreiber A, Stefanelli M. A case study in ontology library construction. Art Intell in Med 1995;7:227-255.

22. Moorman P, van Ginneken A, van der Lei J, van Bemmel J. A model for structured data entry based on explicit descriptional knowledge. Yearbook of Medical Informatics 1995:195-204.
23. Campbell K, Das A, Musen M. A logical foundation for representation of clinical data. J Am Med Inform Assoc 1994;1:218-232.
24. Studer R, Eriksson H, Gennari J, Tu S, Fensel D, Musen M. Ontologies and the Configuration of Problem-solving Methods. In: Gains B, Musen M, eds. 10th Banff Knowledge Acquisition for Knowledge-Based Systems Workshop. Banff, Canada, 1996.
25. International Classification of Diseases, manual of the International Statistical Classification of diseases, injuries and causes of death: 9th revision. WHO, 1977.
26. International Classification of Diseases, manual of the International Statistical Classification of diseases, injuries and causes of death: 10th revision. WHO, 1993.
27. Read Code File structure version 3.1. NHS Centre for Coding and Classification, 1995.
28. Rothwell D, Coté R. Managing Information with SNOMED: Understanding the model. SCAMC, 1996:80-83.
29. UMLS Knowledge sources - Documentation. National Library of Medicine, 1997.
30. Rector A, Bechhofer S, Goble C, Horrocks I, Nowlan W, Solomon W. The Grail concept modelling language for medical terminology. Artif Intell 1997;9:139-171.
31. Young J, Goldfrat C, Rowan K. The development and testing of a hierarchical method to code the reason for admission to intensive care units: the ICNARC Coding Method. Intens Care Med 1999;(submitted for publication).
32. Booch G, Rumbaugh J, Jacobson I. Unified Modeling Language User Guide. Addison-Wesley, 1998.
33. Rumbaugh J, et al. Object-Oriented Modeling and Design. Prentice-Hall, 1991.
34. Gruber T. A translation approach to portable ontology specifications. Knowledge Acquisition 1993;5:199-220.
35. Donini F, Lenzerini M, Nardi D, Schaerf A. Reasoning in Description Logics. In: Brewka G, ed. Principles of knowledge representation and reasoning. CLSI publications, 1996:193-238.
36. Chute C, Cohn S, Campbell J, structures AHisbvwgatCiwgoca. A framework for comprehensive Health Terminology Systems in the United States: Development guidelines, criteria for selection and public policy implications. JAMIA 1998;5(6):503-510.