Using real world data for modeling a protocol for ICU monitoring

Martin Scholz[1]

Abstract. In this paper a new approach for setting up computer based clinical protocols is presented. The goal is to speed up the development by exploiting recorded patient data from the beginning. Having the protocol represented as a knowledge base in a logical formalism, modeling can be supported in various ways. To empirically evaluate the approach a not yet operational protocol is translated into a logical formalism and some experiments with real world patient data records are presented.

1 BACKGROUND

In the year 2000 the Institute of Medicine published a report [9] stating that avoidable human errors in medicine are one of the eight most frequent causes of death. Especially the prescription of drugs at intensive care units bears a comparably high risk of human failure. The mistakes range from misjudgments of the patient situation, and thus following an inappropriate strategy, to mixing up drugs or by mistake choosing a wrong dose at the bedside.

Physicians base their decisions on experience and the actual data elicited from patients. On modern ICUs the number of frequently measured variables can be a few hundred, while human capabilities are limited to processing up to 7 variables in parallel in decision making. Under these circumstances, information overload has to be considered a serious problem. Modern technologies have been suggested to address this problem. Especially decision support systems are well suited to support physicians at the bedside. Given vital parameters of a patient in electronic form, a decision support system can trigger alarms when a patient is in a critical state, or even check if an intervention is a bad choice in a specific situation.

More holistic approaches even transfer all general control decisions to a standardized protocol, reflecting the best practice cases. These protocols will react properly in most of the situations a patient with a specific disease might encounter. In [1] Morris elaborately motivates the development of such protocols.

It is reasonable to use computers in order to automatically generate suggestions for the physicians at the bedside. The decision support facility becomes even more powerful if the patient's measured data can immediately be read from the clinical devices.

In this paper an approach for speeding up the development of computer based clinical protocols is presented and empirically evaluated. The idea is to use a knowledge acquisition environment and to continously validate the (incomplete) model by means of recorded patient data.

The paper is organized as follows: After a brief overview of related work on building clinical protocols, the following section 3 describes the approach of this paper in more detail. In order to be able to evaluate the approach, a draft of a clinical protocol is used. Because the protocol is given in a paper version, the first step was to set up an equivalent operational knowledge base. A brief description of the protocol and its formal representation is given in section 4. The following section 5 introduces a real world dataset that was used for the evaluation. Some validation techniques and experiments are also presented in this section. The paper concludes with a summary and outlook.

[1] University of Dortmund, Computer Science VIII, D-44221 Dortmund, Germany, email: scholz@ls8.cs.uni-dortmund.de

2 RELATED WORK

For the scope of this paper a representation of protocols as knowledge bases in a logical formalism is well suited, basically because a knowledge acquisition environment can be used to support the modeling process. In the literature many different formalisms are suggested. [4] and [13] give an overview of some common guideline representation formalisms.

The ARDEN syntax is one of the early standards for representing medical knowledge. Its modularity enables an exchange of protocols between institutions. The underlying formalism is a procedural programming language. One of the disadvantages is a poorly supported adaptation, as details on data sources are "hard coded". Further on, a declarative representation of higher level concepts like the patient's state, the actual state of executing a plan, temporal patterns, etc. is not possible, but needs to be coded as procedural knowledge.

Later guideline models make use of more sophisticated formalisms. EON uses an ontology to map patient variables to medical data, for example. One way of representing temporal constraints is the use of flow charts and other graph based approaches. Although this is an intuitive way of modeling, the expressiveness and utility of such approaches is limited. For example, real time planning, exploiting critiquing techniques, and being able to suspend a plan as long as necessary are properties missing in such approaches.

An overview of planning in the context of clinical protocols and some details about the guideline model Asbru can be found in [12]. Asbru is a language with LISP-like syntax, designed for the task of planning and plan management in the medical domain. Some of the involved concepts are intentions, conditions and effects. The execution of plans is continously monitored and plans will be adapted if necessary.

A major drawback of these guideline models is a lack of support for revisions. In contrast representations based on restricted first order logic, like the one used in this paper, offer automatic inference capabilities. This helps to reveal contradictory parts of a model. Because of the formal semantics of first order logic, the knowledge acquisition environment may support the knowledge engineer by offering minimal base revision, whenever necessary. An ontology to map variables from data records to protocol variables can be expressed in this formalism. For protocols that depend on planning capabilities this formalism offers at least the means to build and validate the knowledge modules incorporated.

As related operational clinical protocols, two examples shall be given here. The first one is mentioned in [1]. It addresses the highly complex acute respiratory distress syndrome (ARDS). The effectiveness of the protocol was continously evaluated by clinical trials. The VIE-VENT project [15] yields another example of a knowledge-based monitoring system, aiming at optimizing the ventilation of newborn infants.

A major problem with such protocols is that, although they prove useful in practice, the development is a very tedious task. To build the ARDS protocol took about 25 person years.

Apart from setting up the protocol itself, validation and preprocessing of data automatically derived from sensors is a key issue. Especially for intensive care units it is very useful to recognize outliers and to "repair" data, as much as possible. For this task a variety of mainly statistical approaches has been suggested ([16], [10], [11], [2], [3]). An approach combining a statistical signal-to-symbol method, machine learning for state action rules, and declarative domain knowledge is presented in [7].

3 THE MODELING APPROACH

The approach presented in this paper aims at drastically reducing the necessary efforts for building such protocols in the ICU context, especially in early stages of development. The idea is twofold. First of all, we conceive clinical protocols as expert knowledge, and consequently exploit appropriate tools for modeling *knowledge bases*. For the clinical practice guidelines and protocols different representations have been proposed, as discussed in section 2. In the context of this paper a restricted first order logic representation embedded into the knowledge acquisition environment MOBAL [8] turned out to be a good choice. It fits particularly well because the protocol introduced in this paper is based on flow-charts and propositional logic, which can easily be represented in first order logic. How well the approach applies to more complex protocol structures needs further evaluation. The main advantage of using such a knowledge acquisition environment is that consequences of decisions are calculated and visualized automatically, which helps to reveal contradictory parts. Modeling should be regarded as a cyclic process [6]. Thus, revisions of any kind should be taken into account from the beginning. A belief revision tool is part of MOBAL as well, supporting the user when parts of the knowledge base are found to be inappropriate. In the used representation formalism exceptions from the rule can be given by their extensions. Further on, using the integrated Machine Learning facilities of MOBAL, a rule based characterization of such exceptions can be inferred. Finally uncertain inferences are supported, which is useful for extending the simple architecture presented in this paper.

The second and more important idea is to utilize the vast amounts of data collected by electronic devices at the bedside of modern intensive care units. In early stages of the modeling process, expert knowledge can be validated just using recorded patient data. Recorded data is especially useful when, apart from the measurements of patients' variables, it reflects the physician's interventions. That way the suggestions generated by the protocol can directly be compared to the actual decisions made at the bedside. If a suggestion differs from the according intervention, then both should be investigated more closely. Giving experts a tool to retrieve such critical situations from a database speeds up the development and improvement of operational protocols.

4 OPERATIONALIZING AN ICU PROTOCOL

To prove the feasibility of the method sketched in the previous section, the following sections describe experiences made with an exemplarily operationalized protocol.

Modeling a complete domain is a very complex task, so the work was started using a given draft paper version of an ICU protocol. This protocol was non-operational, still under development, and basically addressed the field of haemodynamic monitoring. The incomplete state of the protocol is an advantage in the scope of this paper, because the objective is to investigate how the iterative process of modeling, validating and revising the model is facilitated by incorporating recorded data. To avoid any misunderstandings regarding the quality of this protocol, its origin is not mentioned here[2]. This section gives a brief description instead.

The protocol reflects the view of a physician at an intensive care unit. The usual concept of ward rounds establishes the overall framework. The control structure is given by a *flow chart*, making use of a semi-formal *propositional logic*. [1] motivates the use of flow charts and gives small examples.

The delay between two runs of the protocol for a single patient depends on the patient's actual state. Depending on which patient variables are critical, an elicitation of further variables is carried out, until it is possible to decide which intervention(s) are necessary. The input of the protocol is given by 35 variables. There are two kinds of proposed interventions, drugs and infusions. Time aspects are handled in various ways. Some events, like low blood pressure, are memorized for long time intervals. Changes of variable values are evaluated by comparing them to the one elicited at the last ward round.

In order to exemplarily validate and refine the protocol the first step was to transfer the protocol into an operational knowledge base. Restricted to a single iteration/ward round the characteristic inference structure of Heuristic Classification [14] was found appropriate to organize the given knowledge[3]. In the sense of Heuristic Classification the task is to classify the set of variables known to the protocol into categories, which are given by possible interventions to be performed. The inference structure is given by two hierarchies. The first one abstracts from the given set of variables, in order to find higher level descriptions, based on more meaningful features. These features reflect domain specific terms. The second hierarchy starts with abstract solutions, referred to as *therapies* in this context. Following the direction of inferences, the solutions get more and more specific. In this context the abstract therapy is refined unless an *intervention* is reached, which is specific enough to be suggested to the clinician. A characteristic of Heuristic Classification is that these hierarchies are connected via heuristic rules. The assignment of therapies in the protocol can be regarded as heuristic, as well, for they are not inferred using formalized causal dependencies.

For each ward round and patient, all most recent variable values are collected from the database and represented as a logical fact. In the next step reference values from the past are identified, and the differences to the most recent values are calculated. The differences are used to keep track of the latest *changes* of variable values.

[2] For copyright reasons it is not possible to show an extract.
[3] Please note that this way of structuring the knowledge base is just an example which fitted well for this specific protocol.

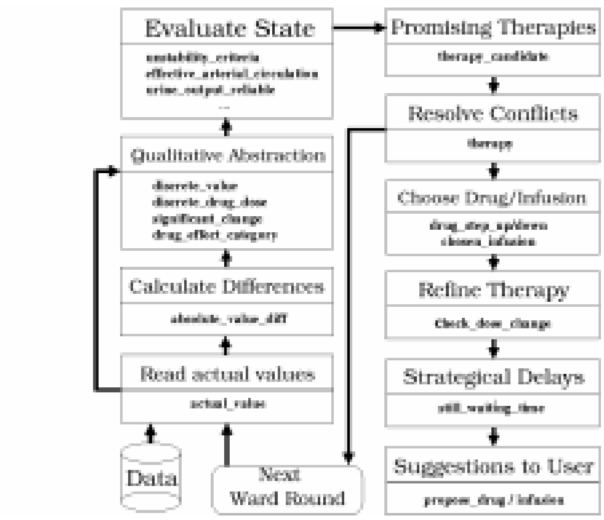

Figure 1. Architecture of the knowledge base. Arcs denote the flow of information.

The most recent values and the changes of variables are qualitatively abstracted by an interval based mapping from numerical values to categories. For most patient variables a normal interval exists, yielding the results "value too low", "value normal" or "value too high". For drugs something similar exists. The lower bound is given by the dose under which the drug does not show any remarkable effects. The upper bound is the highest dose not considered an overdose. Changes of patient variables over time are categorized by distinguishing significant increases and decreases from negligible changes. Drugs are assigned a set of expected effects with respect to the actual dosis.

Basically by aggregating multiple of the symbolic features derived in the last step, a more sophisticated representation of the patient's state is derived. This process is labeled by *Evaluate State* in figure 1. For example the effectiveness of the arterial circulation is evaluated, taking into account that some of the indicative features might be unknown. The ways of aggregating features ranges from a hierarchical "choose first variable available" to complex nested if-then structures, checking a variety of (symbolic) variable values. Further rules conclude about the stability of the patient's state.

Once the state of the patient is described by more meaningful features, promising adjustments of the actual therapies are collected. In some cases the set of therapy candidates is contradictory, for example " increase an inotrope" and "decrease an inotrope". Further on, the protocol forbids to decrease more than one drug at a time[4]. For these reasons rules to resolve conflicts are necessary. Rules defining which therapies are incompatible and a priority based rejection yield a set of "accepted" therapies. The delay until the next ward round is also determined at this point.

Therapies like "increase a pressor" mentioned above are still abstract. Different drugs and infusions share the same category, so the next inference step examines which one is preferable. The domain knowledge is often given by priority lists. For mosts drugs doses are increased up to the valid upper bound or down to zero, respectively, before the successor in the hierarchy is chosen.

The next step finally turns the the therapy into a concrete, applicable intervention by calculating new doses for drugs or total volumes and rates for infusions.

[4] This does not apply for all situations.

Finally a suggestion is inferred if not forbidden by an active waiting time. For each drug, an unsuccessful trial of decreasing it will be remembered for up to 12 hours. This aims at avoiding oscillating dose changes. Another reason for activating a waiting time is to slow down the process of decreasing a drug dose.

5 EXPERIENCES VALIDATING THE MODEL

The usefulness of real life data in the model building process was evaluated using anonymized data of 242 patients from the intensive care unit of the Städtische Kliniken Dortmund, Germany.

After some adjustments to the protocol and mapping of drugs, the operational knowledge base described in section 4 could be applied. The formal model does not only generate suggestions, but the intermediate concepts are transparent, as well. Figure 2 shows some intermediate results of an inference process. All derived logical facts are related to minute 525 of the treatment of patient with identifier 6983. In the first window (`actual_value`) the most recent values for different variables are listed. The result of mapping some of these to discrete categories is given by the predicates `discrete_value` and `discrete_drug_dose`. The information if a variable has changed significantly is summarized by `significant_change`. In `drug_effect_category` the actually given drugs are mapped to the categories, according to the effect the drug is assumed to have in the actual dose. Some present criteria for instability are listed in the lower left. Finally the generated therapy candidates are shown in the lower right. Some successfully applied techniques to exemplarily validate the protocol are introduced in the remainder of this section. First, a straight-forward modification that turns the knowledge base into a critiquing system [5] turned out to be useful. The system compares the suggestions generated by the protocol to the recorded interventions, chosen by a physician at the bedside. This can be done at different levels of abstraction. In my experiments, the level of therapy candidates and chosen therapies was best suited, because of an astonishingly small fraction of cases where protocol and practice chose the same intervention. It was necessary to map the interventions found in the data records to the therapies known to the protocol. For this step the same knowledge could be exploited as was used to refine abstract therapies to concrete dose changes. The therapies proposed and performed were then compared and the results were aggregated into one predicate, stating

- if the therapies are equal,
- if the performed therapy is not even under consideration or
- if the therapies are even clearly contradictory, e.g. *increase pressor* and *decrease pressor*.

Table 1 shows an example of a critiquing result. In order to better exploit the sparse set of interventions found in the data, they were used to trigger the ward rounds. This change increased the effectiveness of the critiquing system and helped to find numerous cases where protocol decision and recorded intervention differed drastically. Such cases were shown to an expert on intensive care medicine and analyzed in detail. Although graph based visualizations of inferences are available, screens like shown in figure 2 were better suited to communicate to the expert in order to find the rule(s) to be refined.

Another straight forward way of changing the protocol for validation is to drop the cyclic ward rounds, in favour of a *monitoring* approach. In this case the protocol changes the proposed interventions immediately whenever the patient's state demands. Such an approach was followed to evaluate if the emerging protocol behavior follows a steady plan, or if the decisions tend to be contradictory within short

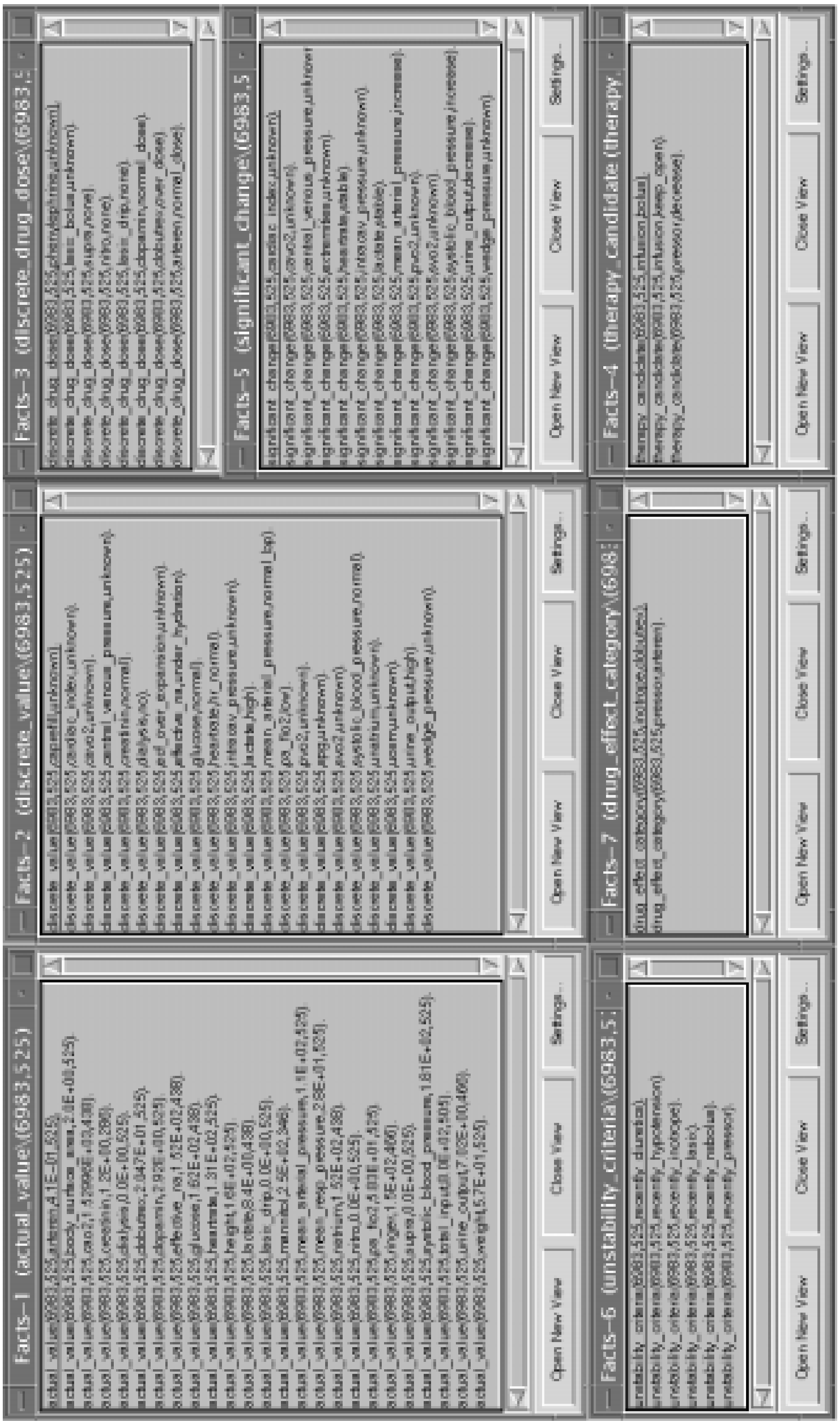

Figure 2. Information for patient 6983, minute 525 at different levels of abstraction.

Table 1. Critiquing system applied to patient 6983

Pat. ID	Time	Status		
6983	1185	same_infusion	bolus	bolus
6983	1192	same_therapy	inotrope	increase
6983	1398	same_infusion	bolus	bolus
6983	1605	performed_is_no_candidate	diuretic	increase
6983	1946	same_infusion	bolus	bolus
6983	2085	performed_is_no_candidate	diuretic	increase
6983	2085	same_infusion	bolus	bolus
6983	2505	different_infusion	bolus	none
6983	2805	same_infusion	bolus	bolus
6983	2904	same_infusion	bolus	bolus
6983	3165	clearly_contradicting	pressor	increase
6983	3165	performed_is_no_candidate	inotrope	increase

time intervals. The latter turned out to be true, especially when no preprocessing of the data was performed. Outliers repeatedly lead to the inverse decision for a single minute within a longer interval with steady therapy decisions. Following the ward round approach, this would lead to the inverse intervention for at least 30 minutes. But even after cleaning the data from most of the outliers[5] the protocol decisions were often contradictory within short time intervals. Table 2 illustrates such an unsteady behavior. During the according time interval there was no intervention in the dataset. Something that is obviously not reasonable is to change the recommendation from increasing to decreasing pressors and vice versa for intervals as short as two minutes. Inspection of the inferences revealed a serious problem, namely the lack of signal to symbol preprocessing that is stable with respect to noise.

Table 2. Therapy for single patient when protocol is evaluated every minute. Continous therapies are aggregated to intervals. Marked rows denote interventions recommended for a very short time interval.

	Start	End	Object	Action
	19815	20727	pressor	decrease
×	20725	20727	infusion	maintenance
×	20727	20728	pressor	increase
×	20727	20728	infusion	bolus
	20728	21085	infusion	maintenance
	20728	21346	pressor	decrease
×	21346	21347	pressor	increase
	21347	21482	pressor	decrease
×	21482	21483	pressor	increase
×	21483	21484	pressor	decrease
×	21484	21485	pressor	increase
	21485	21497	pressor	decrease

The results regarding the protocol as such are not surprising, but it was remarkably easy to adjust the knowledge base in a fashion allowing for the analysis sketched above. The declarative representation made it easy to communicate emerging questions to a domain expert and to refine inadequate parts.

6 SUMMARY / REMAINING WORK

For setting up and maintaining an ICU protocol this paper proposes to integrate available recorded datasets from the beginning. The use of a knowledge acquisition tool and a logical representation eases to find inadequate protocol parts and inconsistencies by automatically processing available data and comparing suggestions to the decisions of physicians, if available as an integral part of the dataset. Further on, refinements of the model are supported.

Once a protocol is available in declarative and operational form, it can easily be changed for validation in many ways. The first change presented was to incorporate automatized critiquing of derived results. Another example is the change of the ward round based iterations into a monitoring like approach. This yields results related to time intervals, which gives a clearer picture of protocol consistency.

Only little efforts were taken until now to evaluate which preprocessing methods are favorable for the given protocol. A plausibility check to remove outliers was successfully applied, still omitting a reasonable signal-to-symbol preprocessing. The impact of several such preprocessing methods should be compared in the future. The advantage of the declarative representation is that it does not only infer suggestions, but the impact can be studied on different levels of abstraction, for example which discrete category is assigned, which higher level concepts are found to be true and which therapies are chosen. The number of contradictory interventions could be one way of measuring the quality of preprocessing.

[5] A simple plausibility check was used.

REFERENCES

[1] A.Morris. Algorithm-Based Decision Making. In M.J.Tobin, editor, *Principles and Practice of Intensive Care Monitoring*, chapter 77, pages 1355 – 1381. McGraw–Hill, 1998.

[2] C.L.Tsien, I.S.Kohane, and N.McIntosh. Multiple signal integration by decision tree induction to detect artifacts in the neonatal intensive care unit. *Artificial Intelligence in Medicine*, 19, 2000.

[3] D.Calvelo, M.-C.Chambrin, D.Pomorski, and P.Ravaux. Towards symbolization using data-driven extraction of local trends for ICU monitoring. *Artificial Intelligence in Medicine*, 19, 2000.

[4] D.Wang, M.Peleg, S.W.Tu, E.H.Shortliffe, and R.A.Greenes. Representation of Clinical Practice Guidelines for Computer-Based Implementations. In *Proc. of the 10th World Congress on Medical Informatics (MedInfo 2001)*, pages 285 – 289, September 2001.

[5] F.Puppe. Meta Knowledge for Extending Diagnostic Consultation to Critiquing Systems. In *Proceedings of EKAW-99 11th European Workshop on Knowledge Acquisition, Modeling and Management*, pages 367 – 372. Springer Verlag, 1999.

[6] K.Morik. Sloppy Modelling. In K.Morik, editor, *Knowledge Representation and Organization in Machine Learning*, pages 107 – 134. Springer Verlag, 1989.

[7] K.Morik, M.Imhoff, P.Brockhausen, T.Joachims, and U.Gather. Knowledge discovery and knowledge validation in intensive care. *Artificial Intelligence in Medicine*, 19, 2000.

[8] K.Morik, S.Wrobel, J.-U.Kietz, and W.Emde. *Knowledge Acquisition and Machine Learning: Theory, Methods, and Applications*. Academic Press, 1993.

[9] L.T.Kohn, J.M.Corrigan, and M.S.Donaldson, editors. *To Err Is Human: Building a Safer Health System*. National Acadamy Press, Washington, 2000.

[10] M.Bauer, U.Gather, and M.Imhoff. The Identification of Multiple Outliers in Online Monitoring data. Technical Report 29, SFB 475, 1999.

[11] M.Imhoff, M.Bauer, U.Gather, and D.Löhlein. Statistical Pattern Detection in Univariate Time Series of Intensive Care On-Line Monitoring Data. *Intensive Care Med*, 24, 1997.

[12] S.Miksch. Plan management in the medical domain. *AI Communications*, 12(4):209–235, 1999.

[13] S.W.Tu and M.A.Musen. Representation Formalisms and Computational Methods for Modeling Guideline-Based Patient Care. http://www-smi.stanford.edu/pubs/SMI_Reports/SMI-2001-0869.pdf, 2001.

[14] W.Clancey. Heuristic Classification. *Artificial Intelligence*, 27, 1985.

[15] W.Horn. VIE-VENT – A Knowledge-Based System for Monitoring and Optimizing the Artificial Ventilation of Newborn Infants. http://www.ai.univie.ac.at/oefai/kbs/vie-vent.html, 2001.

[16] W.Horn, S.Miksch, G.Egghart, C.Popow, and F.Paky. Effective Data Validation of High-Frequency Data: Time-Point-, Time-Interval-, and Trend-Based Methods. *Computer in Biology and Medicine*, 27(5):389–409, 1997.