Available online at www.sciencedirect.com
ScienceDirect
Knowledge-Based Systems xxx (2006) xxx–xxx
Knowledge-Based SYSTEMS
www.elsevier.com/locate/knosys

Combining task execution and background knowledge for the verification of medical guidelines

Arjen Hommersom [a,*], Perry Groot [a], Peter Lucas [a], Michael Balser [b], Jonathan Schmitt [b]

[a] *Institute for Computing and Information Sciences, Radboud University, Nijmegen, The Netherlands*
[b] *Institut für Informatik, Universität Augsburg, Augsburg, Germany*

Received 9 October 2006; accepted 16 November 2006

Abstract

The use of a medical guideline can be seen as the execution of computational tasks, sequentially or in parallel, in the face of patient data. It has been shown that many of such guidelines can be represented as a 'network of tasks', i.e., as a number of steps that have a specific function or goal. To investigate the quality of such guidelines we propose a formalization of criteria for good practice medicine a guideline should comply to. We use this theory in conjunction with medical background knowledge to verify the quality of a guideline dealing with diabetes mellitus type 2 using the interactive theorem prover KIV. Verification using task execution and background knowledge is a novel approach to quality checking of medical guidelines.

Keywords: Medical guidelines; Verification; Logic

1. Introduction

Computer-based decision support in health-care is a field with a long standing tradition, dealing with complex problems in medicine, such as diagnosing disease and prescribing treatment. The trend of the last decades has been to base clinical decision making more and more on sound scientific evidence, i.e., *evidence-based medicine* [15]. In practice this has led medical specialists to develop evidence-based medical guidelines, i.e., structured documents providing detailed steps to be taken by health-care professionals in managing the disease in a patient, for promoting standards of medical care.

Researchers in Artificial Intelligence have picked up on these developments and are working on providing computer-based support for guidelines by designing computer-oriented languages and developing tools for their deployment. In [4,11] the emergence of a new paradigm is acknowledged for modelling complex clinical processes as a 'network of tasks', which model tasks as a number of steps that have a specific function or goal. Examples of languages that support task modelling are PRO*forma* [4] and Asbru [14], which have been evolving since the 1990s. Medical guidelines are considered to be good real-world examples of highly structured documents amenable to formalisation.

However, guidelines should not be considered *static* objects as new scientific knowledge becomes known on a continuous basis. Newly obtained evidence may result in a deterioration of guideline quality, because, for example, new patient management options invalidate the steps recommended by the guideline. Our aim, therefore, *is to provide support for verifying quality criteria of medical guidelines in light of scientific evidence*.

We approach this problem by applying formal methods to quality check medical guidelines. Here, we are mainly concerned with the meta-level approach [7], i.e., verifying general principles of good practice medicine as for example advocated by the General Medical Council [6]. For example, a guideline of good quality should preclude the prescription of redundant drugs, or advise against the

* Corresponding author. Tel.: +31 24 3653147.
E-mail address: arjenh@cs.ru.nl (A. Hommersom).

0950-7051/$ - see front matter © 2006 Published by Elsevier B.V.
doi:10.1016/j.knosys.2006.11.005

prescription of treatment that is less effective than some alternative. For the verification of such quality criteria, the medical knowledge the guideline is based on, i.e., knowledge based on available evidence, is required. We will refer to this knowledge as *background knowledge*.

The structure of this paper is as follows. First, we model the background knowledge concerning the treatment of diabetes mellitus type 2. Then, the advises, given by the guideline as formalised as a 'network of tasks' using the language Asbru, are modelled. Finally, meta-level properties for this model are formalised and verified in KIV, an interactive theorem prover. To the best of our knowledge, verification of a fully formalised guideline, as a network of tasks, using medical background knowledge has not been done before.

2. Medical guidelines

Clinical practice guidelines are systematically developed statements to assist practitioners and patients decisions about appropriate health care in specific clinical circumstances. A fragment of a guideline is shown in Fig. 1, which is part of the guideline for general Dutch practitioners about the treatment of diabetes mellitus type 2 [13], and is used as a running example in this paper. The guideline contains recommendations for the clinical management in daily practice. Each of these recommendations is well founded in terms of scientific evidence obtained from the literature, in conjunction with other considerations such as safety, availability, or cost effectiveness.

The diabetes mellitus type 2 guideline provides practitioners with a clear structure of recommended interventions to be used for the control of the glucose level. This kind of information is typically found in medical guidelines in the sense that medical knowledge is combined with information about order and time of treatment (e.g., a sulfonylurea drug at step 2), and about patients and their environment (e.g., quetelet index lower than or equal to 27).

Although diabetes mellitus type 2 is a complicated disease, the guideline fragment shown in Fig. 1 is not. This indicates that much knowledge concerning diabetes mellitus type 2 is missing from the guideline and that additional knowledge is needed for verifying whether the guideline fulfils some property. The ideas that we use here for verifying quality requirements for medical guidelines are inspired by Hommersom et al. [7], where a distinction was made between the different types of knowledge that are involved in defining quality requirements. We assume that there are at least three types of knowledge involved in detecting the violation of good practice medicine:

1. Knowledge concerning the (patho)physiological mechanisms underlying the disease, and the way treatment influences these mechanisms (*background knowledge*).
2. Knowledge concerning the recommended treatment in each stage of the plan and how the execution of this plan is affected by the state of the patient (*order information from the guideline*).
3. Knowledge concerning good practice in treatment selection (*quality requirements*).

In the following sections we describe these three types of knowledge in more detail, give a formalisation of all three parts, and verify the requirements.

3. Formalisation of medical guidelines

It has been shown previously that the step-wise, possibly iterative, execution of a guideline can be described precisely by means of temporal logic [9]. In this paper we will use the variant of this logic supported by KIV [1], which is based on linear temporal logic. The language used is first-order logic, augmented with the usual modal operators $\Box$ and $\Diamond$. With $\Box\varphi$ being true if φ is true in the current state and all future states, and $\Diamond\varphi$ if φ holds in the current state or in some state in the future. We also use a special operator **last** which is true exactly if there does not exist a future point in time. Additional modal operators are supported by KIV, but they are not used in this article. Algebraic specifications are used in KIV to model the datatypes.

3.1. Background knowledge

In diabetes mellitus type 2 various metabolic control mechanisms are deranged and many different organ systems may be affected. Glucose level control, however, is the most important mechanism. At some stage in the natural history of diabetes mellitus type 2, the level of glucose in the blood is too high (hyperglycaemia) due to decreased production of insulin by the B cells. Oral anti-diabetics either stimulate the B cells in producing more insulin (sulfonylurea) or inhibit the release of glucose from the liver (biguanide). Effectiveness of these oral diabetics is dependent on the condition of the B cells. Finally, as a causal treatment, insulin can be prescribed. The mechanisms have been formalised in terms of a first-order predicate knowledge:

$$\text{knowledge} : \text{patient} \times \text{patient}$$

where patient denotes an algebraic specification of all first-order formulas describing the patient state, e.g., *condition*(*hyperglycaemia*) represents those patients having a

- Step 1: diet.
- Step 2: if quetelet index (QI) ≤ 27, prescribe a sulfonylurea (SU) drug; otherwise, prescribe a biguanide (BG) drug.
- Step 3: combine a sulfonylurea (SU) and biguanide (BG) drug (replace one of these by a α-glucosidase inhibitor if side-effects occur).
- Step 4: one of the following:
  - oral antidiabetic and insulin
  - only insulin

Fig. 1. Guideline fragment on diabetes mellitus type 2 management. If one of the steps k is ineffective, the management moves to step $k+1$.

condition of hyperglycaemia. The postfix function $[\cdot]$ on patients selects the value for a certain variable from the state, e.g., $Patient[\text{'}condition\text{'}] = hyperglycaemia$ if and only if $condition(hyperglycaemia)$ holds for this patient. The predicate knowledge represents the state transitions that may occur between patient states, i.e., the first argument (denoted by pre below) represents the current patient state and the second argument (denoted by post below) represents the next patient state.

The predicate knowledge has been axiomatised with knowledge concerning the mechanism described above. The axiomatisation is a direct translation of an earlier formalisation in temporal logic [7] of which two examples are:

BDM2-1:
knowledge(pre, post) →
(insulin ∈ pre['treatment'] →
post['uptake(liver,glucose)'] = up ∧
post['uptake(peripheral-tissue,glucose)'] = up)

BDM2-8:
knowledge(pre, post) →
(post['uptake(liver,glucose)'] = up ∧
post['uptake(peripheral-tissue,glucose)'] = up) ∧
(pre['capacity(B-cells,insulin)'] = exhausted ∧
pre['condition'] = hyperglycaemia
→ post['condition'] = normoglycaemia)

The axiom BDM2-1 denotes the physiological effects of insulin treatment, i.e., administering insulin results in an increased uptake of glucose by the liver and peripheral tissues. Axiom BDM2-8 phrases under what conditions you may expect the patient to get cured, i.e., when the patient suffers from hyperglycaemia and insulin production of his B cells are exhausted, an increased uptake of glucose by the liver and peripheral tissues results in the patient condition changing to normoglycaemia.

3.2. Medical guidelines in Asbru

Much research has already been devoted to the development of representation languages for medical guidelines. Most of them consider guidelines as a composition of actions, controlled by conditions [10]. However, most of them are not formal enough for the purpose of our research as they often incorporate free-text elements which do not have a clear semantics. Exceptions to this are PRO*forma* [4] and Asbru [14]. The latter has been chosen in our research.

In Asbru, plans are hierarchically organised in which a plan refers to a number of sub-plans. The overall structure of the Asbru model of our running example (Fig. 1), is shown in Fig. 2. The top level plan Treatments_and_Control sequentially executes the four sub-plans Diet, SU_or_BG, SU_and_BG, and Insulin_Treatments, which correspond to the four steps of the guideline fragment in Fig. 1. The sub-plan Insulin_Treatments is further refined by two sub-plans Insulin_and_Antidiabetics and Insulin, which can be executed in any order.

The Asbru specifications of two plans in the hierarchy, namely SU_or_BG and Insulin_Treatments are defined as follows:

plan SU_or_BG
effects
(QI ⩽ 27 →SU ∈ Drugs) ∧
(QI > 27 →BG ∈ Drugs)
abort condition
condition = hyperglycaemia **confirmation required**
complete condition
condition = hypoglycaemia ∨
condition = normoglycaemia

plan Insulin_Treatments
body anyorder wait for one
Insulin_and_Antidiabetics
Insulin

In the case of SU_or_BG there is a relationship between the quetelet index (QI) and the drug administered. If the quetelet index is less or equal than 27 then SU is administered, else BG is administered. The plan SU_or_BG corresponds to step 2 in the guideline fragment of Fig. 1, which completes if the patient condition improves, i.e., the patient no longer has hyperglycaemia. This is represented by the **complete condition**. The plan SU_or_BG aborts when the

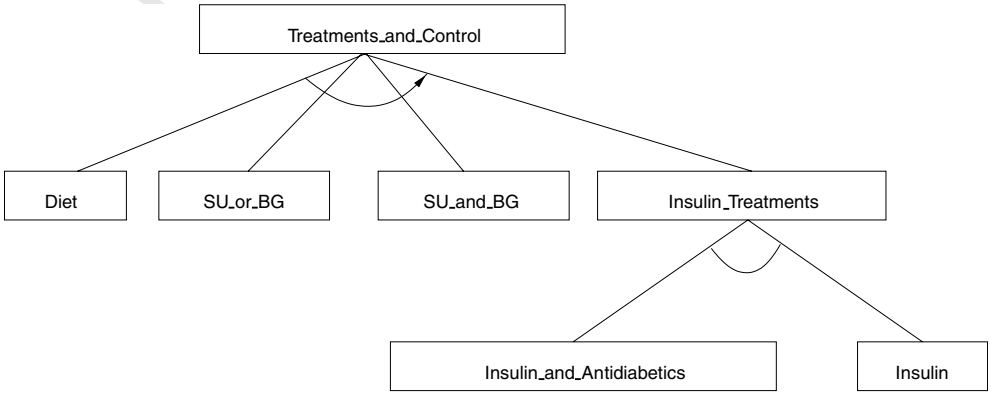

Fig. 2. Asbru plan hierarchy of the diabetes mellitus type 2 guideline.

condition of the patient does not improve, which is represented by the **abort condition**. It requires a manual confirmation to ensure that some time passes for the drugs to have an impact on the patient condition.

The plan Insulin_Treatments consists of two sub-plans, which correspond to the two options of step 4 in the guideline fragment of Fig. 1, i.e., either insulin is administered or insulin and anti-diabetics are administered.

3.3. Quality requirements

Here, we give a formalisation of good practice medicine of medical guidelines. This extends previous work [7], which formalised good practice medicine on the basis of a theory of abductive reasoning of single treatments. The context of the formalisation given here is a fully formalised guideline, which consists, besides a number of treatments, of a control structure that uses patient information to decide on a particular treatment. This contrast with [7], which used a context of a singly chosen treatment.

Firstly, we formalise the notion of a *proper* guideline according to the theory of abductive reasoning. Let $\mathcal{B}$ be medical background knowledge, P be a patient group, N be a collection of intentions, which the physician has to achieve, and M be a medical guideline. Then M is called a *proper* guideline for a patient group P, denoted as $M \in Pr_P$, if:

(M1) $\mathcal{B} \cup M \cup P \not\models \bot$ (the guideline does not have contradictory effects), and
(M2) $\mathcal{B} \cup M \cup P \models \Diamond N$ (the guideline eventually handles all the patient problems intended to be managed)

Secondly, we formalise good practice medicine of guidelines. Let $\preceq_\varphi$ be a reflexive and transitive order denoting a preference relation with $M \preceq_\varphi M'$ meaning that M' is *at least as preferred* to M given criterion φ. With $\prec_\varphi$ we denote the order such that $M \prec_\varphi M'$ if and only if $M \preceq_\varphi M'$ and $M' \not\preceq_\varphi M$. When both $M \preceq_\varphi M'$ and $M' \preceq_\varphi M$ hold or when M and M' are incomparable w.r.t. $\preceq_\varphi$ we say that M and M' are *indifferent*, which is denoted as $M \sim M'$. If in addition to (M1) and (M2) condition (M3) holds, with

(M3) $O_\varphi(M)$ holds, where O_φ is a meta-predicate standing for an optimality criterion or combination of optimality criteria φ defined as: $O_\varphi(M) \equiv \forall M' \in \mathsf{Pr}_P : \neg(M \prec_\varphi M')$,

then the guideline is said to be *in accordance with good practice medicine* w.r.t. criterion φ and patient group P, which is denoted as $\mathsf{Good}_\varphi(M, P)$.

A typical example for O_φ is consistency of the recommended treatment order w.r.t. a preference relation $\preceq_\psi$ *over treatments*, i.e., $O_\varphi(M)$ holds if the guideline M recommends treatment T before treatment T' when $T' \prec_\psi T$ holds. For example, in diabetes mellitus type 2, a preference relation over treatments would be to minimise (1) the number of insulin injections, and (2) the number of drugs involved. This results, among others, in the following preferences: sulfonylurea drug $\sim$ biguanide drug, and insulin $\preceq_\psi$ insulin and anti-diabetic $\preceq_\psi$ sulfonylurea and biguanide drug $\preceq_\psi$ sulfonylurea or biguanide drug $\preceq_\psi$ diet. A guideline M would then be in accordance with good practice medicine if it is consistent with this preference order $\preceq_\psi$, e.g., if M first recommends diet before a sulfonylurea or biguanide drug.

4. Verification using KIV

The formal verification was done with the interactive verification tool KIV [1]. A speciality of KIV is the use of primed and double-primed variables: a primed variable V' represents the value of this variable after a system transition, the double-primed variable V'' is interpreted as the value after an environment transition, where the environment transition models the communication of the system with its environment. System and environment transitions alternate, as shown in Fig. 3, with V'' being equal to V in the successive state.

With the help of KIV, we have verified that the diabetes guideline is proper, i.e., that the guideline satisfies conditions (M1) and (M2), which is discussed in Subsections 4.1 and 4.2. Furthermore, with KIV we have verified various meta-level quality requirements of the diabetes mellitus type 2 guideline. Each meta-level quality requirement is verified using a sequent $\Gamma \vdash \Sigma$ where the succedent Σ is some instantiation of (M3) and the antecedent Γ consists of the initial state of a patient group, the initial state of the guideline, the medical guideline, effects of treatment plans, the background knowledge, and the environment assumptions, which is shown in Fig. 4. The verification of two meta-level requirements are discussed in Sections 4.3 and 4.4.

4.1. Consistency of background knowledge

Property (M1) ensures that the formal model including the Asbru guideline and the background knowledge is consistent. The initial state is – in our case – described as a set of equations and it has been trivial to see that they are consistent. The guideline is given as an Asbru plan. The seman-

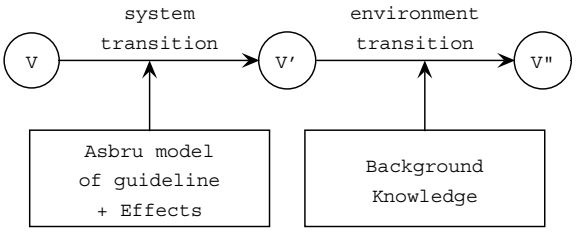

Fig. 3. The relation between unprimed and primed variables as two transitions: the system transition (including the Asbru model and its effects) and the environment transition (including the background knowledge).

$AS[\text{`}tc\text{'}] = inactive, \ldots,$ /* Initial state of guideline */
$[inactive\#(\text{`}tc\text{'}, \text{`}st\text{'}; AS, Patient)],$ /* Asbru plan */
$\Box(AS[\text{`}SU\_or\_BG\text{'}] = activated \leftrightarrow$ /* Effects */
$BG \in Patient'[\text{`}treatment\text{'}] \wedge \ldots),$
$\Box\text{knowledge}(Patient', Patient'')$ /* Background knowledge */
$\Box(AS''[\text{`}tc\text{'}] = AS'[\text{`}tc\text{'}] \wedge \ldots)$ /* Environment assumption */

Fig. 4. Antecedent of proof obligations with tc shorthand for Treatments_and_Control and AS an additional data structure of type asbru-state, which keeps track of all plan states over time, in which initially each plan is set to inactive.

tics of any Asbru plan is defined in a programming language where every program construct ensures that the resulting reactive system is consistent: in every step, the program either terminates or calculates a consistent output for arbitrary input values. The Asbru plan, thus, defines a total function between unprimed and primed variables in every step (Fig. 3). The formula defining the effects maps the output variables of the guideline to input variables of the patient model. Again, it has been trivial to see that this mapping is consistent.

The background knowledge defines our patient model. We consider the patient to be part of the environment which is the relation between the primed and the double primed variables in every step. If the patient model ensures that for an arbitrary primed state there exists a double primed state, the overall system of alternating guideline and environment transitions is consistent: given an initial (unprimed) state, the guideline calculates an output (primed) state; the effects define a link between the variables of the guideline and the variables of the patient model; the patient model reacts to the (primed) output state and gives a (double primed) state which is again input to the Asbru guideline in the next step. In other words, the relation between the unprimed and the double primed state is the complete state transition. The additional environment assumptions referring to the Asbru environment do not destroy consistency as the set of restricted variables of the environment assumption is disjunct to the set of variables of the patient model.

It remains to ensure consistency of the background knowledge which we defined as a predicate knowledge. Consistency can be shown by proving the property

$$\forall pre.\ \exists post.\ \text{knowledge}(pre, post)$$

which ensures that the relation is total. In order to prove that this property holds an example patient has been constructed. Verifying that the example patient is a model of the background knowledge has been fully automatic.

4.2. Successful treatment

In order to verify property (M2), i.e., the guideline eventually manages to control the glucose level in the patient's blood, a proof has been constructed. The verification strategy in KIV is symbolic execution with induction [1]. The plan state model introduced in [2] defines the semantics of the different conditions of a plan and is implemented in KIV by a procedure called asbru, which is symbolically executed. Each plan can be in a certain state, modelled with a variable AS (i.e., inactive, considered, ready, activated, and aborted (or completed)) and a transition to another state depends on its conditions. In the initial state, the top level plan Treatments_and_Control (abbreviated tc) is in inactive state. After executing the first step, the plan is considered, after which execution continues as described in [2]. The execution is visualised in a proof tree (cf. Fig. 5), where the bottom node is the start of the execution and splits if there is a case distinction.

Patients whose capacity of the B cells is normal are cured with diet, while for other patients diet will not be sufficient. In this case, we assume that the doctor eventually aborts the diet treatment. We use induction to reason about the unspecified time period in which diet is applied. As an invariant,

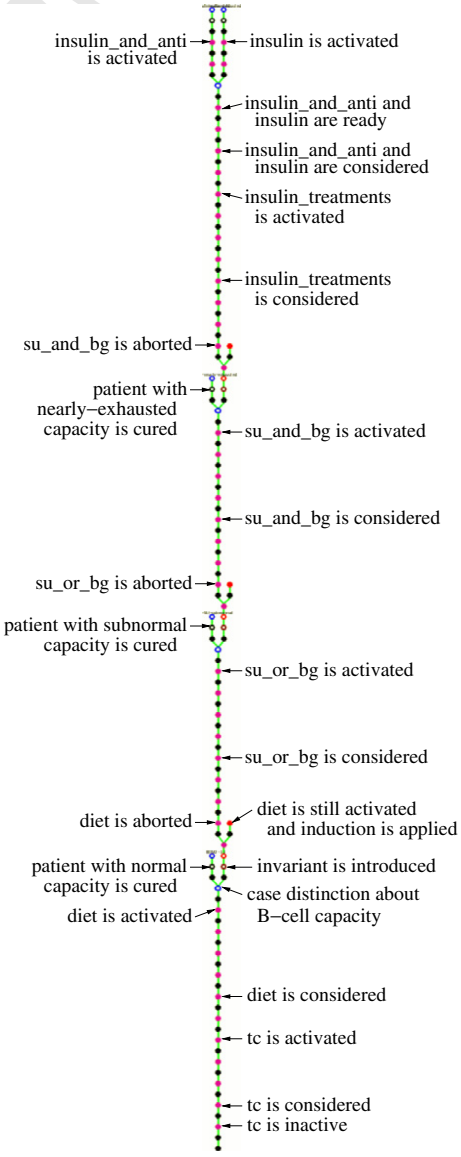

Fig. 5. Overview of the proof that the guideline eventually manages all patient problems, which is explained in Section 4.2.

$$Patient[\text{`}capacity(B\text{-}cells, insulin)\text{'}] \neq normal$$

is used. In the next step, the doctor has either aborted diet or diet is still active. In the second case, induction can be applied. When diet is aborted, tc sequentially executes the next plan, which is SU_or_BG (cf. Fig. 2).

The second treatment SU_or_BG goes, as each Asbru plan, through a sequence of states, i.e., inactive, considered, ready, activated, and aborted, and thus becomes first considered and after some steps becomes activated (cf. Fig. 5). In this case, either SU or BG is prescribed, depending on the quetelet index QI. For a patient whose B cell capacity is subnormal, the background knowledge ensures that the condition of the patient improves. Thus, for the rest of the proof we can additionally assume that

$$Patient[\text{`}capacity(B\text{-}cells, insulin)\text{'}] \neq subnormal$$

After SU_or_BG aborts, the third treatment (SU_and_BG) is executed in similar fashion, where patients with nearly exhausted B cell capacity are cured. Thus, after aborting the first three treatments the precondition concerning the B cell capacity can be strengthened to

$$\begin{aligned} & Patient[\text{`}capacity(B\text{-}cells, insulin)\text{'}] \neq normal \\ \wedge\ & Patient[\text{`}capacity(B\text{-}cells, insulin)\text{'}] \neq subnormal \\ \wedge\ & Patient[\text{`}capacity(B\text{-}cells, insulin)\text{'}] \neq nearly\ exhausted \end{aligned}$$

which, under the assumption that the only possible values of the capacity are normal, subnormal, nearly exhausted, and exhausted, yields:

$$Patient[\text{`}capacity(B-cells, insulin)\text{'}] = exhausted$$

This statement together with the background knowledge ensures that the prescription of insulin, which is prescribed in both final treatments Insulin and Insulin_and_Antidiabetics, finally cures the patient.

4.3. Optimality of treatment

With respect to property (M3), an optimality criterion of the guideline is that no treatments are prescribed that are not in accordance with good practice medicine (Section 3.3), i.e., some preference relation $\preceq$ between treatments exists and the guideline never prescribes a treatment T such that $T \preceq T'$ and T' cures the patient group under consideration.

In our case study the preference for treatments is based on the minimisation of: (1) the number of insulin injections, and (2) the number of drugs involved (cf. Section 3.3). We have defined this using a reflexive, transitive order $\leqslant$ such that for all treatments T, it holds that $insulin \leqslant T$ and $T \leqslant diet$. Furthermore, the treatments prescribing the oral anti-diabetics sulfonylurea and biguanide are incomparable. The proof obligation is then as follows:

$$\Box(\forall_T : Good_{\leqslant}(T, Patient) \rightarrow T \leqslant Patient[\text{`}treatment\text{'}])$$

where $Good_{\leqslant}(T,Patient)$ denotes that T is a treatment according to good practice medicine for *Patient*, as defined in [8]. To prove this, the following axiom was added to the system:

$$\Box Patient[\text{`}QI\text{'}] = Patient''[\text{`}QI\text{'}]$$

i.e., the quetelet index does not change during the run of the protocol. This axiom is needed, because the decision of prescribing a treatment is not exactly at the same time as the application of the treatment and therefore the decision of prescribing this treatment could be based on a patient with a different quetelet index than the patient that actually takes the drugs.

Proving this property in KIV was done in approximately 1 day using several heuristics for the straightforward parts. The theorem was proven using two lemmas for two specific patient groups. In total, it took approximately 500 steps, of which nearly 90% were done automatically, to verify this property.

4.4. Order of treatments

Finally, another instance of (M3) was proven. This property phrases that the order of any two treatments in the protocol is consistent with the order relation as we have defined in Section 3.3. In other words, in case a patient may receive multiple treatments, the less radical treatments are tried first. The formalisation of the property in KIV was done as follows:

$$\Box\forall_T(Tick \wedge T = Patient[\text{`}treatment\text{'}] \rightarrow \Box(\textbf{last} \vee (Tick \rightarrow \neg(T \leqslant Patient[\text{`}treatment\text{'}]))))$$

At each time, the current treatment is bound to a static variable (i.e., unchanged by symbolic execution) T, which can be used to compare against subsequent steps in the protocol. For any future steps, we require that either the protocol completes (**last** holds) or that activated treatments are not more preferred than T. The additional Tick variable is needed in the formalisation to abstract from technical system steps.

This property also had a high degree of automation with roughly 800 steps in total. The reason for this slightly higher number of steps is due to nested temporal operators.

5. Discussion

As the interest in medical guidelines continues to grow, there is a need for criteria to asses the quality of medical guidelines. An important method for the appraisal of medical guidelines was introduced by the AGREE collaboration [3]. A solid foundation for the application of *formal methods* to the quality checking of medical guidelines, using simulation of the guideline [4,12] and theorem proving techniques [9], can also be found in literature.

In [9], logical methods have been used to analyse properties of guidelines, formalised as task networks. In [8], it was shown that the theory of abductive diagnosis can be taken as a foundation for the formalisation of quality

requirements of a medical guideline in temporal logic. This result has been used to verify quality requirements of good practice medicine of treatments [7]. However, in the latter work, the order between treatment depending on the condition of the patient and previous treatments was ignored. In this paper, we consider elements from both approaches by including medical background knowledge in the verification of complete networks of tasks. This required a major change to the previous work with respect to the formulation of quality criteria, because quality is now defined with respect to a complete network of tasks instead of individual treatments as presented in [8].

Compared to previous work concerning the verification of networks of tasks, the meta-level approach we have presented here has a number of advantages. In the meta-level approach, quality is defined independently of domain specific knowledge, and, consequently, proof obligations do not have to be extracted from external sources. One successful attempt of the latter was reported in [5], where quality criteria are formalised on the basis of instruments to monitor the quality of care in practice, i.e., medical indicators. Firstly, the question is whether these indicators, based on compliance with medical guidelines, coincide with the quality of the guideline itself. Secondly, it has been our experience that it is far from easy to find suitable properties in external sources, because these sources may not be completely applicable, e.g., typically, other guidelines may address different problem in the management of the same disease. Thirdly, many useful quality criteria of guidelines are implicit, making this approach fundamentally limiting. In this sense, the meta-level approach provides a more systematic method for the formulation of proof obligations and, thus, verification of medical guidelines.

In summary, in this study we have setup a general framework for the verification of medical guidelines, consisting of a medical guideline, medical background knowledge, and quality requirements. A model for the background knowledge of glucose level control in diabetes mellitus type 2 patients was developed based on a general temporal logic formalisation of (patho)physiological mechanisms and treatment information. Furthermore, we developed a theory for quality requirements of good practice medicine based on the theory of abductive diagnosis. This model of background knowledge and theory of quality requirements were then used in a case study in which we verified several quality criteria of the diabetes mellitus type 2 guideline used by the Dutch general practitioners. In the case study we use Asbru to model the guideline as a network of tasks and KIV for the formal verification.

In the course of our study we have shown that the general framework that we have setup for the formal verification of medical guidelines with medical background knowledge is feasible and that the actual verification of quality criteria can be done with a high degree of automation. We believe both the inclusion of medical background knowledge and task networks to be necessary elements for adequately supporting the development and management of medical guidelines.

An important advantage of using theorem proving compared to alternative techniques such as model checking is that it provides insight in the proof structure. For each case, it is relatively easy to inspect the proof tree and to find out the reason that a certain quality criterion holds. On the other hand, KIV is a tool with a very expressive logic, which may result in an additional overhead when verifying quality criteria of medical guidelines. It is clear that tools for quality checking earlier on in the development process of a guideline, where such an additional overhead is not acceptable, would be useful. Therefore, also techniques such as model checking will be a topic for future research.

References

[1] M. Balser, Verifying Concurrent System with Symbolic Execution – Temporal Reasoning is Symbolic Execution with a Little Induction. PhD thesis, University of Augsburg, Augsburg, Germany, 2005.
[2] M. Balser, C. Duelli, W. Reif, Formal semantics of asbru – an overview, in: Proceedings of the International Conference on Integrated Design and Process Technology, Passadena, Society for Design and Process Science, June 2002.
[3] AGREE Collaboration, Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the agree project, Quality & Safety in Health Care 12 (2003) 18–23.
[4] J. Fox, S. Das, Safe and Sound: Artificial Intelligence in Hazardous Applications, AAAI Press, 2000.
[5] M. van Gendt, A. van Teije, R. Serban, F. van Harmelen, Formalising medical quality indicators to improve guidelines, in: AIMELNAI, vol. 3581, Springer Verlag, 2005, pp. 201–220.
[6] General Medical Council – Protecting patients, guiding doctors, Good Medical Practise, third edition, May 2001. <http://www.gmc-uk.org/>.
[7] A.J. Hommersom, P.J.F. Lucas, M. Balser, Meta-level Verification of the Quality of Medical Guidelines Using Interactive Theorem Proving, in: JELIA'04Lecture Notes in Computer Science, vol. 3229, Springer-Verlag, 2004, pp. 654–666.
[8] P.J.F. Lucas, Quality checking of medical guidelines through logical abduction, in: F. Coenen, A. Preece, A.L. Mackintosh (Eds.), Proc. of AI-2003, vol. XX, Springer, London, 2003, pp. 309–321.
[9] M. Marcos, M. Balser, A. ten Teije, F. van Harmelen, From informal knowledge to formal logic: a realistic case study in medical protocols, in: Proceedings of EKAW, Springer, 2002, pp. 49–64.
[10] S. Miksch, Plan management in the medical domain, AI Communications 12 (4) (1999) 209–235.
[11] M. Peleg et al., Comparing computer-interpretable guideline models: a case-study approach, Journal of the American Medical Informatics Association 10 (1) (2003) 52–68.
[12] S. Quaglini, M. Stefanelli, A. Cavallini, G. Micieli, C. Fassino, C. Mossa, Guideline-based careflow system, Artificial Intelligence in Medicine 20 (1) (2000) 5–22.
[13] G.E.H.M. Rutten, S. Verhoeven, R.J. Heine, W.J.C. de Grauw, P.V.M. Cromme, K. Reenders, NHG-standard diabetes mellitus type 2 (eerste herziening), Huisarts Wet 42 (1999) 67–84.
[14] A. Seyfang, R. Kosara, S. Miksch, Asbru's reference manual, asbru version 7.3. Technical Report Asgaard-TR-20002-1, Vienna University of Technology, Institute of Software Technology, 2002.
[15] S.H. Woolf, Evidence-based medicine and practice guidelines: an overview, Cancer Control 7 (4) (2000) 362–367.