

Quality Checking of Medical Guidelines through Logical Abduction*

Peter Lucas

Institute for Computing and Information Sciences
University of Nijmegen, Toernooiveld 1
6525 ED Nijmegen, The Netherlands
E-mail: peterl@cs.kun.nl

Abstract

Formal methods have been used in the past for the verification of the correctness of formalised versions of medical guidelines. In this paper a second possible application of the use of formal methods is proposed: checking whether a guideline conforms to global medical quality requirements. It is argued that this allows spotting design errors in medical guidelines, which is seen as a useful application for formal methods in medicine. However, this type of verification may require medical knowledge currently not available within the guidelines, i.e. medical background knowledge. In this paper, we propose a method for checking the quality of a treatment for a disorder, based on the theory of abductive diagnosis. We also examine the medical background knowledge required to be able to quality check a guideline. The method is illustrated by the formal analysis of an actual guideline for the management of diabetes mellitus type 2.

1 Introduction

There has been significant progress in the life sciences over the last couple of decades, yielding results, for example in molecular biology and human genetics, which are now gradually reshaping the face of medicine. One consequence of this is now becoming apparent: medicine is becoming more and more complicated at an astonishing rate. On the one hand, the number of different patient management options has risen considerably, whereas, on the other hand, health-care authorities increasingly expect that medical doctors take decisions balancing benefits for the patient against financial costs. The latter will only happen if medical doctors are offered appropriate support in their decision-making process.

A first step into this direction was taken by the medical profession itself by understanding the need of basing clinical decision-making more and more on sound scientific evidence; this trend has become known as *evidence-based medicine* [14, 17]. In practice, evidence-based medicine implies that general practitioners and medical specialists in particular areas are increasingly involved in the development of guidelines, aimed at promoting standards of medical care. Guidelines are documents offering a detailed description of steps that must be taken by health-care professionals in managing a disease in a patient to avoid substandard practices

*Published in: F. Coenen, A. Preece, A.L. Mackintosh (eds.). Proceedings of AI-2003 (Research and Developments in Intelligent Systems XX), Springer, London, pp. 309–321.

or outcomes. Modern guidelines are based on scientific evidence derived from good-quality studies reported in medical journals, rather than on consensus among medical doctors as was the case just a few years ago.

Some researchers in artificial intelligence (AI) have picked up on these developments, and are involved in the design of computer-oriented languages, tools and systems that support the design and deployment of medical guidelines. Examples are PROforma [3, 4], Asbru [15] and GLIF [11]; these languages and their associated tools have been evolving since the 1990s, a process that is now gaining momentum due to the increased interest in guidelines within the medical community. AI researchers see guidelines as good real-world examples of highly structured, systematic documents that are amenable to formalisation.

What the research efforts mentioned above have in common is that medical guidelines are taken as a start and are subsequently translated, sometimes only partially, into a guideline-representation language, so that they can be executed as a program, or at least in principle. Once a guideline has been represented in this way, the resulting program-like representation may also be used as a basis to investigate properties of the original guideline. In some cases, researchers have gone one step further and used logical methods to analyse guidelines [4, 10]. These approaches offer potentially useful ways to exploit formal methods in the context of medical guidelines. However, so far AI researchers have had difficulty in achieving medically relevant results, i.e. the sort of results that could act as input to the improvement of medical guidelines. This is not surprising as the development of medical guidelines is done in a relatively thorough fashion by a team of expert clinicians and guideline developers; even though the resulting documents are in natural language, they rarely contain obvious flaws. The ‘flaws’ that researchers have been able to identify are typical for natural language text; whether they are also typical for guidelines as such is questionable. Examples of detected ‘flaws’ in the text include ambiguities (e.g. use of the term ‘clinical jaundice’ in conjunction with ‘jaundice’, intended to mean the same thing) and incompleteness (e.g. ‘rapid increase in blood glucose levels’, where the adjective ‘rapid’ is not precisely defined) [10]; this likely only indicates that the guideline developers choose to refrain from detail. This approach of translating the text of a guideline into a formal language, and then using tools for the formal analysis of the formal representation is therefore not without problems. We call this the *object-level* approach.

Another approach is to abstract from the actual guideline text and to ask the question what general medical quality criteria need to be fulfilled by the guideline. For example, one would expect that a good-quality medical guideline regarding treatment of a disorder would preclude the prescription of redundant drugs, or advise against the prescription of treatment that is less effective than some alternative. Carrying out such checks could be valuable, in particular during the process of *designing* medical guidelines. This is called the *meta-level* approach.

In this paper, we propose a meta-level approach to quality checking of medical guidelines, based on the theory of *abductive diagnosis* [1, 8, 9, 12]. We also examine the medical background knowledge required to be able to quality check a guideline. Instead of using one of the guideline-representation languages, we explore the use of temporal logic for these purposes, as this allows us to abstract from unnecessary detail, and this language is compatible with program logics, i.e. the interpretation of guidelines as programs. The usefulness of the method is studied by analysing the official guideline on diabetes mellitus type 2 management currently in use by the Dutch general practitioners.

The paper is organised as follows. In the next section, we start by explaining what

medical guidelines are and what purpose they serve in the health-care; in addition, the role formal methods have played so far in checking medical guidelines is described. In Section 3 a method for formalising guidelines and checking them against design decisions based on abductive diagnosis is described. The guideline on the management of diabetes mellitus type 2 that has been used in the case study is given attention to in Section 4. The analysis of this guideline using the method described in Section 3 can be found in Section 5. Finally, Section 6 discusses what has been achieved, offers further motivation on the usefulness of the developed method in comparison to other methods, and suggests some future plans for research.

2 Medical Guidelines

A medical guideline is a structured document, containing detailed advice on the management of a particular disorder or group of disorders, aimed at health-care professionals. As modern guidelines are based on scientific evidence, they contain information about the quality of the evidence on which particular statements are based; e.g. statements at the highest recommendation level are usually obtained from randomised clinical trials [17].

The design of a medical guidelines is far from easy. Firstly, the gathering and classification of the scientific evidence underlying and justifying the recommendations mentioned in a guideline is time consuming, and requires considerable expertise in the medical field concerned. Secondly, medical guidelines are very detailed, and making sure that all the information contained in the guideline is complete for the guideline’s purpose, and based on sound medical principles is hard work. An example of a tiny portion of a guideline is shown in Figure 1; it is part of the guideline for general practitioners about the treatment of diabetes mellitus type 2. This guideline fragment is used in this paper as a running example.

One way to use formal methods in the context of guidelines is to automatically verify whether a medical guideline fulfils particular properties, such as whether it complies with quality *indicators* as proposed by health-care professionals [10]. For example, using particular patient assumptions such as that after treatment the levels of a substance are dangerously high or low, it is possible to check whether this situation does or does not violate the guideline. However, verifying the effects of treatment as well as examining whether a developed medical guideline complies with global criteria, such as that it avoids the prescription of redundant drugs, or the request of tests that are superfluous, is difficult to impossible if only the guideline text is available.

Thus, the capability to check whether a guideline fulfils particular medical objectives may require the availability of more medical knowledge than is actually specified in a medical guideline. How much additional knowledge is required may vary from guideline to guideline. In the development of the theory below it is assumed that at least some medical background knowledge is required; the extent and the purpose of that background knowledge is subsequently established using a real medical guideline. The development and evaluation of a formal method that supports this process is the topic of the remainder of the paper.

3 Formalisation of Good Practice Medicine

Medical guidelines regarding a disorder usually contain particular sections dealing with the diagnosis, treatment and follow-up of a disorder. In this study, we restrict ourselves to the

-
- Step 1: diet
 - Step 2: if Quetelet Index (QI) ≤ 24 , prescribe a sulfonylurea drug; otherwise, prescribe a biguanide drug
 - Step 3: combine a sulfonylurea drug and biguanide (replace one of these by a α -glucosidase inhibitor if side-effects occur)
 - Step 4: one of the following:
 - oral antidiabetics and insulin
 - only insulin
-

Figure 1: Tiny fragment of a clinical guideline on the management of diabetes mellitus type 2. If one of the steps $k = 1, 2, 3$ is ineffective, the management moves to step $k + 1$.

Table 1: Used temporal operators; t stands for a time instance.

Notation	Informal meaning	Formal meaning
$H\varphi$	φ has always been true in the past	$t \models H\varphi \Leftrightarrow \forall t' < t : t' \models \varphi$
$G\varphi$	φ is true at all future times	$t \models G\varphi \Leftrightarrow \forall t' > t : t' \models \varphi$

treatment part, although similar principles may apply to the other parts of a guideline.

3.1 Using Temporal Logic

As medical management is a time-oriented process, diagnostic and treatment actions described in guidelines are performed in a temporal setting. It has been shown previously that the step-wise, possibly iterative, execution of a guideline, such as the example in Figure 1, can be described precisely by means of temporal logic [10]. This is a modal logic, where relationships between worlds in the usual possible-world semantics of modal logic is understood as time order, i.e. formulae are interpreted in a *temporal frame* $\mathcal{F} = (\mathbb{T}, <, I)$, where \mathbb{T} is the set of intervals or time points, $<$ a time ordering, and I an interpretation of the language elements with respect to \mathbb{T} and $<$. Modal operators added to the language of first-order logic include \circ , where $\circ\varphi$ has the operational meaning of an execution step, followed by execution of program part φ , formally: **step**; φ .

In this paper, we abstract from such detailed execution semantics of guidelines, and use more abstract temporal operators. The language of standard logic, with equality and unique names assumption, is therefore augmented with the operators G , H , P and F , where the temporal semantics of the first two operators is defined in Table 1. The last two operators are simply defined in terms of the first two operators:

$$\begin{aligned} \models P\varphi &\leftrightarrow \neg H\neg\varphi && \text{(somewhere in the past)} \\ \models F\varphi &\leftrightarrow \neg G\neg\varphi && \text{(somewhere in the future)} \end{aligned}$$

This logic offers the right abstraction level to cope with the nature of the temporal knowledge in medical guidelines required for our purposes. However, more fine-grained temporal

operators can be added if needed.

In addition it is necessary to specify additional axioms that indicate that progression in time is *linear* (there are other axiomatisations possible, such as branching time, see [16]). The most important of these are:

- (1) *Transitivity*: $\models \text{FF}\varphi \rightarrow \text{F}\varphi$
- (2) *Backward linearity*: $\models (\text{P}\varphi \wedge \text{P}\psi) \rightarrow (\text{P}(\varphi \wedge \psi) \vee \text{P}(\varphi \wedge \text{P}\psi) \vee \text{P}(\text{P}\varphi \wedge \psi))$
- (3) *Forward linearity*: $\models (\text{F}\varphi \wedge \text{F}\psi) \rightarrow (\text{F}(\varphi \wedge \psi) \vee \text{F}(\varphi \wedge \text{F}\psi) \vee \text{F}(\text{F}\varphi \wedge \psi))$

Transitivity ensures that we can move along the time axis from the past into the future; backward and forward linearity ensure that the time axis does not branch. Consider, for example, axiom (3), which says that if there exists a time t when φ is true, and a time t' when ψ holds, then there are three possibilities: φ and ψ hold at the same time, or at some time in the future φ and further away in the future ψ hold; the meaning of the last disjunct is similar. Other useful axioms concern the boundedness of time; assuming that time has no beginning and no end, gives rise to the following axioms: $\models \text{H}\varphi \rightarrow \text{P}\varphi$ and $\models \text{G}\varphi \rightarrow \text{F}\varphi$.

3.2 Application to Medical Knowledge

It is assumed that two types of knowledge are involved in detecting the violation of good medical practice:

- Knowledge concerning the (patho)physiological mechanisms underlying the disease, and the way treatment influences these mechanisms. The knowledge involved could be causal in nature, and is an example of *object-knowledge*.
- Knowledge concerning good practice in treatment selection; this is *meta-knowledge*.

Below we present some ideas on how such knowledge may be formalised using temporal logic (cf. [7] for earlier work).

We are interested in the prescription of drugs, taking into account their mode of action. Abstracting from the dynamics of their pharmacokinetics, this can be formalised in logic as follows:

$$(\text{G}d \wedge r) \rightarrow \text{G}(m_1 \wedge \dots \wedge m_n)$$

where d is the name of a drug or possibly of a group of drugs indicated by a predicate symbol (e.g. $\text{SU}(x)$, where x is universally quantified and ‘SU’ stands for sulfonylurea drugs, such as Tolbutamid), r is a (possibly negative or empty) *requirement* for the drug to take effect, and m_k is a mode of action, such as decrease of release of glucose from the liver, which holds at all future times.

The modes of action m_k can be combined, together with an *intention* n (achieving normoglycaemia, i.e. normal blood glucose levels, for example), a particular patient *condition* c , and *requirements* r_j for the modes of action to be effective:

$$(\text{G}m_{i_1} \wedge \dots \wedge \text{G}m_{i_m} \wedge r_1 \wedge \dots \wedge r_p \wedge \text{H}c) \rightarrow \text{G}n$$

Good practice medicine can then be formalised as follows. Let \mathcal{B} be background knowledge, $T \subseteq \{d_1, \dots, d_p\}$ be a set of drugs, C a collection of patient conditions, R a collection of requirements, and N a collection of intentions which the physician has to achieve. A set of drugs T is a *treatment* according to the theory of abductive reasoning if [12]:

- (1) $\mathcal{B} \cup \mathcal{G}T \cup \mathcal{C} \cup \mathcal{R} \not\models \perp$ (the drugs do not have contradictory effects), and
- (2) $\mathcal{B} \cup \mathcal{G}T \cup \mathcal{C} \cup \mathcal{R} \models N$ (the drugs handle all the patient problems intended to be managed)

If in addition to (1) and (2) condition

- (3) $O_\varphi(T)$ holds, where O_φ is a meta-predicate standing for an optimality criterion or combination of optimality criteria φ , then the treatment is said to be *in accordance with good-practice medicine*.

A typical example of this is subset minimality O_C :

$$O_C(T) \equiv \forall T' \subset T : T' \text{ is not a treatment according to (1) and (2)}$$

i.e. the minimum number of effective drugs are being prescribed. For example, if $\{d_1, d_2, d_3\}$ is a treatment that satisfies condition (3) in addition to (1) and (2), then the subsets $\{d_1, d_2\}$, $\{d_2, d_3\}$, $\{d_1\}$, and so on, do not satisfy conditions (1) and (2). In the context of abductive reasoning, subset minimality is often used in order to distinguish between various solutions; it is also referred to in literature as *Occam's razor*. Another definition of the meta-predicate O_φ is in terms of minimal cost O_c :

$$O_c(T) \equiv \forall T', \text{ with } T' \text{ a treatment: } c(T') \geq c(T)$$

where $c(T) = \sum_{d \in T} \text{cost}(d)$; combining the two definitions also makes sense. For example, one could come up with a definition of $O_{C,c}$ that among two subset-minimal treatments selects the one that is the cheapest in financial or ethical sense.

4 Management of Diabetes Mellitus Type 2

An analysis of the medical guideline on diabetes mellitus type 2 was carried out as a case study. An additional aim was to identify the kind of background knowledge required in order to determine the global quality of the guideline, and so the guideline itself was only formalised so far as required to investigate the usefulness of the theory of quality checking introduced above.

4.1 Initial Analysis

It is well known that diabetes type 2 is a very complicated disease: various metabolic control mechanisms are deranged and many different organ systems, such as the cardiovascular and renal system, may be affected by the disorder. Here we focus on the derangement of glucose metabolism in diabetic patients, and even that is nontrivial. To support non-expert medical doctors in the management of this complicated disease in patients, access to a guideline is really essential.

One would expect that as this disorder is so complicated, the diabetes mellitus type 2 guideline is also complicated. This, however, is not the case, as may already be apparent from the guideline fragment shown in Figure 1. This already indicates that much of the knowledge concerning diabetes mellitus type 2 is missing from the guideline, and that without this background knowledge it will be impossible to spot the sort of flaws we are after. Hence, the conclusion is that a deeper analysis is required; the results of such an analysis are discussed next.

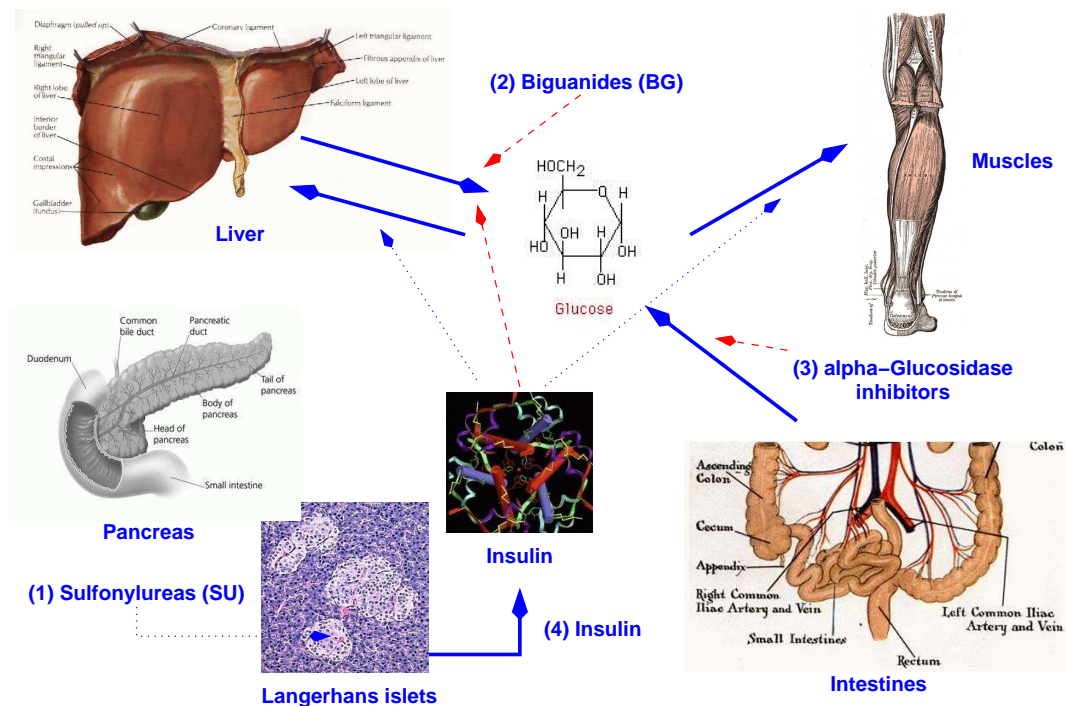


Figure 2: Summary of drugs and mechanisms controlling the blood level of glucose; $- - \rightarrow$: inhibition, $\cdots \rightarrow$: stimulation.

4.2 Diabetes Type 2 Background Knowledge

Figure 2 summarises the most important mechanisms and drugs involved in the control of the blood level of glucose. The protein hormone insulin, which is produced by the *B cells* in the Langerhans islets of the *pancreas*, has the following major effects:

- it increases the uptake of glucose by the liver, where it is stored as glycogen, and inhibits the release of glucose from the liver;
- it increases the uptake of glucose by insulin-dependent tissues, such as muscle and adipose tissue.

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. A popular hypothesis explaining this phenomenon is that target cells have become insulin resistant, which with a delay causes the production of insulin by the B cells to raise. After some time, the B cells become exhausted, and they are no longer capable of meeting the demands for insulin. As a consequence, hyperglycaemia develops.

Treatment of diabetes type 2 consists of:

- Use of *sulfonylurea* (SU) drugs, such as tolbutamid. These drugs stimulate the B cells in producing more insulin, and if the cells are not completely exhausted, the hyperglycaemia can thus be reverted to normoglycaemia (normal blood glucose levels).
- Use of *biguanides* (BG), such as metformin. These drugs inhibit the release of glucose from the liver.

- Use of α -glucosidase inhibitors. These drugs inhibit (or delay) the absorption of glucose from the intestines.
- Injection of *insulin*. This is the ultimate, causal treatment.

As insulin can only be administered by injection, in contrast to the other drugs which are normally taken orally, doctors prefer to delay prescribing insulin as long as possible. Thus, the treatment part of the diabetes type 2 guideline mentions that one should start with prescribing oral antidiabetics (SU or BG, cf. Figure 1). Two of these can also be combined if taking only one has insufficient glucose-level lowering effect. If treatment is still unsatisfactory, the guideline suggests to: (1) either add insulin, or (2) stop with the oral antidiabetics entirely and to start with insulin.

From a medical point of view, advice (1) above is somewhat curious. Since the oral antidiabetics are no longer effective, the conclusion must be that the B cells are exhausted. Under these circumstances, it does not make a lot of sense to prescribe an SU drug. Prescription of a BG (or α -glucosidase inhibitor) is still justified, as by adding such an oral antidiabetic to insulin, the number of necessary injections can be reduced from twice a day to once a day. Finally, when on insulin treatment, patients run the risk of getting hypoglycaemia, which can be seen as a side effect of insulin not mentioned explicitly in the guideline.

In the remainder of the paper, we will carry out an analysis of the diabetes mellitus type 2 guideline adopting the method described in Section 3 to demonstrate its usefulness.

5 Formalisation of the Diabetes Mellitus 2 Quality Check

The formal analysis of the guideline consists of two steps: first the necessary background knowledge, identified in the previous section, is formalised using temporal logic. Subsequently, this knowledge is explored to formally examine the quality of the guideline.

5.1 Formalisation of Background Knowledge

The background knowledge concerning the (patho)physiology of the glucose metabolism as summarised above is formalised using temporal logic, and kept as simple as possible. The specification is denoted by \mathcal{B}_{DM2} :

$$\begin{aligned}
& \text{G Drug}(\textit{insulin}) \rightarrow \text{G} (\textit{uptake}(\textit{liver}, \textit{glucose}) = \textit{up} \wedge \\
& \quad \textit{uptake}(\textit{peripheral-tissues}, \textit{glucose}) = \textit{up}) \\
& \text{G}(\textit{uptake}(\textit{liver}, \textit{glucose}) = \textit{up} \rightarrow \textit{release}(\textit{liver}, \textit{glucose}) = \textit{down}) \\
& (\text{G Drug}(\text{SU}) \wedge \neg \textit{capacity}(\textit{B-cells}, \textit{insulin}) = \textit{exhausted}) \rightarrow \\
& \quad \text{G} \textit{secretion}(\textit{B-cells}, \textit{insulin}) = \textit{up} \\
& \text{G Drug}(\text{BG}) \rightarrow \text{G} \textit{release}(\textit{liver}, \textit{glucose}) = \textit{down} \\
& (\text{G} \textit{secretion}(\textit{B-cell}, \textit{insulin}) = \textit{up} \wedge \\
& \quad \textit{capacity}(\textit{B-cells}, \textit{insulin}) = \textit{subnormal} \wedge \\
& \quad \text{QI} \leq 27 \wedge \text{H Condition}(\textit{hyperglycaemia})) \\
& \rightarrow \text{G Condition}(\textit{normoglycaemia}) \\
& (\text{G} \textit{release}(\textit{liver}, \textit{glucose}) = \textit{down} \wedge
\end{aligned}$$

$$\begin{aligned}
& \text{capacity}(B\text{-cells}, \text{insulin}) = \text{subnormal} \wedge \\
& \text{QI} > 27 \wedge \text{HCondition}(\text{hyperglycaemia}) \\
& \rightarrow \text{GCondition}(\text{normoglycaemia}) \\
& ((\text{Grelease}(\text{liver}, \text{glucose}) = \text{down} \vee \\
& \quad \text{Guptake}(\text{peripheral-tissues}, \text{glucose}) = \text{up}) \wedge \\
& \quad \text{capacity}(B\text{-cells}, \text{insulin}) = \text{nearly-exhausted} \wedge \\
& \quad \text{Gsecretion}(B\text{-cells}, \text{insulin}) = \text{up} \wedge \\
& \quad \text{HCondition}(\text{hyperglycaemia})) \\
& \rightarrow \text{GCondition}(\text{normoglycaemia}) \\
& (\text{Guptake}(\text{liver}, \text{glucose}) = \text{up} \wedge \\
& \quad \text{Guptake}(\text{peripheral-tissues}, \text{glucose}) = \text{up}) \wedge \\
& \quad \text{capacity}(B\text{-cells}, \text{insulin}) = \text{exhausted} \wedge \\
& \quad \text{HCondition}(\text{hyperglycaemia})) \\
& \rightarrow \text{G}(\text{Condition}(\text{normoglycaemia}) \vee \text{Condition}(\text{hypoglycaemia})) \\
& (\text{Condition}(\text{normoglycaemia}) \oplus \text{Condition}(\text{hypoglycaemia}) \\
& \quad \oplus \text{Condition}(\text{hyperglycaemia}))
\end{aligned}$$

where \oplus stands for the exclusive OR. Note that when the B-cells are exhausted, increased uptake of glucose by the tissues may not only result in normoglycaemia but also in hypoglycaemia (something not mentioned in the guideline).

5.2 Quality Check

Subsequently, the consequences of various treatment options were examined using the method introduced above. Hypothetical patients are considered, and treatment is selected according to the guideline fragments given in Figure 1:

- Consider a patient with hyperglycaemia due to nearly exhausted B-cells:

$$\begin{aligned}
& \mathcal{B}_{\text{DM2}} \cup \text{G}T \cup \{\text{capacity}(B\text{-cells}, \text{insulin}) = \text{nearly-exhausted}\} \cup \\
& \quad \{\text{HCondition}(\text{hyperglycaemia})\} \models \text{GCondition}(\text{normoglycaemia})
\end{aligned}$$

holds for $T = \{\text{Drug}(\text{SU}), \text{Drug}(\text{BG})\}$, which also satisfies the minimality condition $O_{\subset}(T)$.

- Prescription of treatment $T = \{\text{Drug}(\text{SU}), \text{Drug}(\text{BG}), \text{Drug}(\text{insulin})\}$ for a patient with exhausted B-cells, as is suggested by the guideline, yields:

$$\begin{aligned}
& \mathcal{B}_{\text{DM2}} \cup \text{G}T \cup \{\text{capacity}(B\text{-cells}, \text{insulin}) = \text{exhausted}\} \cup \\
& \quad \{\text{HCondition}(\text{hyperglycaemia})\} \models \\
& \quad \text{G}(\text{Condition}(\text{normoglycaemia}) \vee \text{Condition}(\text{hypoglycaemia}))
\end{aligned}$$

In the last case, it appears that it is possible that a patient develops hypoglycaemia due to treatment; if this possibility is excluded, then the minimality condition $O_{\subset}(T)$, and also $O_{\subset,c}(T)$, do not hold. In either case, good practice medicine is violated, which is to prescribe as few drugs as possible, taking into account costs and side-effects of drugs. Here, three drugs

are prescribed whereas only two should have been prescribed (BG and insulin, assuming that insulin alone is too costly), and the possible occurrence of hypoglycaemia should have been prevented. This shows that the method is indeed useful in the setting of guideline-quality checking, as both identified problems are medically relevant, and need to be redressed in future versions of the guideline.

6 Discussion

In this paper, we have developed a basis for a theory of quality-checking of medical guidelines, based on the idea that the quality of a guideline design is for the largest part based on its compliance with specific treatment aims and global requirements. To this purpose, use was made of the theory of abductive, diagnostic reasoning, i.e. we proposed to diagnose potential problems with a guideline using logical abduction [1, 8, 9, 12], even though in the example guideline the mentioned treatment was scrutinised. What was diagnosed were problems in the relationship between medical knowledge, suggested treatment actions in the guideline text and treatment effects; this is different from traditional abductive diagnosis, where observed findings are explained in terms of diagnostic hypotheses [9]. To our knowledge it is the first time that the structure of abductive diagnosis has been used in this particular fashion. We think that this approach to quality checking can be practically useful in the development of medical guidelines, as it helps designers to make an explicit distinction between object-knowledge and meta-knowledge, with various levels of detail, and an abductive reasoning system may then assist in finding flaws.

It also became apparent that guideline quality checks may require adding a significant amount of medical knowledge to the guideline knowledge already available. This knowledge may be specified in an abstract fashion, and thus compared to the detailed specification of a medical guideline required for its automated execution, the volume of knowledge to be added may be moderate. Nonetheless, much of this knowledge is not available within an actual guideline, and is therefore genuine background knowledge. Collecting this background knowledge in designing a guideline may be problematic despite the fact that guideline development teams normally include expert physicians in the field concerned: this background knowledge can only be elicited at the cost of some additional effort in the guideline development process.

There are also other ways in which the physiological background knowledge could have been represented. For example, using qualitative reasoning approaches, such as qualitative differential equations as proposed in the theory of qualitative simulation by Kuipers [5, 6], or using the qualitative process theory (QPT) by Forbus [2] could have been considered, as these approaches also allow one to abstract from the numerical detail. Our choice was dictated by the wish to remain close to the program semantics attached to formalised guidelines, which is based on temporal logic [10]. Combining formal verification with the approach to quality checking of guidelines developed in this paper is then possible.

So far, we have not made use of automated means to check the quality of a guideline design. As both object- and meta-knowledge can be represented in a language offered by standard theorem provers, partial automation of the process is possible in principle. This is something we wish to investigate in the future; in particular, an interactive theorem prover like KIV [13] may enable us to actually quality check medical guidelines. This would then complement the earlier work on object-level verification of medical guidelines using KIV [10].

References

- [1] Console L, Theseider Dupré D, and Torasso P. On the relationship between abduction and deduction, *Journal of Logic and Computation* 1991; 1(5): 661–690.
- [2] Forbus KD. Qualitative process theory. *Artificial Intelligence* 1984; 24: 85–168.
- [3] Fox J, Johns N, Lyons C, Rahmanzadeh A, Thomson R, and Wilson P. PROforma: a general technology for clinical decision support systems. *Computer Methods and Programs in Biomedicine* 1997; 54: 59–67.
- [4] Fox J, and Das S. *Safe and Sound: Artificial Intelligence in Hazardous Applications*. Cambridge: MIT Press (jointly published with the American Association of Artificial Intelligence), 2000.
- [5] Kuipers BJ. Qualitative simulation. *Artificial Intelligence* 1986; 29: 289–388.
- [6] Kuipers BJ. *Qualitative Reasoning*. The MIT Press, Cambridge, Massachusetts, 1994.
- [7] Lucas PJF. Logic engineering in medicine. *The Knowledge Engineering Review* 1995; 10(2): 153–179.
- [8] Lucas PJF. Symbolic diagnosis and its formalisation. *The Knowledge Engineering Review* 1997; 12(2): 109–146.
- [9] Lucas PJF. Analysis of notions of diagnosis. *Artificial Intelligence* 1998; 105(1–2): 293–341.
- [10] Marcos M, Balsler M, Ten Teije A, and Van Harmelen F. From informal knowledge to formal logic: a realistic case study in medical protocols. *Proceedings of the 12th EKAW-2002*, 2002.
- [11] Ohno-Machado L, Gennari J, Murphy S, et al. Guideline Interchange Format: a model for representing guidelines. *Journal of the American Medical Informatics Association* 1998; 5(4): 357–372.
- [12] Poole D. A methodology for using a default and abductive reasoning system. *International Journal of Intelligent Systems* 1990, 5(5), 521–548.
- [13] Reif W. The KIV Approach to Software Verification. In: M. Broy and S. Jähnichen (eds.), *KORSO: Methods, Languages, and Tools for the Construction of Correct Software*. Lecture Notes in Computer Science, Volume 1009, Springer-Verlag, Berlin, 1995.
- [14] Sackett DL, Richardson WS, Rosenberg W, and Brian R. *Evidence Based Medicine—How to Practice and Teach EBM*. Haynes Churchill Livingstone, 1996.
- [15] Shahar Y, Miksch S, and Johnson P. The Asgaard project: a task-specific framework for the application and critiquing of time-oriented clinical guidelines. *Artificial Intelligence in Medicine* 1998; 14: 29–51.
- [16] Turner R. *Logics for Artificial Intelligence*. Ellis Horwood, Chichester, 1985.
- [17] Woolf SH. Evidence-based medicine and practice guidelines: an overview. *Cancer Control* 2000; 7(4): 362–367.