Formalising medical quality indicators to improve guidelines

Marjolein van Gendt[1], Annette ten Teije[1], Radu Serban[1], Frank van Harmelen[1]

Vrije Universiteit Amsterdam, Dept. of Artificial Intelligence
De Boelelaan 1081a, 1081HV Amsterdam, Netherlands
{mtvgendt,annette,rserban,frank.van.harmelen}@cs.vu.nl

Abstract. Medical guidelines can significantly improve quality of medical care and reduce costs. But how do we get sound and well-structured guidelines? This paper investigates the use of quality indicators that are formulated by medical institutions to evaluate medical care. The main research questions are (i) whether it is possible to *formalise* those indicators in a specific knowledge representation language for medical guidelines, and (ii) whether it is possible to *verify* whether such guidelines do indeed satisfy these indicators. In a case study on two real-life guidelines (Diabetes and Jaundice) we have studied 35 indicators, that were developped independently from these guidelines. Of these 25 (71%!) suggested anomalies in one of the guidelines in our case study.

1 Introduction

Medical guideline are accepted as an instrument for contributing to a higher quality of care. It is evident that high quality of guidelines is important. [7, 2] present formalisation as a technique for guideline quality improvement. Such formalisation points out anomalies. Analysing these anomalies can result in improvement of a guideline. Given a formalisation, verifying the guidelines against particular properties could improve the guideline even further. The question is then *which* properties are useful to verify.

This paper evaluates the use of medical quality indicators as properties to verify. Medical guidelines prescribe the actions medical practitioners should undertake. The medical quality indicators are designed to judge the execution or performance of the care. These quality indicators are systematically engineered by medical experts and therefore give us new insights into what properties the medical care and thus the medical guidelines should satisfy. The motivation for the research question is that the guideline prescribes the care beforehand, whereas the indicators judge the care afterwards, so one would expect these two to correspond (also suggested in e.g. [3]).

Earlier work. [7] already used an indicator to formally verify a medical guideline, but the formalisation of the indicator was done in an ad hoc fashion. In this

This work has been partially supported by the European Commission's IST program, under contract number IST-FP6-508794 Protocure-II.

paper, we study the systematic formalisation of such indicators, and their use in guideline improvement. In this research we build that bridge by interpreting indicators as properties or requirements of the medical guidelines.

Our main research question therefore is: *Can medical guidelines be improved using medical quality indicators?* This leads to the two sub-questions: (i)whether the medical quality indicators can be *formalised*, and (ii) whether they can be used for *verification*.

Structure of this paper. Section 2 describes the general approach of using medical indicators for quality improvement of guidelines. Section 3 describes the results of modelling the indicators and thus answers the *formalisation* question. Section 4 studies the *verification* question. Section 5 shows the overall conclusions from the preceding two sections. This section also shows related and future work.

2 Approach

This research consists of a number of steps. At first the indicators from the relevant medical areas are selected. These indicators must be formalised in the same language as the guidelines, in order to be able to verify their satisfaction by the guideline. During this formalisation anomalies can surface, for instance that an indicator describes a medical action which is not present in the guideline. In such a case the indicator cannot be formalised. The indicators that can be formalised are used for a manual verification against the appropriate guideline. During this verification it might turn out the guideline does not comply with the indicator. Both these types of anomalies point at possible improvements of the guidelines. This section explains these steps in more detail.

Choice of representation language. We have chosen Asbru as modelling language [14]. It is a modelling language constructed specifically for guidelines. Asbru is a task-specific, intention-based and time-oriented language for representing skeletal plans. Its main characteristics are the hierarchy of plans, the possibility to define timing aspects and to define dependencies between plans in a rich control structure. Another important characteristic of Asbru are the constructs for modelling intentions of a plan. In this research these intention constructs are used for the formalisation of the indicators. When a guideline and indicator are modelled in the same language a check on similarities or discrepancies can be performed.

Choice of guidelines. We have used two medical guidelines that were already formalised by the Protocure project[1]. These guidelines concerned Diabetes [13] and Jaundice [1] and this prior work determined the case studies chosen for this research. The original Jaundice guideline is 10 pages of text and is modelled in an Asbru model containing 40 plans (18 pages of Asbru, [9]) while the original Diabetes guideline is 4 pages of text and is modelled in an Asbru model containing 68 plans (58 pages of Asbru, [10]).

Choice of indicators. For Jaundice the MAJIC indicators are used [6] and for Diabetes the CBO quality indicators [15]. A strong point of our case study is

[1] www.protocure.org

that these indicators are developed independently from the respective guidelines, by different organizations. This avoids the possibility that the guidelines will trivially satisfy the indicators because are both based on the knowledge and experience from the same experts. The results of this research are not limited to only the MAJIC and the CBO diabetes indicators, because a small search revealed many other indicators for both diabetes [12, 4, 8] and for other diseases [11] which were similar to the indicators studied here.

Translating indicators as goals. Indicators usually are measurements of the number of people for which the care has been performed as it should have been. In order to regard indicators as goals or intentions of a care-performance, we need to rephrase the indicators. For instance if an indicator states "*Percentage of people with diabetes suffering retinopathy*", the corresponding goal becomes "*minimize the number of people with diabetes with retinopathy*". In other words the indicators must first be translated into goals to be achieved during guideline execution to be useful for verification purposes.

Modelling the indicators. For the actual modelling of the indicators in Asbru a step-wise translation is used. The natural language indicators are first divided into parts that map onto concepts used in Asbru (such as the time annotation). The second step involves the actual formal translation. The results from the formalisation are described in section 3.

Verifying the indicators. After the modelling, the formalised indicators can be used for verification. The verification has to be done manually, because no automated techniques are available yet. This manual verification is quite labour intensive, but produces good results, as can be seen in section 4. Manual verification consists of a walk through the plans in a guideline during which conditions are checked as to whether a required action will be performed under the required circumstances.

3 Modelling

In this section, we discuss the modelling of the indicators and the problems that we encountered. We illustrate them by a concrete example. Furthermore we give an overview of the anomalies that we found during this phase.

Categories of indicators. During modelling, we divided the indicators in a number of categories, due to their differing characteristics.

1. Result vs. process indicators
2. Maximizing vs. minimizing the number of people the indicator applies to
3. Time-related vs. time-unrelated indicators
4. Quantitative vs. qualitative indicators

During the modelling it turned out that the first dimension is the most important. Result indicators refer to *situations or states* that should be true or avoided, whereas process indicators refer to *actions* that should be undertaken or avoided.

Example result indicator (Diabetes)	
original	Percentage of people with diabetes with a diastolic blood pressure smaller than or equal to 90 mmHG
intermediate	maximise (people with diabetes with (a diastolic blood pressure smaller than or equal to 90 mmHG))
formal	Intermediate-state (Achieve (context glucose-evaluation = DMT2) lower-blood-pressure $\leq$ 90)

Fig. 1. An example indicator modelled in Asbru

How to model indicators in Asbru. Intentions would seem to be the most appropriate Asbru-construction to model indicators. That is because indicators are seen as the goals that must be obtained during or after guideline execution. Plans are the central and essential part of Asbru. Their structure captures the sequence of and the relationship between all the actions. An Asbru intention defines the rationale of a plan i.e. it indicates what purpose a plan has. Intentions are attached to a specific plan and consist of three components. The first component is a Verb. This can be Achieve, Maintain or Avoid. Second there is the indication whether the statement should hold at some time during (Intermediate) or at the successful completion (Overall) of the plan's life cycle. The third building block indicates whether the intention concerns a State (parameter evaluation) or an Action (execution of a plan). In addition to these building blocks an intention consists of a temporal pattern and optionally a time annotation. The time annotation specifies the time period when the parameter proposition used in the intention should hold or be tested. The temporal pattern is the core of every intention, because it describes the situation or action the intention aims for. Within the temporal pattern it is also possible to define a context for an intention. The intention will only be evaluated if the parameter values match the values specified in the context description.

It is interesting to note that the ontology for medical goals as proposed in [5] is very close to the notion of intentions from Asbru. The proposal in [5] consists of the following five components: context, intention verb, target function, temporal constraint, and priority of the goal. This proposal maps rather well to the components of Asbru's intentions, suggesting some concensus among researchers on the elements required for goal modelling.

Example of an indicator in Asbru. Figure 1 shows an example of an indicator that is modelled in Asbru. The example should apply to people with diabetes, so the context for the formalised indicator is a glucose-evaluation of DMT2 (Diabetes Mellitus Type 2). The expressions occuring in the indicator are restricted to those concepts that already occur in the guideline. In this example, both the glucose-evaluation parameter and the diastolic blood pressure already were present in the formalised guideline, so could be used to express the indicator.

The type of the indicator in terms of the four categories mentioned at the beginning of this section, has consequences for the formalisation. The result

indicators usually map onto state-intentions, whereas the process indicators map onto action-intentions. If the indicator must be maximized, the verb used is either Maintain or Achieve. For minimization the Avoid-verb is appropriate. If an indicator is time-related a time-annotation is needed and otherwise this can be ignored. The fourth category has no effects on formalisation.

Different types of anomalies. The potential anomalies in the guideline or in the indicator, can be divided in the following groups:

1. Type mismatch: Parameters in the guideline have a value that is of different type or has a range mismatch to the value mentioned in the indicator. For instance, one CBO indicator mentioned "*Percentage of people with diabetes and microalbumin > 30 mg/hr or blood pressure above 150/85 mmHg who get anti-hypertensive medication*". However, the microalbumin parameter in the guideline ('microalbuminuria') did not match the 'microalbumin' from the indicator, since the one in the guideline is a Boolean value, whereas the one in the indicator is measured in mg/hr.
2. Missing Parameter: The parameter referred to in the indicator is not used at all in the guideline. An example is the following indicator: "*Percentage of people with diabetes who have had a laboratory test for HbA1c during the last 12 months*", while the parameter HbA1c does not occur in the diabetes guideline.
3. Missing Action: The action required by the indicator is not covered by the guideline. An example is the following indicator: "*Percentage of people with diabetes with just diagnosed or worsening proliferating retinopathy who have undergone vitrectomy or laser coagulation during the last 3 months*". The indicator presumes much more detail regarding this than the guideline contains, since the actions vitrectomy and laser coagulation are not mentioned in the guideline.
4. Missing medical knowledge: An example from Diabetes is the following indicator: "*Percentage of people with diabetes and angina pectoris who get anti-angina medication*". The phrase "anti-angina medication" does not occur in the guideline. If it is clear what this medication exactly is, it might be possible to still model the indicator. In that case one could search for other parameters indicating the same substances. This problem can be overcome by consulting a medical expert. In this case there turned out to be no parameters matching 'anti-angina medication' present in the guideline.

Empirical results. Figure 2 summarizes our modelling results[2]. For Diabetes we investigated 21 indicators, of which 12 process indicators and 9 result indicators. We found 10 anomalies in total, which are mainly incompleteness anomalies. The remaining 11 indicators were successfully modelled in Asbru's intentions, for which we used 4 different intention patterns (notice that in principal 12 different intention patterns are possible in Asbru). For Jaundice, we modelled 14

[2] In the table the numbers sometimes do not add up for two reasons. The failure to formalise an indicator can have multiple causes and in one case an indicator is formalised in two parts ('achieve intermediate action' as well as 'achieve intermediate state').

Diabetes					
# indicators	21	**# anomalies**	10	**# modelled**	11
process	12	Incomplete	8	avoid intermediate state	2
result	9	- missing parameter	5	achieve intermediate state	5
		- missing action	3	maintain intermediate state	3
		Type mismatch	3	achieve intermediate action	2
		Missing medical knowledge	1		

Fig. 2. Indicators and anomalies obtained from the modelling phase

indicators (all of which were process indicators), which resulted in 3 anomalies, and 11 succesfully modelled indicators. Again only a few (namely 4) intention patterns were needed for modelling 11 process indicators.

In total for the two guidelines, we studied 35 indicators, of which no fewer than 13 (37%) gave rise to potential anomalies.

Especially the indicators that resulted in an "incompleteness" anomaly are of use for guideline improvement. These modelling failures show differences between the guideline and the indicator content, suggesting that either of these would have to be changed.

The above shows that the formalisation of indicators by itself is already a useful contribution to guideline improvement. In the next section we will check whether the modelled intentions do actually hold for the given Asbru models.

4 Verification

The formalisation of the indicators allows us in principle to formally verify whether a guideline complies with the indicator. Such a verification is preferably done automatically. In our case study we perform this verification process manually, for lack of suitable software tools. We have performed this verification for both the Diabetes and the Jaundice guidelines and their corresponding indicators. Our verification is limited to process indicators, and excludes result indicators, because we do not have access to patient data (ie. the outcomes of actions).

Two types of indicators for verification. We distinguish two types of indicators which require different approaches to the verification process: achievement indicators that intend to maintain or achieve an action, and avoidance indicators that intend to avoid an action. The only two conditions that can make a plan stop are the abort and complete conditions. Thus, when an action must be avoided, the abort and complete conditions of the plans must be checked, because if those conditions are met the plan stops and thus the action-leaf will not be reached. If an indicator-action must be achieved or maintained, all conditions and the control structure must match the situation sketched in the indicator. The verification of this type of indicator requires much more work than the avoidance indicators.

Achievement indicator (Diabetes)	
original	Percentage of people with diabetes with a known albumin value measured during the last 12 months
intermediate	maximize [people with diabetes] with [a known albumin value] measured [during the last 12 months]
formal	Intermediate-state Achieve (or ((context glucose-evaluation = DMT2) albumin-in-urine = known [[0,_][_,1Y][_,_],*now*]) ((context glucose-evaluation = DMT2) albumin-creatinin-ratio-in-urine = known [[0,_][_,1Y][_,_],*now*]))

Fig. 3. Original text, intermediate and formal version of an achievement indicator for the Diabetes guideline

4.1 Example verification

We give an example of the verification of an achievement indicator. Figure 4 shows the relevant part of the Asbru plan hierarchy for the Diabetes guideline; figure 3 shows the original text, the intermediate version and the formal version of an achievement indicator for the Diabetes guideline.

We start with the main plan Diabetes-Mellitus-Type-2. This plan contains two sub-plans, Diagnostics and Policy. Due to the continuation-specification[3] only one of the two sub-plans must necessarily be performed. The wait-for-optional-subplans condition specifies that although only one of the sub-plans must be performed, both of them will be tried and executed if the conditions are met.

Of these two, we first investigate the Diagnostics plan. Again, its wait-for-optional-subplans value dictates that all of its sub-plans must be tried, one of which is the Risk-inventory plan[4]. The filter-precondition of Risk-inventory says: glucose-evaluation = known and glucose-evaluation = DMT2, implying that the plan only applies to people with diabetes. This is indeed also the case for the indicator (figure 3), yielding our first (positive) verification result.
Continuing with the recursive descent down the plan-hierarchy, we arrive at (among others) the Albumin-test subplan of Risk-inventory. The Albumin-test plan has two optional sub-plans: Albumin-test-manual and Albumin-creatinin-ratio-in-urine. These plans aim to obtain just the two values needed in the formal version of the indicator from figure 3: albumin-in-urine and albumin-creatin-ratio-in-urine. Because the values occur in an OR-statement, at least one of these two subplan must be executed. However, neither of these subplans is decorated with a time-annotation that guarantees the requirement stated in the indicator that these two values must date from within the last year. As a result, this aspect of the indicator is not fulfilled by this branch of the diabetes guideline.

Another option for the fulfilment of this indicator is via the Annual-control-plan, reachable from the top-plan via a recursive descent not discussed here (see fig. 4). This plan does indeed guarantee the required 1-year period from the

[3] The detailed conditions of these plans are not shown in figure 4 for space reasons.
[4] In the full verification the other subplans of Diagnostics are also traced, but we skip these here for space reasons

Indicators	Diabetes	Jaundice	percentage
Total Attempted	11	11	100%
Succesfully Verified	4	0	18%
Not Succesfully Verified	3	9	55%
- *provably untrue*	*3*	*6*	
- *additional assumptions needed*	*0*	*3*	
Unverifiable	4	2	27%
- *need patient data*	*4*	*0*	
- *too complicated for manual effort*	*0*	*2*	

Fig. 5. Total verification results for both guidelines

indicator, because it is repeated every 46-50 weeks[5]. Unfortunately, the body of Annual-control-plan only calls the required Albumin-test-plan under the condition if age < 50. Consequently, this path of the overall plan also does not verify the indicator in all cases.

Concluding, there is no single path through the Diabetes-Mellitus-Type-2-plan which completely fulfills all the conditions imposed by the indicator from figure 3. In summary, one branch of the plans misses the time-annotations needed to ensure the required timeless of some parameter-readings, while another branch, while ensuring the required timeless of the parameter-readings, only applies to a subset of all patients. This discrepancy can be caused by a mistake either in the guideline or in the indicator. When consulted, a medical expert from CBO, the institution that produced the diabetes indicators, clarified that in this case the indicator was not precise enough, and should have been refined to limit its applicability to patients under 50 (exactly as stated in the guideline).

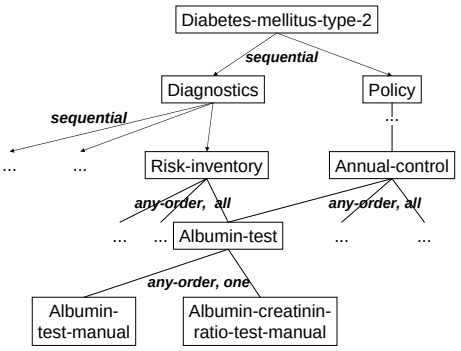

Fig. 4. Part of Asbru plan for Diabetes guideline

4.2 Overall verification results

Figure 5 summarizes our verification results for both guidelines. In total we verified 11 indicators for each guideline, namely exactly the indicators that were indicated as succesfully modelled in figure 2. Only a surprisingly small number of the indicators could be successfully verified (a mere 18%). Of the 12 indicators (3+9, 55%) that could not be successfully verified, 3 could be verified, but only under additional assumptions which should be discussed with a medical

[5] Again, body of the plan not shown

expert. This leaves a signficant number of 9 indicators (6+3) that were provably incompatible with the guideline. For a remaining category of 6 indicators (27%) we were not able to perform a verification, because either they would require patient data (the result indicators discussed in section 3), or because the verification proof was too complicated to complete by hand.

Especially the category of "not succesfully verified" indicators (12 indicators, 55%) can be of use for improving the guideline, because they point at differences of opinion between the creators of the guidelines and of the indicators.

5 Conclusion and future work

Medical quality indicators can be used to improve medical guidelines and thus medical care. The guideline and the indicator must be formalised to do this. In this case study we have used Asbru as a modelling language to formalise guidelines and indicators in the Diabetes and Jaundice domains. Our main findings are as follows:

• A translation of the indicators into a formal guideline modelling language is possible. This translation already reveals a number of potential anomalies, which can be divided into 4 different groups (type mismatch, missing parameter, missing action and missing medical knowledge).

• The formal modelling of the indicators for Diabetes and Jaundice resulted in a significant number of potential anomalies. For the Diabetes guidelines almost 50% of the indicators suggested some kind of anomaly in the guideline. For Jaundice, this percentage was much lower (20%), but still significant.

• The formalisation of indicators enabled the verification of compliance between guidelines and indicators. Only a disappointing 18% of the indicators could be succesfully verified. For a suprisingly high number of indicators (55%) the verification revealed non-compliance of the guideline with the indicator. For the remaining 27% the result remains undecided for a variety of practical reasons.

• Overall, we have started with 35 independently developped indicators for two guidelines. Of these, 13 suggested a possible anomaly during the modelling phase. Of the remaining 22, a further 12 suggested anomalies during the verification phase. In other words 25 out of 35 indicators, a staggering 71%, suggested anomalies in one of the guidelines in our case study.

Future work.

• An obvious next step in this research is the validation of our results with medical experts, to evaluate which changes should be incorporated into the guidelines.

• Another issue is the enhancement of the formalisation process. The manual verification should be replaced by automatic verification and the process must become more formal, for example by using an interactive theorem prover [7].

• Verification of the result indicators should be done on the basis of patient data, which was unavailable at the time of our case study.

• A possible resolution of the mismatch-anomalies could be to use an ontology for semi-automatically bridging the mismatches between guideline and indicator.

References

1. American Academy of Pediatrics, Provisional Committee for Quality Improvement and Subcommittee on Hyperbilirubinemia. Practice parameter: management of hyperbilirubinemia in the healthy term newborn. *Pediatrics*, 94:558–565, 1994.
2. M. Balser, O. Coltell, J. van Croonenborg, C. Duelli, F. van Harmelen, A. Jovell, P. Lucas, M. Marcos, S. Miksch, W. Reif, K. Rosenbrand, A. Seyfang, and A. ten Teije. Protocure: Supporting the development of medical protocols through formal methods. *SCPG-04*, number 101 in Studies in health technology and informatics, pages 103–108, IOS press, 2004.
3. A. Casparie, *et al. Ontwikkeling van indicatoren op basis van evidence-based richtlijnen*. Van Zuiden Communications, Alphen aan den Rijn, 2002.
4. B. Fleming, S. Greenfield, M. Engelgau, L. Pogach, S. Clauser, M. Parrott. Diabetes care. *The Diabetes Quality Improvement Project*, 24:1815–1820, 2001.
5. J. Fox, A. Alabassi, E. Black, C. Hurt, and T. Rose. Modelling clinical goals: a corpus of examples and a tentatove ontology. In *SCGP-04*, pages 31–45. IOS Press, 2004.
6. MAJIC. MAJIC Steering Committee Meets. *MAJIC Newsletter*, 1(2), 1998.
7. M. Marcos, M. Balser, A. ten Teije, F. van Harmelen, and C. Duelli. Experiences in the formalisation and verification of medical protocols. In *AIME-2003*, pgs 132–141. Springer, 2003.
8. L. Nyberg and M. Lawrence. Overall quality indicators in health care and medical services, 2001.
9. Protocure I Project. Asbru protocol for the management of hyperbilirubinemia in the healthy term newborn, August 2002. Technical Report, URL: http://www.protocure.org.
10. Protocure I Project. Asbru protocol for the management of diabetes mellitus type 2, August 2002. Technical Report, URL: http://www.protocure.org.
11. T. C. of State, t. A. o. S. Territorial Epidemiologists (CSTE), T. C. D. P. D. (ASTCDPD), the National Center for Chronic Disease Prevention, H. P. C. for Disease Control, and Prevention. Indicators for chronic disease surveillance, 2001. http://cdi.hmc.psu.edu/pathways/ breastcancer.html.
12. Rockville. AHRQ quality indicators - guide to prevention quality indicators: Hospital admission for ambulatory care sensitive conditions. *Agency for Healthcare Research and Quality*, AHRQ Pub, no. 02-R0203, 2004. revision 3.
13. G. Rutten, S. Verhoeven, R. Heine, W. de Grauw, P. Cromme, K. Reenders, E. van Ballegooie, and T. Wiersma. NHG-Standaard Diabetes Mellitus Type 2 (eerste herziening). *Huisarts en Wetenschap*, 42(2):67–84, 1999. First revision.
14. Y. Shahar, S. Miksch, and P. Johnson. The Asgaard project: a task-specific framework for the application and critiquing of time-oriented clinical guidelines. *AIM*, 14:29–51, 1998.
15. G. Storms, P. ten Have, and R. Dijkstra. *Indicatoren voor verbetering van de diabeteszorg*. Van Zuiden Communications, Alphen aan den Rijn, 2002.