Using Critiquing for Improving Medical Protocols: Harder than It Seems

Mar Marcos[1], Geert Berger[1], Frank van Harmelen[1], Annette ten Teije[2], Hugo Roomans[1], and Silvia Miksch[3]

[1] Vrije Universiteit Amsterdam, Dept. of Artificial Intelligence
De Boelelaan 1081a, 1081HV Amsterdam, Netherlands
[2] Universiteit Utrecht, Institute of Information and Computing Sciences
P.O. Box 80.089, 3508TB Utrecht, Netherlands
[3] Vienna University of Technology, Institute of Software Technology
Favoritestraße 9-11/188, A-1040 Vienna, Austria

Abstract. Medical protocols are widely recognised to provide clinicians with high-quality and up-to-date recommendations. A critical condition for this is of course that the protocols themselves are of high quality. In this paper we investigate the use of critiquing for improving the quality of medical protocols. We constructed a detailed formal model of the jaundice protocol of the American Association of Pediatrics in the Asbru representation language. We recorded the actions performed by a pediatrician while solving a set of test cases. We then compared these expert actions with the steps recommended by the formalised protocol, and analysed the differences that we observed. Even our relatively small test set of 7 cases revealed many mismatches between the actions performed by the expert and the protocol recommendations, which suggest improvements of the protocol. A major problem in our case study was to establish a mapping between the actions performed by the expert and the steps suggested by the protocol. We discuss the reasons for this difficulty, and assess its consequences for the automation of the critiquing process.

1 Introduction and Motivation

During the last years a high number of medical practice guidelines or protocols have been produced from systematic evidence-based reviews [9]. Protocols developed in this way provide clinicians with valid and up-to-date recommendations, and thus have a beneficial influence on quality and costs of medical care [2]. A critical condition for this is that the protocols are clearly and correctly defined. However, medical protocols are often ambiguous or incomplete [5]. If ambiguity or incompleteness (intended or not), or even inconsistency, occur in protocols the risks of being neglected in the delivery of care will be high. We are concerned with methods to uncover such apparent anomalies.

The critiquing approach has been adopted in many systems since Miller's first work [4]. Their common feature is that a problem specification and a solution proposed by the user are given as input, and a series of comments aimed at improving the solution is generated as output. Critiquing has proved to be a very suitable approach for decision support in domains like Medicine, where there is a high variability on what can be considered an acceptable solution. It usually implies identifying points where the proposed

S. Quaglini, P. Barahona, and S. Andreassen (Eds.): AIME 2001, LNAI 2101, pp. 431–441, 2001.

solution is suboptimal and studying what are the explanations, if there are any, so as to produce an appropriate critique of unjustified choices. This is the usual view of critiquing, in which standards are used to give feedback on the user's behaviour. However, another possibility is to reverse these two roles, and to use the user's solution and its justification to refine the standards.

This paper presents the work we have done to explore this possibility in the framework of protocol quality improvement. It has consisted in: (1) modelling a protocol using a specific protocol language, (2) manually critiquing the solutions given by an expert to a set of predefined cases by comparing them against the prescriptions of the protocol, and (3) drawing conclusions about the utility of critiquing for protocol improvement.

As working example we have selected the jaundice protocol of the American Association of Pediatrics (AAP). We will briefly give its background in section 2. The protocol language that we have chosen, Asbru, is briefly explained in section 3. The rest of the paper presents the Asbru model of the jaundice protocol, the critiquing exercise and its results, and our conclusions on the utility of critiquing for the improvement of medical protocols.

2 The Jaundice Protocol

The illness. Jaundice (or hyperbilirubinemia) is a common disease in new-born babies. Under certain circumstances, elevated bilirubin levels may have detrimental neurological effects. In many cases jaundice disappears without treatment but sometimes phototherapy is needed to reduce the levels of total serum bilirubin (TSB), which indicates the presence and severity of jaundice. In a few cases, however, jaundice is a sign of a severe disease.

The protocol. The jaundice protocol of the AAP [1] is intended for the management of the disease in healthy term[1] new-born babies. The reasons for choosing this protocol were that, firstly, it is considered a high-quality protocol[2] and that, secondly, data corresponding to a set of test cases were available[3].

The protocol consists of an evaluation (or diagnosis) part and a treatment part, to be performed in sequence. During the application of the protocol, as soon as the possibility of a more serious disease is uncovered, the recommendation is to abort without any further action. The rationale behind this is that the protocol is exclusively intended for the management of jaundice in healthy new-borns. An important part of the protocol is the table used to determine the adequate treatment from the TSB value and the age of the infant.

The test cases. The data from the test cases were limited to a set of initial facts and lab test results, and they did not include a concrete solution to the case they illustrate. Therefore, we had to resort to a pediatrics expert to provide us with the necessary input for our critiquing exercise.

[1] Defined as 37 completed weeks of gestation.

[2] The jaundice protocol of the AAP is included in the repository of the National Guideline Clearinghouse (see `http://www.guideline.gov/`).

[3] In `http://www.ukl.uni-heidelberg.de/mi/research/dss/sisyphus/`

3 The Asbru Language

Asbru is a semi-formal language intended to support the tasks necessary for protocol-based care [8]. It is a time-oriented, intention-based, skeletal plan specification language. Asbru is an interesting language in a case-study on critiquing because it is much more detailed and formalised than other protocol representation languages, such as PROforma [3] or GLIF [6]. This makes Asbru in principle an attractive candidate for protocol improvement because the additional details in an Asbru representation will help to expose potential problems in the original protocol.

In Asbru, protocols are expressed as plan schemata defined at various levels of detail, precise enough to capture the essence of clinical procedures but leaving space for flexibility during their application. Major features of Asbru are:

- explicit intentions and preferences can be stated for plans.
- intentions, conditions, and states are temporal patterns.
- uncertainty in temporal patterns and parameters can be flexibly expressed.
- plans can be executed in different compositions, e.g. in parallel (TOGETHER), in sequence (SEQUENCE), in any order (ANYORDER), or every certain time (CYCLICAL); it can also be defined whether all the steps should be executed (ALL) or not (SOME).
- conditions can be defined to control plan execution, e.g. to set applicability conditions (FILTER) or to determine when execution should be interrupted (ABORT).

Some of the above elements need some explanation for the purposes of this paper, namely intentions and temporal patterns. Intentions are high-level goals that support special tasks such as critiquing and modification. They are patterns of states or actions to be achieved, maintained, or avoided (ACHIEVE, MAINTAIN or AVOID), during or after the execution of the plan (INTERMEDIATE or OVERALL).

Temporal patterns are crucial in Asbru. The time annotations used in Asbru allow the representation of uncertainty in starting time, ending time, and duration with intervals, as well as the use of multiple reference points. A time annotation is written in the form ([EarliestStarting,LatestStarting] [EarliestFinishing,LatestFinishing] [MinDuration,MaxDuration] REFERENCE). Thus, the temporal pattern (TSB-decrease=yes any [4 hours,_] [_,6 hours] [_,_] *self*) is a proposition becoming true if there is a decrease of TSB, in any context, between 4 and 6 hours after the activation of the current plan, *self*.

4 The Jaundice Protocol in Asbru

We have used Asbru to model the AAP jaundice protocol. In the model we have tried to stay as close as possible to the algorithm in the AAP protocol. Thus, for instance, the exit messages in the flowcharts have been directly translated into Asbru Displays. Likewise the AAP protocol, the Asbru version has as main components a diagnosis plan and a treatment plan. It is made up of more than 30 subplans and has a length of 16 pages. Table 1 lists the most important subplans. Finally, the protocol makes extensive use of

Table 1. Main subplans in the Asbru version of the AAP jaundice protocol. The first subplans correspond to continuous monitoring tasks to be executed in parallel. The main components are the subplans for the diagnosis and the treatment. The grouping and numbering of the plans gives an idea of their hierarchical structure. Note that several simple plans have been omitted for simplicity.

Hierarchical number	**Plan name**
1	Hyperbilirubinemia
1.1	Check-for-rapid-TSB-increase
1.2	Check-for-jaundice>2-weeks
1.3	Check-for-jaundice>3-weeks
DIAGNOSTIC PLANS	
1.4	Diagnostics-hyperbilirubinemia
1.4.1	Anamnesis-abnormal-signs
1.4.2	Blood-tests
1.4.2.1	Check-blood-test-mother
1.4.2.2	Perform-blood-test-child
1.4.3	Anamnesis-hemolytic-disease
1.4.3.1	Evaluation-risk-factors-hemolytic-disease
1.4.4	Jaundice-determination
TREATMENT PLANS	
1.5	Treatment-hyperbilirubinemia
1.5.1	Regular-treatments
1.5.1.1	Feeding-alternatives
1.5.1.2	Phototherapy-intensive
1.5.1.3	Phototherapy-normal-prescription
1.5.1.4	Phototherapy-normal-recommendation
1.5.1.5	Observation
1.5.2	Exchange-transfusion

many Asbru features, in particular for the combination and interleaving of diagnosis and treatment [7].

The Regular-treatments plan in table 2 illustrates the complexity of some of the protocol plans. This plan, which is fundamental in the treatment part, tries to reduce the bilirubin levels without resorting to the application of an exchange transfusion (see table 2 for details on the procedure).

5 The Critiquing Exercise

Our critiquing exercise is based on both the AAP protocol and the Asbru version we have modeled. First, we have presented a subset of the test cases to the doctor, who was trained to solve jaundice cases in a manner similar to the AAP protocol. Second, we have used the Asbru protocol to produce a recommendation for the same cases. Finally, we have tried to establish a mapping between the expert's solution and the Asbru recommendation, and to find and explain the differences. Note that, unlike traditional critiquing work, these may turn out to be problems in either the expert's solution or the Asbru protocol. Figure 1 illustrates the approach.

Table 2. Regular-treatments plan. This plan consists of two parallel parts: the study of feeding alternatives and the different therapies. The therapies can be tried in any order, one at a time. The intentions are both avoiding toxic bilirubin levels during the execution of the plan and attaining normal (observation) ones at the end. The plan completes when the feeding alternatives and the therapies complete. The latter in turn depends on the completion of observation. It aborts when either bilirubin raises to transfusion levels or intensive phototherapy fails to reduce them sufficiently, pointing to a pathologic reason.

PLAN 1.5.1	**Regular-treatments**	
INTENTIONS	AVOID INTERMEDIATE STATE: bilirubin=transfusion ACHIEVE OVERALL STATE: bilirubin=observation	
CONDITIONS	**FILTER:**	(bilirubin≠transfusion any [_,_] [0,_] [_,_] *now*)
	ABORT:	(bilirubin=transfusion any [_,_] [0,_] [_,_] *now*) or (pathologic-reason=yes any [_,_] [0,_] [_,_] *now*)

PLAN BODY	DO-ALL-TOGETHER {
	Feeding-alternatives
	DO-SOME-ANYORDER {
	Retry aborted children
	Continuation specification: **(Observation)**
	Phototherapy-intensive
	Phototherapy-normal-prescription
	Phototherapy-normal-recommendation
	Observation } }	

Acquisition of expert solutions to test cases. The critiquing exercise uses the solutions given by our expert to a series of test cases. With the aim of obtaining a varied set of cases, we have carried out a *selection of data*, i.e. we have occasionally suppressed some initial lab data so as to force the expert to perform the actions requesting for them.

Then, in order to determine whether our expert's practice was coherent with the recommendations in the protocol, and hence that the critiquing made sense, we presented the protocol to her some time before the acquisition of solutions—*priming of the expert*. This was done long beforehand so that the protocol did not influence the behaviour of the expert at the time of solving the test cases.

Finally, during the acquisition we have considered both the actions of the expert and the intentions behind them—*acquisition of actions and intentions*. However, except for few points in the solutions, we have not strictly regarded the sequencing of her actions.

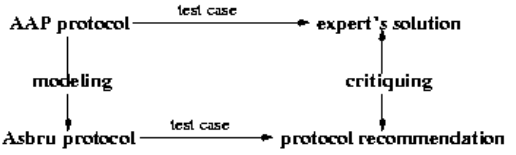

Fig. 1. Critiquing of expert solutions using the recommendations of the Asbru protocol.

Critiquing of expert solutions using the protocol. We have compared the expert's solution to a series of cases with the recommendations of the Asbru protocol to solve them. Henceforth, we will use the term action when referring to expert's actions and the terms plan or statement when making reference to elements of the Asbru protocol. First, we have studied which of her actions could be matched with (sub)plans in the protocol, and which could not. Afterwards we have studied the mismatches, i.e. the protocol recommendations which were not considered in the solution as well as the expert's actions that were not included in the protocol indications. Then, for actions that matched we have analysed their appropriateness according to conditions, time annotations and intentions as stated in the protocol. In the rest of the section we present the critiquing exercise we have carried out. For space reasons, only the critiquing of test case 1 is presented in detail.

5.1 Critiquing Test Case 1

The data provided to the expert to solve test case 1 were:

> Admission of a term icteric[4] newborn. Gestational age is 40+2 weeks. Spontaneous delivery. On day 2 after birth laboratory data show a significantly elevated total bilirubin of 17,8 mg/dl.

In addition to this, the following lab test results were available on request: blood group and rhesus factor of the mother: O positive; antibodies: negative; blood group and rhesus factor of the child: A positive; direct Coombs test: negative. Table 3 shows the actions proposed by our expert to solve the case and the intentions behind them, as she reported. Note that she did not used the Asbru terminology for expressing intentions, which implies that it has been necessary to interpret them in terms of the ACHIEVE, MAINTAIN and AVOID in Asbru.

Matching the expert's actions and the protocol. The first step is studying which of the expert's actions are included in the protocol and which are not. This matching turned out to be very difficult. The results, that also appear in table 3, show that most of times there was no direct correspondence between an action and an Asbru plan (only for action 10), but rather a variety of other situations: actions that correspond to Ask statements in an Asbru plan (e.g. action 1), or to Display statements (e.g. action 5), or even to filter conditions (action 9). Besides, it is frequent to find situations in which several actions are mapped onto the same plan (e.g. actions 5 and 6). **Conclusion**. After some difficulties, most expert's actions appear in the protocol in some form, although not always as plans in their own right.

Studying the protocol recommendations not considered in the solution. Figure 2 illustrates the actions recommended by the protocol in this particular case, with indications of the actions that were (or were not) performed by the expert. It also indicates the actions for which the question was not decidable, mainly because of unavailable data. It is important to note that conditions, time annotations and intentions of the Asbru protocol have not been considered here.

[4] Jaundiced.

Table 3. Solution of the expert to test case 1. This table lists the actions advised by the expert, their intentions, as she reported them, and their correspondence with elements of the Asbru protocol.

	Action	**Intention**	**Is in protocol**
1	Measurement of TSB value	(1) Determine degree of hyperbilirubinemia and (2) Determine interim therapy	in Ask of plan 1.5
2	Blood group tests: **(a)** mother, **(b)** child	Determine possibility of blood group antagonism	**(a)** no **(b)** in Ask of plan 1.4.2.2
3	Rhesus factor tests: **(a)** mother, **(b)** child	Determine possibility of rhesus conflict	**(a)** no **(b)** in Ask of plan 1.4.2.2
4	Child Coombs test	Determine (im)possibility of blood group antagonism/rhesus conflict	in Ask of plan 1.4.2.2
5	Hemoglobin test	(1) Determine presence/severity of hemolysis and (2) Determine need of blood transfusion	in Display of plan 1.4.3.1
6	Reticulocyte count	Search for clues of hemolysis	in Display of plan 1.4.3.1
7	Ratio direct/indirect bilirubin	(1) Determine first hypothesis concerning the cause of hyperbilirubinemia and (2) Determine further diagnosis/therapy	no
8	Question about geographic origin: parents, child	Determine chance of G6PD deficiency	in Ask of plan 1.4.3
9	Looking up TSB value in table	Choose correct therapy	in filter conditions of plans 1.5.1.1–5 and 1.5.2
10	Prescription of phototherapy	Prevent harmful effects of high bilirubin levels	in plan 1.5.1.3
11	New measurement of TSB value (within 6 to 12h)	(1) Determine effect of phototherapy, (2) Determine presence/severity of hyperbilirubinemia, and (3) Determine further therapy	in Ask of plan 1.5

Apart from undecidable questions, we find several points where the physician did not follow the protocol recommendations: different anamnesis questions, determination of jaundice and study of feeding alternatives. For the first two situations, we did not conclude that the expert was deviating from the protocol since we do not deem likely that she forgot to consider those aspects. For the last situation, skipping the study of feeding alternatives turned out to be a common practice in Netherlands. **Conclusion**. Summarising, except for a regional deviation of the protocol, the expert roughly followed the recommendations in the AAP protocol.

Studying the expert's actions not recommended in the protocol. Some of the expert's actions were not recommended by the protocol to solve the current case —actions 2a, 3a and 7. Even more, they are not even included in the protocol. First, the blood typing of the mother (actions 2a and 3a) is simply not considered in the protocol. When these data are not available, the diagnosis goes on with the blood typing and the Coombs test of the child. The intention of the expert, which is determining the possibility of blood group/rhesus incompatibility, causing hemolysis, is certainly relevant for the case. Since

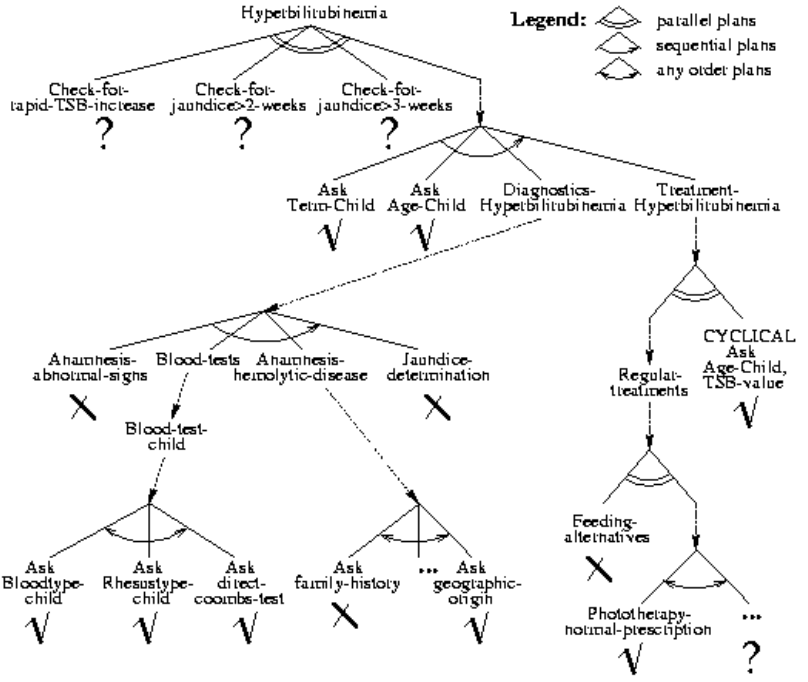

Fig. 2. Recommendations of the protocol to test case 1. Solution recommended by the protocol given the initial data of case 1, together with indications of the actions that were included in the expert's solution (signalled with $\surd$) and the ones that were not (signalled with $\times$), as well as the actions for which the question was not decidable (with ?).

it can be fulfilled by means of a Coombs test of the child, it should be asked which is the optimal action in this case. Second, the determination of the ratio direct/indirect bilirubin is not considered either in the protocol. The intention here is to obtain additional clues concerning the cause of jaundice. This is obviously relevant to the diagnosis part of the protocol, and therefore it should be asked whether the action is optimal in the current context. According to the answers to the previous questions, the protocol can be modified to consider these alternatives. **Conclusion**. These examples show how the protocol can be improved by studying deviating actions together with their justifications.

Studying the appropriateness of expert's actions. For this analysis we grouped the actions that were mapped onto the same plan. The main reason is that intentions are in Asbru ascribed to plans rather than to individual statements of plans. Then we checked for every action group whether the conditions and time annotations of the associated plan enabled it, and whether the intentions of the plan were similar to the intentions behind the action.

Almost all expert's actions were appropriate according to conditions and times. The only untimely actions were the two TSB measurements, which were both matched against the Ask reading a new value of plan 1.5. This was wrong for the initial measurement,

but it was the only possibility given that in the protocol there is no explicit sentence for reading the initial TSB value. Regarding the second TSB measurement, which was matched correctly, it is considered untimely due to the divergence in reading intervals. **Conclusion**. These examples show how the study of actions considered untimely can also contribute to improve the protocol, in this case making explicit the initial reading of the TSB value and refining the time annotation of the second (and further) readings.

However, most expert's actions were not appropriate according to intentions. We ended up considering intentions similar in most cases, but we could only conclude this after using additional knowledge/reasoning (implicit intentions, medical knowledge or reasoning about inheritance of intentions). For instance, the intention of action 8 "determine chance of G6PD deficiency" was considered equivalent to one of plan 1.4.3 "determine possibility of hemolytic disease", because G6PD is a cause of hemolytic disease. Still, intentions are just similar, which means that only part of them match, with the possibility of subsumption. **Conclusion**. We cannot strictly rely on intentions for studying the appropriateness of expert's actions, not even in the crucial step of matching expert's actions against Asbru elements, as we would have liked to.

5.2 Critiquing Other Test Cases

We have performed similar critiquing checks with other six cases—test cases 2 through 8, except test case 7, which was discarded because the expert had problems with the terminology. Next we briefly summarise the main results of critiquing.

Matching the expert's actions and the protocol. The results of test case 1 in this regard are quite representative. In addition, there is a high number of actions corresponding to Display statements, and in some situations it is not clear whether a correspondence can be established at all. For instance, in test case 2 it was not clear whether the haptoglobin test could be considered part of the Display of plan 1.4.3.1 due to its formulation: "Perform appropriate laboratory assessment of infant including (but not limited to):. . . ". The protocol can be improved by making this Display more explicit. Another interesting finding is that sometimes extra medical knowledge is needed for action matching. For instance, the action "determination of shape of erythocytes" in test cases 2 and 4 was matched with the Display of plan 1.4.3.1, which includes a differential smear, because this test provides information about the shape of blood cells.

Studying the protocol recommendations not considered in the solution. In the rest of the cases the expert did not perform either of the actions she skipped in test case 1: several anamnesis questions, the determination of jaundice and the study of feeding alternatives. Our conclusion is again that these deviations from the protocol are either irrelevant or legal. Other discrepancies were found in cases 3, 5 and 8. In test case 5 the expert skipped the blood typing and Coombs test of the child. A plausible explanation is that she thinks jaundice is caused by the infant's cephalic hematoma, and that hence no other causes need to be considered. However, we cannot rely on our own interpretation of the expert's intentions to justify her actions. Discrepancies of cases 3 and 8 are different. The application of the protocol should be interrupted in both cases due to, respectively, a preterm infant and a positive Coombs test. In case 3 the expert only prescribed observation, but the actions to solve case 8 suggest a completely different approach. The latter situations are further discussed in the next paragraph.

Studying the expert's actions not recommended in the protocol. Again, some of the expert's actions were not recommended by the protocol to solve the cases, and they were not even included in the protocol. We found that the blood typing of the mother and the determination of the ratio direct/indirect bilirubin of test case 1, and the above mentioned haptoglobin test, appear as a recurrent pattern in the rest of the cases. Like in test case 1, it should be studied whether these actions are relevant to the cases and optimal in their context, in order to modify the protocol if necessary. The situations referred to in the previous paragraph are different. Here the discrepancies cannot be justified since they directly involve the applicability conditions of the whole protocol.
Studying the appropriateness of expert's actions. The results of test case 1 concerning this aspect are also representative. Again, almost all expert's actions were appropriate according to conditions and times, but not when considering intentions.

6 Results of Critiquing

Our critiquing exercise, in spite of the difficulties we have encountered, shows the utility of critiquing for quality improvement of medical protocols. The results of this exercise can be summarised as follows. *First*, we can say that most expert's actions are included in the protocol.

Second, we consider that the expert did not comply with the protocol in three of the cases.

Third, whenever the expert performs actions that are not recommended by the protocol, this occurs throughout all the cases, which means that the expert relies on a pattern of actions not considered in the protocol, therefore suggesting a gap. This impression is strengthened by the fact that these actions do not appear anywhere in the protocol, i.e. not even in subplans that are used in the particular case. We have seen that a study of the relevance and optimality of these actions can help to determine if the protocol should be modified to include them as valid alternatives.

Last, most expert's actions are appropriate according to conditions and times, but not according to intentions. The reason of this general mismatch is that the intentions we modelled in the protocol and the ones acquired from the expert when giving a solution are quite different. Since intentions are not explicitly stated in the protocol, they are very hard to model. As a result, they vary considerably in the amount of details, the level of abstraction, etc, which precludes their use even for the matching of expert's actions with Asbru plans.

7 Conclusions

Critiquing, even manually, is a difficult task. It is hard to match expert's actions with parts of the Asbru protocol. The actions do not always appear in the protocol as plans, and therefore there is no 1:1 correspondence (sometimes there is not even a correspondence) between expert's actions and protocol plans, but rather a variety of relationships. For this purpose we can think of using the intentions of both actions and plans, but, as we have already pointed out, this is also difficult. In fact, the direct use of intentions for critiquing, either for matching of actions or for studying their appropriateness, is not possible due to

the difficulties inherent to their modelling/acquisition. The intentions of the protocol are not stated explicitly, which makes them very hard to model, and the intentions reported by our expert almost always differ. It is not only a problem of vocabulary, but also a matter of differences in the degree of detail, abstraction level, etc. This problem can be partly solved by using additional knowledge to fill in the gaps. Lastly, critiquing needs a lot of specific knowledge. Not only medical knowledge and vocabularies to fill the gaps between intentions, but also knowledge about the actions requiring a critique and about regional policies, for an effective critiquing.

All the above implies that the automation of critiquing will be very hard, at least in the way we have approached it. Many of the previous problems have been avoided in the critiquing literature by using *ad-hoc* solutions, basically forcing the expert to use a series of predefined actions and intentions—the ones in the system. We believe that this is not a good way of solving the natural differences that exist between the protocol and the individual practices. Besides, it assumes that the protocol constitutes the golden standard, whilst our study has shown that even a high quality protocol might have to be adjusted after systematic exposure to analysis based on test cases.

Acknowledgements. We want to thank Frea Kruizinga, the pediatrics expert without whom the critiquing exercise would not have been possible, and Andreas Seyfang and Robert Kosara, for their support in using Asbru and their feedback on this work.

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. P. Clayton and G. Hripsak. Decision support in healthcare. *Int. J. of Biomedical Computing*, 39:59–66, 1995.
3. J. Fox, N. Johns, C. Lyons, A. Rahmanzadeh, R. Thomson, and P. Wilson. PROforma: a general technology for clinical decision support systems. *Computer Methods and Programs in Biomedicine*, 54:59–67, 1997.
4. P. M. Miller. *A Critiquing Approach to Expert Computer Advice: ATTENDING*. Research Notes in Artificial Intelligence. Pitman Publishing, 1984.
5. M. Musen, J. Rohn, L. Fagan, and E. Shortliffe. Knowledge engineering for a clinical trial advice system: Uncovering errors in protocol specification. *Bulletin du Cancer*, 74:291–296, 1987.
6. L. Ohno-Machado, J. Gennari, S. Murphy, N. Jain, S. Tu, D. Oliver, E. Pattison-Gordon, R. Greenes, E. Shortliffe, and G. Octo Barnett. Guideline Interchange Format: a model for representing guidelines. *J. of the American Medical Informatics Association*, 5(4):357–372, 1998.
7. A. Seyfang, S. Miksch, and M. Marcos. Combining Diagnosis and Treatment using Asbru, 2001. Congress of the Int. Medical Informatics Association (MEDINFO-2001).
8. Y. Shahar, S. Miksch, and P. Johnson. The Asgaard project: a task-specific framework for the application and critiquing of time-oriented clinical guidelines. *Artificial Intelligence in Medicine*, 14:29–51, 1998.
9. S. Weingarten. Using Practice Guideline Compendiums To Provide Better Preventive Care. *Annals of Internal Medicine*, 130(5):454–458, 1999.