Comparison of Three Knowledge Representation Formalisms for Encoding the NCEP Cholesterol Guidelines

Justin Starren and Guochun Xie

Center for Medical Informatics
Columbia University College of Physicians and Surgeons
New York, New York 10032

Although many Knowledge Representation (KR) formalisms have been used to encode care guidelines, there are few direct comparisons among different formalisms. In order to compare their suitability for encoding care guidelines, three different KR formalisms were used to encode the National Cholesterol Education Panel (NCEP) guideline. PROLOG, a First Order Logic system, CLASSIC, a frame-based representation system, and CLIPS, a production rule system, were used in the comparison. All three representations allowed accurate encoding of the guideline. PROLOG produced the most compact representation, but proved the most difficult to debug. The lack of arbitrary disjunction in CLASSIC greatly increased the complexity of the encoding. Overall, the CLIPS representation was the most intuitive and easiest to use.

INTRODUCTION

There are now over 1000 different care guidelines, and more are added or changed almost monthly. Despite documented improvements in outcomes [1], guidelines have had limited penetration into general clinical practice. One reason for this is that no clinician can possibly remember all of the guidelines. Using computerized decision support systems to implement the guidelines has been suggested as a solution. Computerized medical logic has already been used to implement a number of care guidelines and guideline-like protocols. [2]

Several different knowledge representation (KR) formalisms have been used to encode care guideline knowledge. Although studies have compared different systems for the development of treatment recommendations [3], there has been little work directly comparing different schema for the representation of the same care guideline. In order to compare the suitability of different paradigms for representing a typical care guideline, we selected three of the KR systems at Columbia to encode the National Cholesterol Education Panel (NCEP) guideline:

- PROLOG: a First Order Logic based system
- CLASSIC: a frame-based representation system
- CLIPS: a production rule system.

The differences in ease of implementation and intuitiveness of representation will be discussed.

METHODS

The NCEP Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults is now in its second version [4]. Like the first version, it presents its basic recommendations as block diagrams or tables. One of these is shown in Figure 1. An earlier version of the NCEP guidelines has been encoded by two other groups [5,6]. The NCEP guideline is one of the older and better known of the national care guidelines. The NCEP is quite specific about the thresholds and targets for dietary and drug therapy, but is less specific about the details. In particular, it does not state the exact sequence of drugs to be tried, or precisely when to determine that a particular trial of therapy is unsuccessful. Accordingly, only the unambiguous portions of the NCEP guideline was captured in the three KR representations.

Any extensions to the care guideline logic were limited to those which are clearly implicit in the guideline. For example, the guideline recommends testing cholesterol levels every 5 years. It also gives treatment recommendations based on cholesterol levels. This clearly implies that no treatment should be recommended based on a cholesterol level which is over 5 years old. Instead of using an out-of-date value, the system should stop and request that a new value be obtained. A similar expansion was required for each piece of data required by the careplan logic.

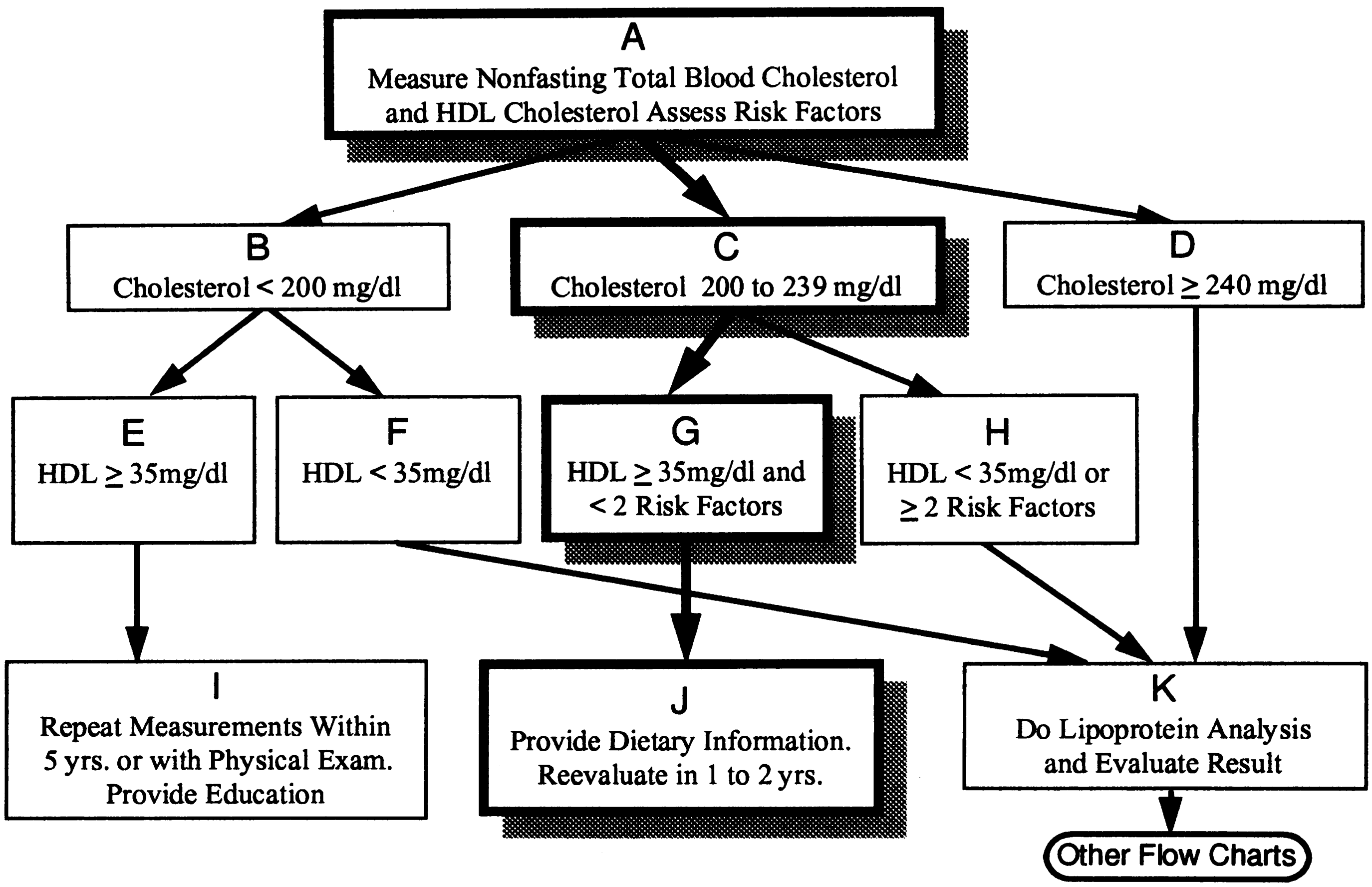

Figure 1. Part of the NCEP cholesterol guideline demonstrating initial evaluation of an uncomplicated patient. Letters were added for clarity. The highlighted path (A-C-G-J) will be used to show the differences between the representations. Also of significance is the fact that several decision paths diverge and then reconverge at box K.

The representation of time in computerized care guidelines has been the subject of considerable study. [5]. For the NCEP guideline, time considerations are limited largely to age of laboratory results and the duration of treatments. For that reason we selected a simple event-based time representation, rather than an interval based one.

For each implementation the patient data were encoded by hand in a form native to the representation. The specific representation syntaxes will be discussed under each of the systems. Data encoded included: patient identifier; laboratory results with times; risk factors; and, therapies with time of start. Correct implementation of the guideline (defined as the ability of each implementation to produce recommendations identical to ones generated manually from the published guideline) was considered a requirement for each implementation. Therefore, the accuracy of the systems was identical, by definition. The major criteria for comparison among the three representations was not the ability to correctly encode the guideline, but rather the difficulty in doing so.

A test set of patients was created for development purposes which were represented the eight different basic outcomes of the NCEP guideline. Additional patients were selected from the practice booklet developed by the NCEP. Since this booklet is based on the 1988 recommendations rather than on the new 1993 recommendations, the treatment recommendations were updated based on the new guideline. For this initial evaluation, no actual clinical cases were used. After initial development of each representation, these cases were used to test the system. In all cases, the initial implementations needed to be corrected to handle the practice cases. However, the representation varied significantly in the number of errors and the amount of effort required to correct them. For each of the implementations, an example patient is shown. The logic flow for this patient is highlighted as path A-C-G-J in Figure 1.

PROLOG IMPLEMENTATION

Quintas PROLOG was utilized for the implementation. Knowledge in PROLOG is represented as clauses. Inference is performed by pattern matching and resolution using Horn clauses.

Patient history data were stored as clauses which contain tuples of patient number, attribute, value, date. The example patient was represented as:

```
pmh(pat2, sex, male,date(94,2,1)).
pmh(pat2, age, 35,date(94,2,1)).
pmh(pat2, h_chd, 0,date(94,2,1)).
pmh(pat2, smoking, 0,date(94,2,1)).
pmh(pat2, htn, 1,date(94,2,1)).
pmh(pat2, dm, 0,date(94,2,1)).
pmh(pat2, chd, 0,date(94,2,1)).
lab(pat2, hdl, 70,date(94,2,1)).
lab(pat2, cholesterol, 236,date(94,2,1)).
```

Treatments were stored as therapy clauses which include the therapy, its status and the date of status change. Only diet and drug therapy are recorded. No distinctions between levels of diet and types of medication were included.

```
therapy(pat1,diet,on,date(92,1,1)).
```

A rule was written for each treatment outcome. Within the actual rules, information was accessed using CHECK_LAB and CHECK_PMH functions. The rule for the example patient is:

```
rule_j(PID) :-
 check_lab(PID, hdl, HDL,_),!,
 HDL >= 35,
 total_risk(PID, Risk),!,
 Risk < 2,
 check_lab(PID, cholesterol, C,_),
 C >= 200,
 C =< 239,
 print_rule_j.
```

Dates were represented using an internal PROLOG format. For time interval computations, this format is converted a second representation as a pair of integers corresponding to UNIX system time.

CLASSIC IMPLEMENTATION

CLASSIC is a frame-based description logic system. It is a special set of functions implemented in Lisp. Knowledge is represented as a collection of concepts, individuals, roles, rules, and the relationships among them. Concepts are defined as descriptions that are applied to individuals. Individuals are instances of concepts. Since the classification of the individuals occurred as they were instantiated, there was nothing to "run" in the typical sense. Once the patients were classified according to the automatic subsumption rules, the value of the role "treatment-plan" was filled automatically. The example patient is represented as:

```
(CL-CREATE-IND
 'PAT2
 '(AND PATIENT
   (FILLS SEX MALE) (FILLS AGE 35)
   (FILLS H-CHD NO) (FILLS SMOKING NO)
   (FILLS HTN YES) (FILLS DM NO)
   (FILLS CHD NO) (FILLS HDL 70)
   (FILLS CHOL 236)))
```

Treatments are handled like any other role filler. The treatment recommendation is determined by checking to see whether the treatment role has a filler.

In CLASSIC encoding of the guideline is treated as a subsumption problem. A patient is matched to the most specific subsuming concept, and the treatment recommendation determined on that basis. For example, in order for a patient to satisfy the criteria for box G it has to also satisfy box C, and therefore, box A. Thus, the concepts to classify the example patient to box J were.

```
(CL-DEFINE-CONCEPT 'OKLAB-PATIENT
 '(AND TESTED-PATIENT
     (FILLS TREATMENT-PLAN " ")))
(CL-DEFINE-CONCEPT
   'UNTREATED-PATIENT
   '(AND PATIENT
     (FILLS TREATMENT-PLAN " ")))
(CL-DEFINE-CONCEPT 'A-PATIENT
 '(AND UNTREATED-PATIENT
     OKLAB-PATIENT
     (FILLS CHD NO)))
(CL-DEFINE-CONCEPT 'C-PATIENT
 '(AND A-PATIENT
    (ALL CHOL
     (AND INTEGER
       (MIN 200) (MAX 239)))))
(CL-DEFINE-CONCEPT 'G-PATIENT
 '(AND C-PATIENT LOW-RISK-PATIENT
   (ALL HDL (AND INTEGER (MIN 35)))))
```

The handling of "or" relations in CLASSIC was complex, and required the creation of artificial branches and artificial roles. The representation of the path convergence for box K of Figure 1 is shown in Figure 2.

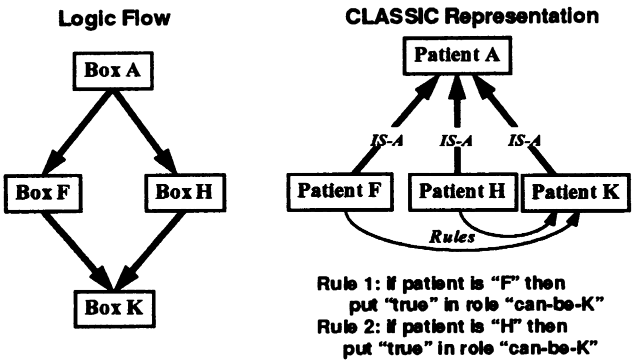

Figure 2. Representation of an "OR" convergence in CLASSIC.

CLIPS IMPLEMENTATION

CLIPS 6.1 is a forward chaining production rule system written in ANSI C by NASA. The knowledge in CLIPS is encoded in rules and functions. CLIPS also supports objects and individuals. Since it does not support automatic classification, we used templates for the creation of patient data structures. The CLIPS inference engine includes truth maintenance, dynamic rule addition and customizable conflict resolution strategies. However, these were not required.

Individual patients are stored as patterns of the type:

```
(patient-type(attribute1 VALUE1)
        (attribute2 VALUE2) ...)
```

This template was used for both history attributes such as "smoking" as well as numeric attributes such as "chol". Because CLIPS has a well developed default value system, it was used for the handling of missing history elements. Missing laboratory values are represented as -1. If the missing value is needed for a rule, the caregiver was prompted. The example patient was represented as:

```
(assert (uncalculated-patient
    (sex male) (name pat2) (age 35)
    (h-chd no) (smoking no)(dm no)
    (htn yes) (chd no) (hdl 70) (chol 236)))
```

For processing of the individual patients, the (uncalculated-patient (...)) data structure is converted to a (calculated-patient (...)) template which includes a numerically computed risk from the history attributes.

Unlike CLASSIC and PROLOG, CLIPS has no intrinsic time or date functions. In order to support date, a system call was executed to perform a UNIX date operation and store the date in a file. CLIPS then read the file and converted the value into a single integer. In a production implementation of this system, an external function could be written to provide date functions within CLIPS.

The rule based structure of CLIPS lent itself to a creation of explicitly defined states which were named based on their transitions among the boxes of the diagram (Figure 1). In addition, a "done" flag was added to force the logic to terminate when a treatment recommendation was generated. The final rule for the example patient was:

```
(defrule C2G2J "Rules to reach box J"
   ?f1 <- (calculated-patient (state c)
      (done no) (hdl ?hdl) (name ?name))
   (test (>= ?hdl 35))
  =>
   (printout t crlf "Patient " ?name
      " needs the following treatement:"crlf)
   (printout t ?treatment-j crlf)
   (modify ?f1 (done yes) (state j)))
```

Although the states were similar, in some ways, to the CLASSIC concepts, there was a fundamental difference. CLIPS allows states to be specified by rules with disjunctive conditions. Consequently, encoding path convergence in the guideline was straightforward.

DISCUSSION

PROLOG provided for the easiest construction of the patient database. In addition, it allowed for the storage of multiple sequential values for each patient and easy retrieval of the appropriate value. The inclusion of a human-readable date format was also quite convenient. The ability to specify arbitrary disjunctions allowed easy encoding of the convergence points in the guideline.

A well-known difficulty with PROLOG is with the control of backtracking. Because each lab result was a separate fact, it was possible for PROLOG backtrack from partial matches of current data into old data and, thereby, into incorrect conclusions. "Cuts" are a method in PROLOG for limiting such undesired backtracking. Unfortunately, the behavior of cuts is notoriously complex, and they were responsible for most of the errors encountered.

Although the automatic classification of CLASSIC seemed to be a natural representation for the block structure of the NCEP guideline, the difficulty of use overshadowed this. CLASSIC was by far, the hardest of the three systems to use, in large part because it required detailed instantiation of every possible outcome. There was no easy way to classify by exclusion. For example, every patient is either treated or untreated, (never both and never neither). Such a dichotomy was easy to encode in PROLOG and CLIPS, but both sides of the dichotomy had to be explicitly defined in CLASSIC. Fortunately, the ability to print the entire lineage of each individual greatly simplified the debugging.

Overall, CLIPS was the most useful of the three systems evaluated. The CLIPS approach, of states and rules, was the easiest to conceptualize. Patient attributes were easy to specify. No awkward manipulations were required to capture the logic. The only real disadvantage was the lack of built-in time and date functions. CLIPS and PROLOG required a similar number of program statements, while CLASSIC required almost three times as many.

Most important, all three representations proved adequate for encoding the guideline. Unlike the problems encountered by computerized diagnostic systems, care guidelines are specifically designed to be clear and unambiguous. Care guidelines for human use are often phrased in terms of "If symptom X, then give treatment Y." This type of explicit linkage between antecedents and consequents is ideally suited to both the rule structure of production-rule systems, as well and the Horn clause structure of PROLOG. There is no need to evaluate the relative merits of competing hypotheses. For a production-rule system, this means that there is no need for resolution of large conflict-sets. In addition, the NCEP guidelines is explicit that all information is either already available or it is requested. Therefore, there is no need to reason with incomplete data. Given all this, it is not suprising that, for this particular guideline, the resolution-based and the production-rule representations were a more natural fit to the problem. No hybrid representations were evaluated, and it is possible that one of these could prove superior to the systems tested here.

Acknowledgement

This work was funded in part by a National Library of Medicine Extramural Training Grant.

References

1. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. Lancet 1993; 342:1317-22.

2. Johnston ME, Langton KB, Haynes RB, Mathieu A. Effects of Computer-based Clinical Decision Support Systems on Clinician Performance and Patient Outcome. Ann Intern Med 1994; 120:135-142.

3. Long WJ, Griffith JL, Selker HP, D'Agostino RB. A comparison of logistic regression to decision-tree induction in a medical domain. Comp. Biomed. Res. 1993; 26, 74-79.

4. Expert Panel. Summary of the Second Report of the National Cholesterol Education Program (NCEP) Expert Panel on the Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel II) JAMA 1993;269:3015-3023.

5. Rucker DW, Maron DJ, Shortliffe EH. Temporal representation of clinical algorithms using expert-system and database tools. Comput Biomed Res. 1990;23(3):222-39.

6. Weissfeld JL, Weissfeld LA, Holloway JJ, Bernard AM. A mathematical representation of the expert panel's guidelines for high blood cholesterol case-finding and treatment. Med Decis Making. 1990;10(2):135-46.