Guideline Formalization and Knowledge Representation for Clinical Decision Support

Tiago Oliveira[a], Paulo Novais[a], José Neves[a]

[a] Department of Informatics, University of Minho, Braga, Portugal

KEYWORD	ABSTRACT

Computer-Interpretable Guidelines
Incomplete Information
Quality of Information
Decision Support

The prevalence of situations of medical error and defensive medicine in healthcare institutions is a great concern of the medical community. Clinical Practice Guidelines are regarded by most researchers as a way to mitigate these occurrences; however, there is a need to make them interactive, easier to update and to deploy. This paper provides a model for Computer-Interpretable Guidelines based on the generic tasks of the clinical process, devised to be included in the framework of a Clinical Decision Support System. Aiming to represent medical recommendations in a simple and intuitive way. Hence, this work proposes a knowledge representation formalism that uses an Extension to Logic Programming to handle incomplete information. This model is used to represent different cases of missing, conflicting and inexact information with the aid of a method to quantify its quality. The integration of the guideline model with the knowledge representation formalism yields a clinical decision model that relies on the development of multiple information scenarios and the exploration of different clinical hypotheses.

1 Introduction

Given the stressful conditions that healthcare professionals are subjected to, it is not surprising for situations of medical error and defensive medicine to occur. On one hand, the excessive workload of healthcare institutions may drive practitioners to commit errors of execution or errors of planning that are responsible for the occurrence of adverse events in patients (e.g., injuries, death) [KALRA, 2004]. On the other hand, they may choose to deliberately refuse a patient or order treatments, tests and procedures that are not really necessary, to avoid lawsuits and accountability for the assessments and decisions they make based on the health condition of a patient, a practice that goes by the name of defensive medicine [CHAWLA & GUNDERMAN, 2008]. The frequency of these events in healthcare should not be neglected. For instance, in London hospitals, the rate of adverse events is about 10.8% and in Australian hospitals this number rises to 16.6% [KALRA, 2004], also, in the United States of America (USA) it is estimated that nearly 44 000 people die annually as a result of medical errors [BRENNAN, 2000]. These numbers are significant mainly because they reflect how a poorly conducted medical practice can affect a patient's life and the worldwide concern with this aspect. Defensive medicine is also among the primary concerns of the medical community, given the excessive spending that spawns from this practice and the increasing rate of false positives from excessive testing [CHAWLA & GUNDERMAN, 2008]. Both situations undermine the confidence of patients in healthcare professionals. To reduce their occurrence it is necessary an increased compliance with Clinical Practice Guidelines (CPGs) [ROSENBRAND et al., 2008].

CPGs are documents based on scientific evidence that contain recommendations for healthcare professionals to deal with specific clinical cases. Their structure varies according to the organizations and countries that produce them, but normally they con-

sist of sets of instructions displayed along extensive textual documents [ROSENBRAND et al., 2008]. However, this format does not facilitate the dissemination of CPGs and is difficult to consult, since it is hard for healthcare professionals to extract the relevant information they need to manage patients [ROSENBRAND et al., 2008]. Furthermore, this type of format is not machine-readable. Therefore, it is impractical for the integration in Clinical Decision Support Systems (CDSSs). A machine-readable format would enable the encoding of the instructions in the computer system and thus the CDSS would be able to provide detailed instructions in real time to healthcare professionals in a truly interactive experience. Moreover, it would facilitate the development of automated mechanisms to update CPGs which would solve another current problem of the text format. Updating guidelines is a complex task and it usually involves the adaptation of whole documents. Given the difficulty of the task, most guidelines are only updated every five years, which is not suited for documents that intend to deliver the latest developments in medical science.

Another problem that healthcare professionals are faced with is the existence of incomplete information (incoherent or missing) during the clinical process [STRASZECKA, 2006]. Frequently, these cases block the information flow during the clinical process and may impair the application of CPGs. The current CDSSs (most of which are still in an academic development phase) do not cope with the issue of incomplete information in clinical careflows.

The objective of the work presented in this paper is the development of a machine-readable representation model for CPGs that may be integrated in a CDSS and is able to encompass all the dimensions of the information in a clinical process, including the incomplete one. In order to present the work we will start by a resumed state of the art in section 2, then we will present the CPG model in section 3. The details of the knowledge representation formalism and the management of incomplete information will be presented in section 4, followed by details of the assessment of the quality of information, in section 5. In section 6, a clinical decision model based on the current developments will be presented. Finally, in the last section of the paper we will conclude about the utility of the model and point the future directions of our research.

2 Related Work about Computer-Interpretable Guidelines

Computer-Interpretable Guidelines (CIGs) are computer representations of CPGs that are machine readable and may be integrated in CDSSs, thus allowing immediate appliance. A digital support for CPGs may be a game changer in all the aspects that revolve around them, especially development, dissemination, implementation and execution. Researchers became interested in CIGs in the end of the 1980s. Currently, the most relevant approaches to the representation of CIGs are Arden Syntax [SAMWALD et al., 2012], Guideline Interchange Format (GLIF) [PATEL et al., 1998], PROforma [FOX et al., 1998], Asbru [SHAHAR et al., 1998] and Standards-Based Sharable Active Guideline Environment (SAGE) [TU et al., 2007].

Arden Syntax [SAMWALD et al., 2012] was initially presented in 1989 and is now a standard of Health Level 7 (HL7). It focuses on providing a simple representation for CPGs. In fact, an Arden Syntax file is only capable of accommodating knowledge for one decision. This excessive simplicity is among the criticisms made to this approach along with the absence of a standard support for the files containing the guidelines. GLIF [PATEL et al., 1998], and its current version, GLIF3, is a model that can accommodate a higher degree of complexity than Arden Syntax. Following the Task Network Model (TNM) paradigm, in which a guideline is represented as a set of steps. Each step corresponds to an instance of a primitive of the model. However, this model requires the use of a subset of Arden Syntax to express clinical decisions and a subset of Asbru to express temporal constraints on the execution of the steps. So, part of the criticisms that are made to Arden Syntax can also be made to GLIF3. Asbru [SHAHAR et al., 1998] is a model that is mainly concerned with the temporal aspects of clinical guidelines. Thus, Asbru defines temporal annotations, which specify four points in time for the execution of plans and the verification of conditions. With the particularity of enabling the expression of uncertainty in starting time, ending time and duration of a time interval, but this temporal representation is often regarded as too complex for healthcare professionals to get familiar with. The PROforma [FOX et al., 1998] model was developed in 1998 and presents an abstract view of guide-

lines as flowcharts where the nodes are instances of pre-defined classes of tasks. The syntax of PROforma, similarly to Arden Syntax, was defined in Backus-Naur Form in an ASCII file, so like its predecessor, since it is not supported by a standard format it is necessary to develop applications to query specifically these files, which is not practical for a guideline format intended to be used by different organizations and software applications. SAGE is the latest of the presented models and depicts CPGs as recommendation sets displayed as a graph of nodes: context, action, decision and routing. Being a recent project, SAGE [TU et al., 2007] is involved with organizations that promote standards for the representation of clinical information (e.g., HL7).

The application of these models in real practice is very limited and it is normally oriented to a specific health condition, since most of them require a certain degree of customization to deal with particular health conditions.

3 Guideline Depiction Model

The proposed model represents guidelines as flowcharts and it is depicted in Fig. 1. The aim is to offer an intuitive representation of clinical concepts and procedures. The model was defined using the Ontology Web Language (OWL) and tasks are its basic unit, every procedure of the clinical process is viewed as a task. So, to represent different types of tasks, some types of primitives were created, namely Plan, Action, Question, Decision and Aggregation Module [OLIVEIRA et al., 2011].

Each Plan models a guideline or a subguideline. It is a collection of tasks that may contain any number of instances of the other primitives. A Plan possesses administrative attributes with information about the guideline and attributes that express its clinical goal.

An Action represents a task that must be performed by healtcare professionals, such as clinical procedures, medication prescription, observations and simple recommendations. This primitive class possesses attributes that enable the expression of the scope of the Action and its detailed description.

The following task is Question, which is used to obtain information about the state of a patient, more specifically, the values of the clinical parameters necessary in order to execute the guideline. Each question has attributes that specify the clinical
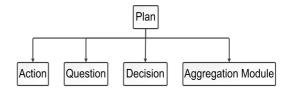

Fig. 1. Schematic view of the primitives of the representation model, where each primitive denotes a task of the clinical process.

parameters to be obtained and the type of value that is expected in return, i.e, if they are numeric or qualitative values and what units they are in.

A Decision task is a bifurcation point in the clinical process. It requires the choice between two or more options, determined by rules. This task has attributes for the options and the conditions that compose the rules, with comparison operators and relational boolean operators.

The last task is Aggregation Module and, as hinted by the name, it aggregates other tasks, forming a block bound by the same execution constraints. These constraints may be defined in terms of clinical goals or in terms of temporal restrictions such as periodicity, cyles, duration and repetitions.

There are common attributes to all of the tasks, namely those that provide scheduling constraints and establish a connection between them, thus providing the order by which they should be executed.

The case study depicted in Fig. 2, corresponds to a fragment of the Detection of High Blood Cholesterol in Children and Adults guideline, developed by the National Heart Lung and Blood Institute of the USA. It refers to the selection of the appropriate treatment for high levels of cholesterol, based on the presence of the following clinical parameters: coronary heart disease, risk factors and the ten year risk of coronary heart disease.

Going from the top to the bottom, the guideline is represented as Plan 1. This plan initially contains Question 1, which is used to obtain the values of coronary heart disease, risk factors and ten year risk. After the input of these values by a healthcare professional, the next task, which is a Decision, is executed. The Decision task contains the rules for the establishment of the Low Density Lipoprotein (LDL) therapy goal, based on the data that was previously collected. The three available options are 100, 130 and 160 mg/dL. After de decision about the therapy goal

is made, the next task will be one of the three Aggregation Modules in the figure, depending on the trigger condition that matches the outcome of the previous task. Each Aggregation Module contains the necessary actions in order to achieve the goal that is expressed in its attributes. The execution of these tasks will stop only if the specified LDL blood level is achieved.

cal approach and an extension to Logic Programming. The representation of the state of a patient resorts to Extended Logic Programming (ELP) [NEVES et al., 2012; NOVAIS et al., 2011]. ELP is one of the few non monotonic logic formalisms capable of representing the classic cases of incomplete information, such as uncertainty, incompleteness, inaccuracy, and incoherence. ELP uses default negation
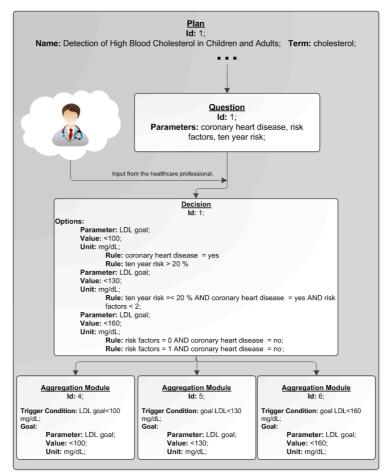

Fig. 2. Representation of a fragment of the Detection of High Blood Cholesterol in Children and Adults guideline, developed by the National Heart Lung and Blood Institute, in the proposed model.

4 Knowledge Representation and Incomplete Information

The Logic Programming paradigm has been used by many approaches for knowledge representation and reasoning, namely by Model Theory and Proof Theory. In this work it is followed the proof theoreti-

(*not p*) along with classic negation ($\neg p$), to explicitly represent negative information, which is useful to distinguish what is false because it cannot be proven, from what is false because its negation can be proven [NEVES, 1984]. An ELP program possesses a finite set of clauses in the form:

$$q \leftarrow p_1 \wedge \ldots \wedge p_n \wedge \text{not } q_1 \wedge \ldots \wedge \text{not } q_m$$
$$? \; p_1 \wedge \ldots \wedge p_n \wedge \text{not } q_1 \wedge \ldots \wedge \text{not } q_m \; (n, m \geq 0)$$

patients		
id	*first_name*	*last_name*
1	James	Smith
2	Paul	Davies
….		

clinical parameters			
id	*parameter*	*type*	*unit*
1	coronary heart disease	string	no_unit
2	risk factors for coronary heart disease	integer	no_unit
3	ten year risk of coronary heart disease	double	%

one-to-many   *one-to-many*

patient_state			
id	*patient_id*	*parameter_id*	*value*
1	1	1	no
2	1	2	{1, 2, 3}
3	1	3	{19, 21}
4	2	1	⊥
5	2	2	1
6	2	3	18
…			

where *?* is a domain atom denoting falsity, the p_i, q_j,

Fig. 3. Representation of a relational model that describes the state of patients.

and *p* are classical ground literals. In this representation formalism, every program is associated with a set of abducibles, given here in the form of exceptions to the extensions of the predicates that make the program. To reason about the knowledge base presented in a particular set, that considers incomplete information on the base of the formalism referred to above, let us consider a procedure given in terms of the extension of a predicate denoted as demo. This meta predicate is given by the signature *demo:T,V → {true, false, unknown}*, according to the following set of terms:

```
demo(T, true) ← T.
demo(T, false) ← ¬T.
demo(T, unknown) ← not T, not ¬T.
```

The first clause establishes that a theorem to be proved is put to a knowledge base of positive information returning the truth-value *true*. The second clause denotes that the theorem to be proved recurred to the negative information presented in the knowledge base, returning the truth-value *false*. The third clause stands for itself, introducing another possibility besides true or false in the universe of discourse, the truth value *unknown*. As an example, let us consider the case showed in Fig. 3, with respect to two patients, James and Paul. James does not have

coronary heart disease. However, the clinical staff is unable to detect exactly how many risk factors he shows, but they believe the number to be in the set {1, 2, 3}. The ten year risk of coronary heart disease is also impossible to determine with accuracy, given the uncertainty about the risk factors, yet the healthcare professionals know it is either 19% or 21%. The health condition of Paul is significantly different, the healthcare professionals are unable to determine if he has coronary heart disease or not, but the number of risk factors and the ten year risk of coronary heart disease are known to be 1 and 18 %, respectively. In the context of ELP, the information about the presence of coronary heart disease, the number of risk factors and the ten year risk of coronary heart disease can be represented with the following extensions of predicates *chd*, *risk_factors* and *ten_year_risk*, in the forms of Programs 1 and 2.

```
¬chd(X,Y) ← not chd(X,Y), not abducble_chd(X,Y).
abducible_chd(X,Y) :- chd(⊥,Y).
chd(no,no_unit).

¬risk_factors(X,Y) ← not risk_factors(X,Y),
not abducible_risk_factors(X,Y).
```

Program 1. Extensions of the predicates that model the health condition of patient James.

```
abducible_risk_factors(X,Y) :- risk_factors(⊥,Y).
abducible_risk_factors(1,no_unit).
```

```
abducible_risk_factors(2,no_unit).
abducible_risk_factors(3,no_unit).

¬ten_year_risk(X,Y) ← not ten_year_risk(X,Y),
not abducible_ten_year_risk(X,Y).
abducible_ten_year_risk(X,Y) :- ten_year_risk(⊥,Y).
abducible_ten_year_risk(19,%).
abducible_ten_year_risk(21,%).
? (ten_year_risk(X₁,Y₁) ∨
ten_year_risk(X₂,Y₂))∧ ¬ (ten_year_risk(X₁,Y₁)
∧ ten_year_risk(X₂,Y₂)).
```

Program 2. Extensions of the predicates that model the
health condition of patient Paul.

```
¬chd(X,Y) ← not chd(X,Y), not abduci-
ble_chd(X,Y).
abducible_chd(X,Y) :- chd(⊥,Y).
chd(⊥,no_unit).

¬risk_factors(X,Y) ← not risk_factors(X,Y),
not abducible_risk_factors(X,Y).
abducible_risk_factors(X,Y) :- risk_factors(⊥,Y).
risk_factors(1,no_unit).

¬ten_year_risk(X,Y) ← not ten_year_risk(X,Y),
not abducible_ten_year_risk(X,Y).
abducible_ten_year_risk(X,Y) :- ten_year_risk(⊥,Y).
ten_year_risk(18,%).
? (ten_year_risk(X₁,Y₁) ∨
ten_year_risk(X₂,Y₂))∧ ¬ (ten_year_risk(X₁,Y₁)
∧ ten_year_risk(X₂,Y₂)).
```

In both programs, the first clause of each exten-
sion denotes its closure, i.e., the application of the
Closed World Assumption to the predicate, meaning
that a statement is considered to be false if it is nei-
ther represented as negative information in the
knowledge base nor as an abducible statement. The
second clause of each predicate means that a state-
ment is abducible if it is represented as positive in-
formation with the symbol '⊥' that stands for a null
value, meaning that the variable X may take any val-
ue of the X domain. In Program 1, the third clause of
predicate *chd* is a case of positive information in
which the value of this parameter is no. In the same
program, the last three clauses of *risk_factors* denote
that the number of risk factors is in the set
{1,2,3,{1,2},{1,3},{2,3},{1,2,3}}, since there is not a
clause stating that the values are disjoint. Clauses

four and five of *ten_year_risk* put the value of this
parameter in the set {19, 21} and the sixth clause is
an invariant that denotes that the value of the ten year
risk is either 19% or 21%, but not both. The infor-
mation cases presented in Program 2 are similar to
the ones presented in Program 1, with the exception
of the value of *chd*, which is represented as positive
information with the symbol '⊥'. Now it is possible
to build different information scenarios, like those of
Table 1, with the cases of incomplete information and
present them to healthcare professionals along with
the conclusions drawn from the application of the
recommendations contained in the digital guideline,
developed in the previous section, to each scenario.

5 Providing a measure of the Quality-of-Information

We have already covered two essential aspects of
the development of CDSSs, the representation of
clinical recommendations and the representation of
the information about the health condition of patients
(including incomplete information). Now, based on
the recommendations of the CPGs and the state of a
patient it is necessary to make decisions about the
tasks to be performed after the decision point in the
context of a clinical process. To do that with incom-
plete information, it becomes necessary to measure
the reliability of the available information. ELP ap-
pears here associated with an evaluation method
called the Quality-of-Information (QoI) [NEVES et
al., 2012; NOVAIS et al., 2011]. The QoI of an ex-
tension of a predicate *i* is defined in terms of truth
values in the interval [0,1], i.e., if the information is
known (positive) or false (negative), the QoI for the
extension of predicate *i* is 1. For cases in which the
information is unknown (like *chd* in Program 2) the
QoI is given by Equation 1.

$$QoI_i = \lim_{N \to \infty} 1/N = 0(N >> 0) \qquad (1)$$

Information Scenarios	QoI	Inferable Options and Scores
From Program 1		
1 $\neg chd(QOI,X,Y) \leftarrow not\ chd(QOI,X,Y).$ $chd(1,no,no\_unit).$ $\neg risk\_factors(QOI,X,Y) \leftarrow not\ risk\_factors(QOI,X,Y)$ $risk\_factors(0.143,1,no\_unit).$ $\neg ten\_year\_risk(QOI,X,Y) \leftarrow not\ ten\_year\_risk((QOI,X,Y)).$ $ten\_year\_risk(0.500,19,\%).$	$QoI_{chd}= 1.000$ $QoI_{risk\_factors}=1/(2^3-1)= 0.143$ $QoI_{ten\_year\_risk}= \frac{1}{2}= 0.500$	$V_{ldl\_goal < 160}= 0.5*1.00+0.5*0.143= 0.573$
2 $\neg chd(QOI,X,y) \leftarrow not\ chd(QOI,X,Y).$ $chd(1,no,no\_unit).$ $\neg risk\_factors(QOI,X,Y) \leftarrow not\ risk\_factors(QOI,X,Y).$ $risk\_factors(0.143,1,no\_unit).$ $\neg ten\_year\_risk(QOI,X,Y) \leftarrow not\ ten\_year\_risk(QOI,X,Y).$ $ten\_year\_risk(0.500,21,\%).$	$QoI_{chd}= 1.000$ $QoI_{risk\_factors}= 0.143$ $QoI_{ten\_year\_risk}=0.500$	$V_{ldl\_goal < 100}=1*0.500=0.500$ $V_{ldl\_goal < 160}=0.573$
3 $\neg chd(QOI,X,Y) \leftarrow not\ chd(X,Y).$ $chd(1,no,no\_unit).$ $\neg risk\_factors(X,Y) \leftarrow not\ risk\_factors(X,Y).$ $risk\_factors(0.143,2,no\_unit).$ $\neg ten\_year\_risk(X,Y) \leftarrow not\ ten\_year\_risk(X,Y).$ $ten\_year\_risk(0.500,19,\%).$	$QoI_{chd}= 1.000$ $QoI_{risk\_factors}= 0.143$ $QoI_{ten\_year\_risk}=0.500$	$V_{ldl\_goal < 130}=0.33*1+0.33*0.143*0.33*0,500$ $= 0.542$
4 $\neg chd(QOI,X,Y) \leftarrow not\ chd(QOI,X,Y).$ $chd(1,no,no\_unit).$ $\neg risk\_factors(QOI,X,Y) \leftarrow not\ risk\_factors(QOI,X,Y).$ $risk\_factors(0.143,2,no\_unit).$ $\neg ten\_year\_risk(QOI,X,Y) \leftarrow not\ ten\_year\_risk(QOI,X,Y).$ $ten\_year\_risk(0.500,21,\%).$	$QoI_{chd}= 1.000$ $QoI_{risk\_factors}= 0.143$ $QoI_{ten\_year\_risk}=0.500$	$V_{ldl\_goal < 100}=1*0.500=0.5$
5 $\neg chd(QOI,X,Y) \leftarrow not\ chd(QOI,X,Y).$ $chd(1,no,no\_unit).$ $\neg risk\_factors(QOI,X,Y) \leftarrow not\ risk\_factors(QOI,X,Y).$ $risk\_factors(0.143,3,no\_unit).$ $\neg ten\_year\_risk(QOI,X,Y) \leftarrow not\ ten\_year\_risk(QOI,X,Y).$ $ten\_year\_risk(0.500,19,\%).$	$QoI_{chd}= 1.000$ $QoI_{risk\_factors}= 0.143$ $QoI_{ten\_year\_risk}=0.500$	$V_{ldl\_goal < 130}=0.33*1+0.33*0.143+0.33*0,500$ $= 0.542$
6 $\neg chd(QOI,X,Y) \leftarrow not\ chd(QOI,X,Y).$ $chd(1,no,no\_unit).$ $\neg risk\_factors(QOI,X,Y) \leftarrow not\ risk\_factors(QOI,X,Y).$ $risk\_factors(0.143,3,no\_unit).$ $\neg ten\_year\_risk(QOI,X,Y) \leftarrow not\ ten\_year\_risk(QOI,X,Y).$ $ten\_year\_risk(0.500,21,\%).$	$QoI_{chd}= 1.000$ $QoI_{risk\_factors}= 0.143$ $QoI_{ten\_year\_risk}=0.500$	$V_{ldl\_goal < 100}=1*0.5=0.5$
7 $\neg chd(QOI,X,Y) \leftarrow not\ chd(QOI,X,Y).$ $chd(1,no,no\_unit).$ $\neg risk\_factors(QOI,X,Y) \leftarrow not\ risk\_factors(QOI,X,Y).$ $risk\_factors(0.072,1,no\_unit).$ $risk\_factors(0.072,2,no\_unit).$ $\neg ten\_year\_risk(QOI,X,Y) \leftarrow not\ ten\_year\_risk(QOI,X,Y).$ $ten\_year\_risk(0.500,21,\%).$	$QoI_{chd}= 1.000$ $QoI_{risk\_factors}= 0.143/2= 0.072$ $QoI_{ten\_year\_risk}=0.500$	$V_{ldl\_goal < 130}=0.33*1+0.33*0.072+0.33*0,500$ $= 0.519$ $V_{ldl\_goal < 160}= 0.5*1.00+0.5*0.072= =0.536$
From Program 2		
1 $\neg chd(QOI,X,Y) \leftarrow not\ chd(QOI,X,Y).$ $chd(0,unknown,no\_unit).$ $\neg risk\_factors(QOI,X,Y) \leftarrow not\ risk\_factors(QOI,X,Y)$ $risk\_factors(1,1,no\_unit).$ $\neg ten\_year\_risk(QOI,X,Y) \leftarrow not\ ten\_year\_risk((QOI,X,Y)).$ $ten\_year\_risk(1,18,\%).$	$QoI_{chd}= 0$ $QoI_{risk\_factors}=1$ $QoI_{ten\_year\_risk}= 1$	$V_{ldl\_goal < 100}= 0$ $V_{ldl\_goal < 160}= 0.5*0+0.5*1=0.5$

Table. 1. Description of the information scenarios generated by Programs 1 and 2.
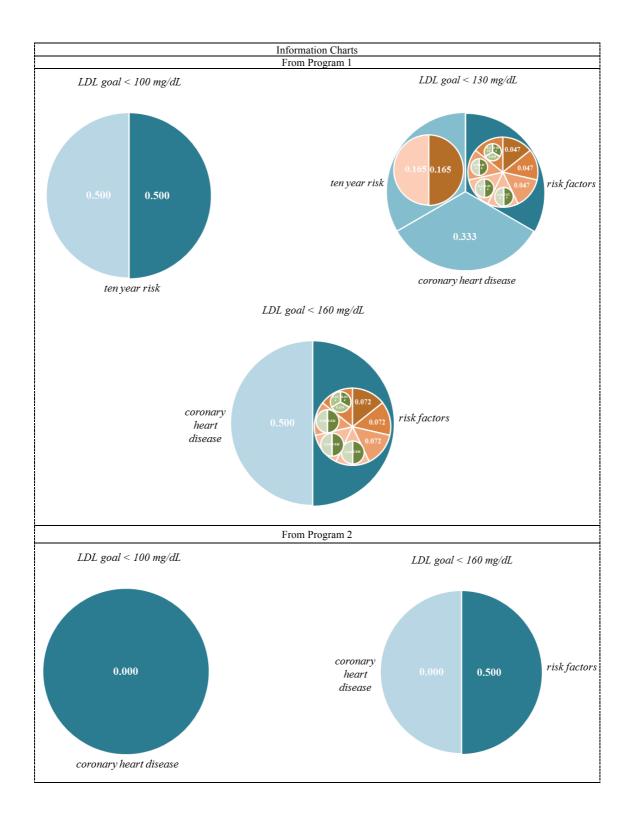
Table. 2. Information charts about the inferable options of each scenario from Programs 1 and 2, with the partial scores for the clinical parameters.

In Equation 1, *N* represents the cardinality of the set of terms and clauses of the extension of predicate *i*, so basically the number the number of abducible values is so great that *N* tends to infinite. For cases where the extension of i is unknown but can be taken from a set of values, the QoI is given by Equation 2.
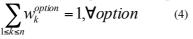
$$QoI_i = 1/Card \qquad (2)$$

The *Card* variable in Equation 2 denotes the cardinality of the set of abducibles for *i*. This equation is used only if the set of abducibles is disjoint. If the set of abducibles is not disjoint, then the QoI is calculated with Equation 3.

$$QoI_i = 1/(C_1^{Card} + ... + C_{Card}^{Card}) = $$
$$1/(\sum_{j=0}^{Card} C_j^{Card} - 1) = 1/(2^{Card} - 1) \qquad (3)$$

The denominator of Equation 3 corresponds to the sum of the possible combinations of the values of the set of abducibles with size *Card*. When transitioning from the original programs to the scenarios, the predicates acquire the following form:

   predicate(QoI, X, Y).

As it was seen in section 3, the selection of the options at a decision task determines the selection of the following task in the clinical process. The choice is determined by rules contained in the instance of the decision task and these rules comprise conditions relative to the parameters that are part of the patient's state. The possible information scenarios are depicted in Table 1 with the QoI of each clinical parameter (represented as a predicate) thus enabling the application of the rules to these scenarios. The next element of the model is a scoring system that assigns to each option a score (V_{option}) based on the QoI of the clinical parameters that act as conditions for its validation. As so, it is necessary to assert the relative importance (w_k^{option}) of clinical parameter *k* in *option*. Equation 4 illustrates this relationship, by stating that the sum of the values of relative importance of the *n* clinical parameters that condition an option must be equal to 1.

$$\sum_{1\leq k\leq n} w_k^{option} = 1, \forall option \qquad (4)$$

It is now possible to define the scoring function for each option according to Equation 5. The option with the highest score is the one to be selected in each decision task. The number of scenarios generated by Program 1 should be 14, but Table 1 only depicts 7. This is due to the low values of QoI and thus lower scores of the remaining scenarios, generated from the combination of the values for the risk factors.

$$V_{option} = \sum_{1\leq k\leq n} w_k^{option} \times QoI_k \qquad (5)$$

In the case of patient James, the appropriate scenarios would be scenario 1 and 4, since they are the scenarios with the option that has the highest score ($V_{option}=0,573$) and the selected option is an LDL goal of less than 160 mg/dL. For patient Paul there is only one available scenario with two inferable options. The option with the highest score is, again, the one that establishes an LDL goal for the therapy of 160 mg/dL. The circular charts of Table 2 provide another perspective of the state of the information that could be helpful to a healthcare professional in order to assess the correct course of the clinical process.

6 The Decision Model

The computational model for the representation of CPGs provides a basis for the automatic processing (creation and execution) of medical recommendations and the integration of this model with a knowledge representation formalism of the clinical parameters of a patient's stat will provide a decision framework capable of reproducing the decision process of healthcare professionals. To do so, the CDSS must have an inference engine that mimics the decision stages of clinicians, extracting knowledge from the recommendations of CPGs and the database of the patient´s state.

The decision model proposed in this paper is depicted in Fig. 4, and it is comprised of four stages with different objectives. The first stage of the model is the *Formulation of Clinical Hypotheses* and it consists of a survey of the available options of a *Decision* task. In the case study presented previously in this paper, the *Decision* task has the following options for consideration in the form of LDL blood levels: 100 mg/dL, 130 mg/dL and 160 mg/dL. These levels are goals for the cholesterol lowering therapy.

As seen in Fig. 2, each option is associated with a set of rules. These rules are the focus of the *Voting* stage, which is divided in two sub-stages: *Evaluation of Conditions* and *Evaluation of the QoI*. In the Evaluation of Conditions a verification of the conditions of each rule is performed, i.e., the engine checks if the conditions hold true. A condition is considered to hold true if it is represented either as positive information or as an abducible. Once it is determined which conditions hold true and which options are inferable, in each scenario, the engine moves to the next stage, the *Evaluation of the QoI*. In this stage, the inference engine calculates the QoI for the clinical parameters and computes the scores for the options in each scenario. As a whole, the *Voting* stage is used for the assessment of the state of the available information and to provide confidence measures in the possible outcomes. At the end of this stage, one should obtain the contents displayed in Tables 1 and 2.

The next stage is the *Selection of the Clinical Option* in which the option that has the highest score is selected, thus being regarded as the most reliable one. Based on the data of Table 1, the scenarios selected for patient James are scenarios 1 and 4 which determine that the LDL goal for the therapy is 160 mg/dL. For patient Paul, the selected option is also 160 mg/dL.

Now, based on the options selected in *the Decision* task, the next task of the careflow is selected.

According to Fig. 2 the next task is Aggregation Module 6, for both individuals, because it is the one activated by the trigger condition stating that the LDL goal should be inferior to 160 mg/dL.

7 Conclusions

This paper introduces a new way to represent medical recommendations extracted from CPGs and information that composes a patient's state, for use in CDSSs.

The representation model for CPGs mimics a model of human task execution to provide intuitive an intuitive of representing medical knowledge that is easily understood and, at the same time, has enough expressivity to represent all the types of information contained in these medical recommendations. Also, since the model was created using OWL, it is achieved the representation of guidelines in a portable format, that enables people to explore the collected knowledge with a wide variety of applications.

The introduction of ELP and the QoI in health information systems would be useful to represent cases of missing, conflicting and inexact information, thus helping the decision process by assuring the continuous flow of information. It offers mechanisms to reason with incomplete information and to measure the confidence in that information, with the capacity to distinguish different cases of uncertainty.
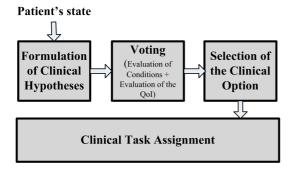
8 Acknowledgments

Patient's state

Fig. 4. Schematic representation of the clinical decision model with the its four stages in sequence: Formulation of Clinical Hypotheses, Voting, Selection of the Clinical Option and Clinical Task Assignment.

9 References

[BRENNAN, 2000]   BRENNAN, T. A., The Institute of Medicine report on medical errors-could it do harm? The New England Journal of Medicine 96(6)(2000) 1123–1125

[CHAWLA & GUNDERMAN, 2008]   CHAWLA, GUNDERMAN, R. B., Defensive medicine: prevalence, implications, and recommendations. Academic radiology 15(7) (2008) 948–9

[FOX et al., 1998]   Fox, J., Johns, N., Rahmanzadeh, A., Disseminating medical knowledge: the PROforma approach. Artificial Intelligence in Medicine 14(2) (1998)157-182

[KALRA, 2004]   KALRA, J., Medical errors: an introduction to concepts. Clinical biochemistry 37 (12) (2004) 1043–51

[NEVES, 1984]   NEVES, J., A Logic Interpreter to handle Time and Negation in Logic Databases, in: ACM '84 Proceedings of the 1984 annual conference of the ACM on The fifth generation challenge, New York, 1984, pp. 50-54

[NEVES et al., 2012]   NEVES, J., RIBEIRO, J., PEREIRA, P., et al., Evolutionary intelligence in asphalt pavement modeling and quality-of-information. Progress in Artificial Intelligence 1(1) (2012) 1–17

[NOVAIS et al., 2011]   NOVAIS, P., SALAZAR, M., RIBEIRO, J., ANALIDE, C., NEVES, J., Decision Making and Quality-of-Information, in Arroyo, A., Corchado, E., Tricio, V. (Eds), Soft Computing Models in Industrial and Environmental Applications, 5th International Workshop (SOCO 2010), 2010, pp. 187–195

[OLIVEIRA et al., 2011]   OLIVEIRA, T., COSTA, A., NEVES, J., NOVAIS, P., Digital Clinical Guidelines Modelling, in Omatu S, Paz Santana J. F., González S. R., Molina J. M., Bernardos A. M., Corchado Rodríguez J. M. (Eds.), Modelling and Simulation 2011, EUROSIS, ISBN: 978-9077381-66-3, 2011, pp. 392-398

[PATEL et al., 1998]   Patel, V. L., Allen, V. G., Arocha, J. F., Shortliffe, E. H., Representing Clinical Guidelines in GLIF. Journal of the American Medical Informatics Association 5(5) (1998) 467–483

[ROSENBRAND et al., 2008]   ROSENBRAND, K., VAN CROONENBORG, J., WITTENBERG, J., Guideline Development, in A. ten Teije, S. Miksch, P. Lucas (Eds.), Computer-based Medical Guidelines and Protocols: A Primer and Current Trends, 2008, pp. 3–22.

[SAMWALD et al., 2012]   SAMWALD, M., FEHRE, K., DE BRUIN, J., ADLASSNIG, K. P., The Arden Syntax standard for clinical decision support: Experiences and directions. Journal of Biomedical Informatics 45(4) (2012) 711-8

[SHAHAR et al.,1998]   Shahar, Y., Miksch, S., Johnson, P. , The Asgaard project: a task-specific framework for the application and critiquing of time-oriented clinical guidelines, Artificial intelligence in Medicine 14(2) (1998) 29–51

[STRASZECKA, 2006]   STRASZECKA, E., Combining uncertainty and imprecision in models of medical diagnosis. Information Sciences 176(20) (2006) 3026-3059

[TU et al., 2007]   TU, S. W., CAMPBELL J. R., GLASGOW, J., et al., The SAGE Guideline Model: achievements and overview. JAMIA 14(5) (2007) 589–98