Noname manuscript No.
(will be inserted by the editor)

ST-Audit: Guideline-based automatic auditing of electronic patient records

Cleo Zanella Billa · Jacques Wainer · Claudia Barsottini

Received: date / Accepted: date

Abstract This work presents the ST-Audit system that audits a patient record for conformance to a particular clinical guideline. The system uses ST-Guide which models a guideline as a set of states and transitions. The audit system tries to find a path in the state/transition diagram that corresponds to the actions taken by the physician, taking into consideration issues such as the unavailability of all the data needed to evaluate the transitions. The system was used to audit an outpatient clinic regarding their procedure for hypertension treatment (using the VI JNC guideline for hypertension), and the results of the number of non-compliant actions were presented and discussed. A follow up auditing showed a small but statistically significant reduction on the number of non-compliant actions for patients treated after the first audit.

Keywords Electronic patient record · automatic auditing · clinical guidelines

1 Background

The issue of the quality of health care, in general, and the quality of a physician's clinical decisions, in particular, are topics of interest for many stakeholders, including patients, of course. There is an increasing support for the implementation and adoption of electronic patient records (EPRs) fueled in part by the hope that these systems will

C. Billa
Institute of Computing, University of Campinas, Campinas, 13083-970, SP, Brazil
E-mail: cleo.billa@gmail.com

J. Wainer *Corresponding author*
Institute of Computing, University of Campinas, Campinas, 13083-970, SP, Brazil
Phone: +55 (19) 3521-5871
Fax: +55 (19) 3521-5847
E-mail: wainer@ic.unicamp.br

C. Barsottini
Department of Health Informatics, Federal University of Sao Paulo, Sao Paulo, 13083-970, SP, Brazil
E-mail: claudia.barsottini@unifesp.br

improve health care quality through faster and more accurate clinical data analysis [1–5].

Among the many methods for quality improvement, clinical auditing is usually considered a first step. Clinical audit [6] is defined as follows.

> Clinical audit is a quality improvement process that seeks to improve patient care and outcomes through systematic review of care against explicit criteria and the implementation of change. Aspects of the structure, processes, and outcomes of care are selected and systematically evaluated against explicit criteria. Where indicated, changes are implemented at an individual, team, or service level and further monitoring is used to confirm improvement in health care delivery.

This paper describes a system that automatically extracts from the EPR actions and decisions taken by the physician that, at a first level of analysis, are not in accordance with a clinical guideline. In this research, the clinical guidelines are represented in the ST-Guide language. Of course, those actions and decisions need to be further reviewed by both the physician responsible and the quality assurance team, but they are a complete set of events in which the physician's actions did not follow the prescribed actions, and thus, by using the system one can concentrate the audit in these events.

Section 1.1 briefly discusses automated clinical guidelines, section 1.2 describe previous research on automatic auditing of patient records, mostly using some clinical guideline as the correct course of actions, and section 1.3 describes the characteristics of the ST-Guide representation language for clinical guidelines. Section 2 describes the basic algorithm of ST-Audit, section 3 discusses 2 applications of the system, and section 4 discusses the contribution and limits of ST-Audit.

1.1 Automated clinical guidelines

Clinical guidelines are potential tools to standardize health care, to improve its quality and cost effectiveness. The general idea behind clinical guidelines is that groups of experts, based on solid scientific evidence, describe a set of effective and efficient procedures to treat or manage patients with a particular disease. The use of clinical guidelines improve the quality of health care [7, 8].

An automatic guideline is a computational system that models a clinical guideline. Usually, it is connected to an electronic patient record (EPR) and it helps the physician on his choices regarding tests and treatment options.

There is a large body of work regarding automatic clinical guidelines, such as Arden Syntax [9], PROforma [10], Asbru [11], GLIF [12, 13], etc. There has been some literature that compare and discuss these and other guideline formalisms [14–17].

1.2 Automatic audit tools

Computer support for clinical auditing falls into two main classes: tools that support the auditing process (also referred to semi-automated tools) and automatic tools. Auditmaker [18] is an example of the former system - it is a generic tool that takes the physicians through the audit process, from planning and data collection, to simple data

analysis. [19] describes a large set of programs that support the clinic audit process, most of them in the same line as Auditmaker. [20] describes both a methodology and a tool to set up and to follow through a clinical audit.

Automatic tools can be divided into two main classes. Older research, for example, HyperCritic [21,22] and its continuation AsthmaCritic [23–25] were systems specific to a particular disease. In these systems, the knowledge of the "correct" treatment were embedded into the system, as well as specific forms in which the real treatment could disagree with the correct one. The critiquing knowledge is actually a set of tasks that are triggered by an event discovered in the patient's record. For example, the event of starting a drug starts a set of verification tasks that collects all other medications and diagnostics, and access the medical knowledge to verify if there are interactions with the new drug, and, if so, warn the physician.

Later projects, similar to the one described herein, followed the boom in guideline research. Since guidelines are a representation of the "correct treatment", such systems would read in a particular guideline, and thus do not need to embed that knowledge into the program. In general, guideline based critiquing systems try to map the particular set or sequences of actions in the EPR to one of the possible treatments paths or plans specified by the guideline. A particular treatment is correct is if it considered as one of the "acceptable variations" of one of the specified paths or plans, and incorrect otherwise.

The definition above is unclear in what is the "mapping" process between the EPR recorded actions and the possible paths in the guidelines and what are the "acceptable" variations. Different systems embed different heuristics to implement these concepts.

A common problem of these systems is the terminological gap between the actions recorded in the EPR and the actions specified in the guideline: terms in the EPR not always are the same as the ones in the guideline. A guideline may specify that a diuretic should be prescribed to the patient, but the EPR states that brand drug Higroton with dosage of 25mg/day was prescribed. Thus, the guideline may refer to diuretics, whereas the EPR refer to particular medicine brands and dosages.

Another important variation among guideline representation languages and therefore the guideline-based critiquing systems is what is considered acceptable variations in these guidelines. For example Asbru, as a representation language, assumes that the person following the guideline must agree with the particular intentions or goals but there is an acceptable variation regarding the ways to achieve that goal. To use an example from [11], if the physician prescribed to a insulin-dependent diabetic patient with an episode of hyperglycemia around bedtime to reduce the carbohydrates intake at dinner, that is considered an "acceptable variation" to the guideline prescribed action of increasing the insulin dosage before dinner, because the two actions achieve the same goal of decreasing the blood-glucose level at night.

An important part of what is considered acceptable variation between the actions taken and the actions prescribed by the guideline are defined by the guideline language itself. In the example above, in order to consider the actions taken as an acceptable variation, the guideline must represent the goals or intentions of each action or set of actions. In particular Asbru is the guideline representation language that more clearly expresses intentions, and thus all Asbru based critiquing we are aware of assume the wide definition of acceptable variation as agreement regarding intentions as described above. Another component of the acceptable variability are the presuppositions the different systems make on how data is represented, what data may be missing, and what extraneous data may be present in the EPR. A common assumption is that the

EPR will contain other actions that are not related to the ones in the guideline because the patient may be treating or diagnosing other condition besides the condition specific to the guideline. Thus, usually the presence of these extraneous actions are among the acceptable variability.

We are aware of three other main lines of research on guideline-based critiquing (or auditing) systems, besides the one described in this paper: Asbru-based systems [11, 26–28], model-checking systems [29,30], and case-based reasoning [31].

The central aspect of Asbru based systems is to take into consideration the intentions of the physician when evaluating if the actions were "correct". Of course, the physician intentions are not explicit in the EPR, and thus such systems must perform a plan-recognition in order to verify if the actions taken can be seen as a plan that achieve the intentions explicitly stated in the guideline. In practice however, plan recognition as implemented in the systems described in [11,26–28] are not yet sophisticated enough to recognize the insulin-dependent patient case described above.

In [28] plan recognition is really just verifying it the actions taken are one of the alternative plan executions explicitly represented in the guideline. For that system, order is also not enforced, that is, actions in different order than the one stated in the guideline are accepted as correct.

Model-checking based systems make use of recent advances in symbolic model checking technology [32,33] which can verify complex properties (expressed as formulas in temporal logic) of finite state machines. There are only a few, proof of concepts systems in this category.

In [29] the guideline is represented as a finite state machine, manually constructed. The actions in the EPR were manually translated to the terms used in the guideline. Let us assume that the translated actions are represented by the sequence $\langle a_1, a_2, a_3, a_4 \rangle$. The critiquing is done by verifying (using a symbolic model checker) if there is a path of execution of the form $\langle a_1, \ldots, a_2, \ldots a_3 \ldots a_4 \rangle$. That is, the acceptable variations include not executing some of the actions prescribed in the guideline. If there is no path of the form above in the guideline, the system would try to find paths in which the orders of a_1, a_2, a_3, and a_4 are shuffled. Thus, also among the acceptable variations are changes in the order in which the actions were executed.

Finally, the case-based reasoning system [31] was not developed as a critiquing system as proposed here, but as a system that once detected a deviation from the guideline would find similar deviations stored in the case data base and show them to the physician as a suggestion for a plan of action. Deviation in that research is defined as not following exactly the order in which the actions were specified in the guideline, thus there is no "acceptable variation" between the patient record and the guideline.

1.3 ST-Guide guideline representation language

[34,35] propose the **ST-Guide** model, a formalism similar to Prodigy [36], which targets the representation of primary and secondary care guidelines. These guidelines assumes that the patient periodically visits the health care worker who evaluates the patient's condition and performs one or more of the following actions: ask for new laboratory or image exams, prescribe new drugs or maintain the previous drug therapy, refer the patient to a specialist, propose changes in the patient's life-style, schedule a new appointment, or release the patient.

In ST-Guide, a guideline is defined as a state diagram. The states represent stages in the patient treatment/management and the transitions among states are evaluated based on physical examination data, results from laboratory exams, and answers to the questions that must be asked to the patient. If a transition is true, the guideline changes to a different state, which represent a different stage in the patient's treatment, for example, the use of a different medicine or a different combination of drugs. In ST-Guide, the state diagram must be *unambiguous*, that is at most one transition will be true at a time, but it does not need to be *compete*, that is, for any combination of conditions, some transition will be true. ST-Guide assumes that if no transition is correct, the patient must remain in the same state.

Figure 1 shows a state diagram example, where the ellipses represent the states, and the arrows represent the transitions.

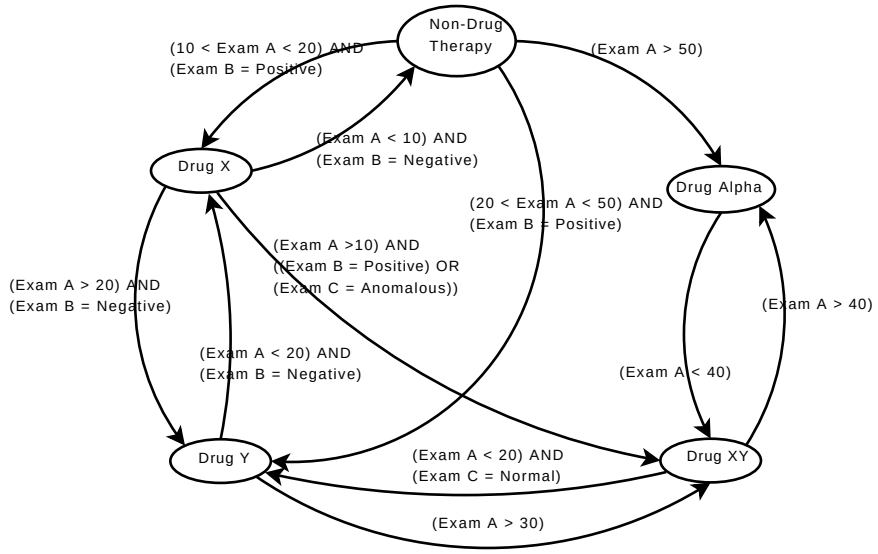

Fig. 1 State diagram of an hypothetical guideline

In the ST-Guide model, a state has a set of attributes that describes all the actions and decisions that must be made during the patient's consultation to the physician. Each state has the following attributes:

Physical Exams: defines which physical exams must be performed in that state.
Laboratory Exams: defines which laboratory exams must be performed in the state.
Questions: describes the questions that must be asked to the patient.
Transitions: defines what are the conditions for a patient to move to another state. The conditions are based on the results of physical exams, results of laboratory exams and answers to questions.
Medications: defines the medications that should be prescribed to the patient.
Suggestions: describes information that are to be presented to the patient. Includes, for example, suggestions of life-style modifications or may refer the patient to an expert.

Next Appointment: defines the periodicity of the next appointments.

In ST-Guide, a patient consultation with the health care provider is modeled as follows:

1. the system retrieves the patient's previous state, or if it is a new patient, it defines that the patient is in the guideline entry state.
2. the system retrieves the lab exams requested in the previous consultation and retrieves the data regarding these lab exams (either by accessing the data in the patient's EPR, or by asking the physician to enter the data). The system then retrieves the physical exams and questions associated to the state and gather or request the corresponding data. For ST-Guide the difference between physical examination and questions are that while physical examination are data that may be available in the patient's record, questions are never retrieved from the medical records - they must be asked by the physician at this time. Example questions asks for the presence of side effects from the medication, investigate if the patient are really taking the prescribed medication, ask if the patient is pregnant, or has some other relevant condition.
3. with the lab exams, the physical exams, and the question data, the ST-Guide evaluate all possible transitions for the state. If none is true, the system remains in the same state. In a well designed guideline, at most one transition will be satisfied and the system makes the change to that next state.
4. in the post-transition state (which may be the original or a new state depending on whether no or one transition was satisfied) the system computes the appropriate suggestions and/or drug treatment. The system also computes the lab exams needed in this new state, and the next appointment.

In summary, the system has two different behaviors: the pre-transition phase (steps (1) and (2) above) is responsible for gathering the appropriate data, and post-transition phase (step (4) above) is responsible for the proposing actions (exam requests, suggestions, drug prescription, next appointment) to be taken by the physician towards the patient.

2 ST-Audit

2.1 Design Objectives

The main goal of the auditor mode in ST-Guide, or ST-Audit, is to compare the actions performed by the physician with the actions suggested in the guideline represented in ST-Guide. Since ST-Guide is based on consultations, the ST-Audit will try to determine a sequence of states in the guideline that best explain the sequence of actions taken in the various consultations.

ST-Auditor makes the following assumptions regarding the EPR:

- the EPR contains the data regarding drugs prescribed, and lab exams requested for each consultation. Thus, there is no missing data regarding these actions and the consultations in which they were taken is explicit.
- the EPR may contain extraneous actions (prescriptions and exam request) because the patient may be under treatment/diagnosis for other conditions besides the one related to the guideline.

– the EPR may or may not contain all of the data regarding physical examinations and questions.

To deal with the terminological gap between the guideline and the EPR we manually constructed ontology that maps particular brand drugs to their generic names and then to the major classes of drugs referred to by the guideline. For example the ontology maps the drug Higroton to chlorthalidone to a diuretic. The ontology also has information on the dosages and in particular the maximum dosage of each generic drug.

Section 2.2 will describe the basic mechanism by which ST-Auditor matches a sequence of consultations with states in an ST-Guide guideline, and how, given the match, it determines that exams (and medications) were missing or were prescribed but were unnecessary (from the point of view of the guideline). Section 2.3 describes how the system proceeds when it does not find a state to match a particular consultation, and section 2.4 explains how the system deals with incomplete information. All these processes are described in the full ST-Auditor algorithm in Algorithm 1.

ST-Auditor is not an interactive system that is supposed to be used by the physician as he/she is performing the actions. In other words, ST-Audit is not a suggestion system. We developed the system to be used by review or audit boards in batch mode. The difference is critical. A interactive system has to be much more careful in what it points out as a deviation from the guideline, since such deviations would cause some alert or the generation of some suggestion to the physician. Incorrect alerts would not only be disrupting but would also corrupt the physician confidence in the system. On the other hand, a review board can simply discard from the report all actions it believe can be explained away.

2.2 A simple example

For the moment let us assume that the EPR is complete, that is, it has all information about the clinical history of the patient. For example, let us consider that figure 1 represents the state diagram of an hypothetical guideline. The states are represented as ellipses. This simple guideline has no lab exams or questions, and only three physical exams named A, B, and C, which are preformed in the pre-transition phase of a consultation. The only post-transition action is to prescribe the drugs X, Y or Alpha. Let us also assume that the frequency of consultations is irrelevant. Thus each state is labeled with the drug that is prescribed when the patient is in the state. Below are the physical examinations that must be performed in each state:

– `Non-drug therapy`: Physical exams A and B
– `Drug X`: Physical exams A, B, and C
– `Drug Y`: Physical exams A and B
– `Drug XY`: Physical exams A and C
– `Drug Alpha`: Physical exam A.

Non-drug therapy is the initial state.

Let us now consider the following sequence of consultations:

– First consultation: Physician performed physical exam A (measured as 8) and exam B (resulted negative). The physician did not prescribe any medication.

Algorithm 1

{Preprocessing to eliminate all questions from the conditions}
$NewTransition \leftarrow \emptyset$
for all $\langle InitialState, FinalState, Conditions\rangle \in Transitions$ **do**
  $Conditions' \leftarrow Conditions$ with all questions substituted by TRUE
  $NewTransition \leftarrow NewTransition \cup \{\langle InitialState, FinalState, Conditions'\rangle\}$
end for

{The main algorithm}
$States \leftarrow$ initial states of the guideline.
$v \leftarrow$ first consultation
repeat
  $Meds(v) \leftarrow$ medications prescribed in consultation v
  $Exams(v) \leftarrow$ physical exams performed in consultation v
  $Labs(v) \leftarrow$ labs exams requested in consultation v
  $ExamsMustDo \leftarrow$ physical exams that are in all $S \in States$
  $ExamsCanDo \leftarrow$ physical exams that are in some $S \in States$
  $ExamsMissing \leftarrow ExamsMustDo - Exams(v)$
  $ExamsUnnecessary \leftarrow Exams(v) - ExamsCanDo$
  $NewStates \leftarrow \emptyset$
  for all $\langle InitialState, FinalState, Conditions\rangle \in Transitions'$ **do**
    if ($InitialState \in States$) AND ($Conditions$ is TRUE) **then**
      $NewStates \leftarrow NewStates \bigcup FinalState$
    end if
  end for
  if $NewStates = \emptyset$ **then**
    Show "Warning!! - Treatment is NOT in agreement with the guideline."
    $NewState \leftarrow$ Re-synchronize using physical exams, laboratory exams, questions and medication.
  end if
  $States \leftarrow NewStates$
  $LabsMustDo \leftarrow$ lab that are in all $S \in States$
  $LabsCanDo \leftarrow$ labs that are in some $S \in States$
  $LabsMissing \leftarrow LabsMustDo - Labs(v)$
  $LabsUnnecessary \leftarrow Labs(v) - LabsCanDo$
  $MedicationDifference \leftarrow$ Difference between the medication prescribed and the one indicated in the guideline.
  Show $MedicationDifference$
  Show $ExamsUnnecessary$
  Show $LabsUnnecessary$
  Show $ExamsMissing$
  Show $LabsMissing$
  $v \leftarrow$ next consultation
until $v \neq NULL$

- Second consultation: Physician performed exam A (= 12), exam B (resulted positive), exam F (resulted positive) and prescribed drug X.
- Third consultation: exam A (= 23), exam B (negative), performed. The physician changed the medication to drug Y.
- Fourth consultation: exam A (= 18), and exam B (negative) performed. The physician prescribed drug X.
- Fifth consultation: exam A (= 19), exam B (positive), and exam C (normal) performed. The physician prescribed drugs X and Y.

In the example above, the auditor will find a path in the state diagram that matches the actions performed by the physician. The patient starts in the initial state `Non-Drug Therapy`, and in the first consultation he remains in that state. In the second consul-

tation, the patient starts at `Non-Drug Therapy` (the previous state), but since exam A has increased, and exam B is positive, the patient should move to state `Drug X`. The physician actions corresponded to the ones specified at state `Drug X` so the system assumes that the transition from `Non-drug Therapy` to `Drug X` happened during the second consultation. Since exam F was performed and it is not one of the required exams in state `Non-drug Therapy`, which was the state at the beginning of the consultation, the system will generate a warning indicating that this exam was not necessary at this point of the treatment.

On the third consultation, the patient starts at state `Drug X`, and the exams A and B are preformed as required, but not exam C. The system warns of a missing exam. The conditions for the transition to state `Drug Y` are true, and the prescription action corresponds to the one required at that state. Thus, at the end of the third consultation, the patient is in state `Drug Y`.

The exams performed at the fourth consultation are in agreement with that state, and the transition was to the state `Drug X`. Finally the fifth consultation marks a transition to the state `Drug XY`. Figure 2 shows the path of the consultations in the state diagram.

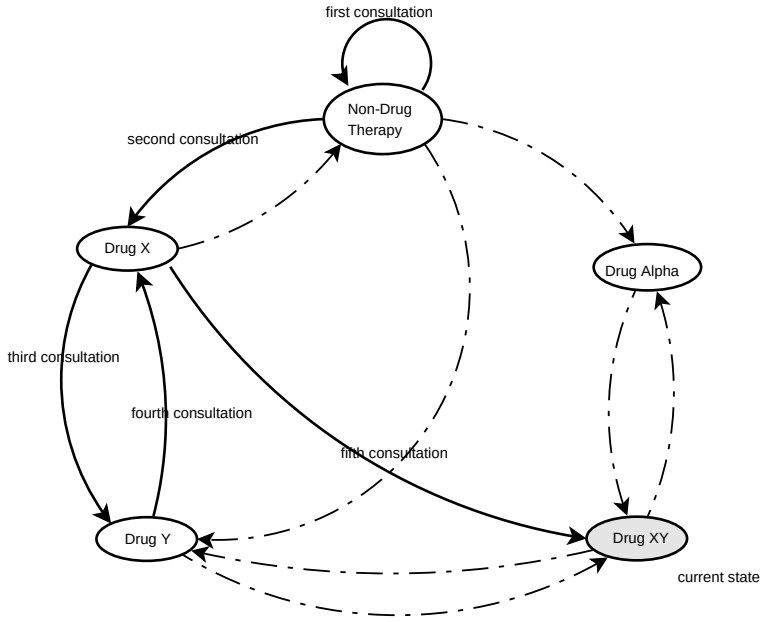

Fig. 2 Example of a path found by the auditor

2.3 Re-synchronization

Let us now address an example in which it is not possible to find a path in the state diagram that matches the actions taken by the physician. In this case, the auditor

generates a warning, saying that the treatment does not match the treatment described in the guideline, and makes a re-synchronization. The auditor will look up for a state that best matches the current actions performed by the physician. To perform the re-synchronization it will consider the medications prescribed and the laboratory exams ordered. Let us suppose another example with the following sequence of consultations:

- First consultation: exam A is measured as 12, exam B is positive, and the physician prescribes drug X.
- Second Consultation: exam A is 19, exam B is positive, and exam C is anomalous. The physician changes the medication to XY.
- Third Consultation: exam A is 8, exame C is normal, and the physician removes the medication.
- Fourth Consultation: exam A is 22 and exam B is positive. The physician prescribes Y.

In this case, in the first consultation, the patient begins in `Non-Drug Therapy`, and go to the state `Drug X`. At the second consultation, from `Drug X` to `Drug XY`. At the third consultation, exam A was less than 20, but the patient did not change to state `Drug Y`, because that drug was not prescribed. Thus the system alerts that an invalid transition was taken, and tries to re-synchronize. In this case the only action that can be used to discover to which state the patient moved is the drug prescription. Only the `Non-drug Therapy` state allows for not prescribing any medication and thus the system resynchronizes to it. In the fourth consultation, the system starts the patient at the `Non-drug Therapy` state and transitions normally to the `Drug Y` state. Figure 3 shows the path in the state diagram corresponding to this example.

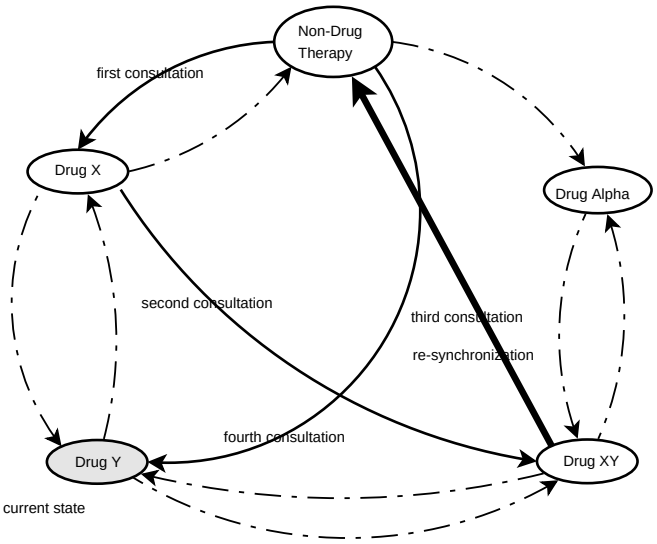

Fig. 3 Example of the auditor re-synchronization

The method for choosing the state to which the system should resynchronize is heuristic. The system looks at the actions taken in the consultation where the synchronization was necessary. The post-transition actions are to prescribe drugs, to request

laboratory exams, and to schedule the next consultation. Only the first two are used to determine the best state to resynchronize to. For each state s the system calculates the mismatch between the actions taken by the physician and the required actions in state s as:

$$mismatch(s) = w_d \frac{|\text{drugs}(s) \,\Delta\, \text{drugs\_prescribed}|}{|\text{drugs}(s) \cup \text{drugs\_prescribed}|} + w_l \frac{|\text{labs}(s) \,\Delta\, \text{labs\_requested}|}{|\text{labs}(s) \cup \text{labs\_requested}|}$$

where $A \Delta B$ is the symmetric difference between the sets A and B, that is $A \Delta B \equiv (A \cup B) - (A \cap B)$. w_d and w_l are coefficients that balance the relative importance of the drugs prescribed and the lab exams requested in calculating the mismatch.

The mismatch of a state s will be zero if the set of drugs prescribed is the same as the ones required in s and the laboratory exams ordered also fully agree with the ones required in s. The mismatch increases if there are more or less drugs prescribed than required, or more or less lab exams ordered than required in the state. The auditor will select the state with lowest mismatch to re-synchronize.

Other formulas for calculating the mismatch are possible, for example one that weights differently the presence or absence of different drugs or lab exams. The discrepancies regarding the time for the next consultation can also be taken into consideration.

2.4 Incomplete data

The two examples above represent situations where all information about the patient is available in the record. But, in practice the patient record does not have all information necessary. In fact, our definition of questions is exactly the data that is not likely to be available in the records. To solve this problem, when a piece of data is not found in the record, the auditor consider all possible values to the data, and thus manage all possible states to which the state diagram could transition. Thus, when dealing with incomplete data, ST-Audit really keeps a set as the possible states of the patient.

To deal with all possible values of the missing data, ST-Audit simply removes from all conditions, all references to that data. Thus if one transition would be true if (exam B > 20 AND side-effects = false), and other will be true if (exam B > 20 AND side-effects = true), and if the question of whether a drug has caused side effects on the patient is unlikely to be documented in the EPR, ST-Audit would remove the side-effect clause from the conditions, resulting that both transitions would be true if exam B > 20. Thus, to deal with incomplete data, ST-Audit may transform an unambiguous ST-Guide guideline into an ambiguous one.

To see how it works, let us consider a guideline similar to Figure 1, but let us assume that the state `Drug XY` recommends the prescription of drug Y alone, and let us rename that state to `Drug Y'`. Figure 4 shows the new state diagram, with all conditions that refer to exam B removed.

Let us assume the following data

- First Consultation: exam A is 12 and the physician prescribes X.
- Second Consultation: exam A is 23, C is normal, and the physician prescribes Y.
- Third Consultation: exam A is 25, exam C is Normal The physician maintains the drug Y.

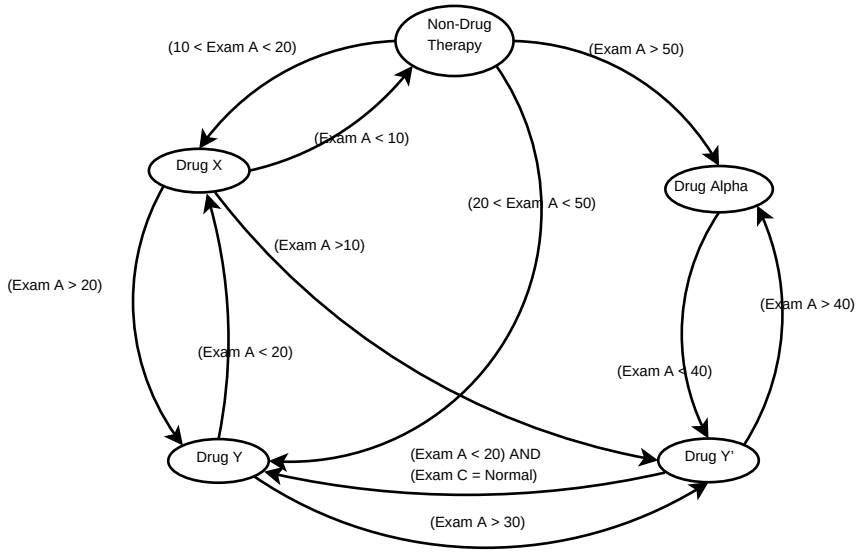

Fig. 4 A variation of the guideline

– Fourth Consultation: exam A is 10, and the physician prescribes X.

At the first consultation, the auditor assumes that at the end of the consultation the patient is in state `Drug X`. In the second consultation, both the states `Drug Y` and `Drug Y'` are possible as the resulting state, and consistent with the prescription. Thus the system will assume both states as initial state of the third consultation. If the patient is in state `Drug Y`, then because of the value of exam A, the patient must remain in the state, and exam C was an unnecessary exam. If the patient is in state `Drug Y'`, then because of the value of exam A, the patient must also remain in state `Drug Y'`, but in this case, exam C was not unnecessary. Both `Drug Y` and `Drug Y'` are consistent with the prescription of drug Y. Thus, at the end of the third consultation, the system still has two states as the possible initial state of the fourth consultation.

In the case where the system tracks multiple states, determining which exams were missing and which unnecessary exams were performed is not straightforward. The auditor will assume a forgiving point of view: if an exam was required in any of the possible states for the consultation, then if performed, it will not be considered an unnecessary exam, but it will not force the system to assume that the patient is in one of the states that require the exam. Thus exam C will not be flagged as unnecessary in the analysis of the third consultation above.

Similarly, only exams necessary in all states are considered necessary. If labs(s) is the set of laboratory exams required for state s, labs_requested is the set of lab exams requested, and S is the set of possible states for the consultation, then a laboratory exam $l \in$ labs_requested will be considered unnecessary if $l \notin \bigcup labs(s)$ for all $s \in S$. And a required laboratory exam m will be considered missing if $m \notin$ labs_requested and $m \in \bigcap labs(s)$ for all $s \in S$. Similarly for drugs prescribed.

For the fourth consultation, if the patient started in state `Drug Y`, he must transition to state `Drug X`, which is consistent with the prescription, and exam C was unnecessary. But if the patient started in state `Drug Y'`, the value of exam A and the lack of value

for exam C will force the system to remain in the state `Drug Y'`, which is not consistent with the prescription. If this was the single path being considered, that would generate the warning and the re-synchronization (section 2.3). But in this case, where multiple paths are being considered, the system will stop processing this path at all. Thus, at the end of the fourth consultation, the system is in state `Drug X`.

Notice that if we assume that the physician did follow the guideline as much as possible (which is the forgiving point of view discussed above), we can conclude that the patient was never in state `Drug Y'`. Another conclusion is that the exam C performed in the third consultation was indeed unnecessary, since in fact the patient was in the `Drug Y` state at the beginning of that consultation. The system will `not` perform such reasoning and section 4 discusses why this is an important aspect of ST-Audit.

3 Case Studies

The system as described above was implemented in Prolog. The basic architecture of the system is shown in figure 5. The small rectangles are components of the system. The `Converter` converts the external representation of ST-Guide guidelines from XML to an internal representation. The `low level EPR` interface was manually constructed and embodies all specificities of the particular EPR system, or data base. It returns a data structure which represents the sequence of consultations, and the physical exams, lab exams, and medications for each consultations. This data structure is further transformed by the ontology, which converts medication brand or generic names into the classes of medications referred to by the guideline. ST-audit receive both the internal guideline representation and the sequence of consultations representation and outputs a report.

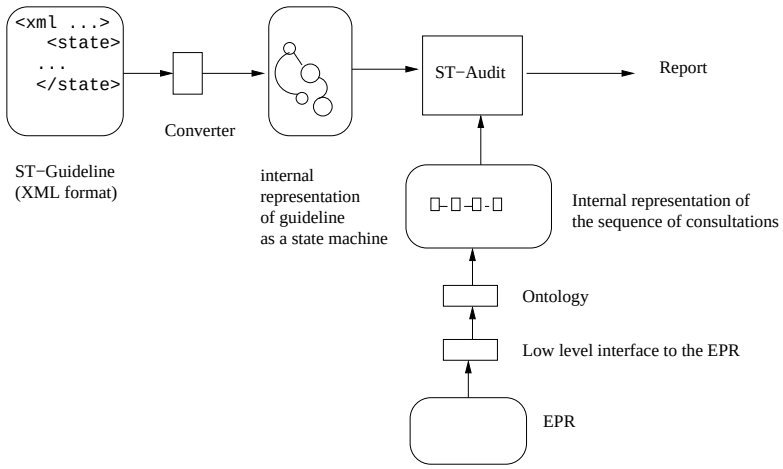

Fig. 5 The architecture of the ST-Auditor system

To test the system we used the Sixth Joint National Committee Guideline for Hypertension (VI JNC) [37], and we audited all hypertension patients in a free outpatient

clinic at UNIFESP (Universidade Federal de Sao Paulo). The free clinic uses the EPR system "Clinic Manager" developed at UNIFESP.

The JNC VI guideline was implemented in ST-Guide, resulting in the state diagram displayed in figure 6. The model is complex but in general, each state corresponds to a particular combination of drug therapies taken from: diuretics, alpha and beta blockers, ACE inhibitors, calcium antagonists, and angiotensin II receptor blockers. Figure 7 is a magnification of the most relevant part of the full diagram. The state `drugTherapy` is the initial state, and depending on different conditions the patient would start a different, single drug therapy. The conditions for the transitions in figure 7 are shown in table 1.

	Initial State	**Final State**	**Conditions**
C1	drugTherapy	diuretic	sbp $\geq$ 140 AND dbp $\geq$ 100.
C2	drugTherapy	beta	sbp $\geq$ 140 AND dbp $\geq$ 100.
C3	drugTherapy	diureticAce	sbp $\geq$ 140 AND dbp $\geq$ 100 AND heart_failure=yes.
C4	drugTherapy	betaCalcium	sbp $\geq$ 140 AND dbp $\geq$ 100 AND angina=yes.
C5	drugTherapy	betaAce	sbp $\geq$ 140 AND dbp $\geq$ 100 AND myocardial_infarction=yes.
C6	drugTherapy	calcium	sbp $\geq$ 140 AND dbp $\geq$ 100.
C7	drugTherapy	ace	sbp $\geq$ 140 AND dbp $\geq$ 100.
C8	diureticAce	diureticAngioII	sbp $\geq$ 140 AND dbp $\geq$ 100 AND side_effects=yes.
C9	diureticAce	diureticAceAngioII	sbp $\geq$ 140 AND dbp $\geq$ 100 AND side_effects=no AND dosage=max.
C10	calcium	calciumAce	sbp $\geq$ 140 AND dbp $\geq$ 100 AND side_effects=no AND dosage=max AND renovascular_disease=no AND pregnancy=no.
C11	calciumAce	calciumAceAngioII	sbp $\geq$ 140 AND dbp $\geq$ 100 AND side_effects=no AND dosage=max.
C12	calciumAce	calciumAngioIIumber	sbp $\geq$ 140 AND dbp $\geq$ 100 AND side_effects=yes.

Table 1 Conditions of the Hypertension Guideline State Diagram

The JNC guideline was changed slightly to adapt it to the realities of the outpatient clinic current practices. The most important adaptation was that there was no need for a second consultation in order to confirm the hypertension diagnostic. The clinic patients usually miss their work and travel some distance for their appointments, and thus the current practice is to accept the diagnostic of hypertension after a single consultation, and to anticipate the treatment (which may or not be using medication).

We used $w_l = 0$ for the mismatch formula, that is, to find the state to re-synchronize we use only the medication data, since medication uniquely define the states in the hypertension model used (figure 6). The auditor also takes into consideration the dosage of the drugs, according to the JNC guideline.

3.1 Auditing a single case

Below is the output of the auditor when auditing a single patient from the outpatient clinic. In the output, `sbp` abbreviates systolic blood pressure and `dbp` stands for diastolic blood pressure.

```
------------------------------------------------------------
  Day: 37265
  Exams:
    -> sbp  : 170
    -> dbp  : 102
  Medication:
    -> none
  Initial patient state:
    -> Drug Therapy
  Current patient state::
    -> Drug Therapy

*** OK!! -  Treatment is in agreement with the guideline.

  Possible unnecessary exams:
     -> none
  Possible missing exams:
     -> none
------------------------------------------------------------
   Day: 37293
    -> sbp: 174
    -> dbp: 114
  Medication:
    -> Ace Inhibitors: Capoten 25mg(2) => Dosage: 50mg/day
    -> Diuretics:      Clorana(1) => Dosage: 1mg/day
  Initial patient state:
    -> Drug Therapy
  Current patient state::
    -> DiureticAce

*** Warning!! -  Treatment is NOT in agreement
                 with the guideline.

  Possible unnecessary exams:
     -> none
  Possible missing exams:
     -> none
------------------------------------------------------------
  Day: 37372
  Exams:
    -> sbp: 140
    -> dbp: 100
  Medication:
    -> Ace Inhibitors: Capoten 25mg(2) => Dosage: 50mg/day
    -> Diuretics:      Higroton 25mg(1) => Dosage: 25mg/day
  Initial patient state:
    -> DiureticAce
  Current patient state::
    -> DiureticAce

*** OK!! -  Treatment is in agreement with the guideline.

  Possible unnecessary exams:
     -> Urinary Ultrasonography
     -> Urine
     -> Creatine
  Possible missing exams:
     -> none
------------------------------------------------------------
  Day: 37428
  Exams:
    -> sbp: 146
    -> dbp: 104
  Medication:
    -> Ace Inhibitors: Capoten 25mg(2) => Dosage: 50mg/day
    -> Diuretics:      Higroton 25mg(1) => Dosage: 25mg/day
    -> Beta-Blockers:  Inderal 40mg(2) => Dosage: 80mg/day
  Initial patient state:
    -> DiureticAce
  Current patient state::
    -> DiureticsBetaAce

*** Warning!! - Dosage of previous medication should
                have been increased.

  Possible unnecessary exams:
     -> none
  Possible missing exams:
     -> none
------------------------------------------------------------
  Day: 37428
  Exams:
    -> sbp: 128
```

```
    -> dbp: 90
  Medication:
    -> Ace Inhibitors: Capoten 25mg(2) => Dosage: 50mg/day
    -> Diuretics:      Higroton 25mg(1) => Dosage: 25mg/day
    -> Beta-Blockers:  Inderal 40mg(2) => Dosage: 80mg/day
  Initial patient state:
    -> DiureticsBetaAce
  Current patient state::
    -> DiureticsBetaAce

*** OK!! -  Treatment is in agreement with the guideline.

  Possible unnecessary exams:
     -> none
  Possible missing exams:
     -> none
------------------------------------------------------------
```

In the first consultation, the patient was diagnosticated with hypertension but the physician decides to not start the drug therapy immediately, which is a possible course of actions according to the JNC guideline. The physician probably suggested to the patient to change some of his/her life style.

On the next consultation, the physician checks the patient blood pressure again and discover that it is still high, so he decides to initiate a drug therapy with two drugs - a diuretic and an ACE inhibitor. The guideline suggests that the initial drug therapy should be only one medication, normally from the diuretic group. The system discovers this difference and generates a warning.

In the third consultation, the blood pressure is still high. The physician decides to keep the current treatment, increasing the dosage of the diuretic. In this consultation, other exams were requested, none of them directly connected to hypertension, so the system flag them as possibly unnecessary exams. This illustrates the problem of reporting unnecessary exams. Therefore, it is possible turn off the warnings regarding unnecessary exams.

56 days later, the patient returns to a new consultation, and still presents a high blood pressure. The physician introduces a new drug from the beta-blockers group. At this point, the system generates another warning, because the guideline states that another medication should only be added to the drug therapy when the current medication is at its maximum dosage, which is not the case.

The last consultation shows the patient blood pressure under control, and thus the system (which has re-synchronized to the three drug state) agrees with the physician decision to keep the treatment.

3.2 Service auditing

The previous section analyzed the output of the auditor system for a single patient. That mode of auditing is appropriate for discussing particular cases. Another mode just summarizes all warnings for all patients' records. This mode is more appropriate to evaluate the service as a whole.

At the end of 2004, we ran the auditor for all patients with hypertension in the clinic's database. The system analyzed 796 patients with hypertension. For 346 patients (43%) there were discrepancies regarding the drug treatment, and for 120 patients (15%) there were exams that should have been performed but were not. If we consider consultations instead of patients, there were 2483 consultations, 427 of which (or 18%) presented some discrepancy regarding the medications, and 152 of which (6%) had missing exams. These results are shown in table 2.

	Period	
	start to 12/2004	1/2005 to 8/2006
Number of patients analyzed	796	613
Number of patients with warning on medication	346 (43%)	232 (38%)
Number of patients with missing exams	120 (15%)	10 (2%)
Number of consultations analyzed	2482	1750
Number of consultations with warning on medication	457 (18%)	276 (16%)
Number of consultations with missing exams	152 (6%)	10 (0.6%)

Table 2 Results for all patients with hypertension.

Although the authors had no mandate to perform a service auditing, we informally presented the results to the physicians responsible for the clinic in a group meeting. Some physicians argued that some of the JNC guidelines suggestions are not in accordance to local practices, and some privately commented with us that they did not know the details of the guideline. But, since this was not an official auditing or quality control activity no organizational decisions were made.

In August 2006 we ran the auditor again, for all consultations that happened from January 2005 to August 2006, to verify if there were changes to the clinic's practices regarding the hypertension treatment. The system analyzed 613 patients, 38% of which had warnings regarding the drug prescribed, and 2% had missing exam warnings. Or, in terms of consultations, 1750 were analyzed, and 16% had drug warnings, and 0.6% had missing exams warnings. These results are summarized in table 2, third column.

Using a chi-square test we determined that the improvement regarding missing exams was significant with 95% for both the patient data (X-squared = 61.8813, df = 1, p-value = 3.648e-15) and for consultations (X-squared = 78.9596, df = 1, p-value ¡ 2.2e-16), but were not significant for the warnings on drugs (X-squared = 1.7665, df = 1, p-value = 0.1838 for patient data, and X-squared = 3.386, df = 1, p-value = 0.06575 for consultation data).

Of course, we cannot claim that the changes were due the discussions regarding the service auditing data, and not due to other competing hypothesis. But the non significance of the drug warnings seems to indicate that either further education efforts regarding the guideline on prescribing anti-hypertensive drugs are necessary, or that the VI JNC hypertension guideline must be further adapted to the local practices, beyond the changes discussed in section 3.

4 Discussion and conclusion

As discussed in section 1.3, ST-Guide is appropriate for representing primary care guidelines in which the patient periodically consults a health care provider, who prescribes medication, requests and reviews lab exams, and performs physical examinations. ST-Guide is not appropriate to model sets or sequences of actions performed in an ambulatorial environment. ST-Guide has been used to model, for example, hypertension and hepatitis C management.

ST-Audit is a critiquing system to be used in batch mode to analyze large volumes of data, for example the service auditing described in section 3.2. This has two consequences. The first one is that the system has to use some heuristics and simplifications to speed the processing. The second one is that it is somewhat less important that the system should be very "intelligent" in determining a non conformance between the

actions recorded and the ones specified in the guideline - the measure of the quality of the system is not find non-conformance on a single, complex example, but to detect generalized non-conformance within a service itself.

ST-Audit tries to follow a path in the guideline that explains the actions taken in each consultation, but this search for a path is not global - ST-Audit does not try to compute a distance between the sequence of actions taken and all possible paths, in order to discover the "closest" one. Instead, it will construct the path incrementally by looking at the data for each consultation. When the path cannot be constructed, the system marks the deviation and perform the resynchronization as described in section 2.3. The resynchronization is an heuristic that avoids the global computation of the "best path," since that computation could be exponential on the number of consultations.

Since resynchronization is an heuristic, the particular formula we used for resynchronization is not *correct* - there is no proof that resynchronization using that formula generates the "best" explanation for the actions taken. The formula is *appropriate* at least for the example discussed in this paper - the modeling of the JNC guideline.

The way the system deals with incomplete information by keeping track of a set of possible states for each consultation (section 2.4) is not exponential either on the number of states of the guideline or on the number of consultations. But this solution is also a simplification, but in this case the simplification adds some error to what is flagged as a deviation. Let us suppose that the system considers possible that the patient (or better the patient's treatment) is in either state A or state B. State A requires that only exam A should be performed and state B requires only exam B. Now, suppose that the physician ordered both exams A and B. The system would not flag that as an unnecessary exam (because both exams are contained in the union of the required exams for the set of states being considered), but clearly one of the exams was unnecessary.[1] To be able to discover that there is an unnecessary exam without determining which, the system would have to perform very sophisticated and computationally complex logical reasoning. Fortunately, unnecessary exams are a form of non compliance that is of low importance given the presuppositions of the system. The EPR will likely contain actions that were taken to diagnose or treat or maintain other diseases besides the one whose guideline is being used, and these actions would be deemed unnecessary by the system, which only focus on one disease.

Our solution to the incomplete data, of assuming that it should take any possible value, also deserve some discussion. Another possible alternative is to assume that the missing data has some special "no value", which will make all comparisons false. In this case, using the previous example, if one transition has the condition (exam B $>$ 20 AND side-effects = false) and the other, (exam B $>$ 20 AND side-effects = true), and if data on side effect is not available in the EPR, then both transitions will be false. In fact the no value solution is the correct one if the guideline is being used to guide the physician decisions in a decision support system. The physician should not make any of the two decisions without asking if there were side effects to the treatment! In fact, in [34,35] ST-Guide is used as a decision support system, and the conditions are evaluated in a 3-value logic, to guarantee that no condition that make reference to missing data are evaluated to true. But for an auditing system, where one knows that there will be data collected in the consultation which was not recorded or cannot

[1] We would like to thank an anonymous reviewer for pointing out this example.

retrieved (because it was written down as free text, for example) it is safer to follow the "all possible values" approach.

ST-Audit can be compared with the Asbru-based systems in many dimensions. Firstly, most of the examples of guidelines for Asbru based systems are ambulatorial. We believe that intentions are less important in primary care (although the example used in this paper was in primary care) for which ST-Guide is tailored. In fact it is not even clear how to add intentions in ST-Guide. The second dimension is scale. ST-Audit was designed for batch review of large number of records, and tested with over 1300 patients. In batch review, the speed of computation is important, and there is less importance attached to minimizing false positive cases of detecting non-compliance. Although not explicitly stated, Asbru based systems seems targeted to a more careful analysis of a single case, possibly as a part of a suggestion system. For example, [28] tested their system in 12 cases; the other systems do not specify how many cases they were applied to.

ST-Audit shares with model checking based systems the basic assumption of modeling the guideline as a finite state machine. In fact, model checkers seems a possible alternative to audit patient records using a ST-Guide guideline. There are two problems that must be explored/solved. The first is speed. Model checking can have very high complexity (up to PSPACE-complete [38]) depending on different aspects. But the fact that the problem has high complexity does not mean that the program will take too long to solve the cases/situations that are common in clinical guideline auditing. Only experimentation can tell us whether using model checking is acceptable in terms of computational speed.

The second problem is that model checkers are "too binary" - they will state whether there is a path in the guideline that is "equal" to the sequence of actions in the records or not. But that is not enough. Even if there is such a path, there may be missing or unnecessary exams in the actions taken, and it is unclear how the model checker would discover them. If there is no path, finding out what was "wrong" with the sequence of actions taken by the physician may require further queries to the model checker, which will increase the computational time.

In conclusion, this work presented ST-Audit, a system that using guideline data (represented in ST-Guide), audits the patient's records for consistency regarding the guideline. St-Audit is targeted for high volume, batch analysis of EPR data and for primary care guidelines. The auditor embodies some of heuristics to find a "reasonable explanation" for the actual treatment the patient received, or in other words, to find a path in the guideline state diagram that matches the actual treatment.

References

1. B. Hazlehurst, D.F. Sittig, V.J. Stevens, K.S. Smith, F.H. Hollis, T.M. Vogt, J.P. Winickoff, R. Glasgo, T.E. Palen, and N.A. Rigotti. Natural language processing in the electronic medical record - assessing clinical adherence to tobacco treatment guidelines. *American Journal of Preventive Medicine*, 29:434–439, 2005.
2. R.S. Dick, E.B. Steen, and D.E. Demer. *The Computer-based Patient Record: an essential technology for health care.* National Academy Press, 1997.
3. J.M. Corrigan, M.S. Donaldson, and L.T. Kohn. *Crossing the Quality Chasm: a new health system for the 21st century.* National Academy Press, 2001.
4. E.C. Schneider, V. Riehl, S. Courte-Winecke, D.M. Eddy, and C. Sennet. Enhancing performance measurement - NCQA's road map for a health information framework. *JAMA*, 282:1184–1190, 1999.

5. B.E. Landon, S.T. Normand, D. Blumenthal, and J. Daley. Physician clinical performance assessment - prospects and barriers. *JAMA*, 290(9):1183–1189, 2003.
6. National Institute for Clinical Excellence, Radcliffe Medical Press. *Principles for Best Practice in Clinical Audit*, 2002. http://www.nice.org.uk/niceMedia/pdf/BestPracticeClinicalAudit.pdf.
7. J.M. Grimshaw and I.T Russell. Effect of clinical guidelines on medical practice: A systematic review of rigorous evaluations. *Lancet*, 342(8883):1317–1322, 1993.
8. S.H. Woolf, R. Grol, A. Hutchinson, M. Eccles, and J Grimshaw. Clinical guidelines. potential benefits, limitations, and harms of clinical guidelines. *BMJ*, 318(7182):527–530, 1999.
9. G. Hripcsak, P. Ludemann, T. A. Pryor, O. B. Wigertz, and P. D. Clayton. Rationale for the Arden syntax. *Comput Biomed Res*, 27(4):291–324, 1994.
10. J. Fox, N. Johns, and A. Rahmanzadeh. Disseminating medical knowledge: the PROforma approach. *Artificial Intelligence in Medicine*, 14(1-2):157–182, 1998.
11. 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(1-2):29–51, 1998.
12. L. Ohno-Machado, J.H. Gennari, S.N. Murphy, N.L. Jain, S.W. Tu, D.E. Oliver, E. Pattison-Gordon, R.A. Greenes, E.H. Shortliffe, and G.O. Barnett. The guideline interchange format. *Journal of the American Medical Informatics Association*, 5:357–372, 1998.
13. A.A. Boxwala, M. Peleg, S. Tu, O. Ogunyemi, Q.T. Zeng, D. Wang, V.L. Patel, R.A. Greenes, and E.H. Shortliffe. GLIF3: a representation format for sharable computer-interpretable clinical practice guidelines. *Journal of Biomedical Informatics*, 37(3):147–161, 2004.
14. R.N. Shiffman, Y. Liaw, C.A. Brandt, and G.J. Corb. Computer-based guideline implementation systems. *Journal of the American Medical Informatics Association*, 6:104–114, 1999.
15. S.W. Tu and M.A. Musen. Representation formalisms and computational methods for modeling guideline-based patient care. In B. Heller, M. Loffler, M. Musen, and M Stefanelli, editors, *Computer-Based Support for Clinical Guidelines and Protocols: Proceedings of EWGLP 2000*, pages 115–132. IOS Press, 2001.
16. D. Wang, M. Peleg, S.W. Tu, A.A. Boxwala, R.A. Greenes, V.L. Patel, and E.H. Shortliffe. Representation primitives, process models and patient data in computer-interpretable clinical practice guidelines: A literature review of guideline representation models. *International Journal of Medical Informatics*, 68(1-3):59–70, 2002.
17. P.A. de Clercq, J.A. Blom, H.H.M. Korsten, and A. Hasman. Approaches for creating computer-interpretable guidelines that facilitate decision support. *Artificial Intelligence in Medicine*, 31(1):1–27, 2004.
18. S. Shakib and P. A. Phillips. The Australian centre for evidence-based clinical practice generic audit tool: Auditmaker for health professionals. *Journal of Evaluation in Clinical Practice*, 9(2):259–263, 2003.
19. S. Kinn, N. Lee, and A. Millman. ABC of medical computing: using computers in clinical audit. *BMJ*, 311:739–742, 1995.
20. N. T. Ellis. C.A.S.M. and C.A.S.T.: A new information based approach to clinical audit. *Computer Methods and Programs in Biomedicine*, 54:29–38, 1997.
21. J. van der Lei, M. A. Musen, E. van der Does, and J. H. Van Bemmel. Review of physician decision making using data from computer-stored medical records. *Lancet*, 338:1504–1508, 1991.
22. J. van der Lei and M. A. Musen. A model for critiquing based on automated medical records. *Computers and Biomedical Research*, 24(4):344–378, 1991.
23. M.M. Kuilboer, J. van der Lei, J.C. de Jongste, S.E. Overbeek, B. Ponsioen, and J.H. van Bemmel. Simulating an Integrated Critiquing System. *Journal of the American Medical Informatics Association*, 5:194–202, 1998.
24. M. M. Kuilboer, M. A. M. van Wijk, M. Mosseveld, and J. van der Lei. AsthmaCritic: Issues in designing a noninquisitive critiquing system for daily practice. *J Am Med Inform Assoc*, 10(5):419–424, 2003.
25. M. M. Kuilboer, M. A. M. van Wijk, M. Mosseveld, E. van der Does, J. C. de Jongste, S. E. Overbeek, B. Ponsioen, and J. van der Lei. Computed critiquing integrated into daily clinical practice affects physicians' behavior-a randomized clinical trial with AsthmaCritic. *Methods Inf Med*, 45(4):447–54, 2006.

26. A. Advani, K. Lo, and Y. Shahar. Intention-based critiquing of guideline-oriented medical care. In *Proceedings of the 1998 AMIA Annual Symposium*, pages 483–487, 1998.
27. A. Advani and Y. Musen. Medical quality assessment by scoring adherence to guideline intentions. *Journal of the American Medical Informatics Association*, 9:92–97, 2002.
28. R. Sips, L. Braun, and N. Roos. Applying intention-based guidelines for critiquing. In *Workshop: AI techniques in healthcare: evidence-based guidelines and protocols*, 2006. http://www.openclinical.org/cgp2006_2.html#critiquing.
29. P. Groot, A. Hommersom, P. Lucas, R. Serban, A. ten Teije, and F. van Harmelen. The role of model checking in critiquing based on clinical guidelines. In *11th Conference on Artificial Intelligence in Medicine, AIME 2007*, volume 4594 of *Lecture Notes in Computer Science*, pages 411–420. Springer, 2007.
30. F. Chesani, E. Lamma, P. Mello, M. Montali, S. Storari, P. Baldazzi, and M. Manfredi. Compliance checking of cancer-screening CareFlows: an approach based on computational logic. In A. Ten Teije et al., editors, *Computer-based medical guidelines and protocols: a primer and current trends: A Primer and Current Trends*, page 183. IOS Press, 2008.
31. S. Montani. Case-based reasoning for managing non-compliance with clinical guidelines. *Proc. Case-Based Reasoning in the Health Sciences Workshop, International Conference on Case Based Reasoning (ICCBR)*, August 2007.
32. K. L. McMillan. *Symbolic Model Checking*. Kluwer, 1993.
33. E. Clarke, O. Grumberg, and D. Peled. *Model checking*. MIT Press, Cambridge, MA, USA, 1999.
34. A. Monteiro and J. Wainer. A conceptual framework to model chronic and long-term diseases. In *8th Conference on Artificial Intelligence in Medicine in Europe - AIME*, volume 2101 of *LNAI*, pages 405–408. Springer, 2001.
35. J. Wainer, C. Billa, M. Paixao, and D. Sigulem. ST-Guide: a framework for the implementation of automatic clinical guidelines. In *23th ACM Symposium on Applied Computing (SAC)*, volume 2, pages 1325–1332, 2008.
36. P.D. Johnson, S.W. Tu, N. Booth, B. Sugden, and I.N. Purves. Using scenarios in chronic disease management guidelines for primary care. *Proc. AMIA Symp.*, pages 389–393, 2000.
37. National Institutes of Health. *The Sixth Report of the Join National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure*, 1997. www.ncbi.nlm.nih.gov/books/bookres.fcgi/hbp/hbp.pdf.
38. P. Schnoebelen. The complexity of temporal logic model checking. *Advances in Modal Logic*, 4:1–44, 2002.

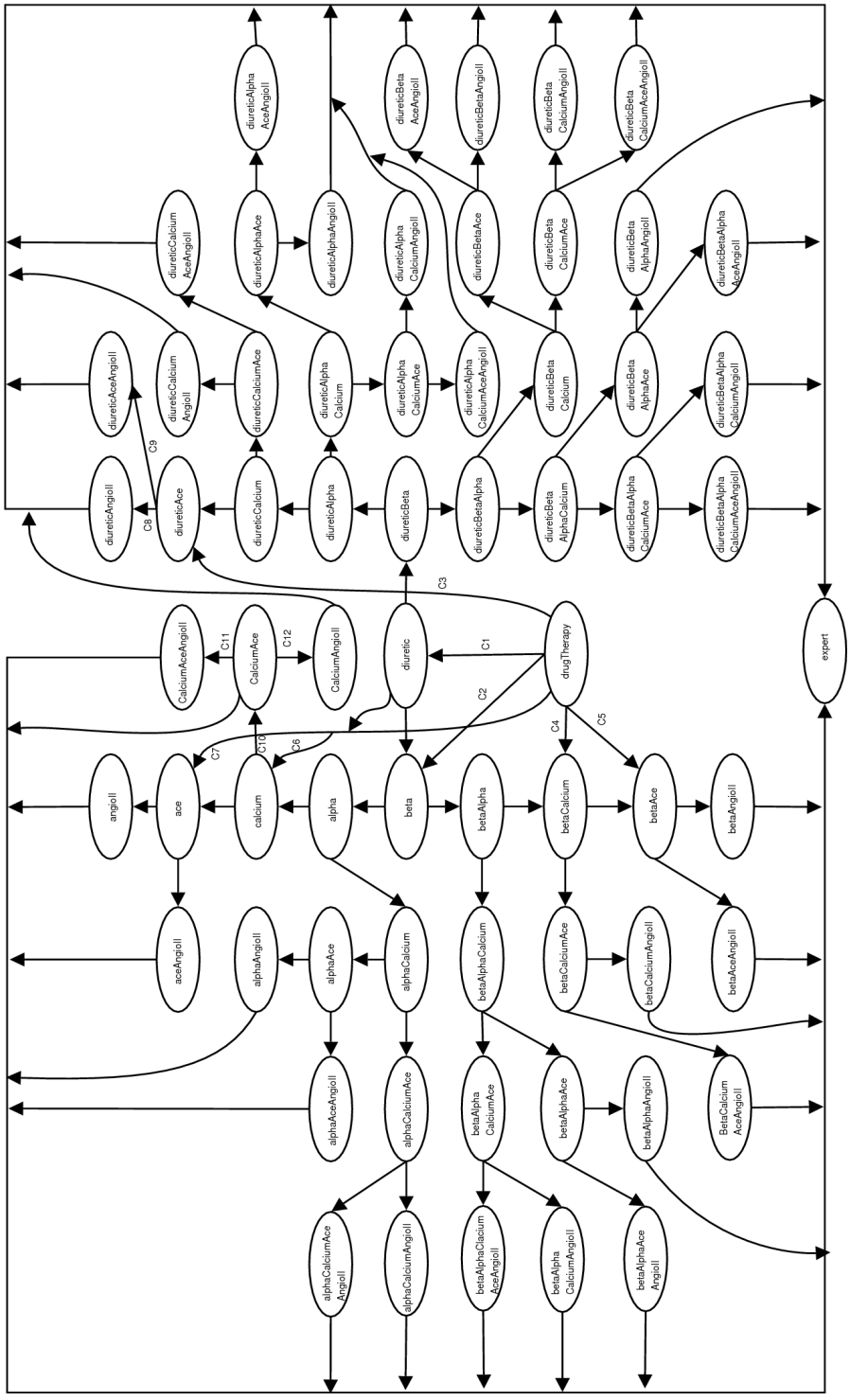

Fig. 6 The full hypertension guideline state diagram

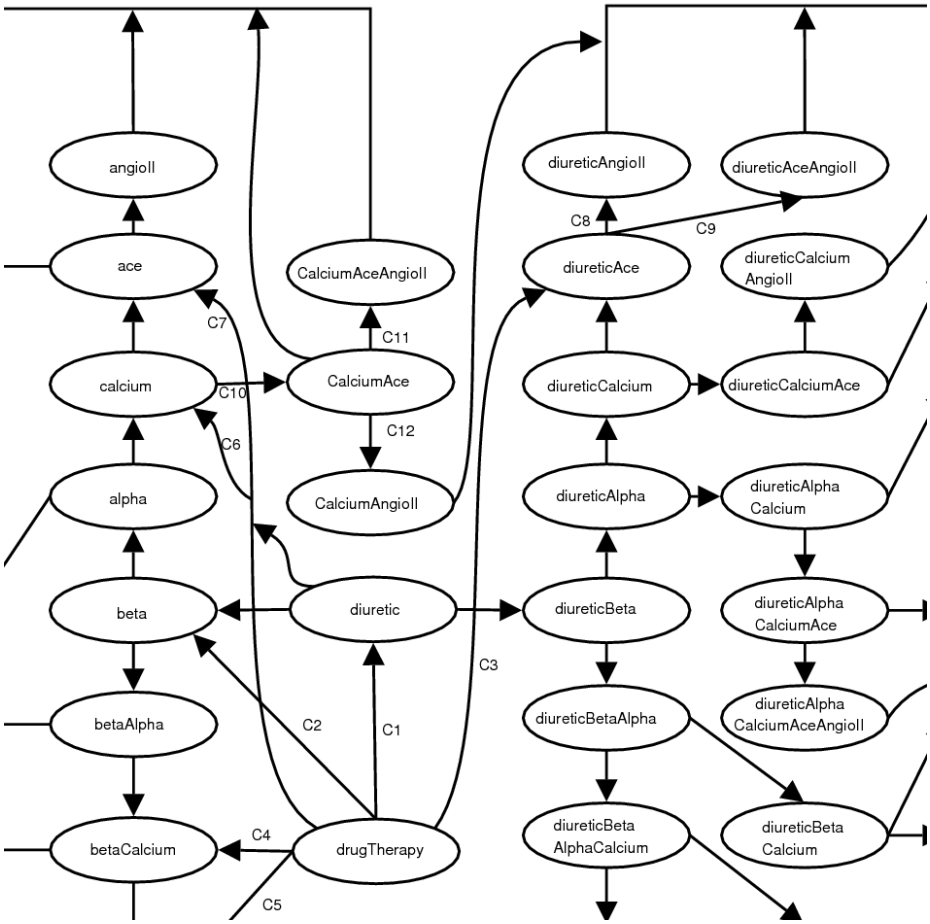

Fig. 7 Detail of the hypertension guideline state diagram