Extending the GuideLine Implementability Appraisal (GLIA) instrument to identify problems in control flow

Mor Peleg, Ph.D.[1], Jeffrey R. Garber, M.D., F.A.C.P., F.A.C.E.[2]

[1]Department of Management Information Systems, University of Haifa, Haifa, Israel
[2]Harvard Medical School, Boston, MA

Clinical guidelines are usually written as text documents that are meant for human consumption. Implementing clinical guidelines as decision-support systems that deliver patient-specific advice at the point of care could increase the effectiveness of clinical guidelines. Several researchers studied the transition from narrative guidelines to computer-interpretable guidelines and have identified specific barriers to guideline implementation. GuideLine Implementability Appraisal (GLIA) is a comprehensive instrument for identifying such barriers, such that they could be revised. We used the GLIA instrument to appraise a historic thyroid nodule guideline that is now being reviewed by the American Association of Clinical Endocrinologists. Our analysis uncovered new guideline implementation barriers related to control-flow that we integrated into GLIA.

1 Introduction

Clinical guidelines are usually written as text documents that are meant for human consumption. Studies have shown that clinician behavior is most effectively influenced through patient-specific advice, particularly if delivered during patient encounters [1]. With this motivation in mind, several groups have been developing computer-interpretable guideline (CIG) formalisms [2]. Clinical guidelines that are specified using those formalisms enable computer-based execution and support automatic inference.

During manual conversion of a narrative guideline into a CIG, guideline encoders encounter problems of lack of completeness of the guideline, unspecified clinical terms and decision criteria, ambiguity, and even contradiction [3]. In order to bridge the guideline implementation barriers, Shiffman et al. developed the GuideLine Implementability Appraisal (GLIA) [4] instrument. It contains questions arranged into categories that can be applied to guideline recommendations in order to identify implementation barriers in a draft guideline and revise them.

This work studies guideline implementation barriers using a case study from the domain of thyroid nodules. The American Association of Clinical Endocrinologists (AACE) and the Associazione Medici Endocrinologi published in 2006 a clinical guideline for the diagnosis and management of thyroid nodules [5]. This guideline is currently being revised. AACE has initiated a task force, joined by the first co-author of this paper (MP), whose aim is to use modern approaches to clinical algorithm creation, with the hope of developing an electronic version of the guideline. These modern approaches include using the GLIA instrument and representing the algorithm using a CIG formalism. AACE already has experience in creating clinical algorithms that adhere to standards for algorithm creation; the second co-author of this paper, who is the President of AACE (JG), published a book chapter on thyroid testing in 1988 [6]. Based on the chapter, a team with the guidance of L. Gottleib created an algorithm [7] that is being used in the current paper. The algorithm was widely distributed, ending up in compendiums and textbooks [7, 8], and was chosen by the Society for Medical Decision-Making as an example for their Proposal for Clinical Algorithm Standards [8].

We used GLIA to appraise the historical thyroid nodule guideline [8]. During our analysis, we found guideline implementation barriers that were related to control flow. Some of the barriers were previously discussed in other works that addressed guideline implementation barriers [9-11]. We integrated these barriers into the GLIA instrument. In this paper we present the added barriers and provide examples for them from the thyroid nodule guideline.

Devising clinical algorithms helps to uncover many problems in the narrative guideline and should thus be done as early as possible in the development of a clinical guideline. Therefore, it is important to include dimensions of implementation barriers that are related not only to individual recommendations, as done in the current GLIA instrument, but also to the way in which the recommendations constitute an algorithm that directs a clinician regarding the appropriate sequence of clinical actions and decisions.

2 Related work

Several researchers [3, 4, 9-11] addressed the process of translating narrative guidelines to computer-interpretable guidelines and have tried to provide recommendations and tools for supporting this transition. Shiffman [3] described a process for translating document-based knowledge into workflow-integrated clinical decision support systems. The process includes 11 steps: (1) selection of a guideline and specific recommendations for implementation, (2) markup of the guideline text, (3) atomization, (4) deabstraction and (5) disambiguation of recommendation concepts, (6) verification of rule set completeness, (7) addition of explanations, (8) building executable statements, (9) specification of origins of decision variables and insertions of recommended actions, (10) definition of action types and selection of associated beneficial services, (11) choice of interface components, and (12) creation of requirement specification. This paper concerns the steps related to making the narrative guideline more clear, unambiguous, correct, and complete, namely, steps 4-6. These steps are addressed by the GLIA [4] instrument. GLIA is a comprehensive instrument that contains 31 items arranged into ten categories (dimensions) that can be applied to guideline recommendations in order to identify barriers to implementation in a draft guideline and revise the recommendations accordingly. The first six dimensions address decidability, executability, global characteristics, presentation and formatting, flexibility, and computability. The other dimensions include measurable outcomes, apparent validity, effect on process of care, and novelty/innovation. Because in this work we are concerned with finding barriers related to the algorithmic logic of the guideline, we are most concerned with the first six dimensions.

Tierney et al. [9] described the problems encountered while they encoded a heart failure guideline. The guideline lacked definitions of terms and branch points, did not focus on errors of commission, and did not account for comorbidities, concurrent drug therapy, or the timing of most interventions and followups. They devised a set of recommendations for avoiding these problems. Peleg et al. [10, 11] studied the process by which a medical expert from the American College of Physicians created algorithms from narrative guidelines and suggested recommendations that go beyond the recommendations of Tierney for improving the quality of clinical algorithms. These recommendations include: (1) verifying that all relevant information is carried from the narrative guideline to all versions of the clinical algorithm, (2) providing information necessary to rank treatment options, and (3) considering different patient scenarios.

Step 2 of the implementation process described above, guideline markup, can also uncover many problems with the guideline. Markup tools are used to mark parts of the narrative guideline text to indicate that they belong to certain structural components of guidelines, according to markup ontologies. Examples of Markup tools include a tool based on the Guideline Elements Model [12], Digital electronic Guideline Library framework [13], and Document Exploration and Linking Tool (Delt/A) [14]. While guideline appraisal tools and recommendations for writing narrative guidelines focus on disambiguating the narrative and making the guideline more complete and error-free, the markup tools focus on translating the guideline text into particular CIG formalisms. If appraisal tools are used to discover barriers to implementation that are subsequently handled then markup tool could be used as a later step in the process of translating narrative guidelines into CIGs.

3 Methods

The informatician on our team (MP) analyzed the clinical algorithm "Evaluation of Thyroid Nodules by Primary Care Providers", the tabular algorithm summary, and annotations A, B, and C of the historical thyroid nodule guideline [8]. We emphasize that the knowledge and approaches (e.g., emergence and role of ultrasound, which in 1988 had a minor role) contained in the historical guideline is not up to date but we use it for illustration purposes.

The informatician translated the guideline into the GLIF3 [15] CIG formalism using the Protégé knowledge modeling tool [16]. GLIF3 specifies guidelines as flowcharts of steps representing clinical actions, decisions, and patient states. The steps' details generate a computable specification enabling logical consistency and inference. During the translation process, the informatician identified potential problems in the guideline pertaining to control flow, inconsistency of terms used throughout the guideline, lack of clarity, and imprecise term definition. The expert endocrinologist (JG) and the informatician went over the problems. The expert clarified the potential problems, identifying which were problems and which were misunderstandings on the part of the informatician. The expert validated the proposed solutions for correcting the problems and we revised the GLIF3 algorithm, the guideline summary, and annotations.

As our goal was to produce a robust instrument that could be used by guideline authors who are less trained in informatics to appraise guideline drafts, identify problems, and correct them, we augmented the GLIA [4] instrument based on the problems that we identified in the historical thyroid nodule guide-

line. First, we integrated into the existing ten dimensions of GLIA recommendation by Tierney et al. [9] and recommendations by Peleg et al. [10, 11] that were discussed in the Related Work section. Then we classified the problems that we found in the thyroid nodule guideline into the existing items in the augmented GLIA instrument. We were left with problems that were not covered by the GLIA instrument. Some of the problems were refinements of existing GLIA items and some problems were not covered by any of the existing dimensions. For these we added a new dimension to the augmented GLIA instrument.

4 Results

We identified three items by Tierney et al. that were not covered by the existing items in GLIA. In addition, three items by Tierney et al. and three items by Peleg et al. that were refinements of GLIA items. We integrated them into the existing GLIA dimensions.

Analyzing the types of problems found in the thyroid guideline, two types were covered by existing GLIA dimensions. Seven problem types were covered by refinements of existing items. Five problem types were not covered by GLIA. We added one item to the Presentation and Formatting dimension to account for one of the problems. The other four problem types

Table 1. Dimensions (marked by the letters A-G) and items (marked by numbers) of the extended GLIA instrument. The type of addition (new dimension, new item, (item) refinement), the source for the additions, and the number of problems found in the thyroid guideline are indicated in the second and third columns; "This" indicates this paper.

Dimensions and Items	Addition type	Source	#problems
A. New dimension: algorithmic control-flow (does the sequence of steps capture the intended flow, are all steps explicitly defined?)	New dimension	This	
32) Examine branching and synchronization points for errors of control-flow	New item	This	1
33) Add implicit action and patient-state steps for clarity and completeness	New item	This	3
34) Add clinical state steps to indicate end of one visit and beginning of next visit	New item	This	3
35) Break a step into multiple steps if that step is not done at one point in time	New item	This	1
B. Global Considerations			
7) Is the guideline internally consistent?			1
7.1) Use one numbering system to number nodes in the algorithm and guideline recommendations (e.g., good examples are found in guidelines by the Institute for Clinical Systems Improvement).	Refinement	This	1
C. Decidability			
8) Are decision conditions clearly defined?			
8.1) Write all guideline rules in a simple "if-then-else" format with all of the parameters strictly defined using routinely collected clinical data	Refinement	[9]	
8.1.1) The parameters (e.g., Thyroid function test, TFT) should also be defined by terms from a controlled terminology	Refinement	This	1
8.2) Make algorithm logic hinge on explicitly defined values of accepted clinical parameters (e.g., hypothyroidism and hyperthyroidism should be defined in terms of thresholds of specific TFT tests)	Refinement	[9]	2
8.3) If a step references time, be precise (e.g., "rapidly" changing thyroid status means 2-4 weeks)	Refinement	This	2
9) Are all reasonable combinations of conditions accounted for?			
9.1) Consider different patient scenarios. When an action (or decision) is based on more than one parameter, consider all combinations of parameter values (e.g., a pregnant woman whose doctor is unsure whether she has a nodule).	Refinement	[10]	1
10) If there are more than one condition in the recommendation, is the logical relationship among all conditions (ANDs and ORs) clear?			
(10.1) Don't confuse AND with OR (or Exclusive OR)	Refinement	[11]	1
D. Executability			
11) Is the recommended action (what to do) stated specifically and unambiguously?			2
11.1) Provide all the information necessary to rank treatment options. Present options according to an order of preference, specifying a level or recommendation and strength of evidence. If no ranking is implied, arrange alphabetically	Refinement	[10]	1
E. Presentation and formatting			
36) Recommendations that relate to several sub-populations of patients, each with different treatments should be broken into parts. The preferred way of displaying such recommendation is a table.	New item	This	1
F. Flexibility			
25) Does the recommendation consider coincident drug therapy and common comorbid conditions?			
25.1) Include rules about errors of commission and omission and consider common comorbidities	Refinement	[9]	
G. Computability			
(37) Expect that local translation of the guidelines will be necessary, and help guide that process	New item	[9]	
(38) Balance the costs of diagnosis and treatment (not only in terms of dollars) and consider the likelihood that individual patients will benefit from the guidelines' recommendations	New Item	[9]	
(39) Evaluate the resulting guidelines using real patients and representatives of users	New item	[9]	

were related to control flow. These were addressed by a single item in GLIA within the General Considerations dimension: "Is it clear in what sequence the recommendations should be applied?". We view algorithmic control flow as central to clinical guidelines which justifies adding the dimension. Table 1 presents the additions to the GLIA instrument, along with examples from the thyroid nodule guideline.

While the previous paragraph discussed the *types* of problems found in the historical thyroid guideline, the last column of Table 1 presents the *number* of problems found in the guideline. Only 3 problems were directly described by existing GLIA items and 9 were described by their refinements, whereas 9 problems were attributed to the items we added to GLIA.

We explain the new items relating to algorithmic control-flow by referring to the original thyroid nodule algorithm shown in Fig. 1 and its revised representation as a GLIF3 algorithm, shown in Fig. 2. One major flaw in the algorithm of Fig. 1 is that when the first examiner observes a nodule (step 1) he proceeds to step 2 that distinguishes treatment based on pregnancy. However, if only the second examiner observed a nodule (step 11) then the question of pregnancy is skipped. This problem in control flow is corrected in Fig. 2. It could have been identified by thinking about different patient scenarios (item 9.1) or by examining branching points (item 32).

Item 33 considers adding implicit action and patient state steps for clarity and completeness. In Fig. 2, Action Step A1 (Refer to second examiner) and Patient State Steps P2 (Thyroid Nodule) and P4 (Benign diffuse disease) were added for clarity.

Guideline authors do not always consider timing of clinical tasks and their partition into visits. Following item 34, we added step P1 (Patient examined by

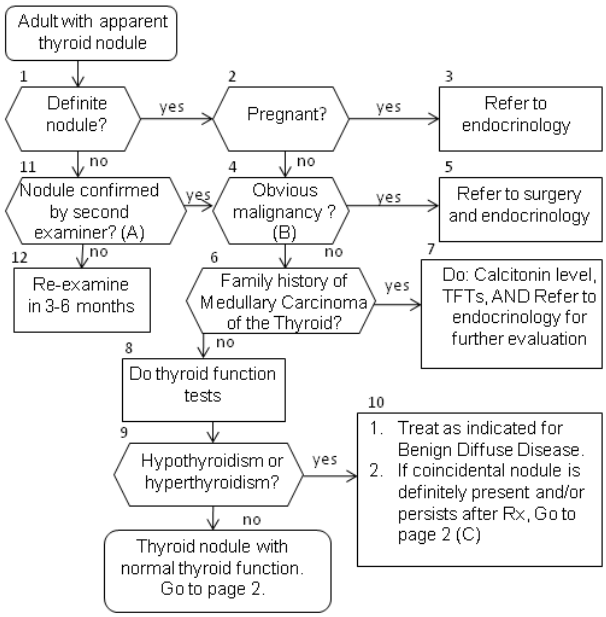

Figure 1. Part of the historical thyroid algorithm [6]

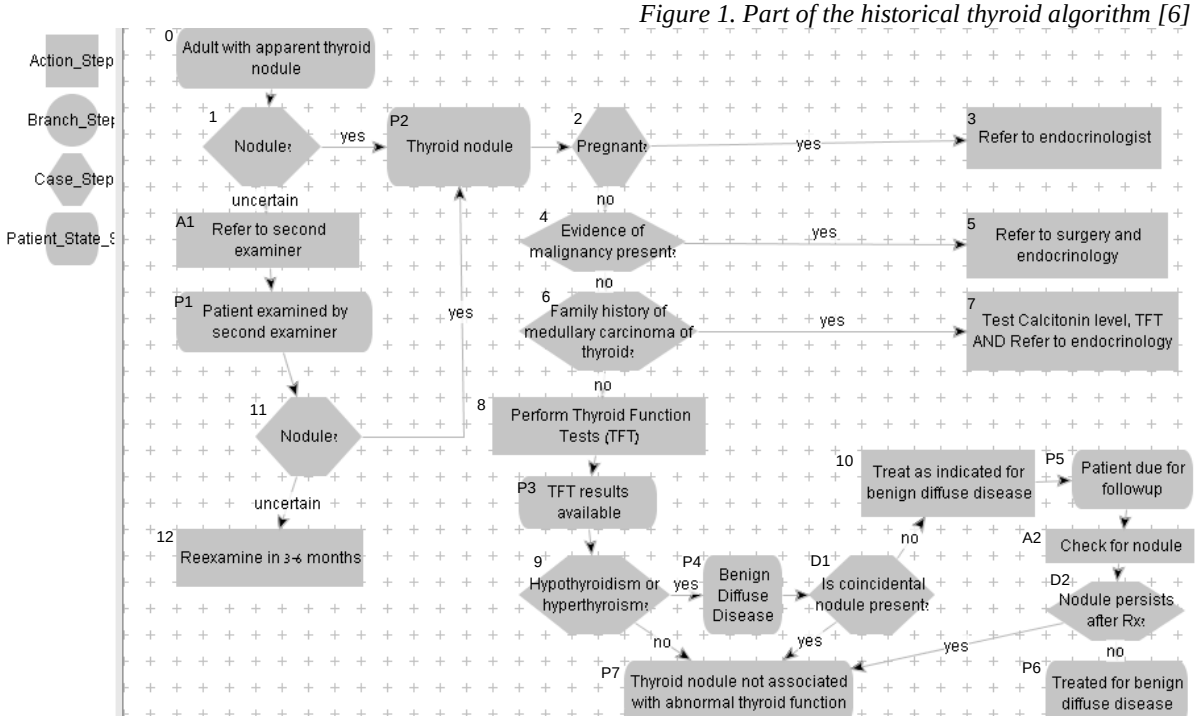

Figure 2. The GLIF3-encoded guideline, revised from the historical thyroid nodule algorithm.

a second examiner), P3 (TFT Test results Available), and P5 (Patient due for followup) to mark visit starts. Following item 35, we broke step 10 into steps P4, D1, 10, P5, A2, D2, and P6.

5 Discussion

We analyzed barriers to implementation of clinical guidelines as they were reflected in a historic thyroid nodule guideline. Despite the fact that the guideline we used for this study is historic, we were still able to draw insights from it which we used to expand the GLIA instrument. Based on our experience, we expect that the problems that we identified in this guideline still occur in more recent guidelines developed by various medical organizations.

While twelve of the problems were covered by existing GLIA items (3) or their extensions (9), nine problems, relating to five items, were not covered by the existing items. Most of these problems were due to errors in control-flow. Since we consider algorithmic control flow as central to guidelines we added a new dimension to account for it.

AACE has been developing guidelines for writing high quality clinical guidelines. One of its aims is to develop guidelines that could easily be the basis for computer-based implementation. The extended GLIA will be used by this organization to appraise the newer version of the thyroid guideline, which is currently under development, in order to find and address barriers to their implementation.

References

[1] Shea S, DuMouchel W, Bahamonde L. A Meta-analysis of 16 Randomized Controlled Trials to Evaluate Computer-based Clinical Reminder Systems for Preventative Care in the Ambulatory Setting. J Am Med Inform Assoc 1996;3(6):399-409.

[2] Peleg M, Tu SW, Bury J, Ciccarese P, Fox J, Greenes RA, et al. Comparing Computer-Interpretable Guideline Models: A Case-Study Approach. J Am Med Inform Assoc 2003;10(1):52-68.

[3] Shiffman RN, Michel G, Essaihi A, et al. Bridging the guideline implementation gap: a systematic, document-centered approach to guideline implementation. J Am Med Inform Assoc 2004;11(5):418-26.

[4] Shiffman RN, Dixon J, Brandt C, Essaihi A, Hsiao A, Michel G, et al. The GuideLine Implementability Appraisal (GLIA): development of an instrument to identify obstacles to guideline implementation. BMC Med Inform Decis Mak 2005(5):23.

[5] American Association of Clinical Endocrinologists and Associazione Medici Endocrinologi. Medical guidelines for clinical practice for the diagnosis and management of thyroid nodules. Endocr Pract 2006;12(1):63-102.

[6] Garber JR. Thyroid Testing. In: Aronson M, Delbanco T, editors. Manual of Clinical Evaluation: Strategies for Cost Effective Care. Boston: Little Brown; 1988. p. 203-211.

[7] Garber JR, Berger B, Herman T, Manson J, Singer D, Seely E, et al. Evaluation of Thyroid Nodules. In: Goldfield N, Pine M, Pine J, editors. Measuring and Managing Health Care Quality: Procedures, Techniques, and Protocols. Gaithersburg, Maryland: Aspen Publishers; 1991. p. 177-186.

[8] Society for Medical Decision Making. Proposal for Clinical Algorithm Standards. Med Decis Making 1992;12:149-54.

[9] Tierney WM, Overhage JM, Takesue BY, Harris LE, Murray MD, Vargo DL, et al. Computerizing guidelines to improve care and patient outcomes: the example of heart failure. J Am Med Inform Assoc 1995;2(5):316-22.

[10] Peleg M, Patel VL, Snow V, Tu S, Motturu-Pilson C, Shortliffe EH, et al. Support for Guideline Development through Error Classification and Constraint Checking. Proc AMIA Symp 2002. p. 607-11.

[11] Peleg M, Gutnik LA, Snow V, Patel VL. Interpreting procedures from descriptive guidelines. J Biomed Inform 2006;39(2):184-95.

[12] Georg G, Seroussi B, Bouaud J. Extending the GEM model to support knowledge extraction from textual guidelines. Intl J Med Inform 2005;74:79-87.

[13] Shahar Y, Young O, Shalom E, et al. DeGeL: A Hybrid, Multiple-Ontology Framework for Specification and Retrieval of Clinical Guidelines. Proc AIME 2003: 122-31.

[14] Miksch S. The Delt/A Project. 2007. http://ieg.ifs.tuwien.ac.at/projects/delta/

[15] Boxwala AA, Peleg M, Tu S, Ogunyemi O, Zeng Q, et al. GLIF3: a representation format for sharable computer-interpretable clinical practice guidelines. J Biom Inform 2004;37(3):147-61.

[16] Grosso WE, Eriksson H, Fergerson R, et al. Knowledge Modeling at the Millennium (The Design and Evolution of Protege-2000). 12th Banff Knowledge Acquisition for Knowledge-Based Systems Workshop 1999:7-4-1-36.