ENDOCRINE PRACTICE Rapid Electronic Article in Press
Rapid Electronic Articles in Press are preprinted manuscripts that have been reviewed and accepted for publication, but have yet to be edited, typeset and finalized. This version of the manuscript will be replaced with the final, published version after it has been published in the print edition of the journal. The final, published version may differ from this proof.
DOI:10.4158/EP13271.OR

Original Article **EP13271.OR**

A COMPUTER-INTERPRETABLE CLINICAL GUIDELINE FOR THE DIAGNOSIS AND MANAGEMENT OF THYROID NODULES

Running title: Computerized thyroid guideline

Mor Peleg, PhD[1]; John Fox, PhD[2,3,4,5]; Vivek Patkar, MD[3,4,5]; David Glasspool, PhD[5]; Ioannis Chronakis, MD[3,5]; Matt South, PhD[5]; Svetlana Nassar, MA[1]; Jason L. Gaglia, MD[6]; Hossein Gharib, MD, MACP, MACE[7]; Enrico Papini, MD, FACE[8], Ralf Paschke, MD[9], Daniel S. Duick, MD, FACP, FACE[10], Roberto Valcavi, MD, FACE[11], Laszlo Hegedüs, MD, DMSc[12,], Jeffrey R. Garber, MD, FACP, FACE[6]

From the [1]Department of Information Systems, University of Haifa, Israel, [2]Oxford University, Oxford, UK; [3]University College London, London, UK, [4]Whittington Health NHS Trust, UK, [5]Deontics Ltd., London, UK, [6]Harvard Medical School, Boston, MA, USA, [7]Mayo Clinic College of Medicine, Rochester, MN, USA, [8]Regina Apostolorum Hospital, Albano and University "La Sapienza", Rome, Italy, [9]University of Leipzig, Leipzig, Germany, [10]Endocrinology Associates, PA, Scottsdale, Arizona, USA, [11]Arcispedale Santa Maria Nuova, Reggio Emilia, Italy, and [12]Odense University Hospital, Odense, Denmark.
Address correspondence to Mor Peleg, PhD, Department of Information Systems, University of Haifa, 3498838, Israel
E-mail: morpeleg@is.haifa.ac.il

ABSTRACT

Objective: Clinical practice guidelines (CPGs) could have more consistent and meaningful impact on clinician behavior if they were delivered as electronic algorithms that provide patient-specific advice during patient-physician encounters. We developed a computer-interpretable algorithm for USA and European users for diagnosis and management of thyroid nodules that is based on "AACE, AME, ETA Medical Guidelines for Clinical Practice for the Diagnosis and Management of Thyroid Nodules" –a narrative, evidence-based CPG.

Methods: We initially employed the guideline modeling language, GLIF3, which emphasizes the organization of a care algorithm into a flowchart. The flowchart specified the sequence of tasks required to evaluate a patient with a thyroid nodule. PRO*forma*, a second guideline modeling language, was then employed to work with data that are not necessarily obtained in a rigid flowchart sequence. Tallis—a user-friendly web-based "enactment tool"—was then used as the “execution engine” (computer program). This tool records and displays tasks that are done and prompts users to perform the next indicated steps. The development process was iteratively performed by clinical experts and knowledge engineers.

Results: We developed an interactive web-based electronic algorithm, which is based on a narrative CPG. This algorithm can be used in a variety of regions, countries, and resource-specific settings.

Conclusion: Electronic guidelines provide patient-specific decision-support that could standardize care and potentially improve the quality of care. The “demonstrator” electronic thyroid nodule guideline that we describe in this report is available at

DOI:10.4158/EP13271.OR

http://demos.deontics.com/trace-review-app[1]. The demonstrator must be more extensively "trialed" before it is recommended for routine use.

Keywords: clinical practice guidelines, computer-interpretable guidelines, thyroid nodules, GLIF3, PROforma, Tallis

Abbreviations:

AACE = American Association of Clinical Endocrinologists; **AME** = Italian Association of Clinical Endocrinologists; **CIG** = computer-interpretable guideline; **EMR** = electronic medical records; **ETA** = European Thyroid Association; **FNA** = fine-needle aspiration; **FT4** = free thyroxine; **GLEE** = GLIF3 execution engine; **GLIA** = guideline implementability appraisal; **GLIF** = guideline interchange format, a CIG modeling language; **MEN2** = multiple endocrine neoplasia type 2; **MNG** = multinodular goiter; **MTC** = medullary thyroid carcinoma; **PRO*forma*** = a CIG modeling language (The name is derived from "PROcess formalization); **Tallis** = a web-based execution engine for PRO*forma* CIGs; **TNM** = task network model; **TPO** = thyroid peroxidase; **TSH** = thyroid stimulating hormone; **US** = ultrasonography

INTRODUCTION

Clinical practice guidelines (CPGs) are "systematically-developed statements to assist practitioner and patient decision making about appropriate healthcare for specific clinical circumstances" (1). They are developed by healthcare organizations and published as articles in professional journals or monographs. They are disseminated using the regular periodical dissemination routes and could additionally be sent to healthcare organizations for local

[1] Username: reviewer; password: tnodule1

distribution to their practitioners. After dissemination, we assume that practitioners read these guidelines, internalize and follow them. Ultimately, guidelines should benefit healthcare by reducing unjustified practice variation, improve healthcare quality, and cut costs. However, given the busy schedule of practitioners and the large number of CPGs that they are expected to read, internalize, and follow, CPGs could have a more consistent and meaningful impact on clinician behavior if they were delivered electronically. This has been demonstrated repeatedly when computerized decision rules were used to deliver to physicians patient-specific alerts (e.g., when laboratory test results were abnormal) and reminders (e.g., to carry out an investigation such as screening or follow-up) (2). By applying the rules directly on patient data recorded in structured format (i.e., not in free text) in electronic medical records (EMRs), the application of alerts and reminders was automated. A standard for specifying such decision rules (the Arden Syntax (3)) has been created by the American Society for Testing and Materials and later by Health Level 7 as early as 1992.

In the second half of the 1990's, in order to utilize CPGs in the care of individual patients, several groups started to develop computer implementations of CPGs: electronic algorithms that specify care processes and which provide patient-specific advice during patient-physician encounters, without busy clinicians having to master the contents of often complex guidelines. Such electronic algorithms are known as Computer-interpretable Guidelines (CIGs) (4, 5) and have been shown to be effective (4, 6-7), in different clinical domains including among others cancer, HIV, genetic counseling (6), and hypertension (8) .

CIGs are formal representations (i.e., representations that have constrained syntax and clear semantics) of narrative CPGs that allow a computer program (called an "execution engine") to apply the formalized representation to patient data and deliver patient-specific guideline

recommendations. While narrative guidelines may have ambiguities (9, 10), CIGs have precise meanings that are attained by a team of knowledge engineers and expert clinicians who disambiguate all decision criteria and actions that are specified in the narrative CPG. CIGs serve as a medium to share guideline knowledge with clear semantics (meanings) and to execute (match) the guideline knowledge with patient data to arrive at patient-specific recommendations. CIGs are represented using guideline modeling languages (5) known as Task Network Models (TNMs). Typical tasks represent data enquiries, clinical actions, and clinical decisions. There are different kinds of guideline modeling languages, each with its own emphasis and abilities. Examples of guideline modeling languages include the GuideLine Interchange Format, version 3 (GLIF3) (11) and PRO*forma* (12), provided in the appendix at http://mis.hevra.haifa.ac.il/~morpeleg/Appendix.html. TNMs are created as both a medium of communication and as a computable representation (i.e., a representation that could be executed by a computer program). Hence, their graphical visualization makes them comprehensible by clinical domain experts who together with knowledge engineers create and validate them (13). To allow execution of CIGs, for each guideline modeling language there is a respective guideline execution engine(s) (14). These software tools can be used to follow a clinical pathway, recording clinical data consistently and accurately, and support the clinician in making patient-specific decisions.

Developing an AACE Electronic CPG

In this paper we report a collaborative process of developing a computer-interpretable algorithm for USA and European users for the diagnosis and management of thyroid nodules. The collaborating team used the narrative, evidence-based *AACE, AME, and ETA Medical Guidelines for Clinical Practice for the Diagnosis and Management of Thyroid Nodules* (15) as

mandated in the AACE guidelines for guidelines publication in 2010 (16). In 2009, the American Association of Clinical Endocrinologists (AACE) created 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 CPGs. AACE already has experience in creating clinical algorithms that adhere to standards for algorithm creation; the last co-author of this paper (JRG), published a book chapter on thyroid testing in 1988 (16). Based on the chapter, a team with the guidance of L. Gottleib created an algorithm (17) that was chosen by the Society for Medical Decision-Making as an example for their *Proposal for Clinical Algorithm Standards* (18). In previous work (9), the first (MP) and last (JRG) authors of this paper have used the *GuideLine Implementability Appraisal* (GLIA) (19) instrument to appraise and identify barriers to guideline implementation in the historic thyroid nodule guideline that was being reviewed at the time by AACE while they were creating a modern version of that clinical practice guideline (15).

The thyroid guideline was selected for computerization for several reasons. It was published shortly after the release of the *American Association of Clinical Endocrinologists Protocol for Standardized Production of Clinical Practice Guidelines—2010 update* (16) which mandated the electronic implementation of guidelines. The marked variation in the diagnosis and management of thyroid nodules established a clear need for guideline-based decision-support for non-expert clinicians. A case in point is occurring in Italy where the national health system is requiring physicians to adhere to CPGs or consensus documents to both guide practitioners and to reduce their risk of malpractice. At Regina Apostolorum Hospital (where author EP is from) and at other hospitals in Rome, Italy, the narrative guideline is currently being used by groups of

four or more primary-care physicians who are being provided with ultrasound machines to evaluate thyroid nodules.

Subsequently, we initiated a thorough process of developing a CIG implementation of this narrative guideline, which we report in this paper. To date the algorithm has been validated with twenty patients whose clinical features were carefully described and whose surgical outcomes were known. System recommendations were compared with the actual management that patients received. Concordance was high (95.7 %; 134/140 decision points). This is a preliminary report. More cases are required to make the algorithm more robust. It is our hope that this report of our early experience will raise awareness about the algorithm and increase its use. Extensive feedback will enable us to refine it and accelerate the path to it becoming a useful clinical and research tool for all those interested in nodular thyroid disease.

METHODS

In Table 1 we describe the general steps in the process of computerizing a CPG. Figure 1 shows the flowchart from the original guideline for diagnosis and management of palpable thyroid nodules (15) to which we have added annotations 1-9. These annotations explain the ambiguities that we discovered in the flowchart by using the *GuideLine Implementability Appraisal* (GLIA) instrument (19) that had been extended with items for flowchart assessment (9). Many ambiguities were identified: annotations (1) through (3) mark split points in the algorithm. However, it is not clear whether the splits indicate concurrent branches done together (AND-Split) as should be the case for annotation (1), or mutually exclusive paths (eXclusive OR-Split) as should be the case for annotation in (2) and (3). Annotation (4) asks whether we always perform ultrasonography (US) or only in the case that TSH is not low. Similarly, annotation (5) asks whether we always perform measurement of free

thyroxine. Annotation (6) indicates that high calcitonin needs to be confirmed by repeating the test. In annotation (7) it is not clear that some of the "suspicious US findings" were more indicative of malignancy than others and annotation (8) indicates that the cutoff for "low TSH" should be defined. Annotation (9) points to incompleteness of the algorithm: for confirmed high Calcitonin in the setting of other negative findings the patient requires surgery, but here a benign FNA would omit surgery.

While the thyroid nodule guideline is a consensus statement for both the USA and Europe, the guideline states different recommendations for the use of scintigraphy in the USA and Europe. Furthermore, recommendation 5.7.2. regarding calcitonin “measurement of basal calcitonin level may be a useful test in the initial evaluation of thyroid nodules” is likely to be interpreted differently in Europe and the USA:

1) In Europe, unlike in the USA, calcitonin measurements are often performed regardless of family history or clinical suspicion of medullary thyroid carcinoma (MTC) or multiple endocrine neoplasia type 2 (MEN2)
2) In iodine-insufficient areas, which are common to some areas in Europe, scintigraphy is performed to look for hot nodules which are common in patients with multinodular goiters (MNG) even if they have normal TSH values. Whereas in the USA, an iodine sufficient region, scintigraphy is not employed to evaluate MNGs for hot nodules since they are uncommon unless TSH is low

Table 2 provides an example where the versions of the algorithm (USA vs. Europe) yield two different recommendations for the same patient.

Other differences exist between the USA and Europe and even between centers on the same continent. How to follow and manage benign thyroid nodules is one notable example. For

example percutaneous ethanol injection of cystic thyroid nodules is more commonly performed in Italy than in other parts of Europe or the United States. However, these differences and options do not change the sequence in the CIG algorithm as it is currently constructed (The GLIF3 model shown in Figure 2 and the PRO*forma* model and its Tallis enactment shown in Figure 3). Customized future iterations of the algorithm could specify when they should be employed rather than simply to list them as options.

RESULTS

As explained in Table 1, Figure 2 shows the GLIF3 clinical algorithm which enforces strict sequencing and Figure 3 presents a screenshot from the Tallis tool showing the physician user's view of the medium-flexibility version of the PRO*forma* algorithm. The screenshot shows what the USA user will see when he will take the decision whether to proceed with or without calcitonin measurement. To support execution by the Tallis engine, detailed specification of rules drawn from the AACE guideline is provided for each decision. Table 3 presents an example of the indications for performing the calcitonin test.

The computerized CPG application demonstrates the practical feasibility of formalizing evidence-based guidance and other recommendations for best practice described within one published AACE guideline, as a basis for deploying clinical decision support services that comply with the guidelines and supporting evidence. The “demonstrator” electronic thyroid nodule guideline that we describe in this report is available at http://demos.deontics.com/trace-review-app.

By the time of publication, the algorithm had been evaluated against data for twenty patients with known surgical outcomes. We extracted EMR data which included all of the information called for by the algorithm together with other items, such as gender and age; in the

future these will enable us to study predictors of thyroid malignancy in different settings and assist us in the development of future iterations.

CONCLUSION

Electronic guidelines provide patient-specific decision-support that can standardize care and potentially improve the quality of care. We have developed an interactive web-based electronic algorithm, which is based on a narrative CPG for thyroid nodules. To the best of our knowledge this is the only real-time interactive algorithm that allows the user to enter any patient case. This is in contrast to a teaching tool that reviews how you performed on a predefined patient case (data set). This algorithm can be used in a variety of regions, countries, and resource-specific settings. Future modifications could be based on what resources are available in a particular region or setting. For example while relatively inexpensive, ultrasonography may not be readily available in some settings. Newer tools such as molecular markers may be routinely used in others. Some users may need ancillary algorithms for how to employ various molecular markers or surgical decision-making.

The "demonstrator" electronic thyroid nodule guideline that we describe in this report is available online (instructions provided online at http://demos.deontics.com/trace-review-app[2]). In order to launch any clinical decision support application such as our guideline a multi-staged evaluation of its performance, usability, safety, and potential clinical impact in clinical use is required and will be carried out during the next phase of our project. To date a pilot evaluation of 20 cases seen in the New England region of the USA has been performed by three authors (JLG, JRG, VP). Concordance was high (95.7 %; 134/140 decision points). Five of the six discordant decisions could be attributed to practice variations such as using scintigraphy to evaluate patients

[2] Username: reviewer; password: tnodule1

with cytology consistent with follicular neoplasms while the sixth one pointed out a flaw in the original narrative guideline; namely, not recommending an FNA on a large nodule (34 mm) without any suspicious ultrasound findings in someone with a normal TSH. An online appendix including details regarding the knowledge-base development process and the executable application, the data set used for evaluation, the cytology and pathology classifications, and additional details regarding the pilot evaluation of the work-flow view of the executable implementation is available at http://mis.hevra.haifa.ac.il/~morpeleg/Appendix.html.

We speculate that various EMRs could be integrated with this electronic guideline. Linking a computer-interpretable guideline to an EMR is a technical issue. It has been done before for various guideline modeling languages and in particular for PRO*forma* (21, 22).

The demonstrator must be more extensively "trialed" before it is recommended for routine use, yet our preliminary evaluation with twenty patient cases indicates that it provides users with evidence-based patient-specific recommendations and explanations for them while at the same time allowing users to deviate from the recommended strict sequencing of indicated actions. Most important, we anticipate continually modifying the electronic version of the guideline as experience with it accrues during the trial and post-trial phase and as new approaches on how to best diagnose and treat thyroid nodules are developed.

ACKNOWLEDGEMENTS

JRG, MP and SN designed the original GLIF3 model of the AACE thyroid nodules guideline as well as the PRO*forma* model in Tallis, corresponding to the GLIF3 model; HG, EP, RP, DSD, RV, and LH participated in the panel of experts who performed repeated validation of the GLIF3 models. Deontics Ltd. enhanced the PRO*forma* knowledge base to support flexible decision making as well as workflow (IC, VP, JF) and designed and built a series of versions

with an advanced clinical interface called TrACE (JF, DG, MS). The version illustrated in the paper is TrACE 3.

We would like to thank AACE, Deontics Ltd., Dove Medical Press and Royal Free Hospital Charity for partial support of this work.

REFERENCES

1. **Field MJ, Lohr KN, eds.** *Guidelines for Clinical Practice: Directions for a New Program.* Washington, DC: The National Academies Press; 1990.
2. **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;3:399-09.
3. **Hripcsak G, Ludemann P, Pryor TA, Wigertz OB, Clayton PD.** Rationale for the Arden Syntax. Comput Biomed Res. 1994;27:291-324.
4. **Latoszek-Berendsen A, Tange H, Herik HJ van den, Hasman A.** From clinical practice guidelines to computer-interpretable guidelines. A literature overview. *Methods Inf Med*. 2010;49:550-570.
5. **Peleg M, Tu SW, Bury J, et al.** Comparing Computer-Interpretable Guideline Models: A Case-Study Approach. *J Am Med Inform Assoc.* 2003;10:52-68.
6. **Fox J, Patkar V, Thomson R, Black E.** PRO*forma* guidelines and care pathways: performance in trials and future plans. In: The biennial European Conference on Artificial Intelligence (ECAI) Workshop: AI techniques in healthcare: computerized guidelines and protocols. Riva del Garda, Italy: 2006.

7. **Peleg M.** Computer-interpretable Clinical Guidelines: a Methodological Review. *J Biomed Inform.* 2013:46:744-763.

8. **Goldstein MK, Coleman RW, Tu SW, et al.** Translating research into practice: organizational issues in implementing automated decision support for hypertension in three medical centers. *J Am Med Inform Assoc.* 2004;11:368-376.

9. **Peleg M, Garber JR.** Extending the GuideLine Implementability Appraisal (GLIA) instrument to identify problems in control flow. *AMIA Symp.* 2010:627-631.

10. **Shiffman RN, Michel G, Essaihi A, Thornquist E.** Bridging the guideline implementation gap: a systematic, document-centered approach to guideline implementation. *J Am Med Inform Assoc.* 2004;11:418-426.

11. **Boxwala AA, Peleg M, Tu S, et al.** GLIF3: a representation format for sharable computer-interpretable clinical practice guidelines. *J Biomed Inform.* 2004;37:147-161.

12. **Fox J, Thomson R.** Decision support and disease management: a logic engineering approach. IEEE transactions on Information Tech in Biomed. 1998;2:1-12.

13. **Shalom E, Bar G, Yarkoni A, et al.** A quantitative assessment of a methodology for collaborative specification and evaluation of clinical guidelines. *J Biomed Inform.* 2008;41:889-903.

14. **Isern D, Moreno A.** Computer-based execution of clinical guidelines: a review. *Int J Med Inform.* 2008;77:787-808.

15. **Gharib H, Papini E, Duick DS, Valcavi R, Hegedüs L, Vitti P.** AACE, ETA, AME. Guidelines for Clinical Practice for the Diagnosis and Management of Thyroid Nodules. *Endocr Pract.* 2010;16(Suppl 1):1-43.

16. **Mechanick JI, Camacho PM, Cobin RH, et al**. American Association of Clinical Endocrinologists Protocol for Standardized Production of Clinical Practice Guidelines--2010 update. *Endocr Pract.* 2010;16:270-283.

17. **Garber JR.** Thyroid Testing. Manual of Clinical Evaluation: Strategies for Cost Effective Care. 1988; 203-211.

18. **Garber JR, Berger B, Herman T, et al.** Evaluation of Thyroid Nodules. Measuring and Managing Health Care Quality: Procedures, Techniques, and Protocols. 1991; 177-186.

19. Society for Medical Decision Making. Proposal for Clinical Algorithm Standards. *Med Decis Making.* 1992;12:149-154.

20. **Shiffman RN, Dixon J, Brandt C, 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.

21. **Acosta D, Patkar V, Keshtgar M, Fox J.** Challenges in Delivering Decision Support Systems: The MATE Experience. D. Riaño et al, eds. KR4HC, Lecture Notes in Artificial Intelligence. 2010:5943; 124-140.

22. **Patkar V, Acosta D, Davidson T, Jones A, Fox J, Keshtgar M.** Using computerised decision support to improve compliance of cancer multidisciplinary meetings with evidence-based guidance. *BMJ Open.* 2012;2;e000439.

LEGENDS

FIG 1. Flowchart indicating a scheme for the diagnosis and management of palpable thyroid nodules, reproduced from (15) with added annotations 1-9 that mark discovered ambiguities, as explained in the text. Associated Key Recommendations shown in parentheses. FNA - fine-

needle aspiration; MNG - multinodular goiter; TSH - Thyroid Stimulating Hormone (thyrotropin); US - ultrasonography.

FIG 2. Version 7 of the European thyroid algorithm in GLIF3. FNA-fine needle aspiration; US-Ultrasound

FIG 3. The Tallis implementation of version 16 of the both USA and European algorithm with medium-flexibility. The screenshot shows what the user will see when he will take the decision step of selecting whether to proceed with/without calcitonin measurement. Data enquiry steps are marked as diamonds; Decision steps are marked as circles; plans are marked as round-corner rectangles; and trapezoids indicate abstract tasks without any specialization as enquiry, decision, action or plan. The Calcitonin Decision is highlighted; it is one of two steps that are currently concurrently active (Calcitonin and Scintigraphy decisions, marked with a bold contour). Steps already performed by the user are indicated by a check mark. For the patient data entered, the candidate of proceeding without calcitonin measurement is recommended (indicated with a green triangle) because there is lack of family history or clinical suspicion of MTC or MEN2 and the clinician has selected USA version of the guideline.

Table 1. Steps in computerizing a clinical guideline

Table 2. The different settings of the USA and European versions of the algorithm yield to different recommendations for the same patient case

Table 3. Indications for performing a Calcitonin test

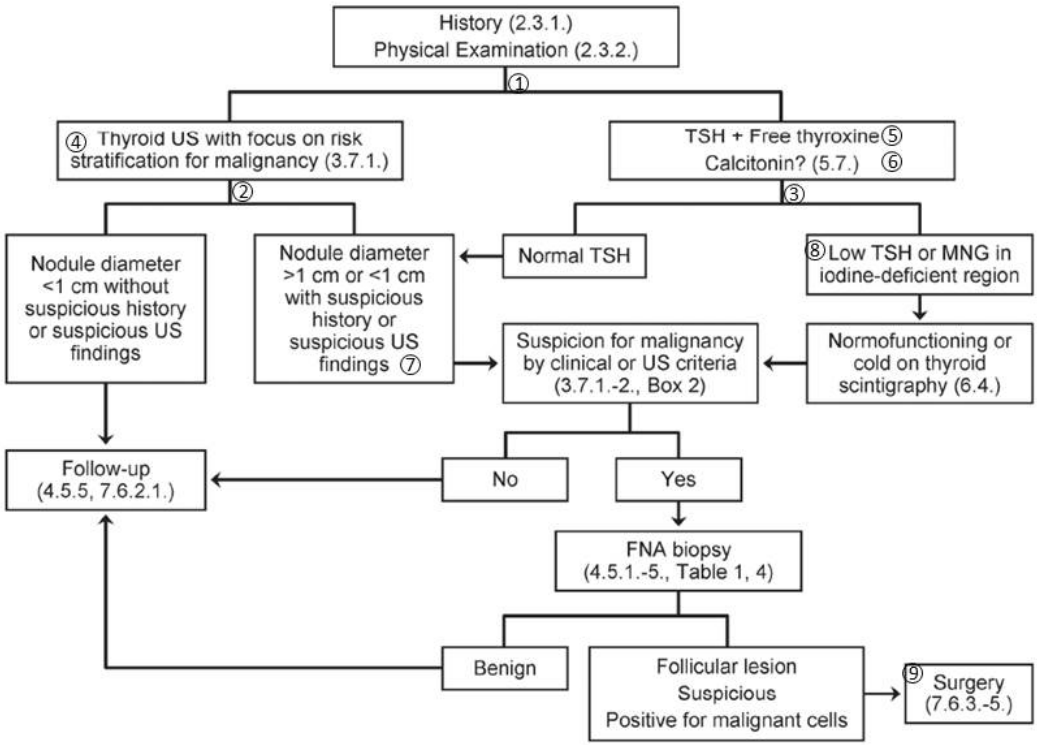

History (2.3.1.)
Physical Examination (2.3.2.)
①
④ Thyroid US with focus on risk stratification for malignancy (3.7.1.)
TSH + Free thyroxine⑤
Calcitonin? (5.7.) ⑥
②
③
Nodule diameter <1 cm without suspicious history or suspicious US findings
Nodule diameter >1 cm or <1 cm with suspicious history or suspicious US findings ⑦
Normal TSH
⑧ Low TSH or MNG in iodine-deficient region
Suspicion for malignancy by clinical or US criteria (3.7.1.-2., Box 2)
Normofunctioning or cold on thyroid scintigraphy (6.4.)
Follow-up (4.5.5, 7.6.2.1.)
No
Yes
FNA biopsy (4.5.1.-5., Table 1, 4)
Benign
Follicular lesion
Suspicious
Positive for malignant cells
⑨ Surgery (7.6.3.-5.)

Figure 1

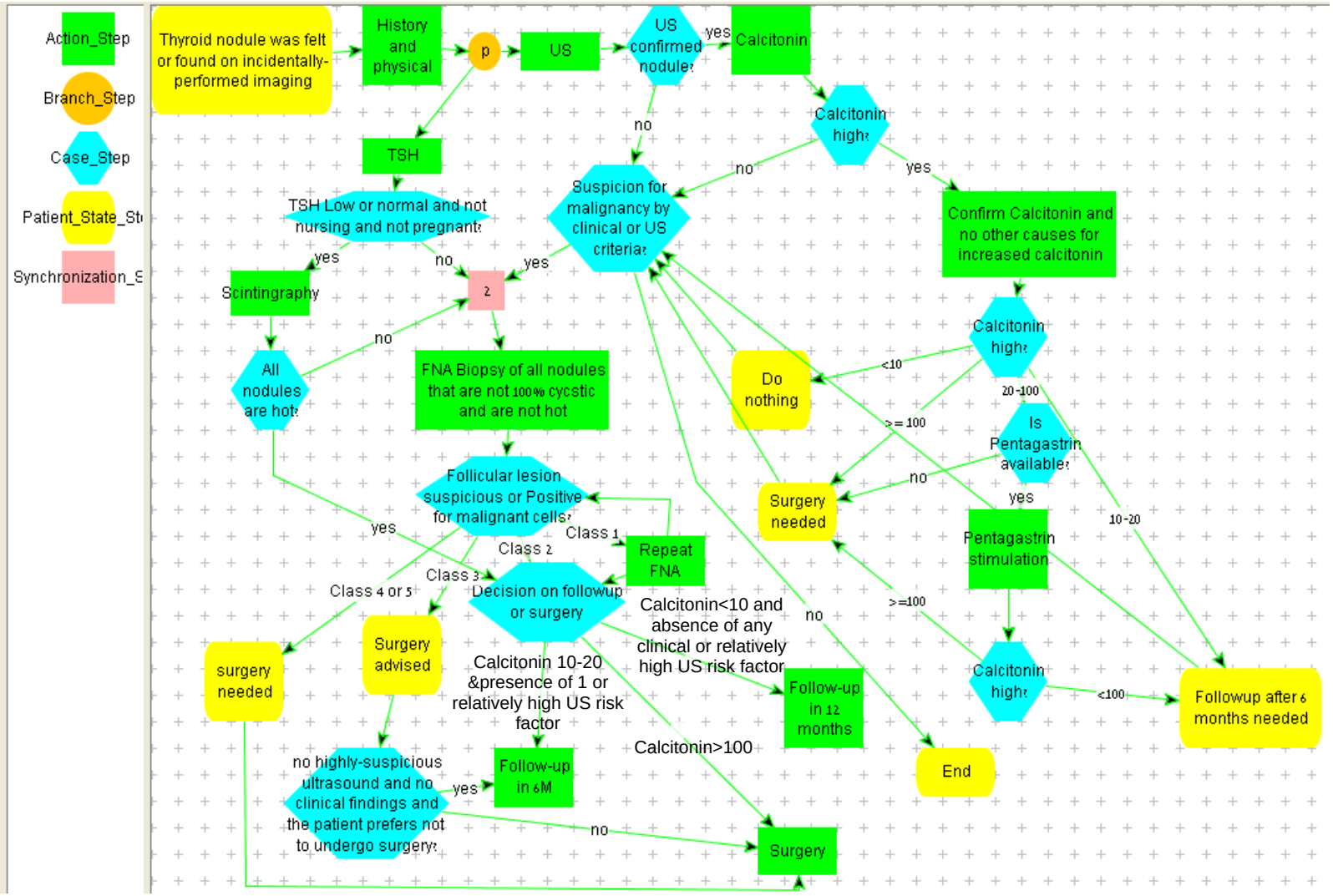
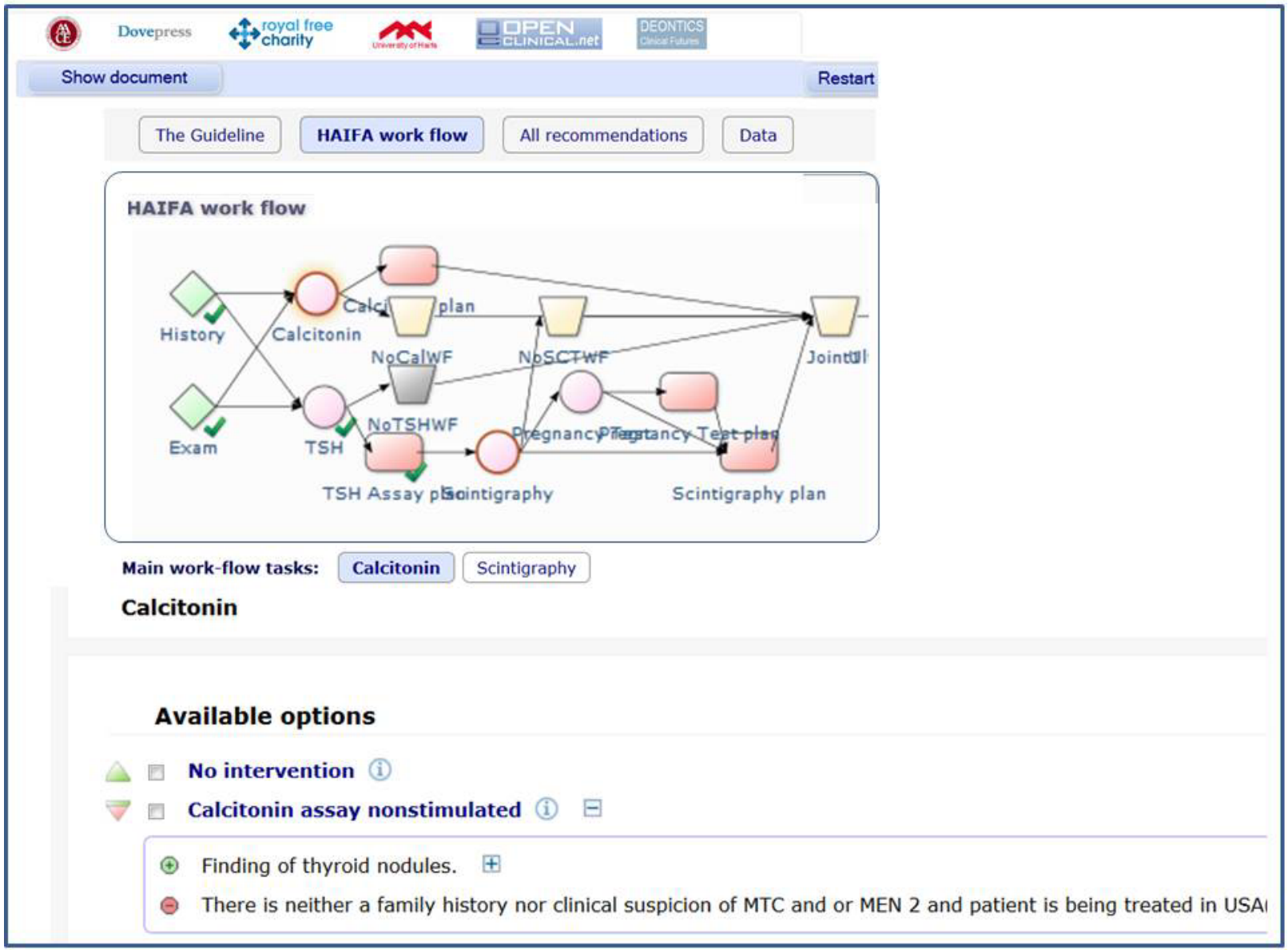


Figure 2. Version 7 of the European thyroid algorithm in GLIF3. FNA-fine needle aspiration; US-Ultrasound

DOI:10.4158/EP13271.OR



DOI:10.4158/EP13271.OR

Figure 3

Table 1.
Steps in Computerizing a Clinical Guideline

General step	Specific step	Comments
1. Choose a narrative guideline for implementation	AACE/AME/ETA guideline (15) selected by expert clinician (JRG)	Thyroid Nodule Guideline
2. Use a guideline appraisal instrument to identify ambiguity and imprecision which are barriers to guideline implementation in the narrative and flowchart	GuideLine Implementability Appraisal (GLIA) (19) extended with items for flowchart assessment (7) employed by knowledge engineer (MP) and expert clinician	Computer-science approach for eliminating ambiguity in the narrative guideline –see **Figure 1**
3. Convert narrative to flowchart using a guideline modeling language and validate it by inspection. Iterate these 2 steps until stable.	GLIF3 (9) modeling language used by the knowledge engineers (MP, SN) to create clinical algorithm, validated by expert clinician The tool used to create the GLIF3 algorithm was Protégé (protégé.stanford.edu)	GLIF3 was selected because it emphasizes the organization of a care algorithm into a flowchart – see **Figure 2**, which makes the task dependencies more clear to clinical experts. Because variations exist in medical practice between different regions and countries, as well as differences in settings (e.g., areas that may be iodine deficient) and in resources (e.g., pentagastrin-stimulation testing vs. calcium-stimulation testing of calcitonin), we decided to prepare two different algorithms: for USA and for Europe. These different versions could lead to different outcomes for the same patient case as shown in **Table 2**.
4. Convert the flowchart into a guideline modeling language that has a user-friendly web-based execution engine that can be used for executing the guideline and	The PRO*forma* (10) guideline modeling language was used by the knowledge engineers The tool used to create the PROforma algorithm and execute it was Tallis (www.cossac.org/tallis). Validation was performed	The PROforma language allows specifying indications for various actions as logical criteria, as shown in **Table 3**. Validation was done by running test cases and evaluating the recommendations offered by the execution engine

validating it with test cases	by the knowledge engineers and clinical experts	
5. Repeat validation by a panel of clinical experts	The panel of experts consisted of the first 6 (of the 7) authors of the original clinical guideline (15), who are also co-authors of this paper (HG, EP, RP, DSD, RV, and LH)	Version 16 of the USA algorithm and version 7 of the European algorithm were approved
6. Convert the flowchart into a less rigid algorithm; improve the user interface and add explanations. This implementation yields a demonstrator application	The PRO*forma* (10) guideline modeling language was used by the UK team (IC, VP, DG, MS, JF)	**Figure 3** shows a screen shot from the Tallis implementation of the version 16 of the USA algorithm represented in PRO*forma* The flexible representation allows the algorithm to work with data that is not necessarily obtained in a rigid flowchart sequence. Note that because the flexible version supports all actions mentioned in the narrative guideline, there is no need to create two different versions for USA and Europe. Explanations for recommendations point to specific paragraphs of the clinical guideline narrative. 3 modes of execution are possible: (a) maximum flexibility: a user can view all the actions and their supporting indications and may select any possible action, even if not indicated (e.g., perform scintigraphy even if TSH is not low and ultrasonography did not reveal multi nodular goiter; (b) strict ordering, as in the GLIF3 algorithm, which guides physicians through the care process, to standardize the care process and base it on the most recommended evidence-based indications; and (c) an in-between medium-flexibility representation that enforces the recommended process flow, while at the same time allowing physician users to deviate from the recommended pathway

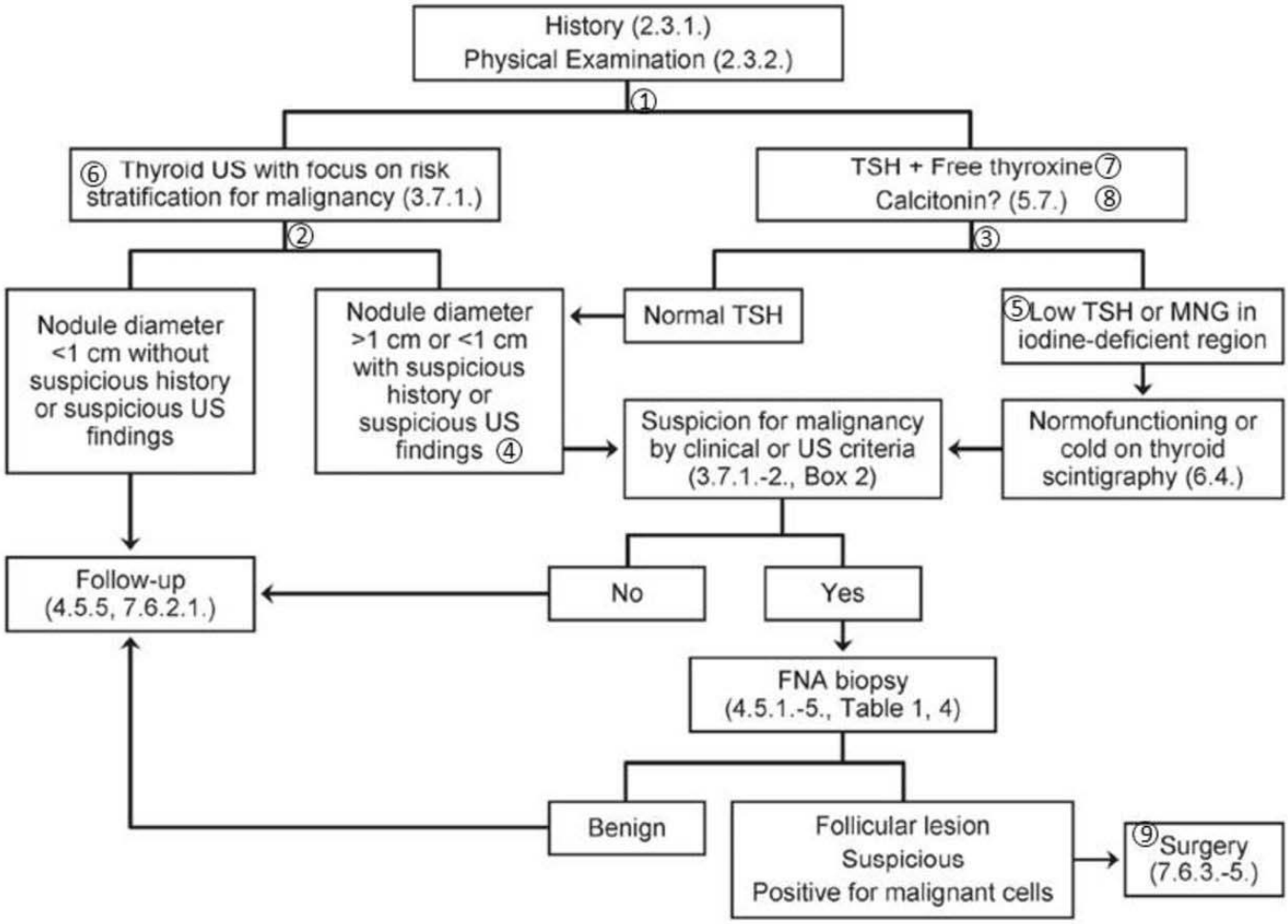
		and select any action mentioned in the guideline.
7. Validate the PRO*forma* representation with patient cases using the Tallis execution engine	Twenty patient cases were prepared by two clinical experts (JRG, JLG) along with the expected recommendations. The knowledge engineer (MP) and PRO*forma* expert (VP) executed the patient cases noting deviations from the proposed workflow (e.g., performing scintigraphy when not indicated) and the final recommendation (surgery or no surgery).	The medium flexibility version was used for evaluation

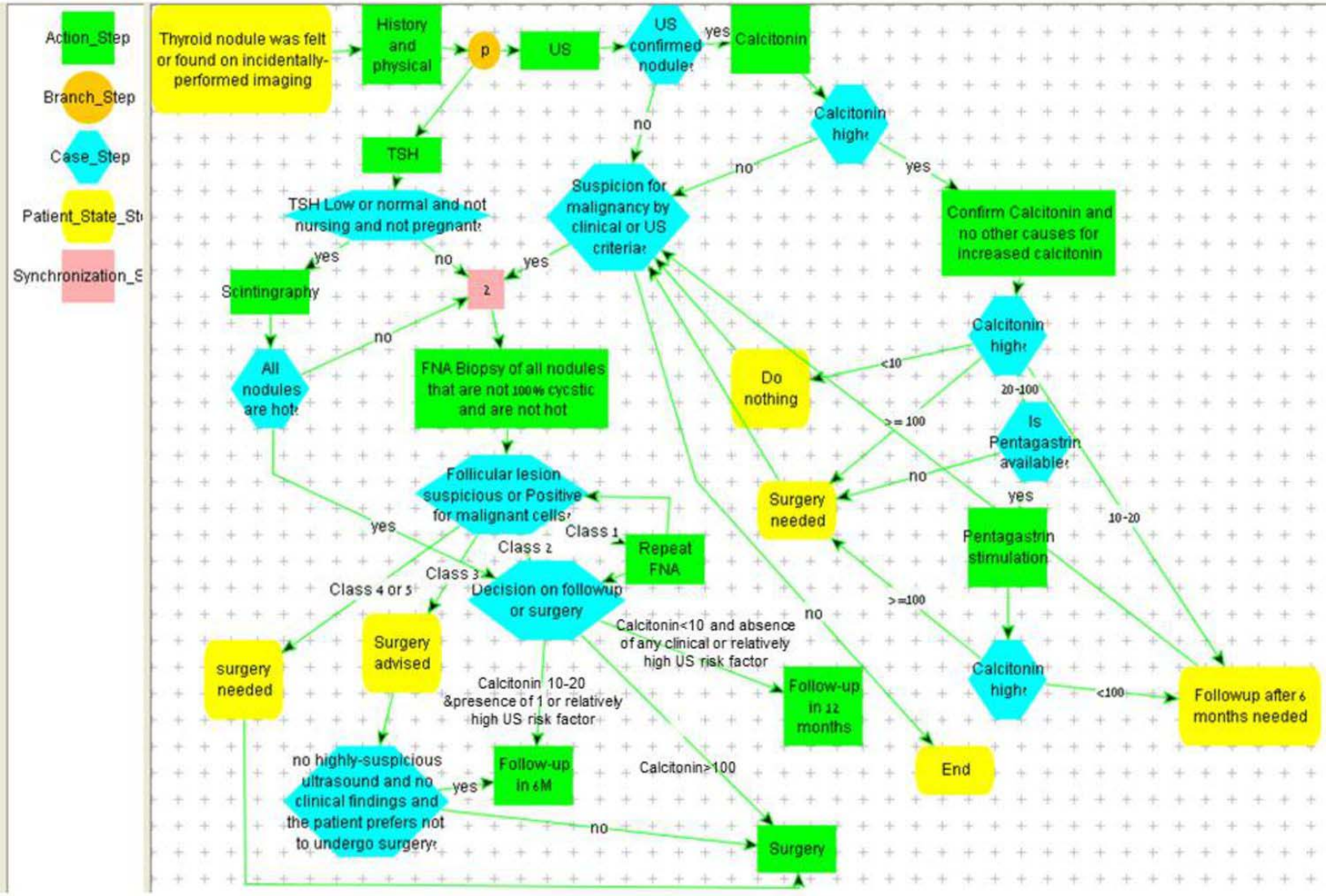
Table 2.
The Different Settings of the USA and European Versions of the Algorithm Yield to Different Recommendations for the Same Patient Case

	USA	Europe
Patient case: setting-independent characteristics	54-Year-Old Woman Referred for consultation after recent exam suggested enlarged thyroid No prior history of thyroid disease and no neck radiation Exam confirmed a small goiter with several nodules TSH 1.3 (not low) /FT4 1.4 (normal)/TPO Antibody negative	
Patient case: characteristics differ for the two settings	Iodine sufficient area Calcitonin not measured	Iodine insufficient Calcitonin 125
Algorithm recommendations and outcomes	Ultrasound: revealed findings suspicious of malignancy FNA biopsy: revealed benign nodules No surgery (just follow-up)	Ultrasound: revealed findings suspicious of malignancy Scintigraphy: not all nodules hot FNA biopsy: revealed benign nodules Although the FNA was benign surgery is indicated because of the high calcitonin level

Table 3.
Indications for Performing Calcitonin Test

Calcitonin_Argument_USA	{Guideline_setting = USA AND Thyroid_nodule_present_ultrasound_= yes AND (Family_history_of_MEN_type_2 = yes OR Family_history_of_MTC = yes OR Family_history_of_thyroid_cancer_type_unknown = yes)}
Calcitonin_Argument_Europe	Guideline_setting = Europe AND Thyroid_nodule_present_ultrasound_= yes

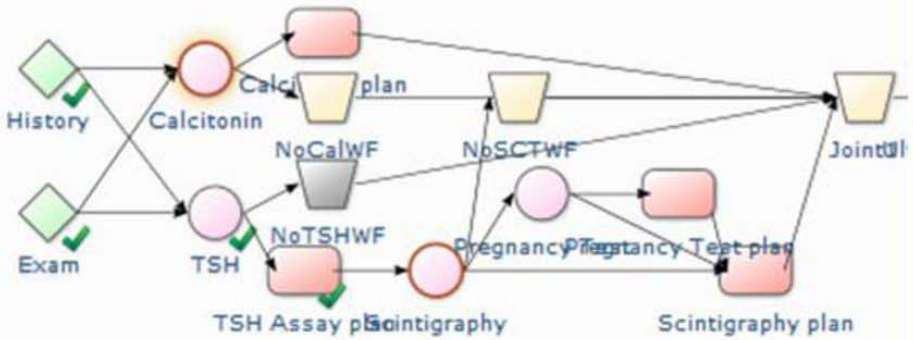

History (2.3.1.)
Physical Examination (2.3.2.)
①
⑥ Thyroid US with focus on risk stratification for malignancy (3.7.1.)
TSH + Free thyroxine⑦
Calcitonin? (5.7.) ⑧
②
③
Nodule diameter <1 cm without suspicious history or suspicious US findings
Nodule diameter >1 cm or <1 cm with suspicious history or suspicious US findings ④
Normal TSH
⑤Low TSH or MNG in iodine-deficient region
Suspicion for malignancy by clinical or US criteria (3.7.1.-2., Box 2)
Normofunctioning or cold on thyroid scintigraphy (6.4.)
Follow-up (4.5.5, 7.6.2.1.)
No
Yes
FNA biopsy (4.5.1.-5., Table 1, 4)
Benign
Follicular lesion
Suspicious
Positive for malignant cells
⑨Surgery (7.6.3.-5.)


Action_Step
Branch_Step
Case_Step
Patient_State_St
Synchronization_S
Thyroid nodule was felt or found on incidentally-performed imaging
History and physical
p
US
US confirmed nodule
yes
Calcitonin
no
Calcitonin high
yes
no
TSH
Suspicion for malignancy by clinical or US criteria
TSH Low or normal and not nursing and not pregnant
yes
no
yes
Confirm Calcitonin and no other causes for increased calcitonin
Scintingraphy
2
Calcitonin high
no
All nodules are hot
FNA Biopsy of all nodules that are not 100% cystic and are not hot
Do nothing
<10
20-100
>= 100
Is Pentagastrin available
Follicular lesion suspicious or Positive for malignant cells
no
Surgery needed
yes
10-20
Class 1
Class 2
Class 3
Repeat FNA
Pentagastrin stimulation
yes
Class 4 or 5
Decision on followup or surgery
>=100
surgery needed
Surgery advised
no
Calcitonin<10 and absence of any clinical or relatively high US risk factor
Calcitonin high
<100
Followup after 6 months needed
Follow-up in 12 months
Calcitonin 10-20 &presence of 1 or relatively high US risk factor
no highly-suspicious ultrasound and no clinical findings and the patient prefers not to undergo surgery
Follow-up in 6M
Calcitonin>100
End
yes
no
Surgery

Show document | Restart

The Guideline | **HAIFA work flow** | All recommendations | Data

HAIFA work flow



Main work-flow tasks: **Calcitonin** | Scintigraphy

Calcitonin

Available options

- **No intervention**
- **Calcitonin assay nonstimulated**
  - Finding of thyroid nodules.
  - There is neither a family history nor clinical suspicion of MTC and or MEN 2 and patient is being treated in USA

Table 1. Steps in computerizing a clinical guideline

General step	Specific step	Comments
1. Choose a narrative guideline for implementation	AACE/AME/ETA guideline (15) selected by expert clinician (JRG)	Thyroid Nodule Guideline
2. Use a guideline appraisal instrument to identify ambiguity and imprecision which are barriers to guideline implementation in the narrative and flowchart	GuideLine Implementability Appraisal (GLIA) (19) extended with items for flowchart assessment (7) employed by knowledge engineer (MP) and expert clinician	Computer-science approach for eliminating ambiguity in the narrative guideline –see **Figure 1**
3. Convert narrative to flowchart using a guideline modeling language and validate it by inspection. Iterate these 2 steps until stable.	GLIF3 (9) modeling language used by the knowledge engineers (MP, SN) to create clinical algorithm, validated by expert clinician The tool used to create the GLIF3 algorithm was Protégé (protégé.stanford.edu)	GLIF3 was selected because it emphasizes the organization of a care algorithm into a flowchart – see **Figure 2**, which makes the task dependencies more clear to clinical experts. Because variations exist in medical practice between different regions and countries, as well as differences in settings (e.g., areas that may be iodine deficient) and in resources (e.g.,

		pentagastrin-stimulation testing vs. calcium-stimulation testing of calcitonin), we decided to prepare two different algorithms: for USA and for Europe. These different versions could lead to different outcomes for the same patient case as shown in **Table 2**.
4. Convert the flowchart into a guideline modeling language that has a user-friendly web-based execution engine that can be used for executing the guideline and validating it with test cases	The PRO*forma* (10) guideline modeling language was used by the knowledge engineers The tool used to create the PROforma algorithm and execute it was Tallis (www.cossac.org/tallis). Validation was performed by the knowledge engineers and clinical experts	The PROforma language allows specifying indications for various actions as logical criteria, as shown in **Table 3**. Validation was done by running test cases and evaluating the recommendations offered by the execution engine
5. Repeat validation by a panel of clinical experts	The panel of experts consisted of the first 6 (of the 7) authors of the original clinical guideline (15), who are also co-authors of this paper (HG, EP, RP, DSD, RV, and LH)	Version 16 of the USA algorithm and version 7 of the European algorithm were approved

6. Convert the flowchart into a less rigid algorithm; improve the user interface and add explanations. This implementation yields a demonstrator application	The PRO*forma* (10) guideline modeling language was used by the PRO*forma* experts (IC, VP, DG, MS, JF)	**Figure 3** shows a screen shot from the Tallis implementation of the version 16 of the USA algorithm represented in PROforma The flexible representation allows the algorithm to work with data that is not necessarily obtained in a rigid flowchart sequence. Note that because the flexible version supports all actions mentioned in the narrative guideline, there is no need to create two different versions for USA and Europe. Explanations for recommendations point to specific paragraphs of the clinical guideline narrative. 3 modes of execution are possible: (a) maximum flexibility: a user can view all the actions and their supporting indications and may select any possible action, even if not indicated (e.g., perform scintigraphy even if TSH is not low and ultrasonography did not reveal multi nodular goiter; (b) strict ordering, as in the GLIF3 algorithm, which guides physicians through the care process, to

		standardize the care process and base it on the most recommended evidence-based indications; and (c) an in-between medium-flexibility representation that enforces the recommended process flow, while at the same time allowing physician users to deviate from the recommended pathway and select any action mentioned in the guideline.
7. Validate the PROforma representation with patient cases using the Tallis execution engine	Twenty patient cases were prepared by two clinical experts (JRG, JLG) along with the expected recommendations. The knowledge engineer (MP) and PRO*forma* expert (VP) executed the patient cases noting deviations from the proposed workflow (e.g., performing scintigraphy when not indicated) and the final recommendation (surgery or no surgery).	The medium flexibility version was used for evaluation

Table 2. The different settings of the USA and European versions of the algorithm yield to different recommendations for the same patient case

	USA	Europe
Patient case: setting-independent characteristics	54-Year-Old Woman Referred for consultation after recent exam suggested enlarged thyroid No prior history of thyroid disease and no neck radiation Exam confirmed a small goiter with several nodules TSH 1.3 (not low) /FT4 1.4 (normal)/TPO Antibody negative	
Patient case: characteristics differ for the two settings	Iodine sufficient area Calcitonin not measured	Iodine insufficient Calcitonin 125
Algorithm recommendations and outcomes	Ultrasound: revealed findings suspicious of malignancy FNA biopsy: revealed benign nodules No surgery (just follow-up)	Ultrasound: revealed findings suspicious of malignancy Scintigraphy: not all nodules hot FNA biopsy: revealed benign nodules Although the FNA was benign surgery is indicated because of the high calcitonin level

Table 3. Indications for performing Calcitonin test

Calcitonin_Argument_USA	{Guideline_setting = USA AND Thyroid_nodule_present_ultrasound_= yes AND (Family_history_of_MEN_type_2 = yes OR Family_history_of_MTC = yes OR Family_history_of_thyroid_cancer_type_unknown = yes)}
Calcitonin_Argument_Europe	Guideline_setting = Europe AND Thyroid_nodule_present_ultrasound_= yes