Can Physicians Structure Clinical Guidelines? Experiments with a Mark-Up-Process Methodology

Erez Shalom[1], Yuval Shahar[1], Meirav Taieb-Maimon[1], Guy Bar[2], Susana B. Martins[3], Ohad Young[1], Laszlo Vaszar[3], Yair Liel[2], Avi Yarkoni[2], Mary K. Goldstein[3], Akiva Leibowitz[2], Tal Marom[4], and Eitan Lunenfeld[2]

[1] The Medical Informatics Research Center, Ben Gurion University, Beer Sheva, Israel
{erezsh,yshahar,ohadyn,meiravta}@bgu.ac.il
http://medinfo.ise.bgu.ac.il/medlab
[2] Soroka Medical Center, Ben Gurion University, Beer Sheva, Israel
{guybar,yarkoni,liel,akival,lunenfld}@bgu.ac.il
[3] Veterans Administration Palo Alto Heath Care System, Palo Alto, CA
{Susana.Martins,Mary.Goldstein}@va.gov, laszlo@vassar.info
[4] E.Wolfson Medical Center, Holon, Israel
maromtal@013.net.il

Abstract. We have previously developed an architecture and a set of tools called the Digital electronic Guideline Library (DeGeL), which includes a web-based tool for structuring (marking-up) free-text clinical guidelines (GLs), namely, the URUZ Mark-up tool. In this study, we developed and evaluated a methodology and a tool for a mark-up-based specification and assessment of the quality of that specification, of procedural and declarative knowledge in clinical GLs. The methodology includes all necessary activities before, during and after the mark-up process, and supports specification and conversion of the GL's free-text representation through semi-structured and semi-formal representations into a machine comprehensible representation. For the evaluation of this methodology, three GLs from different medical disciplines were selected. For each GL, as an indispensable step, an ontology-specific consensus was created, determined by a group of expert physicians and knowledge engineers, based on GL source. For each GL, two mark-ups in a chosen GL ontology (Asbru) were created by a distinct clinical editor; each of the clinical editors created a semi-formal mark-up of the GL using the URUZ tool. To evaluate each mark-up, a gold standard mark-up was created by collaboration of physician and knowledge engineer, and a specialized mark-up-evaluation tool was developed, which enables assessment of completeness, as well as of syntactic and semantic correctness of the mark-up. Subjective and objective measures were defined for qualitative and quantitative evaluation of the correctness (soundness) and completeness of the marked-up knowledge, with encouraging results.

Keywords: Clinical decision support systems, Clinical guidelines, Ontologies, Knowledge acquisition, Evaluation, Mark-up, Completeness, Correctness.

1 Introduction

Clinical guidelines (GLs) have been shown to improve the quality of medical care, and are expected to assist in containment of its costs as well [1]. During the past 20

years, there have been several efforts to support complex GL-based care over time in automated fashion. This kind of automated support requires formal GL-modeling methods. Most of the methods use knowledge acquisition tools for eliciting the medical knowledge needed for the knowledge roles (KRs) of the GL specification *ontology* (key concepts, properties and relations) that each method assumes, in order to specify it in a formal format [2-10]. In most of those tools, however, the process of specification the textual GL into a formal language is not sufficiently smooth and transparent.

The core of the problem is that expert physicians cannot (and need not) program in GL specification languages, while programmers and knowledge engineers do not understand the clinical semantics of the GL. Therefore, we developed an architecture and set of tools, called the Digital electronic Guideline Library (DeGeL) [11], to support GL classification, semantic mark-up, context-sensitive search, browsing, runtime application, and retrospective quality assessment. DeGeL facilitates the process of specification the GL into a formal language, and enables collaboration and inherent iteration between different types of users such as: expert physicians, namely, senior, domain-expert clinicians who assist in formation of a clinical consensus that disambiguates the GL; clinical editors, namely, medically trained editors who mark-up the GL, and knowledge engineers, typically informatics experts who can create a formal GL representation.
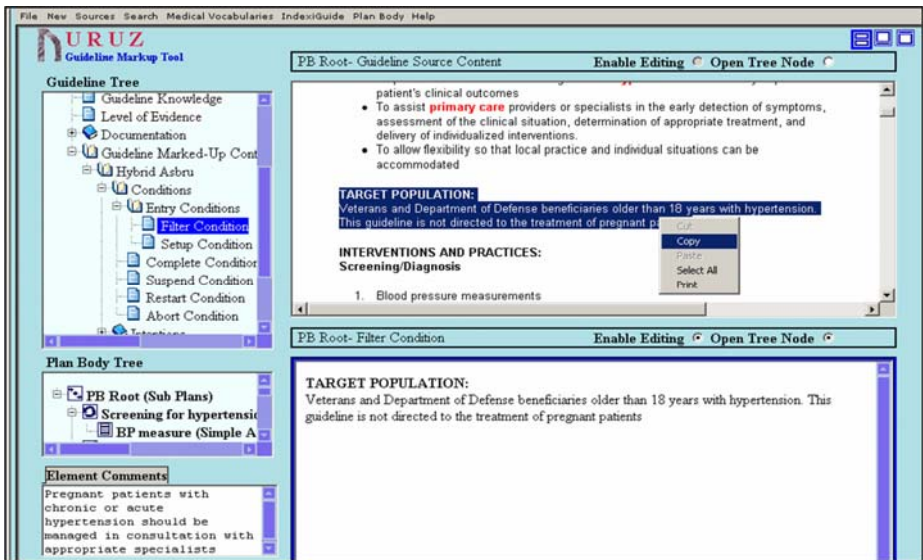

Fig. 1. The Uruz Web-based guideline (GL) mark-up tool in the DeGeL architecture. The tool's basic semi-structuring interface is uniform across all GL ontologies. The target ontology selected by the clinical editor, in this case, Asbru, is displayed in the upper left tree; the GL source is opened in the upper right frame. The clinical editor highlights a portion of the source text (including tables or Figures) and drags it for further modification into the bottom frame's Editing Window tab labeled by a semantic role chosen from the target ontology (here, the Asbru filter condition). Contents can be aggregated from different source location.

In addition, The DeGeL library supports multiple GL ontologies, in each of which, GLs can be represented in a hybrid format. One of the DeGeL framework tools is the web-based URUZ mark-up tool (Figure 1). URUZ uses the infrastructure of DeGeL's hybrid guideline representation model and thus enables expert physicians, clinical editors, and knowledge engineers on different sites to collaborate in the process of GL specification and to mark-up the GL in any representation level: semi-structured (typically performed by the clinical editor), semi-formal (typically performed by the clinical editor in collaboration with the knowledge engineer), and formal representation (typically performed by the knowledge engineer).
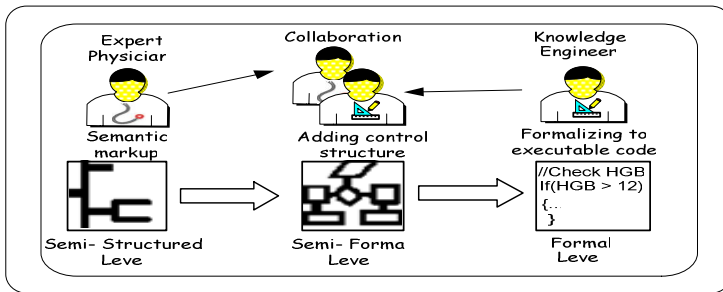

Fig. 2. The incremental specification process in the URUZ mark-up tool

This incremental process is shown in Figure 2: The expert physician indexes and mark-ups the GL (in this study, mark-up means structuring the GL text by labeling portions of text using semantic labels from chosen target GL specification language, sometimes even modifying the text into semi-formal representation), creating semi-structured representation, and in collaboration with a knowledge engineer, creating semi-formal GL representation. Then, knowledge engineers use an ontology-specific tool to add executable expressions in the formal syntax of the target ontology. Thus, each GL is represented in one or more representations levels: free-text, semi-structured text, semi-formal, and a fully structured representation. All of those GL's representation levels co-exist and are organized in the DeGeL library within a unified structure - the hybrid GL representation

2   Objectives

In this study, we aimed to answer three categories of research questions: (i) Can clinical editors actually mark-up a GL into semi-formal level, and if so, at what levels of completeness (coverage) and of soundness, or correctness (semantic and syntactic quality)? (ii) Are there significant differences in the completeness and correctness of the mark-ups between different clinical editors marking-up the same GL? (iii) Is there a significant difference between the correctness levels of the mark-up of different specific aspects of the GL?

3   Methods: The Mark-Up Process and Evaluation Methodology

The activities in the mark-up process include three main phases (Figure 3): 1) Preparations *before* the mark-up activities: choosing the specification language, learning the specification language, selecting a GL for specification, create an Ontology-Specific Consensus, acquiring training in the mark-up tool and making a Gold Standard (GS) mark-up 2) *During* the Mark-up activities: Classifying the GL according to a set of semantic indices (e.g. diagnosis, treatment), and performing the specification process using the tools and consensus, 3) After Mark-up activity: evaluation of the results of GL specification. This methodology implements the incremental specification process described in Figure 2 (particularly the structured-text and semi-formal representation levels).

3.1   Choosing the Specification Language

The first step towards specification of GL into semi-formal mark-up is to select the target GL ontology for specification (activity 1 in Figure 3). As pointed previously [13,14], different ontologies can be used for different proposes: one might select for example PRODIGY[3] for creation of common scenarios of chronic diseases, GLIF[4] or Proforma[5] for general GL modeling, or GEM[7] for documentation. In this study we used the Asbru language [12] as the underlying guideline-representation language due to its expressive procedural structures and its explicit representation: the Asbru specification language includes semantic KRs organized in KR-Classes such as Conditions (e.g., the filter condition, which represents obligatory eligibility criteria, the complete condition, which halts the guideline execution when some predefined criteria is true, and the abort condition, which aborts the guideline execution when some predefined criteria is true); control structures for the GL's Plan-Body KR-Class (e.g., sequential, concurrent, and repeating combinations of actions or sub-guidelines), GL's Intentions KR-Class (e.g. process and outcome intentions), and the Context KR-Class of the activities in the GL (e.g. actors, clinical-context).
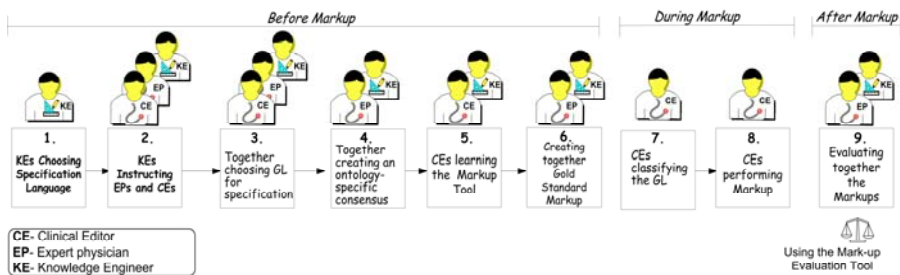

Fig. 3. The three main phases of the methodology before, during and after the mark-up, and the activities in each phase. Note the descriptions under each activity. Activity six (creation of a gold standard) can be performed in start, or in parallel with activities seven and eight (editors mark-up). Note also the participants in each activity. Note that this methodology is implements the structured-text and semi-formal representation levels which are described in Figure 2. Note step 9 is facilitated with the Evaluation mark-up tool (see section 3.8).

3.2 Learning the Specification Framework and Tools

The expert physicians and the clinical editors are instructed by the knowledge engineers about the essential concepts and aspects required for the specification process (activity 2 in Figure 3). This activity includes learning main aspects of the specification language (Asbru in our case), the hybrid model and its representation levels, the overall GL representation framework (in this case DeGeL) and its related tools. Learning the specification language will help, for example, the expert physicians to select a GL appropriate for formal representation considering its semantic aspects, and not only its clinical ones.

3.3 Selecting a Guideline for Specification

Once a decision has been made to automate a set of GLs in some clinical settings, the next step is to decide which GLs are to be formalized (activity 3 in Figure 3). A good candidate GL for specification should be one of a common disease with agreement between the majority of expert physicians on the methods of diagnosis and treatment, and with a clear, well defined clinical pathway. Thus, they might select more than one GL source when, for example, information regarding some directive is defined in more detail in another source. These sources, in addition to their own knowledge and interpretations, will serve as the basis of knowledge for creating the consensus (see Figure 4).

3.4 Creating an Ontology-Specific Consensus

In general, the textual content of the GL is not always complete or self-evident within itself: it might lack of sufficient information, suffer from ambiguousness, or require customization to local settings. Therefore, local, site-specific customizations of the GL (say, motivated by the availability of resources in the local clinical setting, by local practices, or by personal experience) must be specified explicitly to increase the probability of site-specific successful application. In addition, the same free-text GL might sometimes be interpreted differently by several local expert clinicians; in such cases, much discussion can be prevented by an explicit agreement on a common local interpretation. In their recent research Peleg et al.[15] found that in the process of making algorithms from textual GLs, teamwork is crucial for detecting errors, and that the team should include a knowledge engineer. Thus, similarly to Miller et al. [16], we too found the creation of a local clinical consensus regarding the semantics of the GL to be an indispensable, mandatory step before mark-up, and it should include both the expert physicians and the knowledge engineers (activity 4 in Figure 3).

We decomposed the crucial mark-up phase into two steps: In the first step, the local most senior expert physicians first created a clinical consensus, which was independent of any GL-specification ontology. The clinical consensus was created, for each GL, by the local senior expert physicians and a knowledge engineer. The clinical consensus was always a structured document that described in a schematic-only, but explicit, fashion the interpretation of the clinical directives of the GL, as agreed upon by the local expert physicians.

In the second step, we created what we refer to as an *Ontology-Specific Consensus* (OSC), which specifies the consensus, in terms of the chosen target ontology KRs (e.g., entry conditions). The OSC is created by senior expert physicians, who had considerable practical knowledge and experience in the relevant clinical domain, in collaboration with a knowledge engineer, who was familiar with the target specification language. Note that The OSC is a merged result of merging these two steps, and is is independent of a specification tool. An example of a free-text segment of a GL, a part of an OSC generated from that GL is in Figures 4, and 5. A detailed exposition regarding the process of creating an OSC and the evaluation of its effects can be found elsewhere [17].
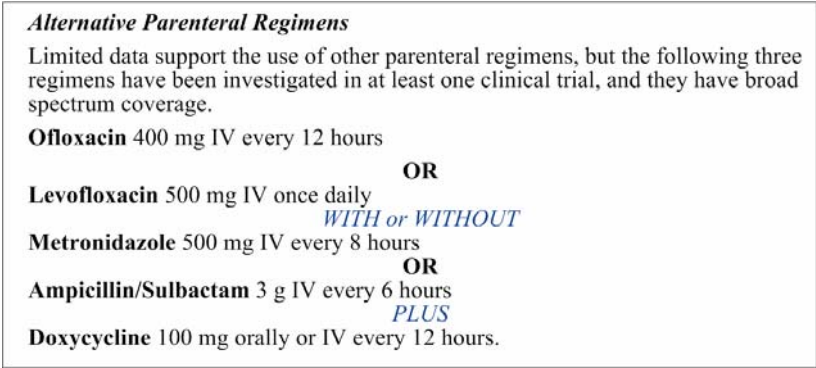

Alternative Parenteral Regimens

Limited data support the use of other parenteral regimens, but the following three regimens have been investigated in at least one clinical trial, and they have broad spectrum coverage.

Ofloxacin 400 mg IV every 12 hours

<div align="center">OR</div>

Levofloxacin 500 mg IV once daily
<div align="right">*WITH or WITHOUT*</div>
Metronidazole 500 mg IV every 8 hours

<div align="center">OR</div>

Ampicillin/Sulbactam 3 g IV every 6 hours
<div align="right">*PLUS*</div>
Doxycycline 100 mg orally or IV every 12 hours.

Fig. 4. A sample of the text of the "Pelvic Inflammatory Disease" guideline's (original) source (demonstrating activity 3 in Figure 3)
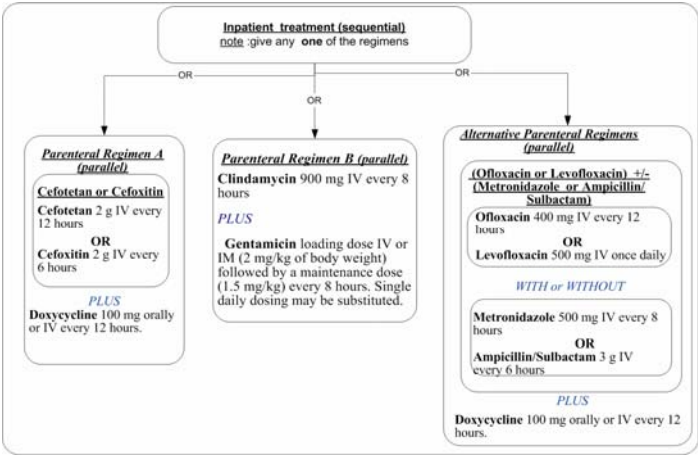

Fig. 5. A sample of the ontology specific consensus created for the "Pelvic Inflammatory Disease" guideline, demonstrating activity 4 in Figure 3. In this case, the consensus is specific to the Asbru ontology and is a small part of the consensus document that refers to the guideline shown in Figure 4.

3.5   Training the Clinical Editors in the Mark-Up Tool

The clinical editors who intend to perform the semi-formal mark-ups are instructed by the knowledge engineer in the specification tool (URUZ mark-up tool in our case) and in the OSC. A help manual or small simulation of marking-up another GL may be used to assist the clinical editor in this process of refreshing the meaning of some concepts or to confirm understanding about a task related to the specification language, the specification tool, or generally about the specification process (activity 5 in Figure 3) .

3.6   Creating the Gold Standard Mark-Up

For each of the GLs, a gold standard is created by a senior expert physician (other expert physician than the one who is participated in the mark-up activity) and a knowledge engineer together, using the OSC and GL source (activity 6 in Figure 3). The gold standard considered to be the best semi-formal mark-up and most detailed from both the clinical and semantic aspects, and therefore is used only for evaluation of the mark-ups. This step can be performed before or in parallel to the activity of performing the mark-ups.

3.7   Performing the Semi-formal Mark-Up

After the clinical editor feels enough confident, he/she can start to specify the GL using the mark-up tool (URUZ, in our case) according to the OSC and the GL sources, and create a semi-formal mark-up. In addition, the clinical editor classifies the GL according to a set of semantic indices (e.g. diagnosis, treatment) using the IndexiGuide Tool [11]. However, the knowledge engineer is not part of this session, and may help the clinical editor in case of technical problems (activities 7 and 8 in Figure 3).

3.8   Evaluation of the Mark-Ups

After the semi-formal mark-ups are completed, they are evaluated by comparing them to the gold standard according to objective completeness and correctness measures. Evaluation of the mark-up is important because it helps to qualify its quality in qualitative and quantitative measures. This stage is done by a senior expert physician and a knowledge engineer using a designated evaluation mark-up tool which enables scoring those measures for each plan, sub-plan and KR of evaluated mark-up (activity 9 in Figure 3, see next section).

3.8.1   The Evaluation Mark-Up Tool
To facilitate the evaluation, we developed a tool specialized designed to produce completeness and correctness scores: The Mark-Up Assessment Tool (MAT) [18]. The MAT is a web-based desktop application which was developed using the Dot.Net technology and enables sharing and collaboration between different sites and users. The MAT enables the expert physician, the knowledge engineer, and other guests to select and browse the desired evaluated mark-up from the DeGeL library. A typical

evaluation session usually includes a physician who is an expert in the clinical domain of the evaluated marked-up GL, and a knowledge engineer who is familiar with Asbru semantics (see Figure 6).

There are two possible working modes in the MAT in each evaluation session: *View* mode and *Evaluation* mode. When one of the *evaluation managers* (usually the knowledge engineer) starts the evaluation session, he opens the MAT in the Evaluation mode, and enters the participants in the appropriate fields in the session: the names of the expert physicians, the knowledge engineers and the optional guests. In the View mode, the users can only view the mark-up. Thus, all other participants at different sites and locations in the session can open the MAT in parallel in View mode. MAT's functionality enables all online participates to see the changes made by the expert physician and by the knowledge engineer who entered in an Evaluation mode, during the evaluation session. When the evaluation manager enters the MAT, he attaches to the marked-up GL a relevant ontology-specific-consensus file. For example, for the Pelvic Inflammatory Disease (PID) mark-up, the ontology-specific-consensus of the PID GL is attached.

Fig. 6. The Login form of the Mark-up Assessment Tool. In the upper right frame the user can select the working mode (Evaluation or View). In the drop down list the user selects the appropriate mark-up to evaluate, and uses the "browse" button to attach an ontology specific consensus file. If the Evaluation mode is selected, the user enters the names of the knowledge engineer and the expert physician who are going to manage the evaluation session. Note the optional fields for guests.
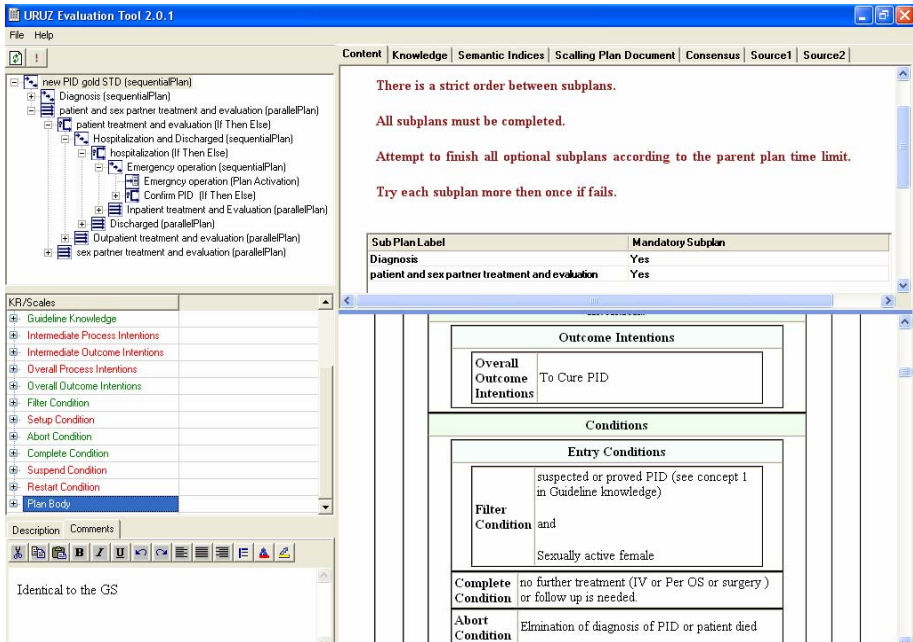
Fig. 7. The main interface of Mark-up Assessment Tool: In the upper left panel, the tree of the plans of the marked-up GL is shown. When the evaluation manager (typically the knowledge engineer) selects a plan, its procedural content (i.e., the type of plan-body, order and plans if the type is subplan, etc.) is loaded to the upper right frame, and the declarative content (i.e., all the plans' knowledge roles and their textual content) is loaded to the bottom right frame. The bottom left panel is used for giving the scores of completeness and correctness to the selected knowledge roles, according to the content quality. In addition, whenever necessary, the evaluation manager, or other participants, can use multiple tabs to view the textual source of the guideline, the ontology-specific consensus, and the guideline knowledge.

3.8.2 Assessment of the Guideline Mark-Ups Using the MAT

In the left bottom panel of the MAT main interface (Figures 7, 8), a list of checkboxes facilitates the scoring of the *completeness* and of the *correctness* measures: for each score of completeness and correctness there is a checkbox; thus, the evaluation manager, who is usually the knowledge engineer, can check the desired score in the relevant checkbox. In an evaluation session, the evaluation manager together with the expert physician opens two instances of the MAT: one for the evaluated mark-up and one for the gold-standard. Then, for each plan, starting with the root plan in the gold-standard instance, they look in the evaluated mark-up instance for a similar plan and decide on its completeness, and check the appropriate completeness checkboxes, indicating whether the plan is missing, exists or redundant, with respect to the gold-standard mark-up of that GL (see section 3.6).

In addition, the *content* of each KR of each plan (existing or redundant) is evaluated according to both syntactic and semantic measures; In this case, according to the Asbru and Clinical correctness measures. For each score [-1,0,1] (i.e, a clear error, a mild

error, and a correct mark-up)of the two correctness measures there is an appropriate checkbox which the evaluation manager can check after discussing it with the expert physician . The definitions for all evaluation scores displayed as well (see Figure 8).

Thus, during the evaluation session the knowledge engineer and the expert physician are collaborating by checking the appropriate checkboxes of the completeness and correctness for each plan and KR. Additional checkboxes list each type of error, enabling them to report the error and its type (clinical semantic or Asbru syntactic error) , and what kind of specific error, again by checking the appropriate checkbox.
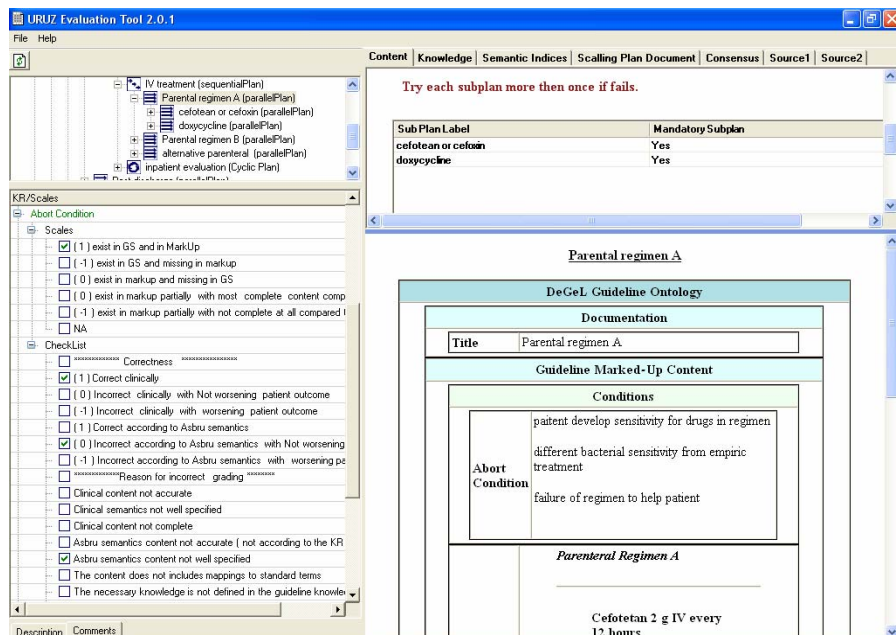

Fig. 8. Evaluation of the mark-up using MAT. Note the selected plan in the left above panel. Note the checkboxes for the completeness and correctness measures in the left bottom panel, in this case of the Abort Condition knowledge role(KR). Note also the content of the procedural content of the KR in the right above panel, and its textual content in the right bottom panel.

4   Evaluation of the Methodology

For the evaluation of this methodology, three GLs, from three distinct medical disciplines, were selected for use as the textual source for mark-ups: Pelvic inflammatory disease (PID)[19], Chronic Obstructive Pulmonary Disease (COPD) [20], and Hypothyroidism [21]. In the first stage, an OSC was created for each GL. After learning the Asbru specification language and the DeGeL framework, and receiving training in the URUZ tool, each of the clinical editors created a semi-formal mark-up using the URUZ and the OSC, and his own knowledge (for each GL two mark-ups were created

by distinct clinical editors). In total six semi-formal mark-ups were created. In order to evaluate each of the mark-ups created by the clinical editors, a gold standard mark-up was created as well. Each of the mark-ups created by the clinical editors was compared to a gold standard mark-up. In addition, subjective and objective measures were defined for qualitative and quantitative evaluation of each plan, sub-plan and KR in each of the mark-ups: the subjective measures included several questionnaires to evaluate the clinical editor's attitude regarding the DeGeL framework, the specification language, the usability of the URUZ tool and their subjective benefits The objective measures were defined in two main categories: a completeness measure of the acquired knowledge, i.e., how much content from the gold standard exists (or not) in each of the semi-formal mark-ups of each clinical editor (for example, a predefined set of plans), and a correctness measure, i.e., accuracy of the acquired knowledge is according to the aspects of 1) Clinical semantics and 2) Asbru semantics. For both aspects, the score was defined according to the trichotomic scale of +1 (correct), 0 (incorrect without causing a worsening of the patient's prognosis), and -1 (incorrect and causing a worsening of the patient's prognosis). Finally, the expert physician and the knowledge engineer collaborated to perform the evaluation, using a designated graphical tool, which enables scoring the measures the mark-ups simultaneously from different sites by different users.

5   Results

A detailed analysis of the results, including the subjective measures and the measures of structuring specific KRs and a comparison of the quality of the mark-ups by distinct clinical editors for each GL is outside of the scope of this paper, which focuses mainly on the evaluation methodology. The detailed results can be found elsewhere [22,23]. However, it can be noted that the evaluation of the methodology was very encouraging when using Asbru as the specification language, DeGeL as the representation framework and URUZ as the mark-up tool. Table 1 summaries the results for all GLs: Overall, there were 196 different plans and sub-plans to be marked-up in total for the three GLs by the clinical editors: 106 plans in the case of PID GL, 59 in the COPD GL and 31 in the Hypothyroidism GL. In addition, a total number of 326 KRs were evaluated and scored in a clinical measure and an Asbru semantic measure for each plan and sub-plans in each mark-up for all GLs: 180 KRs in the case of the PID GL, 97 KRs for the COPD GL, and 49 for the Hypothyroidism GL. The completeness of the specification of all the KRs in all mark-ups was very high, with a mean of 91% ± 0.11 for all the GLs. Regarding correctness, there was quite a variability between the GLs (mean of 0.6 ± 0.7 on a scale of [-1,1]), but the quality of specification for each mark-up was always considerably higher than zero. The levels of clinical and syntactical errors were different in each GL due to the differences in the level of detail in the OSC of each GL: as the OSC was more detailed in the clinical and Asbru aspects, the correctness measures were higher. The quality of acquiring complex procedural KRs (such as cyclical plan or specification of two plans in parallel or sequential) by the clinical editors was slightly lower than acquiring declarative KRs (such as GL's filter condition).

Table 1. A summary of the completeness and correctness measures for all guideline mark-ups. Completeness is presented as a percentage; correctness is presented on a scale of [-1, 1].

KR class	Number	Measure - All GLs	
		Completeness % ± SD	Correctness [-1, 1]. ± SD
Context	47	68 ± 0.01	0.39 ± 0.92
Intentions	26	96 ± 0.04	0.89 ± 0.41
Conditions	60	88 ± 0.1	0.45 ± 0.69
Plan- Body	193	97 ± 0.02	0.66 ± 0.65
All KR Classes	326	91 ± 0.11	0.6 ± 0.7

6 Discussion

This study focused on a methodology for evaluation of a mark-up-based specification of GLs, and describes its main phases before, during and after mark-up, and the activities and participants in each phase. We presented the summarized results of a case study for this methodology, which used subjective and objective measures to compare the mark-up results to a gold standard mark-up (In practice, the gold standard can be replaced by a process of assessing the quality of the mark-up). A relative limitation is that the study used only three guidelines and six mark-up editors; however, the number of plans and subplans (196) and KRs (326) was quite high. Note that the evaluation measures can be used to analyze multiple different cross-sections: for each KR, KR-Class, all KRs in a mark-up, for each GL and for all GLs. It is also possible, for example, to classify a sample of the KRs marked-up by the clinical editors during the mark-up activity, thus supporting an ongoing quality assessment of the mark-up process [22].

7 Conclusions and Future Work

Clinical editors are capable of performing semi-formal mark-up of a GL with high completeness and correctness. However, the collaboration of an expert physician, a clinical editor, and a knowledge engineer is crucial for successful formal specification. In particular, creating an OSC is an indispensable, crucial mandatory step before mark-up. Once an OSC and a textual representation of a GL are provided, any clinical editor can in theory structure its knowledge in a semi-formal representation completely, but to specify it correctly, we should probably select a clinical editor with

good computational skills or at least a solid comprehension of the GL ontology (Asbru in our case) [23]. Once we had realized the significance of the initial results of this research regarding usability and the difficulty of acquisition of procedural knowledge, we had started to develop and evaluated a new graphical interface for GL specification [24].

Acknowledgments

References

1. Grimshaw, J.M., Russel, I.T.: Effect of clinical guidelines on medical practice: A systematic review of rigorous evaluations. Lancet 342, 1317–1322 (1993)
2. Tu, S.W., Musen, M.A.: A flexible approach to guideline modeling. In: Proc. AMIA Symp., pp. 420–424 (1999)
3. Johnson, P.D., Tu, S.W., Booth, N., Sugden, B., Purves, I.N.: Using scenarios in chronic disease management guidelines for primary care. In: Proc. AMIA Annu. Fall Symp., pp. 389–393 (2000)
4. Boxwala, A.A., Peleg, M., Tu, S., et al.: GLIF3: a representation format for sharable computer-interpretable clinical practice guidelines. J. Biomed. Inform. 37(3), 147–161 (2004)
5. Sutton, D.R., Fox, J.: The Syntax and Semantics of the PROforma guideline modelling language. J. Am. Med. Inform. Assoc. 10(5), 433–443 (2003)
6. Quaglini, S., Stefanelli, M., Lanzola, G., Caporusso, V., Panzarasa, S.: Flexible guideline-based patient careflow systems. Artif. Intell. Med. 22, 65–80 (2001)
7. Shiffman, R.N., Karras, B.T., Agrawal, A., Chen, R., Marenco, L., Nath, S.: GEM: a proposal for a more comprehensive guideline document model using XML. J. Am. Med. Inform. Assoc. 7(5), 488–498 (2000)
8. Noy, N.F., Crubezy, M., Fergerson, R.W., et al.: Protégé-2000: An Open-source Ontology-development and Knowledge-acquisition Environment. In: Proc. AMIA Symp., p. 953 (2003)
9. Votruba, P., Miksch, S., Kosara, R.: Facilitating knowledge maintenance of clinical guidelines and protocols. Medinfo. 11(Pt 1), 57–61 (2004)
10. Ruzicka, M., Svatek, V.: Mark-up based analysis of narrative guidelines with the Stepper tool. Stud.
11. Shahar, Y., Young, O., Shalom, E., Galperin, M., Mayaffit, A., Moskovitch, R., Hessing, A.: A Framework for a Distributed, Hybrid, Multiple-Ontology Clinical-Guideline Library and Automated Guideline-Support Tools. Journal of Biomedical Informatics 37(5), 325–344 (2004)
12. Shahar, Y., Miksch, S., Johnson, P.: The Asgaard project: A task-specific framework for the application and critiquing of time-oriented clinical guidelines. Artificial Intelligence in Medicine (14), 29–51 (1998)
13. Peleg, M., Tu, S., Bury, J., Ciccarese, P., Fox, J., Greenes, R.A., et al.: Comparing computer-interpretable guideline models: a case-study approach. J. Am. Med. Inform. Assoc. 10(1), 52–68 (2003)

14. De Clercq, P.A., Blom, J.A., Korsten, H.H., Hasman, A.: Approaches for creating computer-interpretable guidelines that facilitate decision support. Artif. Intell. Med. 31(1), 1–27 (2004)
15. Peleg, M., Gutnik, L., Snow, V., Patel, V.L.: Interpreting procedures from descriptive guidelines. Journal of Biomedical Informatics (2005) (in press)
16. Miller, R.A., Waitman, L.R., Chen, S., Rosenbloom, S.T.: The anatomy of decision support during inpatient care provider order entry (CPOE): Empirical observations from a decade of CPOE experience at Vanderbilt. J. Biomed. Inform. 38(6), 469–485 (2005)
17. Shalom, E., Shahar, Y., Lunenfeld, E., Taieb-Maimon, M., Young, O., Bar, G., et al.: The importance of creating an ontology-specific consensus before a markup-based specification of clinical guidelines. In: The 17th European Conference on Artificial Intelligence (ECAI2006), the workshop on AI techniques in healthcare: evidence-based guidelines and protocols, Trento, Italy, August 29. IOS Press, Amsterdam (2006)
18. Shalom, E., Shahar, Y., Taieb-Maimon, M., Lunenfeld, E.: A Quality Assessment Tool for Markup-Based Clinical Guidelines. In: AMIA Annu. Symp. Proc., Washington, DC, November 6, p. 1127 (2008)
19. Centers for Disease Control and Prevention (CDC) web site. Sexually Transmitted Diseases Treatment Guidelines (2002),
    `http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5106a1.htm`
20. Inpatient Management of COPD: Emergency Room and Hospital Ward Management (B1), Veterans Affairs (VA) web site (2005),
    `http://www.oqp.med.va.gov/cpg/COPD/copd_cpg/content/b1/`
    `annoB1.htm`
21. The American Association of Clinical Endocrinologists (AACE) Web site. American Association of Clinical Endocrinologists medical guidelines for clinical practice for the evaluation and treatment of hyperthyroidism and hypothyroidism (2002),
    `http://www.aace.com/pub/pdf/guidelines/hypo_hyper.pdf`
22. Shalom, E.: An Evaluation of a methodology for Specification of Clinical Guidelines at Multiple Representation Levels. Master Thesis. 2006; Technical report No. 14942, Dpt. of Information Systems Engineering, Faculty of Engineering, Ben Gurion University, Beer Sheva, Israel (2006)
23. Shalom, E., Shahar, Y., Taieb-Maimon, M., Bar, G., Young, O.B., Martins, S., Vaszar, L., Liel, Y., Yarkoni, A.K., Goldstein, M., Leibowitz, A., Marom, T., Lunenfeld, E.: A Quantitative Evaluation of a Methodology for Collaborative Specification of Clinical Guidelines at Multiple Representation Levels. Journal of Biomedical Informatics 41(6), 889–903 (2008)
24. Hatsek, A., Shahar, Y., Taieb-Maimon, M., Shalom, E., Dubi-sobol, A., Bar, G., et al.: A Methodological Specification of a Guideline for Diagnosis and Management of PreEclampsia. In: The 18th European Conference on Artificial Intelligence (ECAI 2008), Workshop on Knowledge Management for HealthCare Processes, Greece, July 21. IOS Press, Amsterdam (2008)