Journal of Biomedical Informatics 41 (2008) 889–903

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

# A quantitative assessment of a methodology for collaborative specification and evaluation of clinical guidelines

Erez Shalom [a,*], Yuval Shahar [a], Meirav Taieb-Maimon [a], Guy Bar [b], Avi Yarkoni [b], Ohad Young [a], Susana B. Martins [c], Laszlo Vaszar [c], Mary K. Goldstein [c], Yair Liel [b], Akiva Leibowitz [b], Tal Marom [d], Eitan Lunenfeld [b]

[a] Medical Informatics Research Center, Department of Information Systems Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel
[b] Soroka Medical Center, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel
[c] Geriatric Research Educational and Clinical Center (GRECC), VA Palo Alto Health Care System, Palo Alto, CA, USA
[d] Edith Wolfson Medical Center, Holon 58100, Israel

ARTICLE INFO

*Article history:*
Received 1 August 2007
Available online 6 May 2008

*Keywords:*
Clinical decision support systems
Clinical guidelines
Ontologies
Knowledge acquisition
Knowledge bases
Evaluation
Mark-up
Structuring
Completeness
Correctness

ABSTRACT

We introduce a three-phase, nine-step methodology for specification of clinical guidelines (GLs) by expert physicians, clinical editors, and knowledge engineers and for quantitative evaluation of the specification's quality. We applied this methodology to a particular framework for incremental GL structuring (mark-up) and to GLs in three clinical domains. A gold-standard mark-up was created, including 196 plans and subplans, and 326 instances of ontological knowledge roles (KRs). A completeness measure of the acquired knowledge revealed that 97% of the plans and 91% of the KR instances of the GLs were recreated by the clinical editors. A correctness measure often revealed high variability within clinical editor pairs structuring each GL, but for all GLs and clinical editors the specification quality was significantly higher than random ($p < 0.01$). Procedural KRs were more difficult to mark-up than declarative KRs. We conclude that given an ontology-specific consensus, clinical editors with mark-up training can structure GL knowledge with high *completeness*, whereas the main demand for *correct* structuring is training in the ontology's semantics.



## 1. Introduction

### 1.1. Clinical guidelines and the importance of their formal representation

Medical practitioners, overloaded with information, do not always have the time, or the computational means, to use the valuable knowledge encoded in clinical guidelines (GLs) during actual patient treatment. Such GLs have the potential both to improve the quality of medical care [1,2] and to contribute to the containment of the costs of care. Although there are thousands of text-based GLs, there is usually no automated support for their specification and application, even though clinicians at the point of care would obviously benefit from such support. Thus, over the past decade, a number of attempts have been made to support complex GL-based care in an automated fashion. Such automated support could assist in a graphical, interactive specification of GLs, search and retrieval of the GLs, patient eligibility determination, runtime application of GLs, and retrospective quality assurance (adherence to GLs). Despite these efforts, most GLs are still text based. Thus, implementing GLs within a computer-based clinical decision support system, i.e., formal GL representation, is fast becoming a critical issue [3].

A recent review [4] has identified the four main areas involved in the development of GL-based decision support systems: (1) GL *modeling and representation*, i.e., the internal format by which a GL is represented in the digital library; (2) GL *specification*, i.e., the act in which an editor creates that representation, typically from a text-based input; (3) GL *verification* and *testing*, i.e., confirming that the GL is in the appropriate format and (potentially) achieves its objectives; and (4) GL *application*, i.e., executing the GL at the point of care. In the current study, we focus on GL specification.

Our recently developed framework for support of GL-based care, the Digital electronic Guideline Library (DeGeL) architecture [5], handles most of the desiderata for GL specification, such as facilitating a gradual, multiple-phase specification process, including mark-up of the GL. (Performing a *mark-up* here means structuring the GL text by labeling portions of the text, using semantic

* Corresponding author. Fax: +972 8 6477161.
*E-mail addresses:* erezsh@bgu.ac.il (E. Shalom), yshahar@bgu.ac.il (Y. Shahar), meiravta@bgu.ac.il (M. Taieb-Maimon), guybar@bgu.ac.il (G. Bar), yarkoni@bgu.ac.il (A. Yarkoni), ohadyn@bgu.ac.il (O. Young), Susana.Martins@va.gov (S.B. Martins), vaszar.laszlo@mayo.edu (L. Vaszar), Mary.Goldstein@va.gov (M.K. Goldstein), liel@bgu.ac.il (Y. Liel), akival@bgu.ac.il (A. Leibowitz), maromtal@013.net.il (T. Marom), lunenfld@bgu.ac.il (E. Lunenfeld).

1532-0464/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jbi.2008.04.009

labels from a chosen target GL-specification language, sometimes even modifying the text.) The gradual-specification process supports 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.

### 1.2. Problems in guideline specification

Despite the considerable work already done in the GL area, the following three challenges have not been considered in sufficient depth, and the current study thus focuses on their clarification:

(1) A comprehensive methodology for the GL-specification process remains to be developed.
(2) Likewise, an *evaluation* methodology to assess the results of the specification is still to be developed.
(3) There are very few quantitative evaluations of GL-specification methodologies in the literature. In particular, there is a lack of appropriate evaluations of GL specification using a gold standard, mainly because of the significant effort required to create such a gold standard and to use it rigorously to evaluate the quality of GL specification. This is especially true in the case of the GL-specification methodology we have evaluated here, since there could be considerable interobserver variability (among different GL-knowledge editors or knowledge engineers) during semantic mark-up [6,7].

With the last challenge in mind, the current study was designed, first, to develop comprehensive, detailed GL specification and evaluation methodologies, and then to answer three specific research questions, defined in the context of these methodological frameworks:

(i) Can clinical editors actually mark-up a GL, and if so, at what quality level?
(ii) Are there differences in the quality of the mark-ups between different clinical editors marking-up the same GL?
(iii) Are there differences in the quality of the mark-ups of different specific aspects of the GL (e.g., eligibility conditions versus objectives)?

To address these three challenges and to answer the three specific research questions raised by assessment of the specification methodology, the current study was performed in three main parts (see Section 3 for details):

(A) We introduced a general methodology for the use of GL-specification tools to specify GLs.
(B) We introduced a general methodology for evaluation of the GL-specification tools.
(C) We then assessed the actual use of that methodology in the case of a particular instance of a GL-specification framework and associated software tool, when used for specification of GLs within three different clinical domains.

## 2. Background

### 2.1. Formal representation and specification of clinical guidelines

Automated support for GL application requires formal GL-modeling methods. During the past decade, a number of research groups have devoted considerable efforts to developing computer-interpretable clinical guidelines (CIGs) to support decision-making during clinical encounters [3,8,9]. Most GL-modeling methods use knowledge acquisition tools for eliciting the medical knowledge needed for the knowledge role (KR) classes and subclasses of the GL-specification ontology (i.e., the key concepts and their properties and interrelations) assumed by each method so as to specify it in a formal, executable format. According to the terminology used in the Stepper tool [10,11], there are two main approaches to GL specification: *model-centric*, i.e., modeling the GL *de novo* using a predefined ontology and computational model and referring to the source text solely for documentation, including multiple projects and related tools, such as the EON and PROforma frameworks and the Protégé and Arrezo tools, respectively [12–23]; and, *document-centric*, i.e., starting from a free-text document and mapping it to a given GL ontology manifested in another set of projects and associated tools, such as the GEM Cutter or Delt/A tools [24–28].

### 2.2. The Asbru guideline-specification language

In this study, we used the Asbru language [21] as the underlying GL-representation language. The Asbru-specification language includes semantic KRs organized into classes including (1) *Conditions* (containing, for example, *the filter condition* subclass, which represents obligatory eligibility criteria, *the complete condition* subclass, which halts the GL execution when some predefined criterion is true, and *the abort condition* subclass, which aborts the GL execution when some predefined criterion is true); (2) control structures for the GL's *Plan-Body* (containing, for example, the *sequential, concurrent*, and *repeating* combinations of actions or subguidelines); (3) the GL's *Intentions* (containing, for example, the *process and outcome* intentions subclasses), and (4) the *Context* class of the activities in the GL (containing, for example, the *actors, and clinical context* subclasses). A detailed description of all Asbru KR classes and their constituent KR subclasses can be found in Appendix A.

The Asbru language enables specification of a GL in terms of a hierarchical procedural structure, consisting of *plans* and *subplans*. The plans and subplans are defined through the very act of an editor marking-up the GL so as to segment it into a hierarchical structure of plans, although, presumably, these plans already exist implicitly within the GL.

### 2.3. The Uruz guideline-specification tool

As discussed above, several challenges relevant to GL specification still require an integrated solution. Thus, to support GL classification, semantic mark-up, context-sensitive search, browsing, run-time application, and retrospective quality assessment, we previously developed the DeGeL architecture and set of tools for classification [5], search, and retrieval [29] and runtime application of the GLs [30]. One of these tools is the document-centric Uruz GL-specification tool (Fig. 1).

Uruz solves a common problem: clinical editors cannot (and need not) program in GL-specification languages, while programmers and knowledge engineers do not always understand the clinical semantics of the GL. One way of addressing this problem is to perform the specification process gradually through several intermediate, semi-structured phases. Uruz enables clinical editors and knowledge engineers to open a text-based GL within it, select a target GL ontology (e.g., Asbru) by which to structure the GL, and drag and drop portions of the text into various nodes and leaves (terminal modes) of the selected GL-ontology's tree, such as into the "entry conditions" and "outcome intentions" knowledge roles. The text is thus implicitly labeled ("*marked-up*") by these semantic tags (this is the "*semi-structured*" representation format). The text