Guideline Classification to Assist Modeling, Authoring, Implementation and Retrieval

Elmer Bernstam, M.D., M.S.E. [1], Nachman Ash, M.D.[2], Mor Peleg, Ph.D.[1], Samson Tu, M.S.[1], Aziz A. Boxwala, M.B.B.S., Ph.D.[2], Peter Mork, B.A. [1], Edward H. Shortliffe, M.D., Ph.D.[1,3], Robert A. Greenes, M.D., Ph.D.[2]

[1]Stanford Medical Informatics, Stanford University School of Medicine, Stanford, CA
[2]Decision Systems Group, Harvard Medical School, Brigham & Women's Hospital, Boston, MA
[3]Department of Medical Informatics, Columbia University, New York, NY

Abstract

The National Guideline Clearinghouse (NGC) and its guideline classification system are significant contributions to the study of clinical practice guidelines (CPGs) and their incorporation into routine clinical care. The NGC classification system is primarily designed to support guideline retrieval. We believe that a guideline classification system should also support identification of features that relate to incorporation of executable CPGs into computer-based applications for sharing and delivering guideline-based advice. We have developed a proposed expansion of the NGC guideline classification for this purpose. The axes of the proposed scheme have implications for designing formal models and structures for representing and authoring CPGs. This scheme also has implications for future research.

Introduction

Clinical practice guidelines (CPGs) promise to improve the efficiency and quality of health care and to reduce practice variability. Professional societies, health care organizations and government agencies have sought to accomplish this by publishing guidelines in the literature or disseminating them via the Web. Nonetheless, CPGs have largely been unsuccessful in effecting beneficial change in practice.[1] There is mounting evidence that guideline-based advice at the point of care is more likely to be successful at influencing clinical behavior.[2,3] One way of providing advice during decision making is to integrate computer-based CPGs into the clinical workflow (*i.e.,* into clinical information systems). As a result, increased attention is focusing on developing decision support-systems capable of executing CPGs that are encoded via a standardized formalism and are therefore well-structured. Because developing executable CPGs takes considerable time and effort, it is desirable for such CPGs to be shareable.

A significant first step toward sharing CPGs is the National Guideline Clearinghouse (NGC), whose explicit mission is to "promote widespread access to clinical practice guidelines through a comprehensive database of clinical practice guidelines".[4] The NCG is maintained by the Agency for Healthcare Research and Quality (AHRQ) and is a collection of guidelines from a variety of sources, not necessarily in executable form. The database went online in January 1999 with 261 guidelines, and grew more than 140% over the next 12 months. As of March 2000, the NGC contained approximately 700 guidelines.[4] Although the NGC database is not the only publicly available guideline repository, it is particularly significant because the NGC actively solicits submission to collect guidelines from a variety of unrelated sources, thereby creating a single publicly available resource.

As a result of the proliferation of CPGs, finding and comparing guidelines of interest can be challenging. This problem will likely worsen as the number of electronically available guidelines increases. To cope with the challenge of maintaining a large and rapidly expanding guideline collection, the NGC has created a classification system based on a variety of guideline characteristics. Table 1 summarizes the NGC classification.

Understandably, the major purpose of this classification scheme is efficient maintenance of a large collection of heterogeneous CPGs (not limited to those that are well-structured or executable). This objective is important for guideline classification, but it is not the only objective. This paper, discusses functional requirements for a guideline classification scheme that addresses other objectives. Specifically, we describe a proposed expansion of the NGC guideline classification scheme to characterize them with respect to their modeling, authoring and executability characteristics. Our project is motivated by the desire to en-

courage electronic sharing and delivery of guideline-based advice.

Table 1. NGC Guideline Classification Axes

Axis	Example
Clinical specialty	Adolescent health allergy and immunology
Disease/condition	Multiple vocabularies via UMLS
Guideline category	Diagnosis, Prevention
Intended users	Dentists, Physicians
Method of review of guideline recommendations	Peer review
Methods used to analyze the evidence	Decision analysis, Review
Methods used to assess the quality and strength of the evidence	Expert consensus Subjective review
Methods used to collect the evidence	Searches of electronic databases
Organization type	Academic institution Medical specialty society
Target population	Infant (1 - 23 months) Female
Treatment/intervention	Multiple vocabularies via UMLS

We have three specific objectives: (1) Our *modeling* objective is to help define functional requirements for guideline representations (*e.g.* GLIF[5,6]) and (2) our *authoring* objective is to provide direction for developing new guidelines. Both of these objectives facilitate the most important objective, which is implementing *executable* CPGs.

New Proposed Guideline Classification Axes

In addition to the eleven axes defined by the NGC, we propose seven new ways to classify guidelines. The combined set of axes is summarized in Table 2. Changes or additions to the NGC classification are in **bold**. Possible values for each category, along with examples, are shown in Table 3. The table shows "sub-axes" to aid in understanding this more complex classification. The seven new ways to classify CPGs are sub-axes within two major new axes. These are *Use of the Guideline* and *Implementation*. Additionally, we modified the *Guideline Category* sub-axis, as described below.

Guideline Category: This sub-axis expands upon the IOM 1992 classification[7] and the NGC classification[4] to address guideline retrieval, modeling and authoring characteristics. We expanded this classification for numerous reasons. For example, specifying a desired guideline category or set of categories is a useful capability for retrieval. For designers of formal schemes for guideline representation, being able to refer to guideline categories when evaluating proposed constructs for inclusion is also useful. The designers must develop a formalism that is sufficiently expressive to represent the scope of currently available guidelines, as well as to anticipate future guideline categories and intended applications.[5]

Although clinical trials are not intended to be part of the NGC, automating their protocols has been a research goal.[8] Thus guideline-representation formalisms should support the expression of clinical trial protocols. Consequently, our axis *Guideline Category* contains the concept of *Clinical Trial*.

Showing that all guidelines can be adequately assigned to one of the chosen categories is difficult. However, if a formalism does not cover one of the categories it is not complete. For example, it is incomplete if it does not allow representation of eligibility criteria for clinical trials.

As guidelines become increasingly detailed and are linked to or incorporate other guidelines, complexity management becomes a major issue. Guideline-representation language developers have responded by supporting constructs such as sub-guidelines (nesting), views and macros.[5,6] Guideline classification can help a CPG author identify potential CPGs to incorporate. For example, a general diabetes guideline may include a sub-guideline regarding ICU management of diabetic ketoacidosis. In this way, guideline authors may use existing modules to create a new CPG. Alternatively, one may divide a large, complex guideline into sub-guidelines that deal with a particular task as defined by *Guideline Category* (*e.g.,* diagnosis of pneumonia vs. management of pneumonia) or *Guideline Use* (see below). These sub-guidelines may be reused in several larger guidelines.

Guideline Use: The *Intended Users* axis is clearly a critical classification axis. However, other parameters are also relevant to guideline use, especially when implementing computer-based decision-support at the point of care. We propose several additional sub-axes, although we recognize that these may not apply to all guidelines. They are:

Table 2. Proposed Classification

Axis Name	Sub-Axis Name	#	Possible Values	Example	Anticipated Purpose
Clinical Area	Clinical Field	[0..n][1]	[Same as NGC]	Cardiology, Dermatology	• CPG authoring • Retrieval
	Disease/Condition	[0..n]	[Same as NGC]	Lower Back Pain, Influenza	
	Treatment/Intervention	[0..n]	[Same as NGC]		
Guideline Category		[0..n]	[See Table 3]	Screening, Clinical Trial	• Defining Guideline Representation Languages (Modeling) • Authoring CPGs • Implementation • Retrieval
Use of the Guideline	Intended Users	[0..n]	[Same as NGC]	Physician Patient	
	Target Population	[0..n]	[Same as NGC]	Child, Male	
	Setting	[0..n]	[Patient Home][Field : Name][ER] [Critical Care (ICU)] [Inpatient Facility] [Extended Care Facility] [Ambulatory Care Facility]	Field : Ambulance	
	Usage Mode	[0..2]	[Within-Encounter] [Outside-of-Encounter]		
	Encounters	[0..2]	[Single][Multiple]		
	Time Frame	[0..3]	[Emergency][Acute][Chronic]		
Development Information	Guideline Name	[1]	[Text]		• Retrieval
	Author Name(s)	[0..n]	[Text]		
	Organization Name(s)	[1..n]	[Text]		
	Organization Type(s)	[1..n]	[Same as NGC]	Academic Institution, Manufacturer	
	Version	[0..1]	[Text]	1.0A	
	Source	[1]	[Original Guideline] [Revised Original] [Adapted from another Source]		
	Last Update Date	[0..1]	[Date]	2/23/99	
	Level of Evidence Based Support[2]		[Same as NGC]		
Implementation	**Format**	[1]	[Text] [Text with Graphics] [Text with Graphics and Flowcharts] [Annotated Text] [Hypertext/HTML] [XML] [Guideline Representation Language: Name]	Guideline Representation Language: GLIF	• Defining Guideline Representation Languages (Modeling) • Authoring CPGs • Implementation • Retrieval
	Distribution by Originator	[0..n]	[Journal Article] [Textbook] [Letter to Practitioners] [Email] [Internal Network] [Internet] [Implemented in Information System]		
	Computability	[1]	[Guiding][Intermediate][Algorithmic]		

Notes: Axes added to the NGC classification and changes within axes are shown in **bold**.
[1] The cardinality of the attribute is shown as [x..y], where x <= y. This represents: "there can be between x and y of these". [1] means that there must be one of these and 0 means that it is possible not to classify along this sub-axis.
[2] The NGC provides four axes to describe the evidence used to create the guideline: Method(s) used to (1) review of guideline recommendations, (2) Analyze the Evidence, (3) Assess the Quality and Strength of the Evidence and (4) Collect the Evidence.

1. The timing of a guideline's use: *within an encounter* or *outside of encounter*. "Encounter" means an interaction between the health-care system and the patient. For example, a reminder automatically generated by a guideline and sent to the patient would be a *within an encounter* event. On the other hand, a guideline meant to be read and applied to practice at a later time would be considered *outside of encounter*. The defining characteristic is whether decision-support is provided at the time that the decision is being made. A preliminary review of a random sample of NGC contents showed that existing included

guidelines are largely intended for *outside of encounter* use.

2. The number of encounters described by the guideline: *single* vs. *multiple.*

3. The setting where the guideline is used (*e.g., ICU* vs. *Ambulatory Care Facility*).

These parameters may also be useful for guideline retrieval. For example, they could help to find guidelines for ICU care of community-acquired pneumonia (CAP) vs. outpatient care of CAP.

Like *Guideline Category*, the *Setting* where a guideline is used may also help guide authors of future guidelines. For example, consider a future guideline defining comprehensive management of CAP. A path through this guideline might include an episode of outpatient management, followed by deterioration leading to hospitalization. A logical separation between parts of such a large guideline can be defined by the transition from the home to the inpatient setting. Thus, guideline authors should explicitly address the issue of where the guideline will be used.

Table 3. Possible Values for the *Guideline Category*

Value	Example
Screening	Screening for anemia
Prevention	Influenza vaccination
Diagnosis or pre-diagnosis management of patients (Workup)	Workup of syncope
Appropriate use of specific interventions and tests in clinical care	Influenza vaccination Indications for CABG Indications for CAT scan
Guidelines for the care of clinical conditions	Management of lung cancer
Clinical trial	CCG95-5941
Quality determinants of a technology	Quality determinants of mammography
Risk assessment	CAD risk assessment

Time Frame refers to a qualitative measure of the time covered by a guideline. *Chronic* problems, like diabetes, are likely to require iteration to manage ongoing problems. *Emergency,* such as cardiac arrest, is a category because this situation requires a different role for guidelines as they are unlikely to be studied and followed during the urgent episode.

Implementation refers to guideline characteristics important for implementation at a given institution. This category includes the way a guideline is distributed by the original authoring organization and in what format.

Format refers to how the guideline is presented to the user. Most current guidelines combine text and graphics (*e.g.,* tables and flowcharts). We anticipate, however, that guidelines will increasingly be packaged in electronic message formats, such as XML, or encoded by guideline representation languages such as the Arden Syntax or GLIF.

Computability refers to the ease with which a guideline can be converted to an algorithm. Many guidelines are not computable in the sense that an algorithm cannot be written to represent their recommendations accurately.[a] For example, a review paper, which is sometimes considered a guideline, would generally be "guiding," whereas a cardiac arrest protocol would be "algorithmic."

This characteristic is not a value judgment. A *guiding* guideline is not necessarily worse than an *algorithmic* one. However, a *guiding* guideline will likely be more difficult to translate into a computable representation.

Discussion

The NGC guideline classification is a significant step forward in making CPGs more effective. In addition to facilitating the maintenance of a large CPG collection, the classification provides a "controlled vocabulary" for discussing guidelines. For example, we can now refer to a defined set of guideline categories. Some axes are defined by explicit enumeration (*e.g., Guideline Category*) and some refer to controlled vocabularies via the UMLS (*e.g., Disease/Condition*).

Goals for Ongoing Research

We are implementing a web accessible *Guideline Warehouse* prototype that gives users a form for retrieving and submitting guidelines. This site could house original guidelines, contain links to other guideline collections or provide a combination of both.

Note that our classification system could be represented as an XML DTD or schema. This system, in combination with an XML parser, could verify that a guideline is classified in the prescribed manner (i.e., well-formed and valid).

We are exploring the comprehensiveness of our classification scheme in covering existing guidelines.

[a] *Computability* in computer science refers to the property of being executable by a machine following an algorithm. In other words, an algorithm can be written to describe the recommendations of the guideline.

There are almost certainly guidelines that our scheme does not reflect adequately. To explore this issue, we are reviewing a large number of representative guidelines from a variety of sources. With the results of this expanded review, we will further refine our classification system.

Another approach to comprehensiveness evaluation will be to undertake a rigorous, systematic analysis of guideline-design patterns. We will examine a corpus of guidelines with the intent of abstracting underlying patterns that define a category of guidelines. The implications of this line of research include, but are certainly not limited to, automated guideline classification.

An inter-observer reliability study of the classification scheme will test whether different users categorize guidelines similarly. Perfect reproducibility seems unlikely given the experience with MeSH. Using a corpus of articles that were doubly indexed at the National Library of Medicine (MEDLINE), Funk and Reid demonstrated that even specially-trained professional indexers did not always agree on the correct way to assign MeSH terms to a given article.[9]

Conclusions

The NGC and its guideline classification are important advances in enhancing the availability of clinical practice guidelines that facilitate sharing and use in clinical practice. In addition to the characteristics indexed by the NGC, however, a guideline classification scheme has other potential attributes relating to guideline representation. These attributes address their applicability to computer-based settings. We have proposed additions to the NGC guideline classification system to support modeling, authoring and implementation of computer-based CPGs.

Acknowledgements

Supported in part by Grant LM06594 from the National Library of Medicine and by the Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Material Command.

References

1. Cabana M, Rand C, Powe N, *et al*. Why don't physicians follow clinical practice guidelines? JAMA 1999 Oct 20. 282(15): 1458-65.
2. Schriger D, Baraff L, Rogers W. Cretin S. Implementation of clinical guidelines using a computer charting system. Effect on the initial care of health care workers exposed to body fluids [see comments]. JAMA 1997 Nov 19. 278(19): 1585-90.
3. Shiffman R, Brandt C, Liaw Y, Corb G. A design model for computer-based guideline implementation based on information management services. JAMIA 1999; 6:99-103.
4. National Guideline Clearinghouse: Agency for Healthcare Research and Quality (AHRQ), URL: http://www.guideline.gov/, classification at http://www.guideline.gov/FRAMESETS/static_fs.asp?view=about.overview.classify
5. Ohno-Machado L, Gennari J, Murphy S, *et al*. The guideline interchange format: a model for representing guidelines. JAMIA 1998; 5:357-72.
6. Peleg M, Boxwala A, Ogunyemi O, *et al*. GLIF3: The Evolution of a Guideline Representation Format. Submitted to Proceedings of the AMIA Fall Symposium, 2000.
7. Field M, Lohr K (eds). Guidelines for clinical practice: From Development to Use. Washington DC: National Academy Press, 1992.
8. Musen M, Carlson R, Fagan L, Deresinski S. T-HELPER: Automated support for community-based clinical research. Sixteenth Annual SCAMC, Washington, D.C., 719-723. 1992.
9. Funk M Reid C. Indexing consistency in MEDLINE. Bul Med Lib Assoc 1983; 2:176-83.