

# Preliminary Evaluation of a Guideline Classification System

Elmer Bernstam, M.D., M.S.E.<sup>1</sup>, Nachman Ash, M.D.<sup>2</sup>, Mor Peleg, Ph.D.<sup>1</sup>, Samson Tu, M.S.<sup>1</sup>, Edward H. Shortliffe, M.D., Ph.D.<sup>3</sup>, Robert A. Greenes, M.D., Ph.D.<sup>2</sup>

<sup>1</sup>Stanford Medical Informatics, Stanford University School of Medicine, Stanford, CA

<sup>2</sup>Decision Systems Group, Harvard Medical School, Brigham & Women's Hospital, Boston, MA

<sup>3</sup>Department of Medical Informatics, Columbia University, New York, NY

**Abstract** In a previous paper, we presented a proposed expansion of the National Guideline Clearinghouse (NGC) classification<sup>1</sup>. We performed a preliminary evaluation of the classification based on 100 guidelines randomly selected from the NGC collection. We found that 89 of the 100 guidelines could be assigned to a single guideline category. To test inter-observer agreement, twenty guidelines were also categorized by a second investigator. Agreement was found to be 40–90% depending on the axis, which compares favorably with agreement among MeSH indexers (30–60%)<sup>2</sup>. We conclude that categorization is feasible. Further research is needed to clarify axes with poor inter-observer agreement.

**Background** The NGC classification is primarily concerned with retrieval of guidelines from a large, heterogeneous collection. In a previous paper, we presented a proposed expansion of the NGC guideline classification scheme that classifies guidelines with respect to their modeling, authoring and executability characteristics in addition to retrieval. Our project is motivated by the desire to encourage electronic sharing and delivery of computer-based guideline advice.

**Objective** To assess the degree to which: (1) guidelines could be categorized using the proposed classification, (2) raters agreed with each other and (3) to identify axes that need further clarification.

**Methods** 100 guidelines were randomly selected from the NGC collection (Oct. 2000). One investigator (EB) categorized these according to the proposed classification system. A second investigator (NA) independently categorized the first twenty guidelines that were available in their entirety to assess inter-observer agreement.

**Results** Of 100 guidelines selected for categorization, 85 were available for review. The remaining 15 could not be obtained in their entirety and were categorized on the basis of the NGC data sheets. Data sheets had been approved by the guideline developer.

Of the 100 guidelines, 89 could be assigned to a sin-

gle *guideline category*. Eight could be assigned to two categories equally well and 3 could not be assigned to any category. The top three *guideline categories* were: Care of clinical condition (35), Appropriate use of a technology (27) and Screening (15).

Inter-Observer Agreement (based on 20 guidelines)

Parameter	% Agree	# / 20	# categories
Organization type	90%	18/20	15 + d
Clinical field	85%	17/20	66 + d
Time frame	65%	13/20	3 + d
Guideline Category	65%	13/20	8 + d
Number of encounters	65%	13/20	2 + d
Intended users	65%	13/20	28 + d
Computability	45%	9/20	3
Setting	40%	8/20	8 + d

(+d = multiple selections allowed, agreement = all terms agree)

**Conclusions** Our evaluation suggests that current guidelines from a variety of clinical fields can be assigned to a small group of guideline categories with reasonable inter-observer agreement. Work remains to better define other axes of our classification (e.g. computability, setting) in order to improve inter-observer agreement and to define the utility of our categorization with respect to modeling, authoring and execution of shareable electronic guidelines.

**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 Materiel Command.

## References

1. Bernstam EV, Ash N, Peleg M, Boxwala A, Mork P, Greenes RA, et al. Guideline classification to assist modeling, authoring, implementation and retrieval. Paper presented at: AMIA Fall Symposium, 2000; Los Angeles, CA.
2. Funk M, Reid C. Indexing consistency in MEDLINE. *Bulletin of the Medical Library Association* 1983;2:176-83.