

# Semantic Processing to Support Clinical Guideline Development

Marcelo Fiszman MD PhD,<sup>1</sup> Eduardo Ortiz MD MPH,<sup>2</sup> Bruce E. Bray MD,<sup>3</sup> Thomas C. Rindfleisch PhD<sup>1</sup>

<sup>1</sup>National Library of Medicine, Bethesda, MD

<sup>2</sup>National Heart Lung and Blood Institute, Bethesda, MD

<sup>3</sup>Department of Biomedical Informatics, University of Utah, Salt Lake City, UT

## Abstract

*Clinical practice guidelines are one of the main resources for communicating evidence-based practice to health professionals. During guideline development, questions that express a knowledge gap are answered by finding relevant citations in MEDLINE and other biomedical databases. Determining citation relevance involves extensive manual review. We propose an automated method for finding relevant citations based on guideline question classification, semantic processing, and rules that match question classes with semantic predications. In this initial study, we focused on a pediatric cardiovascular risk factor guideline. The overall performance of the system was 40% recall, 88% precision ( $F_{0.5}$ -score 0.71), and 98% specificity. We show that relevant and nonrelevant citations have clinically different semantic characteristics and suggest that this method has the potential to improve the efficiency of the literature review process in guideline development.*

## INTRODUCTION

Clinical practice guidelines are “systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances” [1, 2]. Adherence to practice guidelines by clinicians is expected to ensure a consistently acceptable standard of care. Government agencies, medical professional societies, and the research community are increasingly producing such guidelines. For example, the National Heart Lung and Blood Institute publishes guidelines on cardiovascular, pulmonary, and blood disorders for health professionals [3].

The development of clinical practice guidelines proceeds in four steps [4]: converting clinical information needs into questions, acquiring evidence from the medical literature relevant to those questions, appraising the evidence for validity, and applying appraisal results to answer questions. It is important for guideline developers to formulate questions that reflect gaps in current medical

knowledge. For each question, queries are issued to repositories such as MEDLINE to retrieve publications relevant to answering that question. This process is both resource- and time-intensive. Before final critical appraisal is performed by domain experts, the often voluminous numbers of publications retrieved for each question have to be manually assessed for relevance by several nonmedical reviewers.

In order to streamline guideline development, we propose an automated method that discriminates between relevant and nonrelevant MEDLINE citations for questions used during this process. Our methodology is based on semantic natural language processing and has three major components: guideline questions classified semantically, semantic predications produced by SemRep [5, 6], and rules that match guideline question components to semantic predications in MEDLINE citations. For this study, we processed questions provided by the National Heart Lung and Blood Institute that are being used to construct a guideline for pediatric cardiovascular risk reduction.

## BACKGROUND

### *SemRep*

SemRep [5, 6] uses MetaMap [7] and medical domain knowledge in the Unified Medical Language System to identify semantic predications in MEDLINE citations. For example, from the sentence in (1), SemRep extracts the predications in (2), in which the arguments (Physical activity, Obesity, and Child) are concepts from the Metathesaurus, and the relations PREVENTS and PROCESS\_OF are from the Semantic Network.

- (1) Physical activity to prevent obesity in young children
- (2) Physical activity PREVENTS Obesity  
Obesity PROCESS\_OF Child

Semantic predications serve as a normalized representation of document content and provide the core component of our methodology.

### ***Related research***

Although semantic predications have not previously been used to assist the guideline creation process, there is considerable related work. Automatic methods specifically directed at clinical practice guidelines have been based on ontologies [8] or natural language processing [9] used to extract medical knowledge from existing guidelines for dissemination. Other studies have explored methodologies for retrieving documents that contain answers to clinical questions, including concept indexing [10] and retrieval-based feedback [11].

More recent research has classified clinical questions, for example, on a taxonomy for questions that arise at the point-of-care [12] or using the PICO (Problem/Population, Intervention, Comparison, and Outcome) framework from evidence-based practice. Research that exploits the PICO model includes methodological filters [13, 14], a probabilistic search engine [15], and medical question answering [16, 17]. Others (e.g. [18]) have addressed definitional questions such as “What is X?”

We do not use PICO to identify MEDLINE citations relevant to answering questions while creating clinical practice guidelines. The PICO framework is largely concerned with therapy (see [19]), while most of the questions for the guideline on pediatric cardiovascular risk reduction are not therapeutic in nature. The following is typical:

- (3) What is the evidence that atherosclerosis related target organ damage begins in childhood?

In exploiting semantic predications, Mendonça et al. [20] retrieved relevant MEDLINE citations about a patient by matching clinical data with semantic graphs extracted from the biomedical literature. In a recent paper, arguments of SemRep predications were used to retrieve relevant citations for therapy questions of varying degrees of complexity [21]. The work presented here is an extension of that earlier method.

### **METHODS**

In developing an automatic method to select citations relevant to questions used during guideline development, questions were first analyzed semantically and categorized into classes. A crucial aspect of this analysis was identification of core semantic components of the questions. For each question class, MEDLINE citations for training were retrieved and marked as relevant or nonrelevant. SemRep was then applied to all citations and the predications produced were analyzed for patterns

characterizing relevant citations. As a final step we devised rules that match question components to semantic predications and discriminate between relevant and nonrelevant citations for each question class. The method applies to any set of MEDLINE citations, however retrieved.

### ***Categorizing questions***

In this preliminary study we analyzed 30 questions pertinent to pediatric cardiovascular risk factors. All start with “What is the evidence...” Addressing this aspect of the questions requires appraisal of the quality of the study reported in the relevant citation and is beyond the objectives of this study. Such appraisal is performed by medical experts at later stages of the guideline development process. Regarding content, for a specified risk factor (e.g. obesity), questions seek to elucidate the interaction of the risk factor and the disorder with respect to several parameters, including initiation, progression, population, and prospects for reduction, as shown in the following examples.

- (4) What is the evidence that atherosclerosis begins in childhood?
- (5) What is the evidence that the presence of obesity in childhood affects the progression of atherosclerosis into adult life?
- (6) What is the evidence that ethnic background affects obesity status in childhood?
- (7) What is the evidence that obesity in childhood can be decreased?

We isolated several key components which represent question content. These components serve as variables that can be instantiated in questions with the following values: risk factor (obesity, metabolic syndrome, diabetes mellitus); disorder (atherosclerosis, cardiovascular disease, atherosclerosis-related target organ damage); population (children, adults); population attribute (ethnic group, race, geographical location); and action (decrease or prevents, development and progression).

Questions were then sorted into 15 classes; the questions in a class share the same components instantiated to the same values. Questions (8) and (9), for example, belong to the same class because they share values for these components: risk factor (obesity), population (children), action (decrease or prevent).

- (8) What is the evidence that obesity in childhood can be decreased?

- (9) What is the evidence that obesity can be prevented in childhood?

Questions (10) and (11) belonged to another class because they share values different from the previous class for risk factor (obesity), population (children), disorder (atherosclerosis), and action (development and progression).

- (10) What is the evidence that the presence of obesity in childhood affects the development and progression of atherosclerosis during childhood?
- (11) What is the evidence that indicates the importance of obesity on the development and progression of atherosclerosis in childhood?

### *Devising rules*

Since the goal of the method is to find relevant citations for a specified question, and not to provide answers, rules can be written for a class of questions, rather than for individual questions. Rule development was supported by analysis of approximately 100 MEDLINE citations retrieved for a question from each class. Citations were marked as either relevant or nonrelevant and were processed with SemRep to produce semantic predications.

Analysis of the predications in the relevant and nonrelevant citations for each question revealed characteristic patterns, which were used to formulate rules that stipulate which semantic predications must occur in citations relevant to questions in the class.

Question components are used to guide construction of the rules. For example, the risk factor component is associated with the subject of the SemRep predicate PROCESS\_OF, and the population component is associated with the object of this predicate. The value of the action component is associated with particular predicates such as “decrease or prevent” with TREATS.

For example, question (8) above has rule (12) associated with it.

- (12) <Obesity> PROCESS\_OF <Children>  
<Obesity> NOT PROCESS\_OF <Adults>  
X TREATS <Obesity>

Obesity as the subject of PROCESS\_OF in (12) refers to the value of the risk factor component in question (8). Children is the value of the population component in that question and TREATS corresponds to “decrease.”

The arguments in our rules can be interpreted as variables in a schema. “X” can match anything, while

arguments marked with “< >” represent a defined domain of UMLS Metathesaurus concept. For example, <Obesity> matches concepts “Obesity,” “Overweight,” and “Weight Gain,” and <Children> matches “Child,” “Youth,” “Boys,” and “Girls.”

If predications matching the schema in rule (12) are found in the SemRep output of a citation, the citation is marked as being relevant to the associated question (and the class to which the question belongs). Rules such as these were written for each of the fifteen question classes.

### *Evaluation*

We evaluated the ability of our method to discriminate relevant from nonrelevant citations for ten guideline questions, each belonging to a different class. Risk factor was instantiated to obesity and we used citations reserved for testing. The questions and number of citations used in the evaluation are given in Table 1 (N = number of citations). A reference standard annotated by the first three authors marked which citations were relevant and nonrelevant to each of the ten questions. Annotators limited their analysis to titles and abstracts and did not take into account the evidence part of the question. System results were compared to the reference standard, and recall, precision, and specificity were determined. We also calculated a weighted  $F_{0.5}$  score  $(1.25 * P * R) / (0.25 * P + R)$ , which values precision twice as much as recall.

## **RESULTS**

Out of the 596 citations, 148 were marked as being relevant in the reference standard. The system missed 89 of these, resulting in recall of 40%. Sixty-seven citations were marked as being relevant. Of these, 8 were false positives, resulting in precision of 88% and an  $F_{0.5}$  score of 0.71. The system considered 440 citations as nonrelevant out of 448 nonrelevant citations in the standard. Therefore, specificity was 98%. Table 1 shows system performance for each question.

## **DISCUSSION**

Based on precision and specificity, preliminary results suggest that our method has promise in supporting guideline development. The modest recall value was caused by several factors, including SemRep limitations.

The fact that SemRep currently does not resolve anaphora led to a considerable number of false negative errors. We illustrate this problem using question (10) above. The rule to find relevant

Question	N	R	P	F	S
What is the evidence that atherosclerosis begins in childhood?	75	44%	78%	0.67	97%
What is the evidence that the presence of obesity in childhood affects the development and progression of atherosclerosis during childhood?	87	33%	94%	0.69	97%
What is the evidence that the presence of obesity in childhood affects the progression of atherosclerosis into adult life?	83	43%	75%	0.65	99%
What is the evidence that a decrease in obesity in childhood alters the development and progression of atherosclerosis in childhood?	25	43%	100%	0.79	100%
What is the evidence that a decrease in obesity in childhood alters the development and progression of atherosclerosis in adult life?	25	38%	100%	0.75	100%
What is the evidence that atherosclerosis-related target organ damage begins in childhood?	30	50%	100%	0.83	100%
What is the evidence that an increase in obesity in childhood alters the development of clinical cardiovascular disease in childhood or adulthood?	53	26%	75%	0.55	93%
What is the evidence that a decrease in obesity in childhood alters the development of clinical cardiovascular disease in adult life?	53	50%	100%	0.83	100%
What is the evidence that race or ethnic background affect obesity status in childhood?	90	52%	92%	0.80	99%
What is the evidence that obesity in childhood can be decreased?	75	60%	86%	0.74	98%

Table 1 – Questions and performance metrics R = Recall, P = Precision, F =F<sub>0.5</sub> score, S = Specificity.

citations for this question requires the presence of predications “<Obesity> PROCESS\_OF <Children>” and “<Obesity> PREDISPOSES <Atherosclerosis>.” Sentence (13) is in a citation missed by the system that is relevant to this question.

- (13) Obesity in children should no longer be regarded as a variation of normality, but a disease which predicts the development of atherosclerosis.

SemRep produced the predications in (14) for this sentence. Since “Disease,” rather than “Obesity, is the subject of PREDISPOSES as the interpretation of this sentences, the predications do not satisfy the rule, leaving the citation as a false negative for the question.

- (14) Obesity PROCESS\_OF Child  
Disease PREDISPOSES Atherosclerosis

The underlined words in (13) are in an anaphoric relationship, with “disease” meaning “obesity” in this sentence. If SemRep could resolve that relationship, the subject of PREDISPOSES in (14) would be “Obesity,” thus satisfying the rule.

The questions analyzed for this study stipulate the risk factors of interest for atherosclerosis. However, guideline developers are also concerned with gleaning from the research literature previously unknown risk factors that may affect the development and progression of atherosclerosis in childhood. Although we did not include such open questions in

our method, we conducted an informal study to investigate the possibility of accommodating them.

We applied a modified version of the rule associated with question (10) (with the risk factor left unspecified) to the semantic predications from a set of 639 MEDLINE citations on atherosclerosis in childhood. We then retrieved all the concepts that occurred as subject of the predication “PREDISPOSES Atherosclerosis.” After excluding risk factors that already appear in the guideline questions, such as obesity, diabetes, cigarette smoking, cholesterol, apolipoproteins, and many others, several substances remained as potential new risk factors for atherosclerosis in children. These include C-reactive protein, homocysteine, and adiponectin, which are known or debated risk factors for atherosclerosis in adults, and are now being studied in children. Some examples of relevant citations (with PMID) are: “Elevated serum C-reactive protein levels and advanced atherosclerosis in youth” (15802624), “Elevated plasma homocysteine in obese schoolchildren with early atherosclerosis” (16344991), and “Early atherosclerosis in obese juveniles is associated with low serum levels of adiponectin” (15928248).

### Limitations

In this study we concentrated on topical relevance, which is used in the first stages of guideline development. Automated determination of quality of evidence in relevant citations requires additional

research. The evaluation was based on ten questions and was not conducted in the context of actual guideline development. Finally, the method was trained and tested on questions for pediatric cardiovascular risk factor guideline and needs to be extended to guidelines in other areas.

## CONCLUSION

We propose a method for automatically finding relevant citations for questions used for developing clinical practice guidelines and suggest its potential use in facilitating guideline development. The method is based on question classification, semantic processing, and rules that match semantic predications with question classes to find relevant MEDLINE citations. Our results suggest that semantic characteristics of relevant citations are clinically different from nonrelevant citations for each guideline question.

## Acknowledgments

This study was supported in part through an interagency agreement between the National Library of Medicine and the National Heart Lung and Blood Institute and by the Intramural Research Program of the National Institutes of Health, National Library of Medicine.

## References

1. Institute of Medicine. (1990). *Clinical Practice Guidelines: Directions for a New Program*. M.J. Field and K.N. Lohr (eds.) Washington, DC: National Academy Press.
2. Institute of Medicine, Committee on Quality Health Care in America. (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington, DC: National Academy Press.
3. <http://www.nhlbi.nih.gov/guidelines/index.htm>
4. Sackett DL, Straus SE, Richardson WS, et al. *Evidence-Based Medicine: How to Practice and Teach EBM*. Philadelphia, PA, Churchill Livingstone, 2000
5. Rindflesch TC, Fiszman M. The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. *J of Biomed Inf.* 2003 Dec;36(6):462-77.
6. Rindflesch TC, Fiszman M, Libbus B. Semantic interpretation for the biomedical research literature. In Chen, Fuller, Hersh, and Friedman, eds. *Medical informatics: Knowledge management and data mining in biomedicine*. Springer, 2005, pp. 399-422
7. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: The MetaMap program. *Proc AMIA Symp.* 2001;17-21.
8. Shankar RD, Tu SW, Musen MA. A knowledge-acquisition wizard to encode guidelines. *AMIA Annu Symp Proc.* 2003;1007.
9. Serban R, ten Teije A, van Harmelen F, et al. Extraction and use of linguistic patterns for modelling medical guidelines. *Artif Intell Med.* 2007 Feb;39(2):137-49.
10. Hersh W, Hickam DH, Haynes RB, et al. Evaluation of SAPHIRE: an automated approach to indexing and retrieving medical literature. *Proc Annu Symp Comput Appl Med Care.* 1991;808-12.
11. Srinivasan P. Retrieval feedback in MEDLINE. *J Am Med Inform Assoc.* 1996 Mar-Apr; 3(2):157-67.
12. Ely JW, Osheroff JA, Gorman PN, et al. A taxonomy of generic clinical questions: classification study. *BMJ.* 2000 Aug 12;321(7258):429-32.
13. Haynes RB, Wilczynski N, McKibbon KA, et al. Developing optimal search strategies for detecting clinically sound studies in MEDLINE. *J Am Med Inform Assoc.* 1994 Nov-Dec;1 (6):447-58.
14. Schardt C, Adams MB, Owens T, et al. Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Med Inform Decis Mak.* 2007 Jun 15;7:16.
15. Ide NC, Loane RF, Demner-Fushman D. Essie: a concept-based search engine for structured biomedical text. *J Am Med Inform Assoc.* 2007 May-Jun;14(3):253-63.
16. Demner-Fushman D, Lin J. Answering clinical questions with knowledge-based and statistical techniques. *Computational Linguistics.* 2007 Mar; 33 (1):63-103.
17. Niu Y, Zhu X, Hirst G. Using outcome polarity in sentence extraction for medical question-answering. *AMIA Annu Symp Proc.* 2006;599-603.
18. Yu H, Lee M, Kaufman D, et al. Development, implementation, and a cognitive evaluation of a definitional question answering system for physicians. *J Biomed Inform.* 2007 Jun;40(3):236-51.
19. Huang X, Lin J, Demner-Fushman D Evaluation of PICO as a knowledge representation for clinical questions. *AMIA Annu Symp Proc.* 2006;359-63.
20. Mendonça EA, Johnson SB, Seol YH, et al. Analyzing the semantics of patient data to rank records of literature retrieval. *Proceedings of the ACL Workshop on NLP in the Biomedical Domain.*2002;69-76.
21. Sneiderman CA, Demner-Fushman D, Fiszman M, et al. Knowledge-based methods to help clinicians find answers in MEDLINE. *J Am Med Inform Assoc.* 2007 Nov-Dec;14(6):772-80.