A Meta Schema for Evidence Information in Clinical Practice Guidelines as a Basis for Decision-Making

Katharina Kaiser[a], Patrick Martini[a], Silvia Miksch[a,b], Alime Öztürk[a]

[a] *Institute of Software Technology and Interactive Systems, Vienna University of Technology, Austria*

[b] *Department of Information and Knowledge Engineering, Danube University Krems, Austria*

Abstract

Clinical practice guidelines are an important instrument to aid physicians during medical diagnosis and treatment. Currently, different guideline developing organizations try to define and integrate evidence information into such guidelines. However, the coding schemas and taxonomies used for the evidence information differ widely, which makes the use cumbersome and demanding. We explored these various schemas and developed a meta schema for the evidence information, which covers the most important components of the existing ones, is comprehensible, and easy to understand for the users. We developed and assessed the usefulness and applicability of our meta schema with guideline developers and physicians.

Keywords:

Clinical practice guidelines, evidence-based medicine, recommendations, study characteristics, clinical decision support, otolaryngology

Introduction

Evidence-Based Medicine (EBM) is defined as "the integration of best research evidence with clinical expertise and patient value" [1]. EBM advocates the use of up-to-date best scientific evidence from health care research as the basis for making medical decisions. One means to communicate research evidence is to integrate it into clinical practice guidelines (CPGs), which are "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances" [2]. Evidence-based CPGs involve a comprehensive search of the literature, an evaluation of the quality of individual studies, and recommendations that are graded to reflect the quality of the supporting evidence.

Evidence-based recommendations are mostly classified in particular grading schemas to provide a unique format at least for guidelines of the developing organization. Various organizations popularized taxonomy systems for grading the quality of evidence and the strength of recommendations (SoR) (e.g., [3-5]) and developed methodologies to categorize recommendations according to their systems.

In this paper we discuss the importance of a meta schema for levels of evidence (LoEs) and SoRs as a means for comparing, handling, and connecting LoEs and SoRs of different taxonomy systems for supporting the medical decision-making process. Due to the profusion of grading schemas users are often puzzled by the message a grade conveys. The different application of codes (e.g., I, II, III, …; A, B, C, …; 1, 2, 3, …; Ia, Ib, IIa, …) and the different definitions of the levels are not only confusing to users, but also aggravates a comparison and decreases the transparency of the schemas [6]. Table 1 and Table 2 give two examples of LoEs and SoRs from two different guideline-developing organizations.

Table 1 – Levels of Evidence used by the University of Michigan Health System

Level	Definition
A	Randomized controlled trials
B	Controlled trials, no randomization
C	Observational trials
D	Opinion or expert panel

The overall objective of this work is to facilitate the decision-making process on the basis of a systematic representation of the evidence information. A systematic representation is required to handle evidence information in computer-interpretable guideline representation languages (see [7] and [8] for a comprehensible overview). To achieve this objective we meet the following more specific objectives:

1. Development of a meta schema for grading evidence information. This schema should cover the most important components of various rating schemas for LoEs and SoRs.
2. Mapping of different LoEs and SoRs used by different organizations into this meta schema.

For our research we used 21 evidence-based CPGs from the clinical specialty otolaryngology. We selected the clinical specialty otolaryngology, because there are many, well structured guidelines available for our purpose. Based on the different LoEs and SoRs in these CPGs we have developed a

meta schema for both graded and ungraded evidence information and SoRs.

Table 2 – Strength of Recommendations defined by the Scottish Intercollegiate Guidelines Network (SIGN)

Strength	Definition
A	At least one meta analysis, systematic review, or RCT[1] rated as 1++, and directly applicable to the target population; or A systematic review of RCTs or a body of evidence consisting principally of studies rated as 1+, directly applicable to the target population, and demonstrating overall consistency of results
B	A body of evidence including studies rated as 2++, directly applicable to the target population, and demonstrating overall consistency of results; or Extrapolated evidence from studies rated as 1++ or 1+
C	A body of evidence including studies rated as 2+, directly applicable to the target population and demonstrating overall consistency of results; or Extrapolated evidence from studies rated as 2++
D	Evidence level 3 or 4; or Extrapolated evidence from studies rated as 2+

The meta schema connects existing grading systems to provide a means to increase the transparency among the various schemas and to appraise ungraded recommendations. By the direct comparability of various grading systems the communication of the underlying information is quickly and concisely possible.

Furthermore, decision-support is a crucial topic in computer-supported guideline's research. The meta schema will thereby facilitate the integration of evidence information and form a basis to handle the multitude of grading systems on an equal level.

In the following section we describe the process of developing our meta schema and we show the results in the subsequent section. Furthermore, we discuss the outcomes and cease with concluding remarks and future work.

Methods

Guyatt et al. [9] defined several criteria for developing an optimal grading schema. As our intention is not to develop a new system, but a meta schema that connects the existing schemas, other requirements apply. These are to have sufficient categories to cover a good portion of systems, to be consistent with existing systems, and to be simple and transparent to users.

We selected 21 CPGs from the National Guideline Clearinghouse[2], which were developed by nine different organizations (see Table 3). These CPGs cover eight different representations of LoEs and three different representations of SoRs.

Table 3 – Guideline Development Organizations

Organization/Cooperation	Number of Guidelines
American Academy of Family Physicians; American Academy of Otolaryngology – Head and Neck Surgery; American Academy of Pediatrics	1
American Academy of Pediatrics	2
Allergic Rhinitis and its Impact on Asthma Workshop Group	1
Cincinnati Children's Hospital Medical Center	3
Finnish Medical Society Duodecim	1
Institute for Clinical Systems Improvement	6
Practice Guidelines Initiative	1
Scottish Intercollegiate Guidelines Network	3
University of Michigan Health System – Academic Institution	3

After analyzing the different grading schemas of these organizations we decided to use the definitions of LoEs and SoRs of the Scottish Intercollegiate Guidelines Network[3] (SIGN) [10] as a basis for our meta schema. One reason for this decision was that SIGN's representation of evidence information is systematically evaluated and clearly structured and defined. As the SIGN approach does not cover all required information to represent the different LoEs and SoRs, we expanded the definitions for other organizations to cover their representations of the evidence information, too.

The most relevant attributes for developing the meta schema are:

- Code schema of guideline developing organization
- Study design and quality
- Strength of recommendations
- Benefits and harms

During the development process we conducted interviews with various guideline developers and physicians. We discussed the correctness, sensibility, availability, and understandability of the hierarchical structure, the quality of the LoEs and SoRs, the mapping tables, and the balance between benefits and harms. Furthermore, we surveyed the availability of required information and the facilitation of the decision-making process. Moreover, the covering of the meta schema with existing grading systems was verified during the entire process. The remarks and comments were incorporated in our schema altogether.

[1] RCT … Randomized Controlled Trial

[2] http://www.guidelines.gov (last assessed December 3, 2006)

[3] http://www.sign.ac.uk (last assessed December 1, 2006)

Code Schema of the Guideline Developing Organization

This attribute is essential to differentiate between various grading schemas, because a symbol or code communicating a grade can represent different meanings (see also [6]). For example, the *University of Michigan Health System* uses the symbols "*A", "B", "C", "D*" for LoEs, whereas *SIGN* uses these symbols for SoRs. Thus, it is not possible to extract the evidence information from the guidelines and map them to the meta schema, without the information about the developing organization.

Study Design and Quality

The quality of evidence is described by LoEs. They are mostly explicitly represented in the guidelines but different symbols are used to refer to them (see for instance Table 1).

The attribute of the study design covers all study types (i.e., Meta-Analysis, Systematic Reviews, Randomized Controlled Trials (RCTs), Cohort Studies, Case Control Studies, Expert Opinion) used in CPGs.

We represent the LoE on the basis of the study design's attribute, because in that way we get an ordered structure, where *meta-analysis* is on the top of the hierarchy and *no study design* is at the bottom.

The study design attribute plays a significant role by assigning a grade to ungraded evidence information. Often CPGs include information about the study design upon which the recommendations are based, but they do not provide any explicit grades for their evidence (e.g., "*The recommendations are supported by randomized controlled trials. Adverse parasympathetic events were reported by participants in randomized controlled trials, the most frequent and troublesome being increases sweating which occurred in about one-quarter of patients taking 5 mg three times per day and about one-half of patients taking 10 mg*" [11]).

Another attribute to be considered for establishing the LoEs is the study quality. It refers to the detailed study methods and execution. The study quality is thereby the degree to which a study employs measures to minimise biases, focussing on internal validity [12].

Our representation has to address both the study design and the study quality. The levels have to be clearly distinguishable and easily and clearly interpretable.

Strength of Recommendations

For SoRs also different symbols or names are used but they are not always explicitly mentioned in our CPGs. Three out of nine organizations have defined SoRs and only six of the 21 CPGs include explicitly defined SoRs. In 15 CPGs no information about SoRs is included.

The representation of SoRs has to be representable to the different existing SoRs. The requirements for our SoR taxonomy are that

- The strengths have to be clearly distinguishable from each other
- The names of the grades have to be meaningful
- The strengths have to be easily and clearly interpretable
- The number of grades should be limited to ensure an easy understanding and application

The GRADE Working Group[4] developed a system for defining the recommendations based on four factors [4]:

1. The trade-offs, taking into account the estimated size of the effect for the main outcomes, the confidence limits around those estimates, and the relative value placed on each outcome
2. The quality of the evidence
3. Translation of the evidence into practice in a specific setting, taking into consideration important factors that could be expected to modify the size of the expected effects, such as proximity to a hospital or availability of necessary expertise
4. Uncertainty about baseline risk for the population of interest.

Recently, medical associations and organizations adapt the GRADE approach for their needs [9,13]. However, the publication of guidelines using the new grading systems will take time. In our CPGs, this new approach is not implemented yet. The strength of a recommendation is only based on the underlying quality of evidence.

Benefits and Harms

Information about benefits and harms of a particular treatment plays a significant role in the decision-making process. But in our guidelines they are described very briefly and limited. They do not contain information about the trade-off between the benefits and harms either. For embedding information about benefits and harms into the decision-making process, we need them to be represented explicitly. Thus, a schema for the trade-off between benefits and harms is necessary, because this information is essential for the medical staff to assess benefits and harms of a treatment recommendation. For example:

1. In patients with peptic ulcer, drug A reduces acidity. This recommendation is based on RCTs.
2. In patients with cardiac problems, drug A may cause heart attacks and hence is contraindicated. This recommendation is based on case reports.

The second argument is based on lower quality evidence, but defeats the first argument, because of the more important claim (heart attack is worse than having acidity reduced).

Results

The Meta Schema

Based on the considerations described in the previous section we developed a meta schema for representations of the quality of evidence, the strength of recommendations, and benefits and harms.

[4] http://www.gradeworkinggroup.org (last assessed November 30, 2006)

Definition of Levels of Evidence

Our definition of LoEs is based on the *study design* and the *study quality*. The *study design* is essential to get a hierarchical representation and to assign a level to ungraded evidence information in CPGs. The LoEs consist of symbols that cover information about the study design and the quality of the studies. We introduced our own symbols (e.g., I_1, I_2, … II_1, II_2, …) that represent both the study design and quality. The first character describes the study design whereas the number describes the quality. Table 4 shows a part of our LoEs schema.

Table 4 – Part of the meta schema representing Levels of Evidence

Study Design	Evidence Level	Definition
Meta Analysis	I_1	Meta-analysis of RCTs
	I_2	High quality meta-analysis
	I_3	Well-conducted meta-analysis
	I_4	Meta-analysis
Systematic Reviews	II_1	High quality systematic reviews of RCTs with large sample
	II_2	High quality systematic reviews of RCTs with small sample
	II_3	High quality systematic reviews of RCTs with very low risk of bias
	II_4	Systematic reviews of RCTs
	II_5	High quality systematic reviews of cohort studies
	II_6	High quality systematic reviews of case-control studies
	II_7	Systematic reviews
…	…	…

Definition of Strengths of Recommendations

Based on the requirements for our SoRs taxonomy we defined four different strengths, because more than four hierarchical levels are hardly distinguishable, but less do not adequately cover existing systems. The four grades are:

1. *Strong Recommendation*
2. *Recommendation*
3. *Weak Recommendation*
4. *No Recommendation*

They have a unique definition and are easy to differentiate from each other. For example, *Strong Recommendation* is directly applicable to the target population and bases on at least one meta-analysis, systematic review of RCTs, RCTs with very low risk of bias, high quality meta-analysis of observational studies, or high quality systematic reviews of observational studies.

Our aim with these definitions of SoRs is to provide guideline users a proposed recommendation that should only be a direction if there is no explicit representation of SoRs in the CPGs.

Definition of trade-off between benefits and harms

Table 5 shows the definitions of the trade-off between the benefits and harms. Our definitions are based on the descriptions used by the GRADE working group [4], because they have a well defined categorization of the trade-off between the benefits and harms in their grading schema.

Table 5 – Schema for trade-off between benefits and harms

Classification	Benefits and Harms
Clear Benefit	The benefits of the recommended approach clearly exceed the harms.
Benefit	The recommended intervention explicitly does more good than harm or the benefits outweigh the harms.
Unclear Balance	It is unclear whether the recommended intervention does more good than harm. The trade-off between benefits and harms is quite unclear.
No Clear Benefit	The recommended intervention clearly does not do more good than harm.

Mapping the Evidence Information of CPGs

The meta schema should provide a general representation of different classifications of LoEs and SoRs used in CPGs. We connected each individual LoE and SoR taxonomy to our meta schema. Figure 1 shows a mapping between SIGN and our meta schema. The mapping tables provide guideline users a better handling and understanding of the evidence information in CPGs. With this representation the users have a means to compare different LoEs and SoRs.

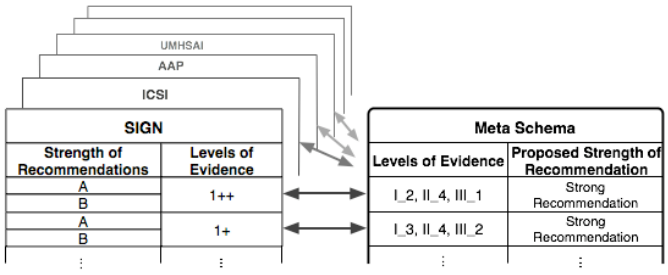

Figure 1 – Mapping table for the meta schema showing mappings from and to the SIGN schema

Discussion

The most significant factor for decision-making is the strength of recommendations. In most taxonomies the following aspects are taken into account:

- The level of evidence of individual studies
- The type of outcomes measured by these studies (patient-oriented or disease-oriented)

- The number, consistency, and coherence of the evidence as a whole
- The relationship between benefits, harms, and costs

In most guidelines this information is not entirely available. Thus, it is only possible to assign a constricted and intermediate SoR. Furthermore, benefits and harms for each recommendation are needed to incorporate them in the constitution of the SoR. But often, benefits and harms are only given for the entire guideline and not individual recommendations.

Conclusion

Our meta schema is an instrument to connect different schemas of LoEs and SoRs. The meta schema is representable to eight different systems defining LoEs and three different systems defining SoRs and incorporate the ideas and concepts of the GRADE Working Group. Furthermore, it is possible to assign a LoE to an ungraded evidence recommendation based on the study design and quality if available. It covers also information about the trade-off between benefits and harms, which are mostly not included in the existing grading schemas. We used the attributes study design and study quality (defined in LoEs), SoRs, the organization's code schema, and benefits and harms, which were significant for the development process.

Furthermore, we think that our meta schema can also support instruments for guideline appraisal (e.g., AGREE [14], GLIA) in terms of providing means to better understand and compare the various existing grading schemes for evidence information.

Acknowledgments

This work is supported by "Fonds zur Förderung der wissenschaftlichen Forschung - FWF" (Austrian Science Fund), grants P15467-N04 and L290-N04.

References

[1] Sackett D. Evidence Based Medicine: What it is and what it is not. *BMJ*, 312:71, 1996.

[2] Field M and Lohr K. Clinical Practice Guidelines: Directions for a New Program. National Academy Press, 1990.

[3] Harbour R and Miller J. A new system for grading recommendations in evidence based guidelines. *BMJ*, 323:334–336, 2001.

[4] GRADE Working Group. Grading Quality of Evidence and Strength of Recommendations. *BMJ*, 328(7454):1490-1498, 2004.

[5] Ebell MH, Siwek J, Weiss BD, Woole SH, Susman J, Ewigman B, and Bowman M. Strength of Recommendation Taxonomy (SORT): A Patient-Centered Approach to Grading Evidence in the Medical Literature, *American Family Physician*, 69(3): 548–556, 2004.

[6] Schünemann HJ, Best D, Vist GE, and Oxman AD for the GRADE Working Group. Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations. *Canadian Medical Association Journal (CMAJ)*, 169(7): 677–680, 2003.

[7] Peleg M, Tu SW, Bury J, Ciccarese P, Fox J, Greenes RA, Hall R, Johnson PD, Jones N, Kumar A, Miksch S, Quaglini S, Seyfang A, Shortliff E, and Stefanelli M. Comparing Computer-Interpretable Guideline Models: A Case-Study Approach, *JAMIA*, 10(1):52–68, 2003.

[8] de Clercq PA, Blom JA, Korsten HHM, and Hasman A. Approaches for Creating Computer-interpretable Guidelines that Facilitate Decision Support, *Artificial Intelligence in Medicine*, 31(1):1–27, 2004.

[9] Guyatt GH, Guttermann D, Baumann MH, Addrizzo-Harris D, Hylek E, Philips B, Raskob G, Zelman Lewis S., and Schünemann HJ. Grading strength of recommendations and quality of evidence in clinical guidelines: report from an American college of chest physicians task force. *Chest*, 129:174–181, 2006.

[10] Scottish Intercollegiate Guidelines Network. SIGN 50: A guideline developers' handbook. SIGN, 2004.

[11] Practice Guideline Initiative. Symptomatic treatment of radiation induced xerostomia in head and neck cancer patients, 2004.

[12] Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, and Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials*, 16:62-73, 1995.

[13] Schünemann HJ, Jaeschke R, Cook DJ, Bria WF, El-Solh AA, Ernst A, Fahy BF, Gould MK, Horan KL, Krishnan JA, Manthous CA, Maurer JR, McNicholas WT, Oxman AD, Rubenfeld G, and Turino GM. An Official ATS Statement: Grading the Quality of Evidence and Strength of Recommendations in ATS Guidelines and Recommendations. *American Journal of Respiratory and Critical Care Medicine*, 174:605-614, 2006.

[14] The AGREE Collaboration. Writing Group: Cluzeau FA, Burgers JS, Brouwers M, Grol R, Mäkelä M, Littlejohns P, Grimshaw J, Hunt C. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Quality and Safety in Health Care* 2003; 12(1): 18–23.

[15] Shiffman RN, Dixon J, Brandt C, Essaihi A, Hsiao A, Michel G, and O'Connell, R. The guideline implementability appraisal (GLIA): development of an instrument to identify obstacles to guideline implementation. *BMC Med Inform Decis Mak*, 5, July 2005.

Address for correspondence

Katharina Kaiser
Vienna University of Technology
Institute of Software Technology & Interactive Systems
Favoritenstrasse 9-11/188, A-1040 Vienna, Austria
kaiser@ifs.tuwien.ac.at