

Die approbierte Originalversion dieser Dissertation ist an der Hauptbibliothek der Technischen Universität Wien aufgestellt (<http://www.ub.tuwien.ac.at>).

The approved original version of this thesis is available at the main library of the Vienna University of Technology (<http://www.ub.tuwien.ac.at/englweb/>).



Ph.D. Thesis

Embedding the Evidence Information in Computer-Supported Guidelines into the Decision-Making Process

Conducted for the purpose of receiving the academic title
'Doktorin der technischen Wissenschaften'

Supervisor

Silvia Miksch

Institute of Software Technology & Interactive Systems [E188]

Submitted at the Vienna University of Technology

Faculty of Informatics

by

Alime Öztürk

9925136

Vienna, May 15, 2007

Acknowledgements

First of all I would like to thank my supervisors **Univ.-Prof. Mag. Dr. Silvia Miksch** and **a.o. Univ.-Prof. Dr. med. univ. Christian Popow** for their support and guidance during my research.

Further, I would like to thank all experts who participated in the evaluation by answering my questionnaire and invested a lot of time to support me during the development process of my meta schema. My thanks also go to my colleagues at the Information Engineering Group at the Vienna University of Technology.

Very special thanks go to my father **Ibrahim Ceyhan** and my mother **Fadime Ceyhan** for their support and motivation during my whole life. Also special thanks go to my children **Muhammed-Kerim** and **Zeynep**, my husband **Ayhan**, my sisters **Dr. med. univ. Keziban Öztürk**, **Asiye Öztürk**, and **Gülcan Kaya** for their patience and support and for being their for me when I needed them.

Also special thanks go to **Dipl.-Ing. Dr. Burcu Yildiz** for being such an invaluable friend during the years of our study.

Abstract

Clinical practice guidelines (CPGs) are widely used to support medical staff in treatment planning and decision-making during health care. To increase the effectiveness and the quality of health care, many research projects deal with computer-based representations and implementations of CPGs. Representation languages to formulate CPGs in a computer-interpretable way are complex, what makes the formulation process difficult and time-consuming. CPGs contain major recommendations about a certain disease that are usually based on clinical studies indicating the level of evidence and hence the strength of the recommendations. However, not all CPGs provide explicit information about the level of evidence or the strength of recommendation (i.e., ungraded evidence information).

In this thesis we propose a methodology that supports guideline users during the decision-making process on the basis of a semi-formal representation of the evidence information that can be found in CPGs. A semi-formal representation is required to handle evidence information in computer-interpretable guideline representation languages. For this purpose, we have developed a meta schema that covers various kinds of grading systems including graded and ungraded evidence information. The classification of different recommendations in CPGs are one of the most important information sources to use. However, there is a lack of consensus amongst guideline developers, regarding those classification schemes. To address this problem, we mapped various kinds of grading systems into our meta schema. Furthermore, we extended two guideline representation languages (Asbru, PROforma) to model our meta schema.

Finally, we present the results of our qualitative study we performed with physicians, guideline developing organizations, and guideline representation language developers to examine the correctness, feasibility, and understandability of our meta schema and language extensions. The results of our evaluation indicate that using a semi-formal representation of the evidence information is of particular importance to facilitate the decision-making process.

Zusammenfassung

Klinische Leitlinien (KL) dienen dem medizinischen Personal zur Unterstützung bei der Planung und der Entscheidungsfindung in der medizinischen Versorgung. KL beinhalten wichtige Empfehlungen für eine bestimmte Krankheit, die normalerweise auf klinische Studien basieren. Diese klinischen Studien weisen auf den Grad der Evidenz und infolgedessen auf die Stärke der Empfehlungen hin. Doch nicht alle KL bieten explizite Informationen über den Grad der Evidenz oder die Stärke der Empfehlung (nicht-gradierte Evidenzinformation).

Viele wissenschaftliche Projekte beschäftigen sich mit der computerbasierten Darstellung und Implementierung von KL um die Effektivität und die Qualität der medizinischen Versorgung zu erhöhen. Für die Computer-interpretierbare Darstellung von KL werden Modellierungssprachen verwendet, die aufgrund ihrer Komplexität den Formulierungsprozess erschweren.

In dieser Arbeit stellen wir eine Methode vor, die auf eine semi-formale Darstellung der Evidenzinformation in den KL basiert und die Nutzer der KL während dem Entscheidungsprozess unterstützt. Eine semi-formale Darstellung ist notwendig, um die Evidenzinformation anhand von computer-interpretierbaren Modellierungssprachen zu verarbeiten. Für diesen Zweck haben wir ein Meta Schema entwickelt, das verschiedene Gradierungssysteme einschliesslich gradierte und nicht-gradierte Evidenzinformationen beinhaltet. Die Klassifikation von verschiedenen Empfehlungen in KL ist einer der wichtigsten Informationsquellen, die wir nutzen. Allerdings existiert ein Mangel an Übereinsimmung zwischen den Entwicklern von KL in Bezug auf diese Klassifikationsschemata. Um dieses Problem zu behandeln bilden wir verschiedene Gradierungssysteme auf unser Meta Schema ab. Ausserdem erweitern wir zwei Modellierungssprachen (Asbru, PROforma) um unser Meta Schema zu modellieren.

Zuletzt präsentieren wir die Ergebnisse von unserer qualitativen Studie, die wir mit Ärzten, Organisationen, die KL entwickeln und Entwicklern von Modellierungssprachen durchgeführt haben, um die Korrektheit, Realisierbarkeit und die Verständlichkeit unseres Meta Schemas und der Spracherweiterungen zu überprüfen. Die Ergebnisse unserer Evaluierung haben deutlich gemacht, dass die Verwendung einer semi-formalen Darstellung der Evidenzinformation besonders wichtig ist, um den medizinischen Entscheidungsprozess zu erleichtern.

Contents

1	Introduction	1
1.1	Problem Domain	2
1.2	Research Question	3
1.3	Our Approach	4
1.4	Publications	5
1.5	Thesis Outline	5
I	State of the Art	7
2	Evidence-Based Medicine	8
2.1	What is Evidence-Based Medicine (EBM)	8
2.2	Procedure of Evidence-Based Medicine (EBM)	9
2.3	Basic Study Types	11
2.4	Methods to Enhance the Quality of Basic Study Types	13
2.4.1	Meta-Analysis	13
2.4.2	Systematic Reviews	16
3	Guidelines and Evidence Information	21
3.1	Clinical Practice Guidelines	21
3.1.1	Guideline Development Process	22
3.1.2	Quality of CPGs	26
3.2	Evidence Information in CPGs	28
3.2.1	Levels of Evidence	28
3.2.2	Strengths of Recommendations (SoRs)	28
3.3	Guideline Developing Organizations	29
3.3.1	AAFP, AAOHNS and AAP	31
3.3.2	American Academy of Pediatrics (AAP)	32
3.3.3	Allergic Rhinitis and its Impact on Asthma Workshop Group (ARIA)	33
3.3.4	Cincinnati Children’s Hospital Medical Center (CCHMC)	35
3.3.5	Finnish Medical Society Duodecim (FMSD)	37
3.3.6	Institute for Clinical Systems Improvement (ICSI)	37
3.3.7	Practice Guidelines Initiative (PGI)	38
3.3.8	Scottish Intercollegiate Guidelines Network (SIGN)	39

3.3.9	University of Michigan Health System - Academic Institution (UMHS)	42
4	Guideline Representation Languages	43
4.1	Asbru	44
4.1.1	Main Components of Asbru	44
4.1.2	Syntax and Semantics of Asbru	47
4.1.3	Application	48
4.2	PROforma	52
4.2.1	Task Properties in PROforma	52
4.2.2	Syntax and Semantics of PROforma	54
4.2.3	Application	55
5	Conclusion	59
II	The EviGuiDe Approach	60
6	Preliminary Work	61
6.1	Guideline Selection	61
6.2	Analyzing the Evidence Information	62
6.3	Basic Grading Systems	63
6.3.1	Grading System of SIGN	65
6.3.2	Grading System of GRADE	65
7	The EviGuiDe Methodology	67
7.1	The Meta Schema	68
7.2	The Way to the Meta Schema	68
7.3	Levels of Evidence	70
7.3.1	Concepts	71
7.3.2	Definitions	73
7.4	Strengths of Recommendation	78
7.4.1	Graded Recommendations	78
7.4.2	Ungraded Recommendations	81
7.5	Trade-off between Benefits and Harms	82
7.6	Costs	86
7.7	Mapping Process	87
7.8	Extending Asbru and PROforma	97
7.8.1	Extending Asbru	97
7.8.2	Extending PROforma	100
8	Evaluation	103
8.1	Questionnaire	103
8.2	Selecting Professionals	104
8.3	Evaluating the Comments	105
9	Supporting the Decision-Making Process	111

10 Summary and Future Work	113
10.1 Summary	113
10.2 Future Work	115
A Evidence-based Clinical Practice Guidelines	117
List of Figures	128
List of Tables	129
Bibliography	130
Curriculum Vitae	138

Chapter 1

Introduction

Evidence-Based-Medicine (EBM) is defined as "the integration of best research evidence with clinical expertise and patient value" [Sackett *et al.*, 1996]. EBM advocates the use of up-to-date best scientific evidence from health care research as the basis of medical decisions. Medical decision-making is a complex task because physicians have to know the facts about a disease (e.g., frequencies, signs, symptoms) and how all these facts are related with the patient's characteristics (e.g., age, sex, family history, risk factors, and other disease).

The number of facts and connections between these facts that a physician has to consider to make a medical decision is extraordinary. A book about a medical specialty that summarizes the most important information is typically more than 2,000 pages long [Eddv, 1986]. In addition to these books, there exist a great number of published biomedical journals and articles. Facing such an information flood, physicians have a hard task in delivering the best possible health care.

Clinical Practice Guidelines (CPGs) represent a way to provide physicians with accurate, relevant and updated decision support by supplying collections of information (i.e., recommendations) to increase the effectiveness and quality of clinical practice [Molino, 1999].

An important factor with this regard is the presence of recommendations in CGPs. They are usually based on certain *Levels of Evidence* (LoEs) and certain *Strengths of Recommendations* (SoRs) that provide physicians various treatment options. Whereas several definitions of LoEs and SoRs exist in the medical literature, the following definitions seem more appropriate for our interests [Ebell *et al.*, 2004]:

- Level of Evidence (LoE): The validity of an individual study is based on an assessment of its study design. According to some methodologies, LoEs can refer not only to individual studies but also to the quality of evidence from multiple studies about a specific question or the quality of evidence supporting a clinical intervention.
- Strength of Recommendation (SoR): The SoR of a clinical practice is based on a body of evidence that takes into account:
 - the LoEs of individual studies

- the type of outcomes measured by these studies
- the number, consistency, and coherence of the evidence as a whole
- the relationship between benefits, harms, and costs

Since they had been introduced, many guidelines have been developed, edited and validated for different purposes (e.g., treatment, diagnosis) and different domains, supporting medical staff in treatment planning and decision-making. Various guideline developing organizations have generated their own grading systems to classify major types of recommendations in CPGs. Today, more than 100 different grading systems are in use in medical publications [Ebell *et al.*, 2004] and the process of grading the evidence information in guidelines and journals are described in several publications (compare e.g., [(SIGN), 2001; Ebell *et al.*, 2004; Guyatt *et al.*, 2006]). For example, Guyatt *et al.* [2006] state the following seven criteria for an optimal grading system, which also facilitate the medical decision-making process:

1. "Separation of grades of recommendations from quality of evidence
2. Simplicity and transparency for clinician consumer
3. Sufficient (but not too many) categories
4. Explicitness of methodology for guideline developers
5. Simplicity for guideline developers
6. Consistent with general trends in grading systems
7. Explicit approach to different levels of evidence for different outcomes"

CPGs are usually published as paper-based documents, and thus do not allow physicians to retrieve the information needed to solve a medical problem easily. Therefore, several systems and representation languages (e.g., Asbru, PROforma) have been developed in order to obtain computer-interpretable representations of CPGs (consider e.g., [Teranziani *et al.*, 2004; Shahar *et al.*, 1998a; Peleg *et al.*, 2000; Fox *et al.*, 1998]). But the final implementation of CPGs in decision support systems depends on several other factors, like the structure, writing mode, formalization process, and last but not least on the physicians' attitude toward CPGs [Séroussi *et al.*, 1999].

In this work, we will focus on the computer-interpretable representation of evidence information that can be found in CPGs.

1.1 Problem Domain

To support physicians during the decision-making process several CPGs and decision support systems for guideline based medical care have been developed. However, CPGs contain in general "imprecise" terms to describe medical knowledge

and therefore are difficult to formalize. Guideline representation languages with a clear syntax and semantics, such as PROforma [Fox *et al.*, 1998] are of particular importance here, because they provide the means to handle various concepts that medical knowledge formalization implies.

In practice, physicians have to select and interpret guideline statements to provide optimal treatment of diseases. Hereby, it is important that these statements are based on some kind of scientific evidence, because evidence-based recommendations are better followed in practice than recommendations that are not based on scientific evidence [Grol *et al.*, 1998]. It is also important to classify and represent several kinds of evidences in a standardized way, so that physicians can extract and embed the represented information into their decision-making. This classification and representation is a challenging task, because various organizations have their own classifications of LoEs and SoRs. The situation worsens in cases where physicians use more than one guideline from different organizations to find the best available treatment for an individual patient and therefore have to deal with different grading systems which use different terminology. In addition to graded recommendations, there exist ungraded recommendations in CPGs as well, where the guidelines do not contain any classification of the LoEs and SoRs. Such misrepresented information in guidelines with regard to the recommendations increases the intricacy of the decision-making process and indicates that the classification of ungraded evidence information is a challenging but necessary task.

Although the recommendations described in the guidelines are the most important factors for decision-making, the LoEs and the SoRs are inadequately treated in existing tools and representation languages. Many of the existing tools and representation languages do not support the formalizing and modeling process of recommendations with regard to the LoEs and SoRs sufficiently. Therefore, a method is required to embed the evidence information in computer-supported guidelines into the decision-making process. This new approach should facilitate the decision-making process during execution by providing:

- A simple method to make the evidence information in guidelines manageable
- A traceable unique representation of various existing grading schemes
- A method for representing ungraded evidence information
- An extension to existing guideline representation languages for representing the evidence information

1.2 Research Question

The idea proposed in this thesis is to facilitate the decision-making process on the basis of a semi-formal representation of the evidence information. In CPGs a semi-formal representation is required to handle evidence information in computer-interpretable guideline representation languages. As a basis for our research we formulated our main research question and divided it into several subquestions.

How can we develop a meta schema for evidence information in guidelines, that, on the one hand, covers different kinds of existing grading systems and, on the other hand, supports the medical decision-making process?

We can divide this main research question into the following subquestions:

- Which influence has the evidence information in CPGs on the medical-decision making process?
- Which components of CPGs are essential to represent the evidence information?
- Is it practicable to model both graded and ungraded evidence information using existing guideline representation languages, so that the evidence information influences the medical decision-making process in executing?
- Is it possible to assign a grade to ungraded evidence information using the meta schema?
- Is it applicable to map various kinds of graded and ungraded evidence information into the newly developed meta schema?
- Which extensions are required to enable guideline representation languages to model the evidence information according to our meta schema?

1.3 Our Approach

To provide answers to our research questions we developed a meta schema called EviGuiDe. EviGuiDe is an acronym for "embedding the *EVID*ence information in clinical practice *GUI*delines into the *DE*cision-making process". The realization of this goal is based on the following more specific objectives:

1. Providing a way to implement ungraded evidence information with our meta schema
2. Developing a meta schema that covers the different kinds of existing grading schemes
3. Mapping of different grading systems from different organizations into our meta schema
4. Extending the guideline representation languages Asbru and PROforma

1.4 Publications

Parts of this thesis have been published in:

K. Kaiser, P. Martini, S. Miksch, A. Öztürk. *A Meta Schema for Evidence Information in Clinical Practice Guidelines as a Basis for Decision-Making*. In K. Kuhn, T. Y. Leong, J. Warren (eds.) 12th World Congress on Health (Medical) Informatics (Medinfo'2007), AMIA, Brisbane, Australia, 2007, forthcoming.

A. Öztürk, K. Kaiser, P. Martini, S. Miksch. *Embedding the Evidence Information in Guideline Representation Languages*. In P. Kokol, T. Welzer-Družovec, D. Mičetić-Turk, M. Zorman (eds.) 20th IEEE International Symposium on Computer-Based Medical Systems (CBMS'07), IEEE Computer Society Press, Maribor, Slovenia, 2007, forthcoming.

1.5 Thesis Outline

This thesis is outlined in two main parts.

Part I: State of the Art

Chapter 2: Evidence-Based Medicine (EBM)- starts with an overview of EBM and Clinical Practice Guidelines (CPGs).

Chapter 3: Guidelines and Evidence Information - contains relevant information and definitions of evidence information in CPGs. These are the Levels of Evidence (LoEs) and the Strengths of Recommendation (SoRs). Further, several guideline developing organizations and their different grading systems are presented.

Chapter 4: Guideline Representation Languages - compares two guideline representation languages (Asbru, PROforma) and their decision-support systems.

Chapter 5: Conclusion - summarizes the main components of Part I

Part II: The EviGuiDe Approach

Chapter 6: Preliminary Work - contains a description of the preprocessing we performed to select the CPGs we used to analyze the evidence information and to define the basic grading systems. These two grading systems are of particular importance for our approach.

Chapter 7: The EviGuiDe Methodology - presents the concepts of our approach and explains our meta schema in detail. Furthermore, it shows the process of embedding and extending Asbru and PROforma with the evidence information with regard to our meta schema.

Chapter 8: Evaluation - contains the results of our evaluation of the meta schema.

Chapter 9: Supporting the Decision-Making Process - shows the benefits of our meta schema and its influence to the medical decision-making process.

Chapter 10: Summary and Future Work - contains our conclusions and outlines our future plans.

Part I
State of the Art

Chapter 2

Evidence-Based Medicine

Knowledge in medicine reduplicates itself every five to ten years. Annually available publications in different areas of medicine have reached a number where it is nearly impossible for medical practitioners to observe the improvement in his/her specific field. Physicians need wide ranging and good information on the effectiveness of a large number of therapeutic interventions. But the explosion in biomedical publishing in the latter half of the 20th century, with more than 30.000 journals and more than 2 millions articles a year, makes it not easier. In any single area of medicine exists a great number of published studies. This makes it difficult to know which studies should be used as the basis for clinical practice. The result of these studies are often unclear, confusing, or contradictory [Evans, 2001]. On the strength of this information flood medical practitioners have a hard task to deliver the best health care without any additional support. Evidence- Based-Medicine (EBM) tries to arrange this chaos with well defined procedures and methods to assign a level of evidence and incorporate such evidence into patient care recommendations.

2.1 What is Evidence-Based Medicine (EBM)

The term "Evidence-Based Medicine" first appeared in medical literature within the work of Guyatt et al. [1992]. After that many definitions followed. For example, Last [1995] defines EBM as:

"the process of finding relevant informations in the medical literature to address a specific clinical problem, the application of simple rules of science and common sense to determine the validity of information; the application of the information to the clinical question. In short, patient care based on evidence derived from the best available ("gold standard") studies"

However, the most used definition of EBM is formulated by Sackett [1996] as follows:

"Evidence-Based Medicine is the *conscientious, explicit* and *judicious* use of current best evidence in making decisions about the care of individual patients. The practice of Evidence-Based Medicine means integrating individual clinical expertise with the best external evidence

from systematic research. By individual clinical expertise we mean the proficiency and judgment that individual clinicians acquire through clinical experience and clinical practice. By best available external evidence we mean clinically relevant research, often from the basic sciences of medicine, but especially from patient centered clinical research.”

In the past the work of physicians relied on their clinical experience, so that the physician decided what to do based on his/her clinical experience or by asking a local expert. However, according to the principles of EBM, a physician has to know that his/her personal clinical experience has only a limited value and that he/she also needs to update his/her knowledge about the evidence that can be found in medical literature. In this way the experts opinion is challenged and controlled with the vast amount of information in the medical literature.

The main idea behind EBM is to assign recommendations regarding patient care with certain levels of evidence indicating how much research supports these recommendations in order to identify and incorporate the evidence into the decision-making processes.

2.2 Procedure of Evidence-Based Medicine (EBM)

Raspe and Stange [1999] considered EBM in terms of the following four steps (compare Figure 2.1):

1. Formulate accurate clinical questions
2. Search for the best evidence in the literature
3. Critically evaluate the found evidence
4. Judge about the adaptability of found and valid evidence for the current clinical situation at hand

The first step is to formulate the clinical problem at hand in terms of clearly answerable questions. Hereby a well formed question consists of four elements, which are referred to "Person in question, Intervention given, Comparison and Outcomes considered" (PICO) [(NHS), 2006]:

- **Person or population in question:** How would I describe a group of patients similar to mine?
- **Intervention:** Which kinds of intervention am I considering?
- **Comparison:** What is the main comparable alternative to the considered interventions?
- **Outcomes:** What can I hope to accomplish?

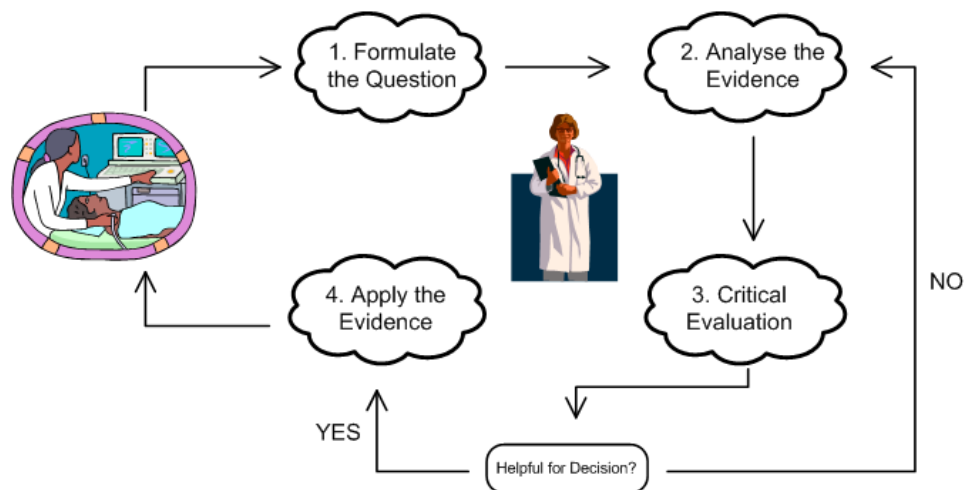


Figure 2.1: The four primary steps a user of EBM should follow (adopted from [Zielinski, 2002])

With formulating the questions using the PICO structure we can see what type of questions we have and more importantly what type of answers we need. This brings us to the next step of using EBM in search for the best evidence that namely could provide the best answers. There are some extensive and easy to use data collections of medical literature (e.g. Medline, Embase). The Cochrane Collaboration, for example, releases four times a year a CD-Rom with a database of systematic reviews, a database of abstracts, a Cochrane controlled trials register with over 150.000 controlled studies with their abstracts, and a Cochrane review methodology database [Raspe and Stange, 1999]. Such collections can be used to select candidate documents related to various clinical questions.

The step of evaluating found evidence critically requires physicians to be accurate. The Oxford Institute of Health Services introduced the "Critical Appraisal Skills Programme"(CASP) which aims to help health personnel to develop skills in appraising evidence about clinical effectiveness of care recommendations. According to this program, the health personnel have to consider three questions while evaluating articles [Belsey and Tony, 2001]:

1. Are the reported results valid?
2. What are the concrete results?
3. Will the reported results help us with our local problem?

This procedure of EBM implies that EBM depends on three properties that a medical practitioner should possess: knowledge, skills, and the correct attitude (see Figure 2.2). The crucial property to correctly use and apply EBM in practice is **knowledge** about internal and external evidence. This means to have sufficient individual clinical expertise and information about the patient, and also access to and knowledge about external evidence sources, such as databases and existing literature. In addition to specific knowledge, certain **skills** are important to find and evaluate information, particularly with regard to modern electronic media. Hereby, it is

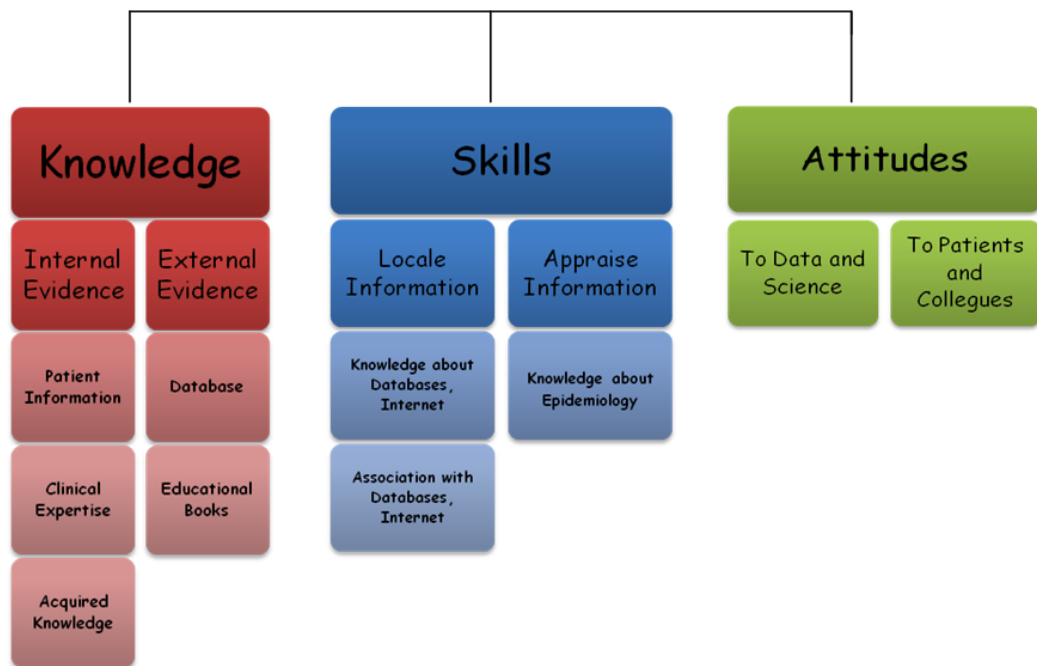


Figure 2.2: The three properties a medical practitioner who uses EBM has to possess [Zielinski, 2002]

important to follow the main procedure of EBM, by defining the problem, constructing and conducting an efficient search to locate best evidence, critically appraising the found evidence, and finally by considering that evidence and its implications in the context of the current patient's circumstances. All of these require a critical and appropriate **attitude** towards data, science and also towards colleagues and patients [Guyatt *et al.*, 2000].

2.3 Basic Study Types

We stated earlier that articles for EBM contain care recommendations for certain kinds of disease based on several types of clinical studies. In this section, we will explain the most common types of clinical studies by stating their definitions, application areas, and their benefits and pitfalls.

Randomized Controlled Trials (RCTs)

To evaluate the effectiveness of a studied intervention the best way is to use Randomized Controlled Trials (RCTs) that is, experiments in which the efficiency of medicines and medical procedures are tested. With this kinds of studies treatments, interventions, or enrollments into different study groups are assigned by random allocation rather than by conscious decisions of clinicians or patients. If the sample size is large enough, this study design avoids problems because of bias and confounding variables by assuming that both known and unknown determinants of reported outcomes are evenly distributed between treatment and control groups [Weil,

1999]. RCTs are considered the "gold standard" for determining effectiveness of health care interventions and they generally held to be the most powerful research design available to assess the effects of mental health care [Collaboration, 2001]. RCTs are the best way to get effective results but are at the same time very time and labor intensive. It must be pointed out that with RCTs it is possible to obtain unbiased distribution and that they facilitate statistical analysis, but such trials are sometimes expensive and sometimes ethically problematic. Like any other research, the results of RCTs should be treated critical [Cochrane, 1989].

Cohort Study

With cohort studies a group of people are studied who potentially share one or more characteristics like common experiences or conditions [Eldredge, 2000]. The study groups are observed over a certain period of time to describe differences between a condition during an earlier period to a condition at a later period. Prospective cohort studies normally begin to measure relevant indicators of variables prior to an exposure or incidence of diseases. Retrospective cohort studies identify the cohort, their exposure, and outcomes afterwards, as a follow-up study [Lichtenstein *et al.*, 1987]. In practice researchers identify two groups, where one group has a particular condition or receives a particular treatment and the other group does not have such a condition or treatment. For a period of time they are observed by the researchers. At the end the outcomes between the two groups are compared and analyzed.

Case-Control Study

In case-control studies persons who have a specific adverse effect or disease ("case") are compared to a group of persons without the specific adverse effect or disease ("controls") [Lichtenstein *et al.*, 1987]. In practice, case-control studies are used to identify possible causes for a certain condition. An existing present condition is analyzed by looking back at the past events to identify causative factors for a disease.

Cross-Sectional Survey

Cross-sectional survey is a study type in which disease and exposure status are measured simultaneously in a given population. Patients are interviewed, examined, and studied to gain answers about the prevalence of acute or chronic conditions in a population. The data of this study is collected at a single time, but they also refer to experiences in the past. In a cross-sectional survey, a particular population is observed at one point in time. The researchers collect information from one particular population and compare this data on specific subgroups. The exposure and the outcome are determined simultaneously. It is useful for looking at the prevalence of a disease, but it is unable to establish a temporal relationship between a presumed cause and an effect [Greenhalgh, 1997].

The advantages of cross sectional surveys are that they are cheap, simple, and ethically safe. Another important advantage is that the results of such studies are

relatively quickly available, because information is collected during a finite time period.

Cross-over Design

Cross-over design is a clinical trial design in which patients receive certain treatments in a specified or random order. In this design, every patient serves as his/her own control. This type of design is useful for studying the differences between individual treatments or sequences of treatments [Senn, 2002].

Case Report and Case Series

A case report describes the medical history of only one patient in a form of an anecdote. A collection of case reports is needed to form a case series. Case series are medical histories of more than one patient with defined conditions and treatments, which are reported [Greenhalgh, 1997].

Depending on the primary hypothesis or the topic of research, the type of the study changes. RCTs are mostly used for treatment whereas cohort or case-control studies are recommended to establish causation. Cross-sectional survey is preferred to establish diagnosis or screening to determine the test values. Unfortunately, sometimes such studies are too small-scaled and therefore not sufficient enough to draw conclusions. Meta-analysis and systematic reviews can be used in such cases to incorporate these studies in order to obtain representative conclusions.

2.4 Methods to Enhance the Quality of Basic Study Types

2.4.1 Meta-Analysis

Clinical studies are often performed on a small scale. Although it would be wrong to say that such studies do not contain any significant information regarding possible results and effects of different treatments, rather it can be said that they are not suitable to make generalizations. In order to enable generalization it is required to consider collections of such individual studies on a given subject. Meta-analysis is a statistical evaluation technique to appraise the results of more than one study and is used to get general results from study collections [Augustin and Fischer, 2002]. In other words, meta-analysis combines the results of several independent clinical trials to improve the potential for uncovering and studying any differences in available scientific material and to provide a basis for explanations [Scuka, 2004].

The basic idea of meta-analysis is to construct a "big study" from many small studies. Each study contains a certain measure of information to obtain an objective combination of the data, which comes from different and independent randomized studies. The difficulty lies in deciding which sets of studies are combinable and which are not. After relevant studies have been identified, decisions must be taken

about which studies are sufficiently well conducted to be worth including. Good meta-analysis use explicit and objective criteria for inclusion or rejection of studies on quality ground [Davies and Cromble, 2001].

The description of unequal data from different kinds is called heterogeneity related to meta-analysis. The main question is to find out the real cause of the heterogeneity. Heterogeneity can be caused by differences of studied populations, by unequal interventions, or by inconsistent study results. Data from smaller studies provide more heterogeneity than large studies [Schneider and Tramèr, 1999].

Meta-analyses offer a systematic and quantitative approach to review important therapeutic questions. The reviewer of meta-analyses looks for relevant studies on the basis of predefined criteria to achieve completeness. Found data is evaluated critically and bias that may exist in the found data is being considered. Data is represented graphically and is combined quantitatively if possible. Health care managers and clinicians are then able to appraise meta-analyses for validity and therefore to decide if they are implementable in their daily work or not [Schneider and Tramèr, 1999].

Strengths

The main advantage of meta-analysis that is relevant to us, is the possibility to obtain a complete picture of the treatment effects from a collection of individual studies. Davies describes four benefits of meta-analysis in the dealing with practical difficulties of individual studies [Davies and Cromble, 2001]:

1. **A clearer picture:** Usually small studies show no statistical difference between the treated and controlled groups. However, the possibility of them reporting a sizable effect cannot be excluded per se. Collecting studies in a systematic and unbiased way obtain a clearer picture, where the question to ask is whether a specific treatment affords significant benefits when used for specific patient groups or not.
2. **Overcoming bias:** Unsystematic methods tend to provide a plenty of scope for bias. Meta-analysis can overcome this problem by providing an unbiased synthesis of the empirical data.
3. **Precision:** The precision assesses the size of any effect that depends on the number of studied patients. Because meta-analysis draw on many trials (i.e. more studied patients), they have more power to detect small but clinically significant effects, and can give more precise assessment of the size of any uncovered effects. This is of particular importance when a physician is looking for beneficial effects for specific subgroups. Individual studies contain mostly too few patients and therefore a systematic aggregation of data from many individual studies are needed to obtain a clearer picture.
4. **Transparency:** The advantage of meta-analyses lie in the openness with which good meta-analyses reveal all the decision that have been taken throughout the process of achieving the collected effect sizes. Good meta-analyses

should allow physicians to decide for themselves how reasonable the decisions taken are and what their likely impact on the final estimate of effect size will be.

Representations of Meta-Analysis

Meta-analysis can be graphically represented in two ways: blobbograms and odds ratio diagrams [Davies and Cromble, 2001].

”Blobbograms: Blobbograms display the findings of each individual study as a blob or square with a horizontal line, representing usually 95% confidence interval, around the main findings. The size of the blob and the small vertical line reflect the amount of information in that individual study and therefore can vary. The length of the horizontal line represents the uncertainty of the estimation of the treatment effect of the study. The aggregated effect size for certain subgroupings and the overall effect size are also usually displayed in the same figure (see Figure 2.3).”

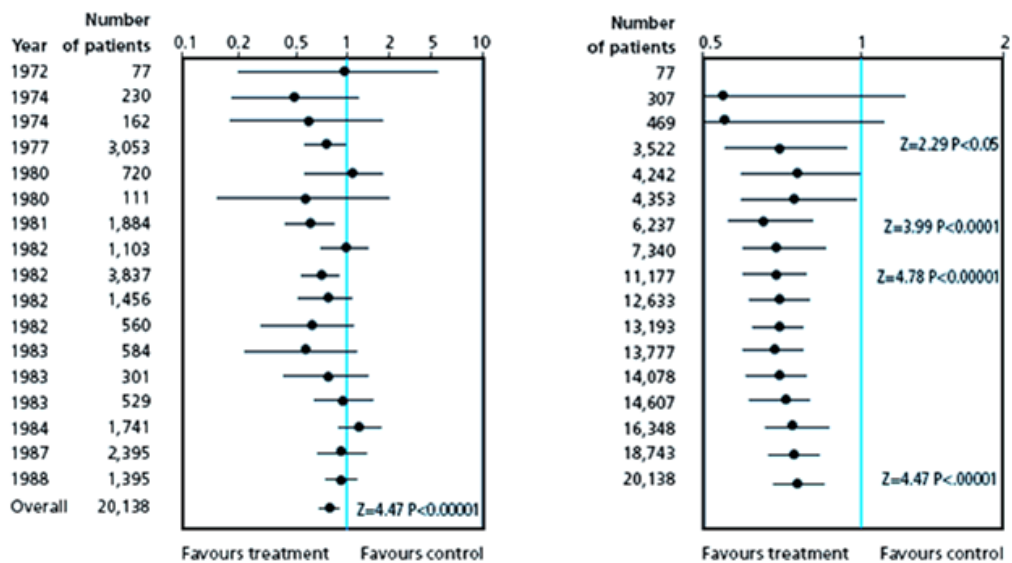


Figure 2.3: Blobbograms display the results from each individual study as a blob or square with a horizontal line (usually representing the 95% confidence interval) around the main results [Davies and Cromble, 2001].

”Odds ratio diagram: The main measure to quantify the effect used for meta-analysis is the odds ratio. It offers some technical advantages when combining data from different studies (see Figure 2.4). For most practical purposes, the odds ratio can be interpreted as a relative risk. For example, an odds ratio of 2 implies that the defined outcome appears about twice as often in the intervention group as in the control group. An odds ratio of 0.5 implies a reduction of around 50% of the represented event in the treated group compared with the control group.”

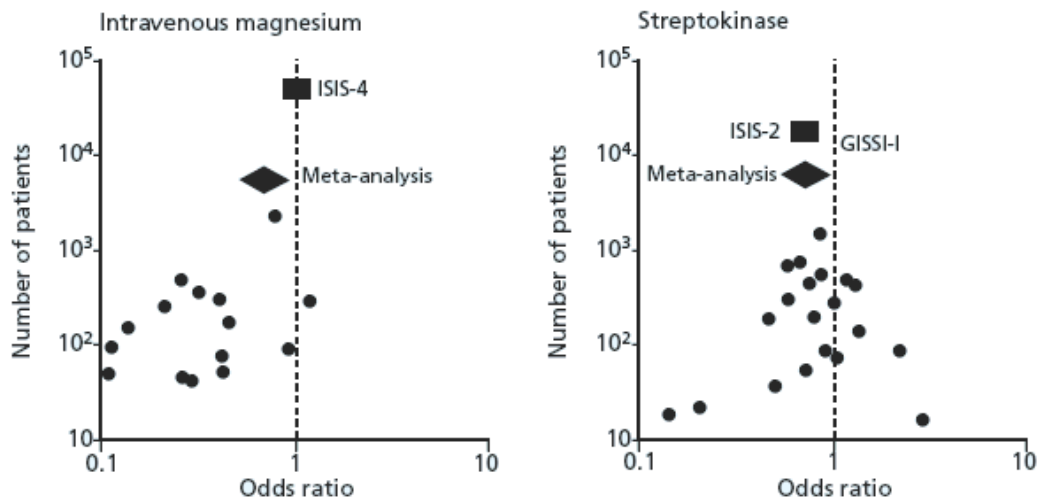


Figure 2.4: Points illustrate odds ratio from small and medium sized trials, diamonds illustrate combined odds ratios with 95% confidence intervals from Meta-Analysis of these trials, and squares illustrate odds ratio with 95% confidence intervals from mega sized trials [Davies and Cromble, 2001].

2.4.2 Systematic Reviews

Reviews have always been a part of the medical domain. Such narrative reviews, are usually written by domain experts and are mostly qualitative summaries of evidence on a given topic. These reviews consists of informal, subjective methods to select and interpret studies, and do not explicitly describe how the reviewers searched, selected or appraised the quality of studies [Pai *et al.*, 2004]. In such cases, small but important effects can be missed, different reviewers can reach different conclusions using the same research base, and the reported findings often have less to do with the underlying evidence [Davies and Crombie, 2001]. In contrast to these narrative reviews, systematic reviews consist of the following main features [Pai *et al.*, 2004]:

- comprehensive, effective, and complete search for primary studies on a focused clinical question
- clear and reproducible selection criteria to select appropriate studies
- critical appraisal of the study quality
- pre-determined and explicit methods to represent results of the review

Evans [2001] describes systematic reviews as "summaries of all past research on a topic of interest. Unlike the traditional approach to reviewing literature, they utilize the same principles and rigor that is expected of primary research. As the name suggests, they are systematic in their approach and use methods that are pre-planned and documented in a systematic review protocol." In short, the critical appraisal and summarization of all available information related to a specific topic is called *Systematic Review*. In this case "systematic" refers to the systematic identification of

all available information about a specific topic. It also refers to systematic critical appraisal of the quality of these selected studies. Such well conducted systematic reviews provide the most secure and precise information (i.e., studies) and therefore appear at the top of the "Hierarchy of the level of evidence" by Evans (see Table 2.1). However, not all systematic reviews are of good quality. Therefore they should be appraised critically.

Grade	Description
Level I	Evidence obtained from a Systematic Review of all relevant randomized controlled trials.
Level II	Evidence obtained from at least one properly designed randomized control trials.
Level III.1	Evidence obtained from well designed controlled trials without randomization.
Level III.2	Evidence obtained from well designed cohort or case control analytic studies preferably from more than one center or research group
Level III.3	Evidence obtained from multiple time series with or without the intervention. Dramatic results in uncontrolled experiments.
Level IV	Opinion of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees.

Table 2.1: Hierarchy of the level of evidence [Evans, 2001]

High-quality systematic reviews take great care to find all relevant published and/or unpublished studies. Reviewers then assess each study, synthesise the findings from each individual study, and present a balanced summary of the findings with due consideration of any flaws in the evidence [Davies and Crombie, 2001].

Process of Systematic Reviews

During the development of systematic reviews several steps have to be followed in order to obtain good quality results (compare Figure 2.5). In the following these steps are described [Davies and Crombie, 2001]:

1. **Defining an appropriate therapeutic question:** Clear information about the intervention of interest, relevant patient groups, as well as appropriate outcomes are required to define an appropriate clinical question. These details will then be used to select studies for inclusion into the review.
2. **Searching the literature:** The use of only published studies may yield to an overestimation of the effects of the intervention. For an unbiased assessment, this search must cover all available literature, including non-English sources, conference proceedings, and company reports. Published and unpublished literature have to be carefully searched to find all reports of controlled trials about a certain intervention.

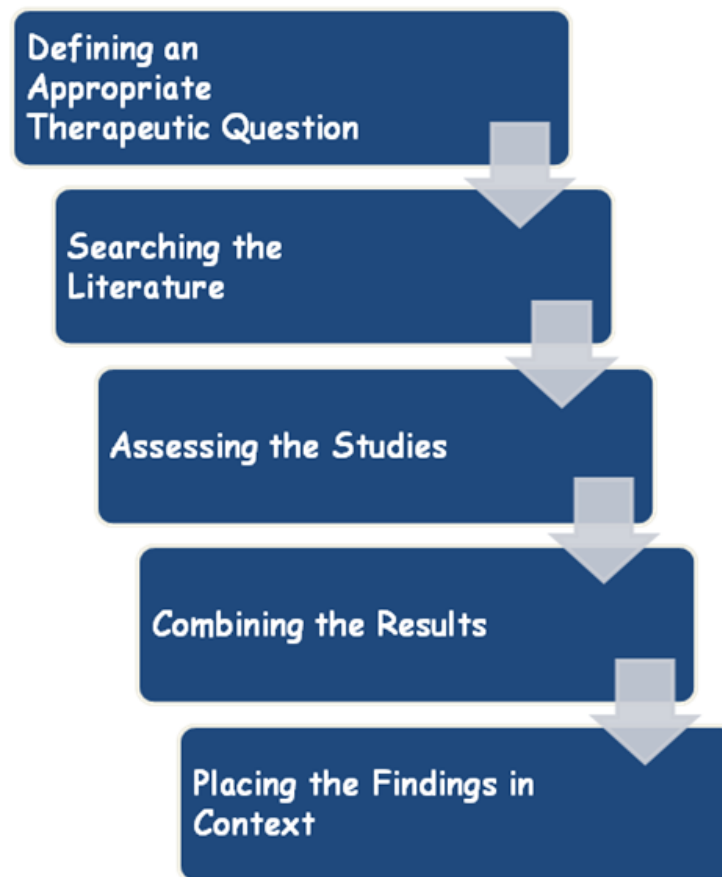


Figure 2.5: Process of Systematic Reviews

3. **Assessing the studies:** Once all possible study reports have been identified, each study has to be evaluated with regard to its appropriateness for inclusion into the review, its study quality, and its reported findings. Hereby, the best way is to involve more than one independent reviewer in the evaluation process.
4. **Combining the results:** To gain an overview regarding the effectiveness of an intervention or a treatment, the results from the included studies have to be combined. Depending on the type of data and the quality of the studies, this can be achieved by applying the meta-analysis method (see Section 2.4.1).
5. **Placing the findings in context:** The findings from this aggregation of an unbiased selection of studies then have to be discussed in order to put them in context. This task will address issues such as the quality and heterogeneity of the included studies, the likely impact of bias and chance, and the applicability of the findings to the current situation. Thus, judgment is not obviated by the rigour of systematic reviews, it has just reduced in terms of its impact and it has been made more explicit.

Critical Appraisal of Systematic Reviews

We stated that not all systematic reviews are of good quality. Therefore, they should be critically appraised by users to decide whether the methods are correct, to assess what the results are actually saying, and to decide whether these results can be applied locally for the current situation at hand [Hill and Spittlehouse, 2001]. The most common questions to be asked during the critical appraisal of systematic reviews are shown in Table 2.2 and are described below [David, 2000]:

Focus	Specific Questions
Question	Is the specific purpose of the review stated? Is the review question clearly and explicitly stated?
Literature Search	Were comprehensive search methods used to locate studies? Was a thorough search done of appropriate database and were other potentially important sources explored?
Study Selection	How were studies selected? Are the inclusion criteria reported?
Critical Appraisal	Was the validity of included studies assessed? Was the validity of studies assessed appropriately? Are the validity criteria reported?
Similarity of Groups and Treatments	Are treatments similar enough to combine? Were reasons for any differences between individual studies explored?
Data Synthesis	Were findings from individual studies combined appropriately? Are the methods used to combine the studies reported?
Methods Documented	Are review methods clearly reported?
Summary of Findings	Is a summary of findings provided? Are specific directives for new research proposed? Were the conclusion supported by the reported data?

Table 2.2: Checklist for Appraising Systematic Reviews [David, 2000]

Review Questions. Systematic reviews has to be based on clearly defined questions. These questions should contain the following elements:

- population of interest and condition
- intervention
- a comparison or control
- the outcome measure that is to be used to determine the effectiveness

These components of questions can be used to determine what type of studies are required to provide appropriate answers.

Literature Search. The critical part of systematic reviews is the location of existing research addressing the topic of interest. The primary goal of the literature search is to determine as much of the conducted research on the topic as possible. Like all steps in the review process, the search strategy should

be documented in detail to allow others to estimate its quality. Usually, systematic review searches include electronic databases, but additionally other specialized databases may also be searched. For example, bibliographies and reference lists of all retrieved articles should be searched as well to increase the likelihood of identifying all relevant studies.

Study Selection. The selection process of various kind of studies are of particular importance to decide which studies should be included in the review and which ones should not. While the review question defines the area of interest, it is the study selection that explicitly records the focus, nature, and limits of the review. This criteria is used to determine if the population, intervention, and outcome measures of a study are consistent with the focus of the systematic review.

Critical Appraisal of Studies. The appraisal of the validity of all identified studies is one of the important parts of the systematic review process. The expectation here is that by excluding lesser quality studies the risk of error and bias in the findings of the review will be reduced.

Similarity of Groups and Treatments. If studies are different in terms of their population, intervention, or how outcomes are measured, it makes no sense to use meta-analysis on them. Therefore, findings of individual studies that differ significantly should not be combined in meta-analysis.

Data Synthesis. The goal of a systematic review is to summarize the outcomes and results from various studies to obtain an overall evaluation of the effectiveness of an intervention or treatment. In this case, meta-analysis is used to support this task by providing a framework for a systematic review where similar measures from comparable studies are listed systematically and the measures of the effects of an intervention are combined. Meta-analysis are valuable when many studies address the same issue and when studies are too small and therefore lack the power to detect treatment effects, as combining studies increases the sample size and therefore the expressive power.

Reporting and Recommendations. The methods used during the development of a review should be documented in sufficient detail to allow reproduction of the review and critical appraisal of the applied processes.

Chapter 3

Guidelines and Evidence Information

Evidence-based CPGs for health care follow a rigorous development process and are based on the best available evidence. They contain information in various formats, like text, tables, flowcharts, graphs, maps, and lists. Therefore, the structure of CPGs can be complex and not easy to interpret. A difficult task of physicians is to correctly interpret the information in CPGs and to combine this information with their own knowledge and experience to derive medical decisions and treatment plans. Despite the mentioned difficulty, CPGs are increasingly used to prescribe how a physician should behave in certain circumstances during the medical treatment [Bosse, 2001]. When discussing the properties of evidence information in clinical practice guidelines (CPGs), we must first look at the methods used to convey information to physicians regarding the levels of evidence that support the major recommendations and the strengths assigned to the recommendations by the guideline developing organizations. Guideline users need to know how much confidence they can place in the recommendations [Atkins *et al.*, 2004a]. Therefore, we discuss and define CPGs in Section 3.1, and the evidence information including the Levels of Evidence (LoEs) and the Strengths of Recommendations (SoRs) that can be found in CPGs and play a major role in EBM in Section 3.2. We mentioned earlier that different guideline developing organizations use different representations for LoEs and SoRs. To give an overview of the most common representations we introduce some of them in Section 3.3.

3.1 Clinical Practice Guidelines

Clinical Practice Guidelines (CPGs) are "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances" [Field and Lohr, 1990]. This section describes the development process of CPGs (Section 3.1.1), because they extent recommendations for physicians and support them during the decision-making process, therefore they become a part of clinical practice and furthermore we discusses guideline properties that are necessary to develop qualitative CPGs (Section 3.1.2).

3.1.1 Guideline Development Process

Before presenting the development process of CPGs as described in [(SIGN), 2001; (NHS), 2006], we have to state that in order to foster the acceptance and usability of guidelines, the development methods have to be transparent to guideline users, because only then they can see that potential biases in guideline development have been addressed adequately and that the recommendations are internally and externally valid. In Figure 3.1 the development process of CPGs is depicted. In the following we will describe the phases of the process in detail.

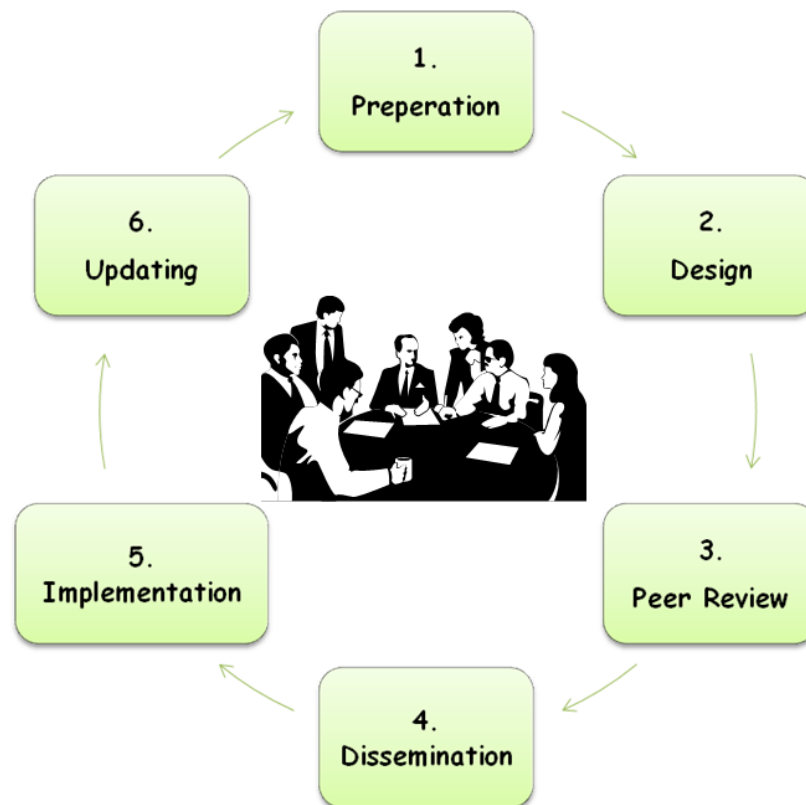


Figure 3.1: Guideline Development Process

Preparation

Topic Selection. The guideline development process begins with the selection of an appropriate topic to define the main areas the guideline should address. The following criteria have to be taken into account in selecting topics for guideline development [(SIGN), 2001]:

- Wide variations in practice and outcome
- High burden of disease
- High health care costs
- High prevalence of morbidity and mortality

- Potential to improve outcomes

Putting together a guideline development group. One of the most important steps while developing CPGs is to convene a guideline development group. This group should be multidisciplinary, including professionals, patients, and other health care providers. A multidisciplinary group is important to provide expertise regarding all stages in the patient's journey of care, to locate and critically evaluate all relevant scientific evidence, and to solve practical problems that will arise while using the guideline [(SIGN), 2001]. The members of the guideline development group are responsible for the formulation of specific clinical questions, grading levels of evidence, and developing the recommendations [(NHS), 2006]. The most important skills that members should possess are [(SIGN), 2001]:

- clinical expertise
- specialist expertise (e.g., health economics, social services)
- practical understanding of problems faced in the delivery of care
- communication and team working skills
- critical appraisal skills

Patient Involvement. Patients play an important role in the development group. The presence of patients is important to ensure that the CPGs reflect their needs and address issues that are significant to patients and carers, because physicians and other experts have a different perspective on health care priorities, processes and outcomes compared to those of the patients [(SIGN), 2001; (NHS), 2006].

Formulation of Specific Clinical Questions. The relevant clinical questions should be formulated as soon as the development group is convened. The questions have to "clearly identify the population concerned, the intervention under investigation, the type of control used, and the outcome measures used to measure the effectiveness of the interventions" [(SIGN), 2001]. They are also decisive for the systematic literature search and the development of recommendations by the development group. Questions about interventions can be developed with the Patient Intervention Comparison and Outcome (PICO) framework (see 3.2). Questions about diagnosis have to identify key issues specifically relevant to test their accuracy, reliability, safety and acceptability by the patient [(NHS), 2006].

Design

This stage of the development process starts with literature search to identify available evidence in order to translate the found evidence into recommendations.

Literature Search. Evidence-based guidelines should be based on systematic reviews, (see Section 2.4.2) therefore the literature search is performed according to an explicit search strategy. Several database are recommended by various guideline development organizations for literature search (e.g., Cochrane Library, Medline, Embase, Guideline International Network, National Guideline Clearinghouse). This literature search have to be followed by the critical appraisal of the identified literature. The literature has to be selected using defined inclusion and exclusion criteria to cover all relevant aspects of the defined clinical questions [(SIGN), 2001; (NHS), 2006].

Identifying the Evidence. There are several steps needed to identify the most appropriate data to answer the defined clinical questions. The identification of the evidence begins with selecting relevant studies, assessing their quality, synthesizing the results and finally grading the evidence [(NHS), 2006]. A key stage in the guideline development process is the assessment of the quality of studies, because the results will affect the level of evidence, which will again influence the strength of the recommendations. In this stage of the development process evidence tables play a significant role. They help to identify the similarities and the differences between studies and give information about the characteristics of the study population, interventions, and outcome measures [(NHS), 2006]. Evidence tables summarize all the validated studies collected with the literature review relating to each clinical question. These tables are important to facilitate comparing results across studies and ensure that the basis of the recommendations is transparent [(SIGN), 2001]. On the basis of studies, levels of evidence help the guideline developers and guideline users to understand the type of the evidence on which the recommendations have been based [(NHS), 2006].

However, forming clear and unambiguous recommendations for any given clinical question is a challenging task. SIGN has described a concept of considered judgment to facilitate this development process [(SIGN), 2001]:

- Quantity, quality, and consistency of evidence
- Generalizability of study findings
- Directness of application to the target population for the guideline
- Clinical impact
- Implementability

By means of this concept the development group can summarize their view of the evidence and assign a level of evidence to considered studies, before forming and grading recommendations.

Forming Recommendations. Guideline recommendations are based on the best available evidence, because they are graded to differentiate between recommendations based on strong evidence and those based on weak evidence. This assessment is made on the basis of critical appraisal of the design and quality

of each study. Recommendations should be clear, unambiguous and easy to translate into clinical practice. Many users have not enough time to read the whole guideline or they are only interested in the recommendations, as they are important for the decision making process in clinical practice. Therefore, they should be clear and transparent. Translating the evidence into recommendations is a challenging task. The National Institute for Health and Clinical Excellence (NHS) mentions several problems that mostly occur during the formulation of recommendations [(NHS), 2006]:

- Literature search resulted in no evidence that answers the clinical questions
- Quality of the found evidence is poor
- Available clinical evidence is conflicting and of similar level
- The evidence is not directly applicable to the population covered by the guideline (e.g., different age group)
- No published estimation of cost-effectiveness that is applicable to the relevant population is available

It was reported that most of the disagreements between the guideline development group members occur at the stage of the development process. Therefore, it should be clearly documented how this disagreements have been resolved by the development group.

Peer Review

All guidelines should be reviewed by independent expert referees prior to publication. The best time for reviewing is during the development process where the draft recommendations of each guideline can be discussed with health care professionals, patient representatives and others interested in the topic covered by the guideline [(SIGN), 2001]. The aim of this part of the development process is to ensure that any risk of bias in the development process has been minimized. The main issue of this development phase is to increase the likelihood of a successful implementation of CPGs into clinical practice for the benefit of patients [(SIGN), 2001].

Dissemination

Dissemination means to bring the guideline to the attention of the guideline users. The guidelines should be published in different forms to reach the target group and all relevant organizations. They can be published and presented in medical journals or by means of the World Wide Web.

Implementation

The guideline implementation phase is also challenging. There exist no "cookbook" how to present and implement guidelines, because each implementation strategy

is effective under certain circumstances. However, SIGN presents six steps towards guideline implementation to provide a direction to overcome several problems (compare [(SIGN), 2001]).

Updating CPGs

Updating CPGs is an important task to provide guideline users with updated and state-of-the-art knowledge. The medical progress results in changes in evidence, benefits and harms, outcomes, available interventions, improvement in current performance, and resources available for health care. Therefore, updated levels of evidence and strengths of recommendations are major quality criteria to judge the validity of CPGs. The most important changes that can occur over time can be described as follows [Shekelle *et al.*, 2001]:

Changes in benefits and harms of interventions. The actual strength of the benefits and harms can make the existing information in CPGs irrelevant.

Changes in outcomes. Since the development of a CPG, new outcomes may have been identified, which had not been recognized earlier.

Changes in available interventions. New preventive, diagnostic, or treatment interventions may have emerged to replace or complete the existing ones.

Changes in evidence that current practice is optimal. Guidelines should provide an ideal clinical practice. Therefore, the gap between the ideal and the current clinical practice has to be narrowed using CPGs. There can be changes over time where a guideline is no longer needed.

Changes in values placed on outcomes. The values of outcomes are placed differently from situation to situation. Currently, in most guidelines costs play not an important role, but they will be considered explicitly in CPGs in the future.

Changes in resources available for health care. The increase of available resources over time have to be included in the CPGs.

Guidelines should be updated when new information becomes available. Currently, CPGs are updated every two to five years, which is a problem, because mostly knowledge in guidelines ages rapidly. A new concept to update CPGs is to develop so called "living guidelines", which are being updated annually [Twaddle, 2005].

3.1.2 Quality of CPGs

The positive effect of CPGs in clinical practice have been demonstrated by several studies [Lobach and Hammond, 1997; Grimshaw and Russel, 1993]. Therefore, the number of guidelines available to physicians grows rapidly. This causes that guideline developers have an increasing responsibility to develop guidelines with

several requirements: validity, reliability, applicability, flexibility, clarity, multidisciplinary, updatability, and usability. These requirements have been defined in [Field and Lohr, 1990] as follows:

Validity. CPGs are valid if they lead to the health and cost outcomes reported in them. The most important factors regarding the validity of CPGs are:

- Relationship between the evidence and the recommendations
- Quality of the scientific and clinical evidence
- Means used to evaluate the evidence
- Outcomes and costs of alternative courses of action

Reliability. One way to weight the reliability of CPGs is, given the same evidence and methods for the guideline development, two different expert panels produce essentially the same statements. An important factor for the reliability is also that the information in the guidelines are interpreted and applied consistently by different physicians under the same circumstances.

Clinical Applicability. CPGs should include information about the target population.

Clinical Flexibility. Identification of the specifically known or generally expected exceptions to recommendations in CPGs are very important.

Clarity. CPGs should be clear, transparent, and easy to understand. Use of unambiguous language, precisely terms, and transparent modes of the presentation are crucial factors. The major recommendations have to be distinguishable, significant on their own, and explicitly interpretable.

Multidisciplinary Process. The guideline development process should include all representatives (e.g., patients, experts).

Scheduled Review. CPGs should include information when a guideline should be updated.

Documentation. Usability of CPGs increases if the guidelines are well structured and include information about the involved participants, levels of evidence, strength of recommendation, development methods, and study design.

In this section we introduced CPGs in general by describing their development process and main requirements to weight their quality. However, as we stated before, our focus lies on the evidence information in CPGs as they are essential for facilitating medical decision-making process. In the next section we describe what we consider as evidence information (i.e., LoEs and SoRs).

3.2 Evidence Information in CPGs

Clinical practice guidelines are not consistent in the form or extent of the evidence information they contain. Some guidelines contain explicit levels of evidence, others do also contain explicit grades to indicate the strength of recommendations, and again others do not contain any evidence information at all. Therefore, we have to state what we mean when talking about "evidence information in CPGs".

3.2.1 Levels of Evidence

The Levels of Evidence (LoEs) identify the similarities and differences between studies and give information about the characteristics of the study population, interventions, or outcome measures [(NHS), 2006]. They help physicians to understand the type of the evidence on which the recommendations are based. LoEs are generally represented in form of evidence tables and are important to facilitate comparing results across studies and to ensure that the basis of recommendations is transparent [(SIGN), 2001]. Making decision about grading the evidence requires assessments of the validity of results of individual studies for important outcomes. Thereby, four key elements are of particular importance [Atkins *et al.*, 2004a]:

Study design. Study design refers to a type of study that can be classified in systematic reviews, meta-analysis, RCTs and observational studies (see Section 2.3 and Section 2.4).

Study quality. The quality of individual studies apply to the detailed study methods and execution of the study. Guideline developers should use appropriate criteria to estimate the study quality for each important outcome.

Consistency. Consistency refers to the similarity of assessments of effects across studies. The confidence in the effect for that outcome decreases in the case of essential unexplained inconsistency in the results.

Directness. Directness relate to the considered extent to which people, interventions, and results in the study are similar to those of interest. For example, uncertainty about the directness of the evidence can occur "if the people of interest are older, sicker, or have more comorbidity than the people considered in the studies." [Atkins *et al.*, 2004a]

The above mentioned descriptions of the four elements have to be considered while grading the LoEs. For practical purposes, judgment should be made in the context of systematic reviews (compare Section 2.4.2). However, judgment about the overall quality of the evidence and recommendations typically requires information beyond the results of a review [Atkins *et al.*, 2004a].

3.2.2 Strengths of Recommendations (SoRs)

As mentioned before, in CPGs recommendations are of particular importance, because they are intended to influence physicians behavior. Guideline recommendations should convey clear, informative, and helpful information to physicians during

the decision-making process. Therefore, they should be distinguishable, significant on their own, and explicitly interpretable. The classification of the recommendations are based on the assessment of the study design they are based on (see Section 2.3) and the quality of each study. The consistency, clinical relevance, and the external validity of the whole body of evidence is of particular importance in order to assign an appropriate strength to the major recommendations in CPGs [(SIGN), 2001]. They are graded to differentiate between recommendations based on strong evidence and those based on weak evidence.

Evidence-based recommendations are mostly classified in particular grading schemes to provide a unique format at least for guidelines of the same developing organization. Therefore, guideline developing organizations have to consider a number of factors when grading recommendations as defined by Guyatt and colleagues [2006]:

- Methodological quality of the evidence supporting estimates of likely benefit, and likely risk, inconvenience, and costs
- Importance of the outcome that treatment prevents
- Magnitude of treatment effect
- Precision of estimate of treatment effect
- Risks associated with therapy
- Burdens of therapy
- Risk of target event
- Costs
- Varying values

To give an overview of how LoEs and SoRs are represented in different CPGs developed by different developing organizations we introduce some of them in the next section.

3.3 Guideline Developing Organizations

In 1979, the Canadian Task Force on the Periodic Health Examination published one of the first efforts to explicitly characterize the LoEs underlying health care recommendations and the SoRs [Atkins *et al.*, 2004b]. Since then, a large number of organizations developed various representations of LoEs and SoRs, which are classified according to the LoEs on which they are based.

In our work we consider eight different guideline developing organizations and one co-operation, and 21 CPGs with the specialty otolaryngology (see Table 3.1). Considering these CPGs, we have to deal with eight different rating schemes of LoEs and three different representations of the SoRs that are used by the developing organizations of the guidelines. This section contains a brief description of these guideline developing organizations and their grading systems.

Organization	Title	Year	Pages
AAFP; AAOHNS; AAP	Otitis media with effusion.	2004	15
AAP	Recommendation for influenza immunization of children.	2004	12
	Reduction of the influenza burden in children.	2002	12
ARIA	Allergic rhinitis and its impact on asthma.	2003	26
CCHMC	Evidence based clinical practice guideline for medical management of acute otitis media in children 2 months to 13 years of age.	2004	20
	Evidence based clinical practice guideline for medical management of otitis media with effusion in children 2 months to 13 years of age.	2004	19
	Evidence based clinical practice guideline for children with acute bacterial sinusitis in children 1 to 18 years.	2001	17
FMSD	Sore throat and tonsillitis.	2005	12
ICSI	Acute sinusitis in adults.	2004	21
	Acute pharyngitis.	2005	20
	Diagnosis and treatment of obstructive sleep apnea.	2005	29
	Diagnosis and treatment of otitis media in children.	2004	21
	Rhinitis.	2003	23
	Viral upper respiratory infection (VURI) in adults and children.	2004	20
PGI	Symptomatic treatment of radiation-induced xerostomia in head and neck cancer patients.	2004	11
SIGN	Diagnosis and management of childhood otitis media in primary care.	2003	16
	Management of obstructive sleep apnoea/hypopnoea syndrome in adults.	2003	17
	Management of patients with stroke: identification and management of dysphagia.	2004	18
UMHS	Allergic rhinitis.	2002	12
	Management of gastroesophageal reflux disease (GERD).	2002	11
	Otitis media.	2002	12

Table 3.1: Evidence-Based Guidelines with the speciality Otolaryngology

3.3.1 AAFP, AAOHNS and AAP

The co-operation of American Academy of Family Physicians (AAFP), American Academy of Otolaryngology - Head and Neck Surgery (AAOHNS), and the American Academy of Pediatrics (AAP) [(AAFP) *et al.*, 2004] use a common representation of LoEs (see Table 3.2) and SoRs. The SoRs and the guideline definitions for the evidence-based statements are defined in this guideline as follows [(AAFP) *et al.*, 2004]:

- **”Strong Recommendation:** The benefits of the proposed recommendations clearly exceed the harms and the quality of the supporting evidence is of **Grade A** or **Grade B**. It is possible that strong recommendations are based on lower evidence if high-quality evidence is impossible to obtain and the anticipated benefits strongly outweigh the harms.
Implication: Clinicians should follow a strong recommendation unless a clear and compelling rationale for an alternative approach is present.
- **Recommendation:** The benefits of the recommended approach exceed the harms, but the quality of the evidence is not stronger than **Grade B** or **Grade C**. In some identified situations, recommendations may be based on lower evidence if the estimated benefits outweigh the harms.
Implication: Clinicians should generally follow a recommendation but they should also stay observant to new information and be sensitive to patient preferences.
- **Option:** An option means that either the quality of evidence that exists is suspect (**Grade D**) or that well-done studies (**Grade A, Grade B, or Grade C**) show little clear advantage to one approach over another.
Implication: Clinicians should be flexible in their decision-making and patient preference should have a substantial influence.
- **No Recommendation:** No recommendation means that there is a lack of pertinent evidence of **Grade D** and an unclear balance between benefits and harms.
Implication: Clinicians should be up-to-date to newly published evidence that clarifies the balance of benefits versus harms. The patient preference should also have a substantial influence.”

Here are some examples of the way evidence ratings may appear in the text of a guideline from the cooperation. The LoE is represented as ”Aggregate evidence quality” and the SoR as a ”Policy level” [(AAFP) *et al.*, 2004]:

- **”Pneumatic Otoscopy:** Clinicians should use pneumatic otoscopy as the primary diagnostic method for otitis media with effusion (OME),....

Level	Definition
Grade A	Well-designed, randomized controlled trials or diagnostic studies performed on a population similar to the guidelines's target population
Grade B	Randomized controlled trials or diagnostic studies with minor limitations; overwhelmingly consistent evidence from observational studies
Grade C	Observational studies (case-control and cohort design)
Grade D	Expert opinion, case reports, or reasoning from first principles (bench research or animal studies)

Table 3.2: Level of Evidence used by the American Academy of Family Physicians, American Academy of Otolaryngology-Head and Neck Surgery and American Academy of Pediatrics [(AAFP) *et al.*, 2004]

(This is a strong recommendation based on systematic review of cohort studies and the preponderance of benefit over harm).

Aggregate evidence quality: A, diagnostic studies in relevant populations
Policy level: strong recommendation"

- **"Tympanometry:** Tympanometry can be used to confirm the diagnosis of OME.
 (This option is based on cohort studies and a balance of benefit and harm)

Aggregate evidence quality: B, diagnostic studies with minor limitations
Policy level: option"

3.3.2 American Academy of Pediatrics (AAP)

For developing guidelines, the American Academy of Pediatrics (AAP) search electronic databases to collect the evidence information. To analyze the quality of the evidence, they use systematic reviews with evidence tables. The major recommendations are graded with the evidence information defined in Table 3.3 and are represented in the guidelines in parenthesis with the designation "evidence grade" [(AAP), 2004; (AAP), 2002]:

"Practitioners should increase their efforts through tracking and recall systems to ensure that children traditionally considered at high risk of

Level	Definition
I	Evidence obtained from at least one properly designed, randomized controlled trial
II-1	Evidence obtained from well-designed controlled trials without randomization
II-2	Evidence obtained from well-designed cohort or case-control analytic studies, preferentially from more than one center or group
II-3	Evidence obtained from multiple time series with or without the intervention, or dramatic results in uncontrolled experiments (such as the results of the introduction of penicillin treatment in the 1940s)
III	Opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees

Table 3.3: Level of Evidence used by the American Academy of Pediatrics

severe disease and complications from influenza receive annual immunization. High-risk children and adolescents who should receive priority for influenza immunization are those with the following (**evidence grade II-3**):

- Asthma or other chronic pulmonary diseases, such as cystic fibrosis
- Hemodynamically significant cardiac disease
- Immunosuppressive disorders or therapy
- Human immunodeficiency virus (HIV) infection
- Sickle cell anemia and other hemoglobinopathies
- Diseases requiring long-term aspirin therapy, such as rheumatoid arthritis or Kawasaki syndrome
- Chronic renal dysfunction
- Chronic metabolic disease, such as diabetes mellitus”

3.3.3 Allergic Rhinitis and its Impact on Asthma Workshop Group (ARIA)

The Allergic Rhinitis and its Impact on Asthma Workshop Group (ARIA) selects evidence information with a manual search of published literature and electronic

Level	Definition
Ia	Evidence for meta-analysis of randomized controlled trials
Ib	Evidence from at least one randomized controlled trial
IIa	Evidence from at least one controlled study without randomization
IIb	Evidence from at least one other type of quasi-experimental study
III	Evidence from non-experimental descriptive studies, such as comparative studies, correlation studies and case-control studies
IV	Evidence from expert committee reports or opinions or clinical experience of respected authorities, or both

Table 3.4: Category of evidence used by the Allergic Rhinitis and its Impact on Asthma Working Group [Bousquet *et al.*, 2001]

databases. They search Medline using PubMed with special keywords to find papers with information about treatment and diagnosis options. They also search in EMBASE to find papers, which are important for treatment options and are not available in Medline. For further information they used the Cochrane Database of Systematic Reviews (CDSR) and the Database of Reviews of Effectiveness (DARE). The quality and the strength of evidence is based on a weighting according to a rating scheme (see Table 3.4) [Bousquet *et al.*, 2001].

Meta-analyses and systematic reviews are used to analyze the evidence information. The literature is reviewed and the treatment options are appraised by an expert committee. The LoEs used by ARIA base on available studies, which are published in papers found in Medline and EMBASE, while the recommendations are formulated reaching full consensus among the involved experts [Bousquet *et al.*, 2001]. Their rating scheme for the SoRs is shown in Table 3.5.

In [Bousquet *et al.*, 2001] LoEs and SoRs are represented in parenthesis after a description of a possible treatment method. The strength of the recommendations include also a definition of the population to which this recommendation apply:

- "The early identification of occupational sensitizers and the removal of sensitised patients from any further exposure are important aspects of the management of occupational rhinitis.
- Prevention of latex allergy is essential." [Bousquet *et al.*, 2001]

Strength of Recommendation	Definition
A	Directly based on category I evidence (A* for recommendations based on double-blind studies without a control group).
B	Directly based on category II evidence or extrapolated recommendation from category I evidence
C	Directly based on category III evidence or extrapolated recommendation from category I or II evidence.
D	Directly based on category IV evidence of extrapolated recommendation from category I, II or III evidence.

Table 3.5: Recommendation grades used by the Allergic Rhinitis and its Impact on Asthma Working Group [Bousquet *et al.*, 2001]

”(Level of evidence = IV; Strength of recommendation = D for adults and children with perennial rhinitis or adults and children with latex allergy.)” [Bousquet *et al.*, 2001]

3.3.4 Cincinnati Children’s Hospital Medical Center (CCHMC)

The guideline development at the Cincinnati Children’s Hospital Medical Center (CCHMC) begins with searches of electronic databases like the Medline, EMBASE, and the Cochrane Database. The CCHMC search with several keywords for papers to answer special clinical questions. They eliminate duplicates, review articles (also non-english articles and older articles) to select the evidence information with high quality and without irrelevant information. For assessing the quality and SoRs they use subjective reviews and a weighting according to a rating scheme. The analysis of the evidence is based on reviews, reviews of published meta-analyses, and systematic reviews. Recommendations are formulated by an interdisciplinary group [(CCHMC), 2004a; (CCHMC), 2004b; (CCHMC), 2006] and use the grading scale shown in Table 3.6.

The LoEs and the SoRs are represented in the guidelines in squared brackets as:

”Persistence of upper respiratory symptoms for greater than 10 days without improvement (Wald et al., 1981 [B]; Wald et al., 1984 [B]; Wald, Chiponis, & Ledesma-Medina, 1986 [B]; Isaacson, 1996 [S]; Brook et al., 2000 [E]; Wald, 1994 [S]).” [(CCHMC), 2006]

Level	Definition
A	Randomized controlled trial: large sample
B	Randomized controlled trial: small sample
C	Prospective trial or large case series
D	Retrospective analysis
E	Expert opinion or consensus
F	Basic laboratory research
S	Review article
M	Meta-analysis
Q	Decision analysis
L	Legal requirement
O	Other evidence
X	No evidence

Table 3.6: Classes of evidence used by the Cincinnati Children's Hospital Medical Center [(CCHMC), 2004a; (CCHMC), 2004b; (CCHMC), 2006]

Level	Definition
A	Strong research-based evidence: several relevant, high-quality scientific studies with homogeneous results
B	Moderate research-based evidence: at least one relevant, high-quality study or multiple adequate studies
C	Limited research-based evidence: at least one adequate scientific study
D	No scientific evidence: expert panel evaluation of other information

Table 3.7: Levels of evidence used by the Finnish Medical Society Duodecim [(FMSD), 2005]

3.3.5 Finnish Medical Society Duodecim (FMSD)

This guideline development group of the Finnish Medical Society Duodecim (FMSD) manually searches published literature and electronic databases (Cochrane Database, Database of Abstracts of Reviews of Effectiveness (DARA)) to select the evidence [(FMSD), 2005]. The quality of the strength of evidence is based on systematic reviews and is weighted according to a rating scheme which is shown in Table 3.7.

The SoRs are represented in squared brackets as:

”Antibiotics shorten the duration of symptoms somewhat (”Antibiotics for sore throat,” 2002) [A] and reduce the risk of rheumatic fever (”Penicillin for secondary prevention of rheumatic fever,” 2002) [C].” [(FMSD), 2005]

3.3.6 Institute for Clinical Systems Improvement (ICSI)

The Institute for Clinical Systems Improvement (ICSI) selects the evidence with searches of electronic databases and analyzes them with systematic reviews, reviews of published meta-analyses, and systematic reviews with evidence tables [(ICSI), 2004a; (ICSI), 2005; (ICSI), 2006; (ICSI), 2004b; (ICSI), 2003; (ICSI), 2004c]. However, there is no information about the methods used to formulate the recommendations and the methods used to select and assess the quality of the evidence. The LoEs are defined as shown in Table 3.8. They appear in [(ICSI), 2004a] with the designation ”Evidence supporting...”:

Antibiotics

Level	Definition
Class A	Randomized, controlled trials
Class B	Cohort study
Class C	Non-randomized trial with concurrent or historical controls Case-control study Study of sensitivity and specificity of a diagnostic test Population-based descriptive study
Class D	Cross-sectional study Case series Case report
Class M	Meta-Analysis Systematic Review Decision analysis Cost-effectiveness study
Class R	Consensus statement Consensus report Narrative review
Class X	Medical Opinion

Table 3.8: Classes of research reports used by the Institute for Clinical Systems Improvement [(ICSI), 2004a; (ICSI), 2005; (ICSI), 2006; (ICSI), 2004b; (ICSI), 2003; (ICSI), 2004c]

- "Amoxicillin: 500 mg tab three times per day (TID) 10 days or 875 mg tab two times per day (BID) 10 days
- For those allergic to amoxicillin: Trimethoprim-sulfamethoxazole (TMP/SMX): one double strength tab BID 10 days"

"Evidence supporting the conclusion on antibiotics is of classes: A, C, M, R"

3.3.7 Practice Guidelines Initiative (PGI)

The Practice Guidelines Initiative (PGI) uses manual searches of published literature and searches in electronic databases (e.g., Medline, Cancerlit, Embase) to select the evidence. The recommendations are supported by randomized controlled trials. They search databases with different search terms to answer their stated clinical questions and use different criteria to select articles and eliminate irrelevant information. The inclusion criteria for the selection of the articles is that they

are English publications and represent RCTs that measure symptomatic relief of radiation-induced xerostomia in head and neck cancer [Head and Neck Cancer Disease Site Group *et al.*, 2004]. The analysis of the evidence is based on meta-analysis of the RCTs and systematic reviews with evidence tables, while the recommendations have been formulated by an expert committee. The major recommendations are not classified and appear in the guidelines as [Head and Neck Cancer Disease Site Group *et al.*, 2004]:

”Adverse events were dose-related. Adverse parasympathetic events were reported by participants in randomized controlled trials, the most frequent and troublesome being increased 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.” [Head and Neck Cancer Disease Site Group *et al.*, 2004]

3.3.8 Scottish Intercollegiate Guidelines Network (SIGN)

The Scottish Intercollegiate Guidelines Network (SIGN) develops evidence-based guidelines designed by multidisciplinary groups. These guidelines are based on a systematic review of scientific evidence and the recommendations are precisely associated to the supporting evidence and are graded considering the strength of that evidence.

In 1998 SIGN decided to improve their existing system for grading guideline recommendations because of several reasons. One of the reasons was that they used Randomized Controlled Trials (RCTs). Also, RCTs were accepted as the most robust study design with the least risk of bias in the results, under certain circumstances, it was not practical or ethical to undertake them. There were common situations where other types of study design provided the best evidence. The misinterpretation of the grading system and the level of evidence represented also a problem. Instead of relating to the strength of the supporting evidence, the guideline users related to the importance of the recommendation. Therefore they failed to give due weight to low grade recommendations [Harbour and Miller, 2001].

After a detailed consultation and international peer review, the SIGN presented their new grading system with several levels of evidence (see Table 3.9) and grades of recommendations (see Table 3.10) in autumn 2000.

The strength of the recommendations appear in their guidelines as:

”**B** - If an antibiotic is to be prescribed, the conventional five day course is recommended at dosage levels indicated in the British National Formulary.

A - Children with acute otitis media should not be prescribed decongestants or antihistamines.” [(SIGN), 2003a]

Level	Definition
1++	High quality meta-analysis, systematic reviews of randomized controlled trials (RCTs), or RCTs with a very low risk of bias
1+	Well-conducted meta-analysis, systematic reviews of RCTs, or RCTs with a low risk of bias
1-	Meta-analysis, systematic reviews of RCTs, or RCTs with a high risk of bias
2++	High quality systematic reviews of case control or cohort studies. High quality case control or cohort studies with a very low risk of confounding or bias and a high probability that the relationship is causal
2+	Well-conducted case control or cohort studies with a low risk of confounding or bias and a moderate probability that the relationship is causal
2-	Case control or cohort studies with a high risk of confounding or bias and a significant risk that the relationship is not causal
3	Non-analytic studies, e.g., case report, case series
4	Expert Opinion

Table 3.9: Levels of evidence used by the Scottish Intercollegiate Guidelines Network [(SIGN), 2003a; (SIGN), 2003b; (SIGN), 2004]

Strength of Recommendation	Definition
A	<p>At least one meta-analysis, systematic review of randomized controlled trials (RCTs), or randomized controlled trial rated as 1++ and directly applicable to the target population; or</p> <p>A body of evidence consisting principally of studies rated as 1+, directly applicable to the target population, and demonstrating overall consistency of results</p>
B	<p>A body of evidence including studies rated as 2++, directly applicable to the target population, and demonstrating overall consistency of results; or</p> <p>Extrapolated evidence from studies rated as 1++ or 1+</p>
C	<p>A body of evidence including studies rated as 2+, directly applicable to the target population, and demonstrating overall consistency of results; or</p> <p>Extrapolated evidence from studies rated as 2++</p>
D	<p>Evidence level 3 or 4; or</p> <p>Extrapolated evidence from studies rated as 2+</p>

Table 3.10: Grades of strengths of recommendations used by the Scottish Intercollegiate Guidelines Network [(SIGN), 2003a; (SIGN), 2003b; (SIGN), 2004]

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

Table 3.11: Levels of evidence used by the University of Michigan Health System - Academic Institution [(UMHS), 2002a; (UMHS), 2002b; (UMHS), 2002c]

3.3.9 University of Michigan Health System - Academic Institution (UMHS)

This guideline developing organization of the University of Michigan Health System (UMHS) use electronic databases for search to select the evidence and assess the quality and strength of the evidence by weighting them according to a rating scheme. For analyzing the evidence they use systematic reviews [(UMHS), 2002a; (UMHS), 2002b; (UMHS), 2002c]. Their rating scheme for the strength of the evidence is described in Table 3.11.

The SoRs are rated on a scale of these evidence-levels from **A** to **D**. These ratings reflect the quality of evidence and involve a trade-off between harms and benefits. How this grading system is used in their guidelines is showed below, where the levels of evidence are placed in squared brackets at the end of the sentence [(UMHS), 2002c]:

”For isolated symptomatic episodes of AOM, the antibiotic of choice is amoxicillin (at a dose of 60 to 90 mg/kg/day, divided dosing twice a day [div b.i.d.] for 5 to 10 days). Treat AOM that is clinically unresponsive to amoxicillin after 72 hours of therapy with high-dose amoxicillin/clavulanate [C]. Patients with persistent symptoms on high-dose amoxicillin/clavulanate should receive 1 to 3 doses of intramuscular (IM) ceftriaxone [C].” [(UMHS), 2002c]

In this chapter we emphasized the need for CPGs as means to increase the quality of clinical practice. The importance of CPGs resulted in several guideline developing organizations developing their own guidelines using their own terminology to represent levels of evidence and based on these the overall strengths of their recommendations. We also introduced some these organizations, because we will propose our method to represent LoEs and SoRs based on their representations. But before that we are going to take a look at main guideline representation languages in order to decide which of them we can use and how we have to adopted them for our purposes.

Chapter 4

Guideline Representation Languages

During the last years, various guideline representation languages and systems have been developed to represent CPGs in a form that is interpretable by computer systems. Such computer interpretable guidelines can be used by computer systems to provide guideline querying, electronic distribution, and automatic decision-support. Based on these tasks, an ideal knowledge model to represent guidelines has to be [Shiffman *et al.*, 2000]:

Comprehensive: Knowledge models for guideline representation should be capable of expressing as much as possible knowledge contained in CPGs.

Expressively adequate: Knowledge models for guideline representation should convey the complexities and nuances of clinical practice while remaining informationally equivalent to the original guideline.

Flexible: Knowledge models for guideline representation should enable the modeling at high and low levels of granularity to interpret guidelines at different levels of abstraction. Therefore, flexibility is the basis for a useful model to deal with the variety of complexity of CPGs.

Comprehensible: Knowledge models for guideline representation should match the stakeholders' usual problem-solving language and so allow domain experts to describe their knowledge with little effort.

Shareable: Knowledge models for guideline representation should be sharable across different institutions.

Reusable: Knowledge models for guideline representation should provide reusability across all phases of the guideline life cycle.

In this section we describe the guideline representation languages Asbru (Section 4.1) and PROforma (Section 4.2), because they have been developed to handle various concepts that care formalization implies and many guideline modeling tools (e.g., AsbruView, Tallis) are based on these languages. We focus on the main features, the syntax and semantics of these languages. Furthermore, we give a brief overview on their methods and on available tools (e.g., AsbruView, DELT/A, Arezzo, Tallis) that are based on these languages and are in use to support guideline users in treatment planing and decision-making.

4.1 Asbru

Based on specific requirements, a time-oriented, intention-based, and sharable guideline representation language, called Asbru [Miksch *et al.*, 1997; Kosara *et al.*, 1998; Miksch *et al.*, 1998], has been developed as part of the Asgaard¹ project [Shahar *et al.*, 1998b] to represent CPGs as time-oriented, skeletal plans, to support both, designers and executors of skeletal plans. Asbru provides designers with a means to represent both, the prescribed actions of a skeletal plan and the knowledge roles required by the problem-solving methods performing the intertwined supporting sub-tasks. The major features of Asbru can be stated as follows [Miksch, 1999]:

- prescribed actions and states can be continuous
- intentions, conditions, and world states are defined as temporal patterns
- uncertainty in temporal scopes as well as in parameters can be represented with bounding intervals
- plans can be executed in several ways (i.e., sequential, parallel, particular order, or periodically)
- particular conditions are used to follow and control the plan's execution
- for each plan explicit intentions and preferences can be defined separately

4.1.1 Main Components of Asbru

Skeletal plans are organized and represented in the *plan-specification library*, where a plan is hierarchically composed of a set of plans with arguments and time annotations. A plan consists of a name, a set of arguments, containing a time annotation and the following components [Miksch *et al.*, 1997; Kosara *et al.*, 1998; Miksch *et al.*, 1998; Seyfang *et al.*, 2002; Miksch, 1999]:

Preferences. express the desired behavior of the plan to achieve a given goal. Hereby we can differentiate between five components [Miksch *et al.*, 1997; Kosara *et al.*, 2002]:

1. **”Strategy:** a general strategy to deal with the problem (e.g., aggressive, normal)
2. **Utility:** a set of utility measures (e.g., minimize cost or inconvenience)
3. **Select-method:** a matching heuristic for the applicability of the whole plan (e.g., exact-fit, roughly-fit)”
4. **Resources:** a specification for prohibited or obligatory resources

¹In Norse mythology, Asgaard was the home of the gods. It was located on the heavens and was accessible only over the rainbow bridge, called Asbru (or Bifrost).

5. **Start-condition:** an indication whether the transition from a ready generic plan to the started state of an actual plan instance is automatic or requires approval of the user”

Intentions. are context-dependent temporal patterns of executing-agent actions and external-world states that should be maintained, achieved, or avoided. Such information are of particular importance to select the right plan and to critique treatment plans. This part of Asbru consists of the following four categories [Miksch *et al.*, 1997]:

1. **”Intermediate state:** the state that should be maintained, achieved, or avoided during the applicability of the plan.
2. **Intermediate action:** the action that should take place during the execution of the plan.
3. **Overall state pattern:** the overall pattern of states that should hold after finishing the plan.
4. **Overall action pattern:** the overall pattern of actions that should hold after finishing the plan.”

Conditions. are temporal patterns needed to hold in order for a plan to be (see Figure 4.1):

- started,
- suspended,
- reactivated,
- aborted, or
- completed.

Different conditions are specified enabling the transition from one plan state into another. An important factor is that a plan is completed only when the completed conditions become true. In other cases the plan’s execution will be suspended or aborted. Different kinds of conditions are defined and illustrated as follows [Miksch *et al.*, 1997]:

1. **”Filter-preconditions:** useful to hold initially if the plan is applicable but can not be achieved, and are necessary for a possible state.
2. **Setup-preconditions:** are needed to be achieved to enable to start and allow the transition from a possible plan to a ready plan.
3. **Suspend-conditions:** define the condition when a started plan has to be suspended.
4. **Abort-conditions:** determine when a started, suspended, or restarted plan has to be aborted.

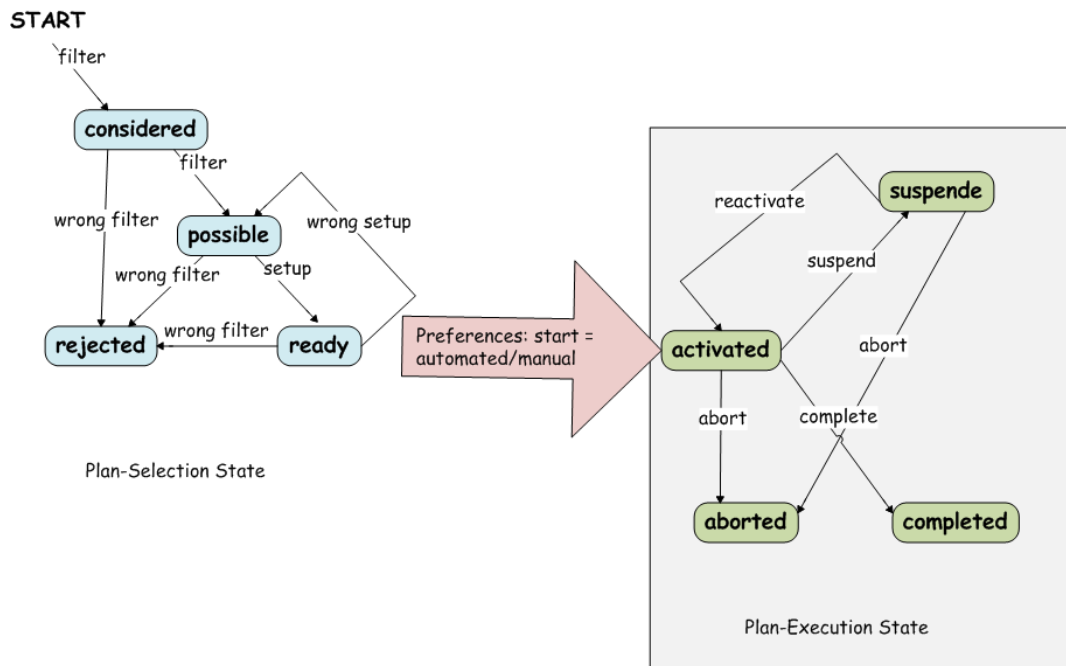


Figure 4.1: Plan execution and selection in Asbru [Miksch *et al.*, 1997]

5. **Complete-conditions:** determine when a started or restarted plan has to be completed successfully.
6. **Restart-conditions:** determine when a suspended plan has to be restarted.”

In cases where a plan is aborted, it means that it has failed to reach its goal, whereas if a plan completes, it means that it has reached its goal and the next plan in the sequence has to be executed.

Effects. define the functional relationship between the plan arguments and measurable parameters using mathematical functions or the overall effect of a plan on parameters. Effects have a likelihood annotation that represents the probability of their occurrence [Kosara *et al.*, 1998].

Plan body. consists of a set of plans or actions to be executed or performed in parallel, sequence, in any order, or in some order when the precondition holds. In summary we differentiate between the following kinds of plans [Kosara and Miksch, 2001; Kaiser, 2005]:

”**Sequential:** means that the set of plans are executed in sequence and can be activated if the preceding plan is finished. The state of the plan has to be completed or aborted.

Parallel: means that plans start at the same time, but they do not need to be finished at the same time.

Any-order: means that only the set of plans to be used are known, but not the order of the plans. However, only one plan can be in the state 'activated' at one time.

Unordered: means that all plans are executed without any synchronization.

Subplans: means that a plan can be decomposed in sub-plans consisting of the same components as the original plan, namely: preferences, intentions, conditions, effects, and the plan body itself. Sub-plans can be activated during their parent's activated or suspended state and can last during their parent's completed or aborted state.

Cyclical: means that a plan can be repeated several times. It is the most difficult plan, because the duration and end time vary over a long time, and additionally the number of applications of their single sub-plans is not known."

Time Annotations. One of the main features of Asbru is to represent temporal patterns. The time annotation part is of particular importance to represent temporal relationships. Figure 4.2 shows a time interval represented in Asbru.

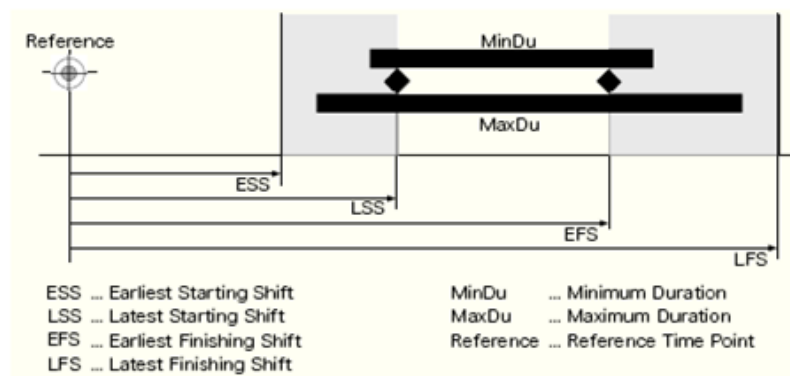


Figure 4.2: Representation of time intervals in Asbru [Kaiser, 2005]

4.1.2 Syntax and Semantics of Asbru

The syntax of Asbru is graphically presented using syntax diagrams in "railroad style" to illustrate the relations between the elements. The syntax description is represented using the XML-format which is described as a Document Type Description (DTD) [Seyfang *et al.*, 2002].

The root element of a plan library in Asbru is 'plan-library' containing a set of domain specifications, value definitions, plan groups, which again contain individual plans. Every single plan specifies a set of actions to be taken to reach a certain goal. These actions are specified in the *plan body*, while the goal is given by the *intentions* and *effects* elements. Furthermore, each plan consists of a set of *conditions* controlling its execution [Seyfang *et al.*, 2002].

We already described the main components of Asbru including the *preferences*, *intentions*, *conditions*, *effects*, *plan body*, and *time annotations* in the previous section in detail. For our purpose, the preference components are of particular importance to embed the evidence information of CPGs into the language definitions

of Asbru, because they are essential for the plan execution. Table 4.1 shows the DTD of the element preferences with its children, whereas the children (resource-constraints, costs) are represented in Table 4.2 and 4.3.

ELEMENT preferences

<i>Child Name</i>	<i>Occurrence</i>		
any-comment	zero or more		
resource-constraint	zero or more		
costs	zero or more		

<i>Attribute Name</i>	<i>Type</i>	<i>Default</i>
strategy	NMTOKEN(String)	optional
responsible-actor	NMTOKEN(String)	optional

Table 4.1: Preferences and their children [Seyfang *et al.*, 2002].

ELEMENT resource-constraint

<i>Child Name</i>	<i>Occurrence</i>		
any-comment	zero or more		
time-annotation	once		

<i>Attribute Name</i>	<i>Type</i>	<i>Default</i>
name	NMTOKEN(String)	required
type	prohibited recommended discouraged obligatory	required

Table 4.2: The child element resource-constraint of preferences [Seyfang *et al.*, 2002].

4.1.3 Application

In the following we will introduce two applications that are based on Asbru. The first, Asbru-View, represents a solution to visualize Asbru based documents, whereas DELT/A try to narrow the gap between the original text and its formal representation.

ELEMENT costs		
<i>Child Name</i>	<i>Occurrence</i>	
any-comment	zero or more	
once		
expression	once	
list-or-set-ref	once	
<i>Attribute Name</i>	<i>Type</i>	<i>Default</i>
label	NMTOKEN(String)	optional
name	NMTOKEN(String)	optional

Table 4.3: The child element costs of preferences [Seyfang *et al.*, 2002].

Asbru-View

A plan representation language like Asbru is very complex and therefore not easy to use by medical domain experts. Therefore, the Asgaard project focused on the visualization of plans and data during the design and execution of CPGs and protocols. For that purpose, they developed visualization and user interface applications for Asbru, called AsbruView. Figure 4.3 consists two main views, which are essential to enhance the understandability and usability of Asbru by the medical staff [Kosara and Miksch, 2001; Miksch *et al.*, 1998; Kosara *et al.*, 1998; Seyfang and Miksch, 2002]:

- Topological view: utilizes the metaphor graphics of "running tracks" and "traffic" and displays the relationship between plans without a time scale.
- Temporal view: concentrates on the temporal dimensions of plans and conditions using glyphs to make the underlying concepts traceable.

In particular, AsbruView supports physicians during treatment planning and execution by providing the following benefits [Kosara *et al.*, 1998]:

- deals with temporal dimensions of plans, conditions, intentions, and effects
- copes with all possible and unpredictable orders of plan execution
- handles exception conditions that might arise
- handles domain-specific features (e.g., plans' intentions)

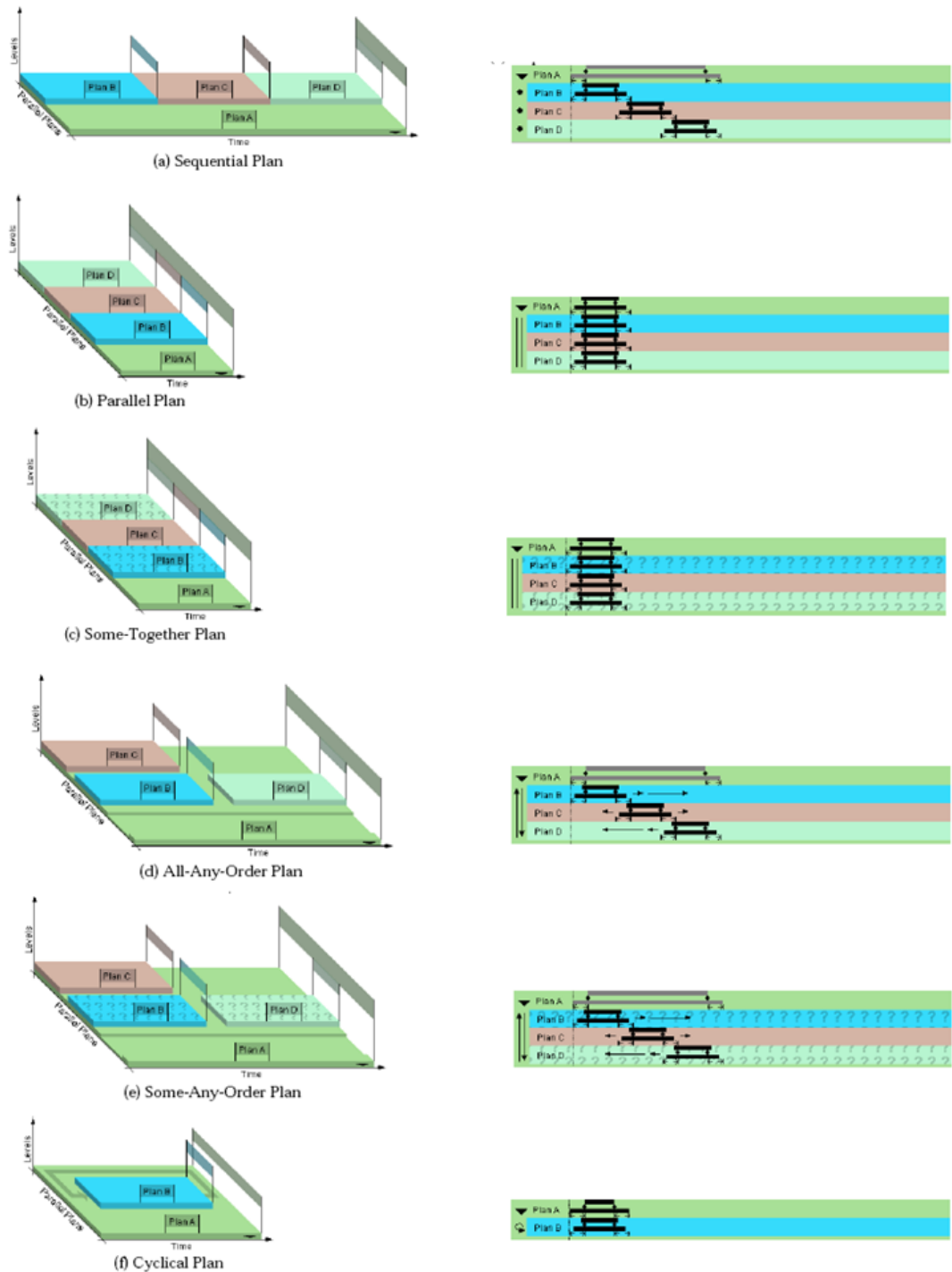


Figure 4.3: AsbruView: The figures on the left side represent the topological view whereas the temporal view is represented with the figures on the right side showing various kinds of plan representations [Kosara and Miksch, 2001].

Document Exploration and Linking Tool / Addons (DELT/A)

The Document Exploration and Linking Tool / Addons (DELT/A) has been developed by the Institute of Software Technology and Interactive Systems at the Vienna University of Technology. The main goal of this tool was to achieve a means to facilitate the translation of free text into a formal XML representation and thus to establish a relationship between the original text and the formal representation. DELT/A consists of the following two main components [Votruba *et al.*, 2004; Votruba *et al.*, 2003]:

- **Links:** provide the possibility of linking between the original CPG and the corresponding Asbru representation. They facilitate the possibility to find out from where the certain value in Asbru notation comes from.
- **Macros:** combine various kinds of Asbru elements to facilitate the creation and extension of Asbru XML files through the usage of common design patterns.

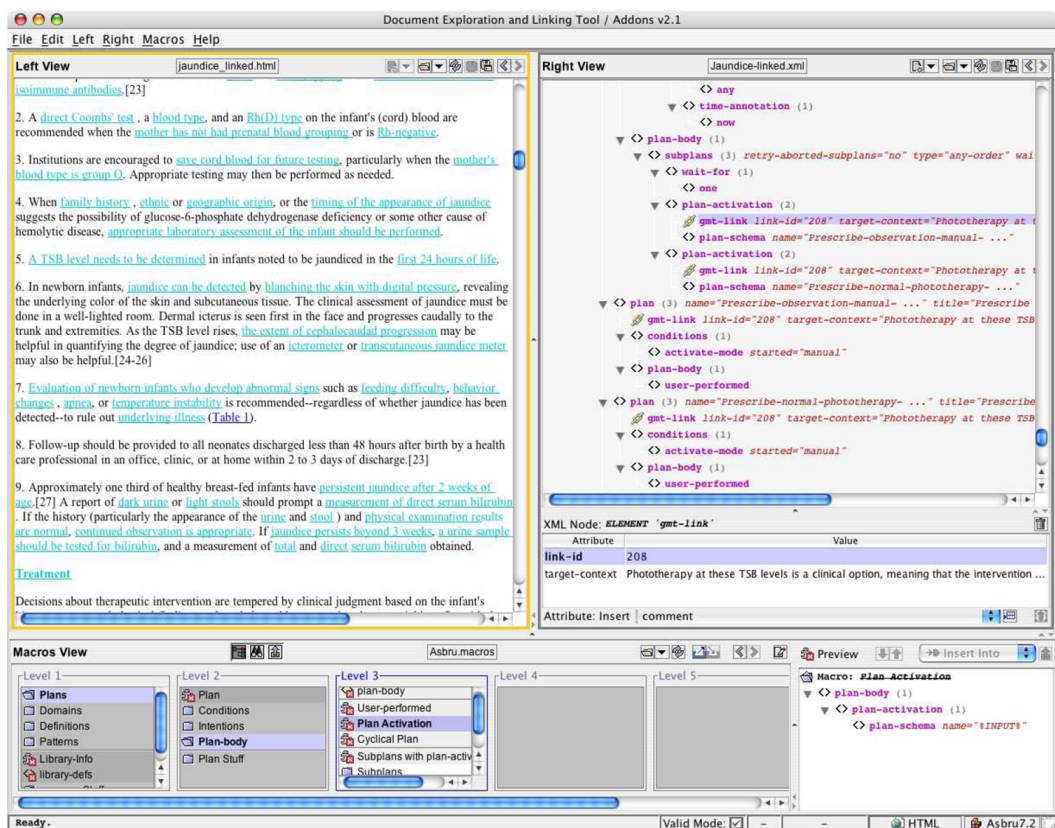


Figure 4.4: The left pane shows a CPG in free text whereas the right pane on the top shows a formal representation of this document. The macros shown in the bottom pane, support the formalization task by templates of various kinds of models [Kaiser, 2005].

Figure 4.4 shows a screenshot of this tool consisting of three main parts: two panes enabling the user to view and edit HTML/XML files and the macros to browse through a macros file at the bottom [Votruba *et al.*, 2004].

Additionally, DELT/A provides the following features [Votruba *et al.*, 2004; Votruba *et al.*, 2003]:

Authoring and augmenting CPGs: by enabling the user to produce a (XML-based) representation using a new CPG in plain text, and to add links to appropriate parts of a CPG to an already existing XML file.

Understanding intermediate representation of CPGs: by facilitating the understandability of the intermediate representation to medical experts. It provides an insight into the translation process of the CPG into the formal representation and enables the user to see where values in the different parts of the formal representation's code come from.

Structuring the intermediate representation: by providing a structured list of Asbru elements, namely macros that are needed to support the authoring of plans.

4.2 PROforma

PROforma is a guideline representation language to support the management of medical procedures as well as decision systems. It provides a basis for a method, a technology and different applications for developing and publishing executable CPGs. It supports the definition of CPGs in terms of a well-defined set of tasks, which can be composed into networks representing plans or procedures to be carried out over time.

4.2.1 Task Properties in PROforma

In PROforma a guideline application is modeled as a set of tasks and data items with the following properties [Bury *et al.*, 2000]:

- **Preconditions:** logical conditions, which must be true when a task is started.
- **Postconditions:** logical conditions, which are assumed to be true after the task has finished.
- **Goal:** logical condition expressing the situation that the task is intended to bring about. A task terminates when its goal is achieved.
- **Description:** describes the task and refers to external information sources if necessary. It justifies and explains the operations defined by a task or a CPG.
- **Trigger condition:** a message that will be passed to a task in order to start it even if its parent plan has not scheduled it to start.

The four basis classes of tasks in PROforma are defined as [Vollebregt *et al.*, 1999; Bury *et al.*, 2000] (see Figure 4.5):

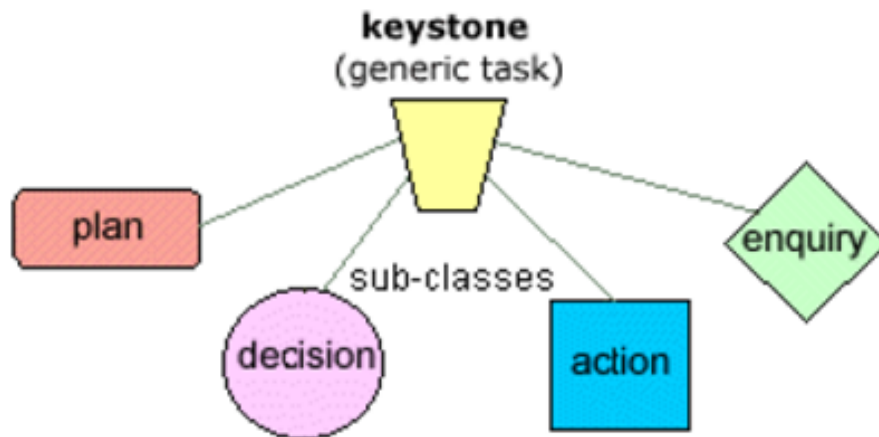


Figure 4.5: The PROforma task ontology [Bury *et al.*, 2000]

Plan: A plan represents a set of tasks to be carried out to achieve a clinical goal. Plans are the basic building blocks of CPGs and may contain any number of tasks of any type, including other plans. With other words, a plan is a sequence of several sub-tasks or components, having an ordering to specify temporal, logical, or source constraints. Additionally, a plan has the following properties [Bury *et al.*, 2000]:

- **Components:** refer to the tasks that the plan contains.
- **Scheduling constraints:** are used to define the order in which the component tasks are executed.
- **Abort conditions:** refer to logical conditions causing the plan to fail if they are true.
- **Termination conditions:** refer to logical conditions that, in case of being true, will result in the termination of the plan whether some of its candidates have not been considered for execution.

Decision: The decision task describes the decision options, relevant information, and a set of argument rules determining the options to be chosen according to current data values. This task is used to choose a candidate from a given set using pro and contra arguments. Therefore, a decision may have values for (see Figure 4.6):

- **Candidates:** these are the options of the decision task. Each candidate is associated with a set of arguments, which are logical conditions with an associated weight. A support value can be estimated for each candidate by adding the weights of true arguments if a decision becomes active. Each candidate can also be associated with a recommendation rule determining if it is considered advisable for that candidate to be chosen.

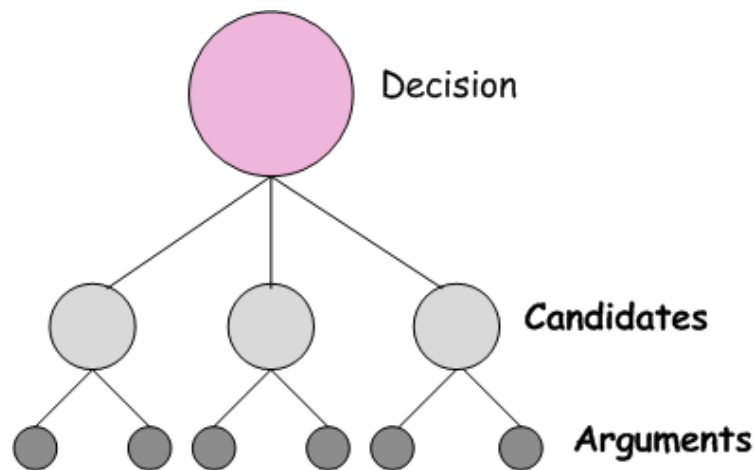


Figure 4.6: Structure of a decision in PROforma (adopted from [PROforma, 2007])

- **Choice mode:** can be single (one candidate) or multiple (many candidates).

Action: This task represents a procedure that has to be executed outside of the computer system.

Enquiry: This task includes a description of the information required and a method to obtain it. Therefore, it demands information needed to execute a certain procedure.

4.2.2 Syntax and Semantics of PROforma

The syntax of PROforma is based on the Backus Naur Form (BNF) consisting expressions defining the forms that PROforma allow logical conditions and mathematical expressions to take. Additionally, the BNF defines how the definitions of tasks and other components represented in CPGs should be arranged and separated [Sutton and Fox, 2003b]. Sutton et al. [2003a] defined the BNF syntax in terms of the following lexical tokens:

Reserved word: any text string appearing in double quotes (e.g., "completed" indicates that the lexical analyzer considers the string *completed* as a reserved word).

Atom: is represented in the BNF notation by the symbol $\langle atom \rangle$ and is described as a text string which consists of one or more underscores or non-whitespace alphanumeric characters (e.g., ali123_PRo). An atom can also be represented as a pair of single quotes enclosing a sequence of zero or more characters containing any character (e.g., 'pRo').

Integer: is defined as an optional minus sign ("-") followed by one or more digits. The BNF symbol for this token is $\langle integer \rangle$.

Float: is defined as an optional minus sign ("-") followed by (1) a sequence of zero or more digits followed by a period (".") followed by one or more digits, or (2) a sequence of one or more digits followed by a period (".") followed by zero or more digits. Floats are represented in the BNF notation with the symbol $\langle float \rangle$.

Double quoted string: is a pair of double quote characters enclosing a sequence of zero or more characters which may contain any character other than an unescaped double quote. They are represented in the BNF notation with the symbol $\langle double_quoted_string \rangle$.

In order to embed the evidence information as proposed in the following chapters into the PROforma syntax we focus on the decision task and its properties (see Figure 4.6), because this task is responsible for the decision-making process during execution. The following descriptions of the properties of the decision task including the definitions of the candidates and arguments are of particular importance for our purpose (see Tables 4.4 to 4.6).

Property Name	Allowed Values	Intended Meaning
<i>candidates</i>	A sequence of Candidate Identifiers	The decision's candidates
<i>sources</i>	A sequence of Source Identifiers	The decision's sources
<i>support_mode</i>	<i>symbolic</i> or <i>numeric</i>	Whether arguments are to be weighed up numerically or symbolically
<i>choice_mode</i>	<i>multiple</i> or <i>single</i>	Whether many candidates may be chosen or only one
<i>result</i>	A sequence of candidate identifiers.	The chosen candidate(s)

Table 4.4: Decision task properties [Sutton and Fox, 2003a]

Property Name	Allowed Values	Intended Meaning
<i>recommendation</i>	Any PROforma expression	Condition that must be true in order for this candidate to be "recommended"
<i>priority</i>	Any integer	Priority of this candidate
<i>arguments</i>	Any sequence of argument identifiers	Arguments associated with this candidate

Table 4.5: Properties of candidates [Sutton and Fox, 2003a]

4.2.3 Application

We already mentioned that PROforma applications have been developed to support the management of medical procedures and clinical decision making at the point

Property Name	Allowed Values	Intended Meaning
<i>value</i>	Any PROforma value	The value that has been assigned to this parameter
<i>expression</i>	Any PROforma expression	An expression that will be evaluated in order to assign a value to this parameter

Table 4.6: Properties of arguments [Sutton and Fox, 2003a]

of care. PROforma software includes a graphical editor to support the authoring process and provides an engine to enact the "proformalised" guideline specification, which has been developed to support the design, testing and execution of CPGs.

Developing an application based on the PROforma language is a two-step process consisting (1) a high level diagram which describes the outline of guideline in terms of the PROforma set of tasks and (2) converts this graphical structure into a database. Additionally, software implementations of the task templates with detailed procedural and medical knowledge is needed to execute CPGs.

In the following we will introduce two applications that are based on PROforma. Both, Arezzo and Tallis are tools for authoring, publishing, and enacting clinical knowledge.

Arezzo

Arezzo is a software developed in the Advanced Computation Laboratory of the Imperial Cancer Research Fund in United Kingdom. The Arezzo application comprises three components (compare Figure 4.7) [InferMed, 2007]:

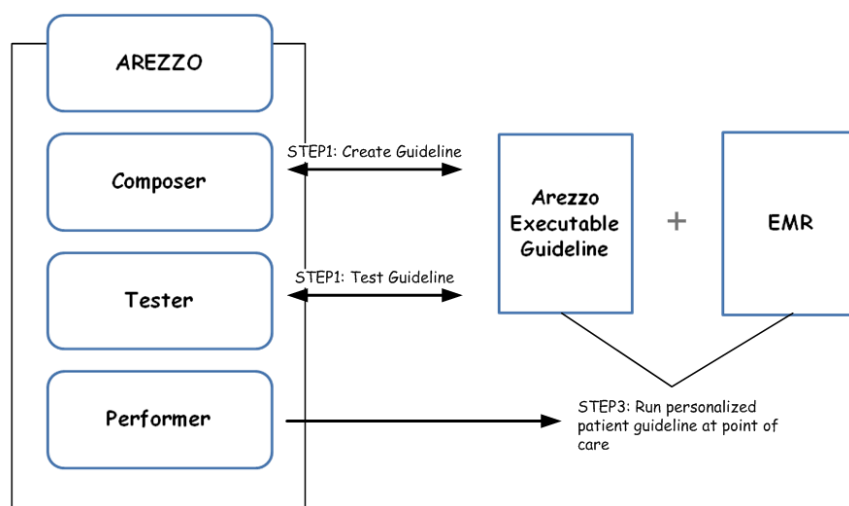


Figure 4.7: Main components of the Arezzo software

1. **Composer:** creates executable CPGs and protocols and ensures the rapid authoring and development of CPGs. It provides a means for specifying CPGs in terms of tasks, data items, and task and data item attributes.

2. **Tester:** tests the CPGs logic before deployment.
3. **Performer:** is a PROforma-compatible inference engine running the Arezzo guidelines and protocols at the point of care.

The Arezzo Composer's user interface consists of three panes. The left pane views a hierarchical tree, the task-authoring tool is shown in the middle pane, and the attributes-authoring tool in the right pane (see Figure 4.8).

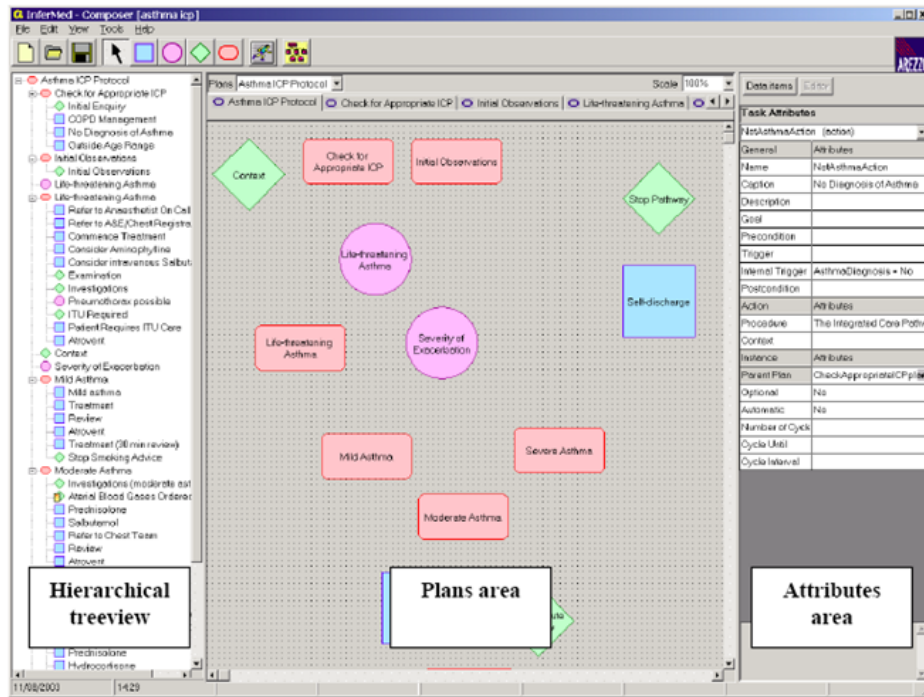


Figure 4.8: Arezzo Composer tool. The left pane shows the hierarchical tree. The right pane shows the attribute area and the middle pane shows the plans area [InferMed, 2007].

Among others, Arezzo provides the following major features [InferMed, 2007]:

- Supporting human experts during the decision-making process by (1) providing options for each decision, (2) producing clearly defined arguments for and against all options, inclusively those options which are not recommended, (3) personalizing the arguments appropriate to the specific circumstances in which a decision is being made.
- Dealing with clinical uncertainty in the decision-making process.
- Facilitating and allowing decisions with the context of executable CPGs and protocols.
- Providing tools to create knowledge bases needed for CPGs and decision-making.
- Minimizing the time spent at the computer by processing the knowledge base and stored guidelines.

- Utilizing a declarative approach so that (1) knowledge and behavior of any CPG is clearly defined and (2) CPGs are flexible and adaptable to the needs of specific users.

Tallis

The Tallis tool was developed at the Advances Computation Laboratory of Cancer Research in UK to provide a possibility for authoring, publishing, and enacting clinical knowledge applications to support the medical staff during the decision-making process. The Tallis application consists of three elements: composer, tester, and engine [Steele and Fox, 2002].

The Tallis Composer is a graphical editor supporting the authoring of PROforma processes. It describes and displays this process in various windows and panes (compare Figure 4.9). On the left side is a hierarchical tree view based on plans, where a plan defines a new level in the hierarchical structure. The top right pane shows a network view represented with a set of PROforma components. The network view is useful to represent the contents of one plan at a time. The bottom pane shows the properties of a specific task as well as a generic task where the user can enter values for the generic attributes that all tasks share [Steele and Fox, 2002]. The Tallis Tester tests the flow of a process-description by enacting it locally. Furthermore, the Tallis Engine also enables the execution of the clinical application over the web [Steele and Fox, 2002].

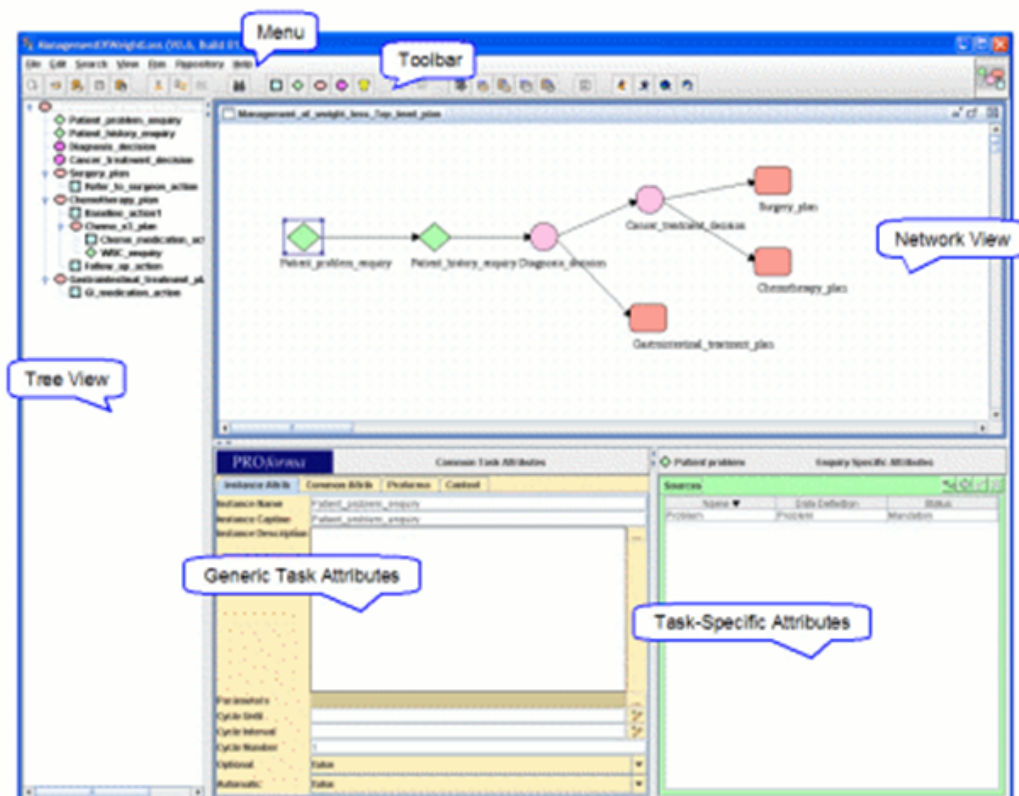


Figure 4.9: Tallis Composer

Chapter 5

Conclusion

In this part of this work we introduced the main components which are of particular importance for our purpose: Evidence-Based Medicine (EBM), Clinical Practice Guidelines (CPGs), evidence information, and guideline representation languages. Now, we will summarize and describe them briefly, to give an overview about the most important facts.

Evidence-Based-Medicine (EBM) means the best medical treatment based on the best available research. EBM uses individual clinical expertise of physicians with the best available external evidence from systematic research [Sackett *et al.*, 1996]. The idea behind EBM is to assign a level of evidence to identify and incorporate such evidence into the patient care recommendations.

Evidence-Based Clinical Practice Guidelines (CPGs) follow a rigorous development process and are based on the best available evidence to support physicians during the decision-making process in specific clinical circumstances. In other words, CPGs are standard means for dissemination of medical knowledge and they are increasingly used to support physicians in decision-making [Peleg *et al.*, 2003].

Evidence information, especially recommendations, described in CPGs are one of the most important information sources to use during decision-making, because they provide physicians various treatment options. They are, in general, based on some kind of evidence, represented by different levels of evidence (LoEs), and on strengths of recommendations (SoRs).

Several guideline representation languages and systems, like Asbru and PROforma, have been developed to facilitate the decision-making process. However, LoEs and SoRs are inadequately treated in guideline representation languages and tools, as they do not support the formalizing and modeling process of recommendations with regard to the LoEs and SoRs sufficiently. In the following part of this work we will describe a methodology that overcomes this problem and embeds the evidence information in CPGs into the decision-making process.

Part II

The EviGuiDe Approach

Chapter 6

Preliminary Work

The overall objective of this thesis is to facilitate the decision-making process by including evidence information into CPGs. For that purpose a semi-formal representation of evidence information is needed to include evidence information in computer-interpretable guideline representation languages (see [Peleg *et al.*, 2003] and [de Clercq *et al.*, 2004] for a comprehensive overview). As several CPGs contain different forms of evidence information and are therefore hard to compare and look through, we decided to develop a meta schema that covers several forms of evidence information from different CPGs and that is representable in a formal way. We called our meta schema EviGuiDe, which is an acronym for "embedding the EVIDence information in clinical practice GUIDelines into the DECision-making process". Since the basis of our meta schema are existing CPGs themselves, we stated several requirements that CPGs have to reconcile in order to be considered for the development of our meta schema.

In this chapter we describe the preprocessing phase, which consists of the guideline selection process (Section 6.1), the description of the evidence information within them (Section 6.2), and the grading systems of the evidence information used by SIGN¹ and GRADE²(Section 6.3).

6.1 Guideline Selection

Guideline features are of particular importance for the selection process, because guidelines have been developed for different goals, different intended users, different applications, and specialties [Kaiser, 2005]. Thus, we formulated several requirements they have to reconcile in order to collect a set of guidelines for our purpose. In the following, these requirements are stated and described, using which we chose a set of CPGs and analyzed them with regard to the LoEs, the SoRs and the major recommendations:

Guideline character. In order to get information about the evidence information we need evidence-based CPGs. They have to follow a rigorous development process and have to be based on the best available evidence.

¹Scottish Intercollegiate Guidelines Network

²The Grades of Recommendation, Assessment, Development, and Evaluation Working Group

Guideline specialty. To obtain a consistent set of guidelines, they should describe treatment and recommendations for similar diseases to support the comparison of the evidence information.

Guideline quality. The relationship between the evidence and the recommendations should be clear and transparent. CPGs should consist of information about the means used to evaluate the evidence, the quality of the scientific and clinical evidence, the outcomes, and the costs. It is not essential that all listed information are explicitly represented in the guidelines (e.g., costs, SoRs, LoEs), though, we need at least information about the study type the guideline is based on.

Guideline developing organization. The evidence information in the guidelines differ from each other depending on the guideline developing organizations, because every organization has its own evidence grading schema. Therefore, we have chosen CPGs from nine different organizations and one co-operation to get a representable design.

Number of guidelines. To gain significant results the number of the guidelines have to be as large as possible.

These requirements provided us with a set of criteria to search for appropriate guidelines from the repository of the National Guideline Clearinghouse (NGC³). We received several guidelines from different clinical specialties (e.g., cancer, otolaryngology). The process of searching and getting appropriate guidelines resulted in a set of 21 CPGs from the medical specialty 'otolaryngology' developed by nine different organizations and one co-operation.

6.2 Analyzing the Evidence Information

In the previous section we described the process of collecting a set of suitable CPGs to build our meta schema for evidence information. After this selection has been done, we had to analyze the evidence information in these guidelines. When discussing the properties of evidence information, we have to look at the features of the LoEs and the SoRs with regard to graded and ungraded evidence information. In our case, evidence information means:

- explicit or implicit representation of Levels of Evidence (LoEs)
- explicit or implicit representation of Strengths of Recommendations (SoRs)
- graded LoEs and SoRs
- ungraded LoEs and SoRs

The results of the analysis of our set of guidelines with respect to these kinds of evidence information can be summarized as follows:

³<http://www.guidelines.gov/>

Levels of Evidence (LoEs). In 20 of 21 guidelines the LoEs are explicitly represented, but use different synonyms to refer to them (see Tables in Section 3.3). Therefore, we introduced our own synonyms for the LoEs (e.g., I_3, II_6) to have a common representation of the study type and the quality of evidence. We will illustrate these synonyms in Section 7.3.2.

Strengths of Recommendations (SoRs). SoRs have to be:

- distinguishable
- significant on their own
- explicitly interpretable
- clear
- unambiguous
- easy to translate into clinical practice

In our guidelines the SoRs are represented with different synonyms, although they are not always explicitly represented. Only five of 21 CPGs include explicitly defined SoRs. In 16 CPGs no information about the SoRs are included.

Graded Evidence Information. Usually, major recommendations represented in the guidelines have an evidence grade with explicit definitions. In cases where only the LoEs are defined, the major recommendations are graded with the LoEs. In cases where the guideline developing organizations have defined SoRs, the major recommendations are supported with the SoRs.

Ungraded Evidence Information. In contrast to graded evidence information the ungraded evidence information is not based on a grading system. In such cases the major recommendations are not graded and provide, usually, only a short information about the study type (see Section 2.3).

Table 6.1 shows the guideline developing organizations and the one co-operation along with the information whether they have a grading schema for the LoEs and SoRs or not. There exist one co-operation of three organizations (AAFP, AAOHS, AAP) and eight different organizations. In addition to the grading system for the LoEs the co-operation (AAFP, AAOHNS, and AAP), ARIA and SIGN have a grading system for the SoRs. Five organizations (AAP, CCHMC, FMSD, ICSI, and UMHS) have only a grading system for the LoEs. One organization (PGI) has no grading system at all. In summary we can say that the CPGs cover eight different LoEs and three different SoRs.

6.3 Basic Grading Systems

After an in-depth analysis of the grading systems of the several guideline developing organizations, we decided to use the grading system of SIGN as a basis for defining

#	Organization and Co-operation	Levels of Evidence	Strengths of Recommendation
1	American Academy of Family Physicians (AAFP), American Academy of Otolaryngology - Head and Neck Surgery (AAOHS) and American Academy of Pediatrics (AAP)	x	x
2	American Academy of Pediatrics (AAP)	x	
3	Allergic Rhinitis and its Impact on Asthma Workshop Group (ARIA)	x	x
4	Cincinnati Children's Hospital Medical Center (CCHMC)	x	
5	Finnish Medical Society Duodecim (FMSD)	x	
6	Institute for Clinical Systems Improvement (ICSI)	x	
7	Practice Guidelines Initiative (PGI)		
8	Scottish Intercollegiate Guidelines Network (SIGN)	x	x
9	University of Michigan Health System - Academic Institution (UMHS)	x	

Table 6.1: Guideline Developing Organizations and their LoEs and/or SoRs

our LoEs and SoRs and the GRADE approach as a basis for defining the trade-off between the benefits and harms. In the following we present these two grading systems, because they include the basic evidence information we used to develop our meta schema.

6.3.1 Grading System of SIGN

The grading system of SIGN relates to the level of the supporting evidence and the predictive power of the study type. The evidence tables of SIGN summarize validated studies identified from the systematic literature review relating to each clinical question. In order to show what course of action should be recommended the evidence tables have to be clear, unambiguous, and transparent. Therefore, SIGN has developed a concept of considered judgment with the following aspects [(SIGN), 2001]:

- Quantity, quality, and consistency of evidence
- Generalizability of study findings
- Directness of application to the target population of the guideline
- Clinical impact
- Implementability

The grading system of SIGN is designed to give greater weight on the quality of the evidence supporting each recommendation. The SoRs provide physicians the likelihood that, in cases where these recommendations are implemented, the predicted outcome will be achieved. For practical purpose, it is intended to give more weight to recommendations supported by good quality observational studies (e.g. cohort studies) where RCTs are not considered for practical or ethical reasons [(SIGN), 2001].

We chose the grading system of SIGN, because of the well structured, systematically reviewed and defined representation of its LoEs and SoRs (see Table 3.9 and 3.10). However, the SIGN approach is insufficient to represent all information about the LoEs, the SoRs, and the trade-off between benefits and harms in our CPGs. Therefore, to design an appropriate meta schema we used grading schemes of the other organizations as well.

6.3.2 Grading System of GRADE

The aim of the GRADE working group is to develop a common representation for grading the LoEs and SoRs in CPGs. Of particular importance is the trade-off between benefits and harms involved in recommendations. They suggest to make explicit judgments about the balance between the main health benefits while considering the costs. The balance between the benefits and harms play a significant role in the decision-making process. Therefore, the GRADE working group recommends the following categorization of trade-offs [Atkins *et al.*, 2004a]:

Net benefits = the intervention clearly does more good than harm.

Trade-offs = there are important trade-offs between the benefits and harms.

Uncertain trade-offs = it is not clear whether the intervention does more good than harm.

No net benefits = the intervention clearly does not do more good than harm.

They also emphasize four main factors, which are important for making recommendations considering the trade-off between benefits and harms [Atkins *et al.*, 2004a]:

- The trade-offs taking into account the estimated size of the effect of the main outcomes, the confidence limits around those estimates, and the relative value placed on each outcome.
- The quality of the evidence.
- 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.
- Uncertainty about baseline risk for the population of interest.

We used the GRADE approach as a basis for defining our meta schema of the trade-off between benefits and harms, because they have a well defined categorization of this trade-off in their grading schema (see [Atkins *et al.*, 2004a]). The adaptation and publication of CPGs that use the GRADE approach will take time, therefore, they are not represented in our chosen set of guidelines.

Taking into account all the described considerations in this chapter we developed our meta schema, which we will explain in the subsequent chapters.

Chapter 7

The EviGuiDe Methodology

Several guideline representation languages and systems have been developed to provide physicians a computer interpretable representation of guidelines to enable automated decision-making support. However, the levels of evidence (LoEs) and the strength of recommendations (SoRs) are inadequately treated in guideline representation languages and tools, as they do not support the formalizing and modeling process of recommendations with regard to the LoEs and SoRs sufficiently. Therefore, a method is required to deal with the multiplicity of existing grading systems and therewith to extend guideline representation languages with the evidence information in computer-supported guidelines.

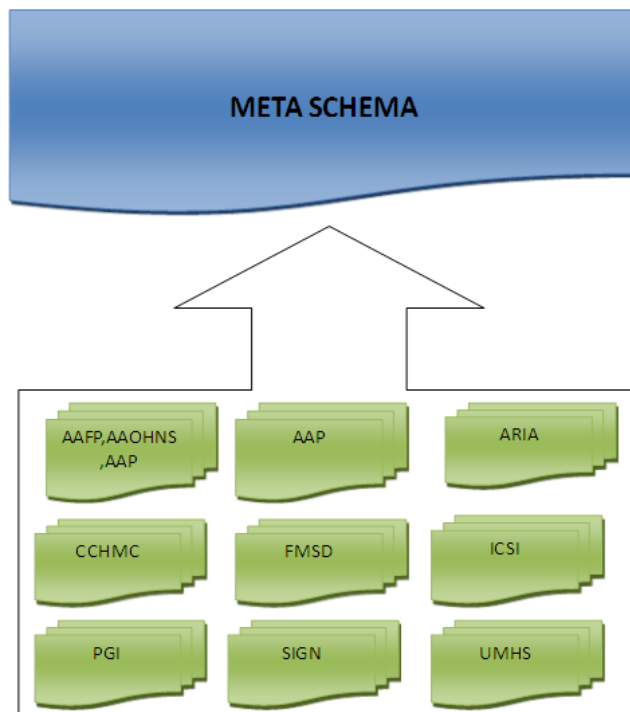


Figure 7.1: An abstract representation of the idea behind the meta schema. The green items symbolize the different guideline development organizations indicating the various existing grading systems. The blue item represents the meta schema connecting these existing grading systems in a unique representation.

Based on these requirements, we decided to develop a meta schema that connects a huge amount of existing grading systems and represents them in a clearly and transparent way (see Figure 7.1). Our objective was not to develop a new grading system, because there are already enough systems with different kinds of definitions and descriptions of evidence.

7.1 The Meta Schema

In this thesis we propose a meta schema that merges existing grading systems to provide a means to increase the transparency and clarity of various grading systems and to classify ungraded LoEs and SoRs. Since we are not experts in the medical domain, our meta schema gives a direction to take if there is no explicit information about the grading of the LoEs and SoRs in CPGs. The direct comparability of various grading systems in our meta schema, makes a quick and clear flow of the underlying information possible. The features of our meta schema can be listed as follows:

- It is a comparable means facilitating the decision-making process by forming a basis to handle the multitude of grading systems on an equal level.
- It consists of detailed definitions of LoEs, SoRs, and the trade-off between benefits and harms.
- It provides a unique representation of different LoEs and SoRs, which can be embedded into existing guideline representation languages.
- it is a representation allowing guideline users and modelers to embed the evidence information in several methods and tools to automatically support the decision-making process.

In the following section we state the main concepts of our approach (Section 7.2) by describing the main components of the meta schema. The subsequent sections describe these components in detail, including the LoEs (Section 7.3.2), the SoRs (Section 7.4), the trade-off between benefits and harms (Section 7.5), and the costs (Section 7.6). The mappings of the various existing grading schemes are discussed in Section 7.7 to present the relations between the existing grading schemes and the meta schema.

7.2 The Way to the Meta Schema

We mentioned before that the overall objective of this work is to facilitate the decision-making process on the basis of a semi-formal representation of the evidence information in CPGs. A semi-formal representation is required to handle evidence information in computer interpretable guideline representation languages (compare [de Clercq *et al.*, 2004; Peleg *et al.*, 2003]). The realization of this objective will be based on the following more specific objectives (see Figure 7.2):

1. Analyzing the evidence information in CPGs

2. Developing a meta schema that covers different existing grading systems
3. Mapping existing grading systems used by different organizations into the meta schema
4. Extending Asbru and PROforma with means (i.e., modeling primitives) to capture evidence information according to the new meta schema.

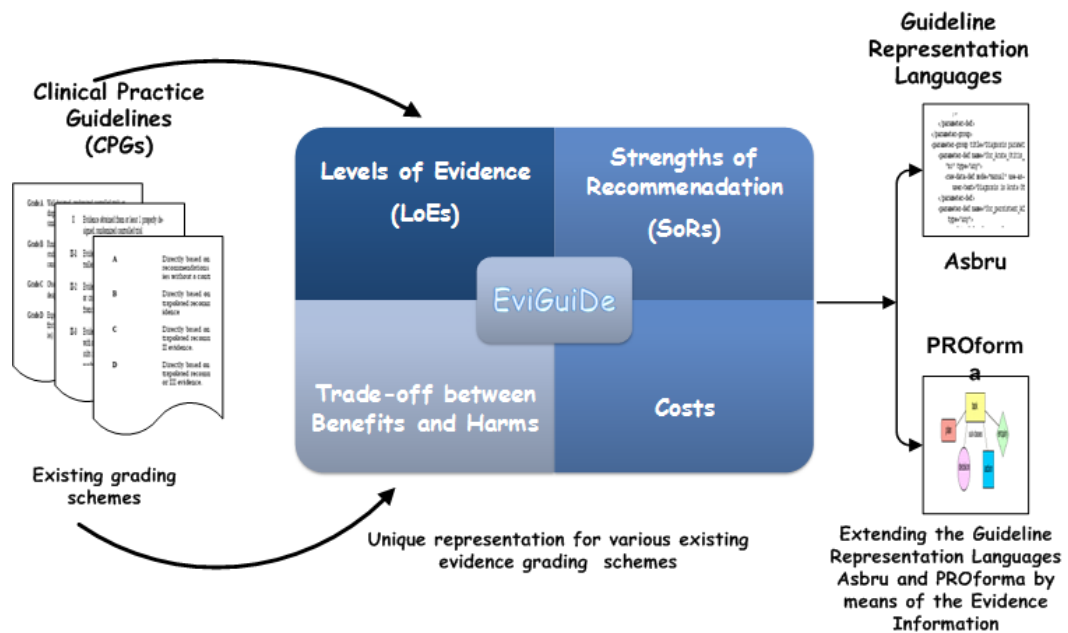


Figure 7.2: Concept of EviGuiDe

To address the first objective, we analyzed 21 CPGs developed by nine different organizations and one co-operation (compare Figure 6.1). Depending on the particular task we manually extracted different kinds of information needed to develop our meta schema and to integrate the evidence information in CPGs into the representation languages Asbru and PROforma. Based on the different grading systems we have developed a meta schema to represent both, graded and ungraded evidence information. This meta schema consists of four components that are of particular importance for our purposes:

1. Levels of Evidence (LoEs)
2. Strengths of Recommendations (SoRs)
3. Trade-off between benefits and harms
4. Costs

These components are appropriate for representing many kinds of evidence information in CPGs as hierarchies to support transparency and clarity. This can be achieved by providing guideline users and modelers with a unique representation of the evidence information to ease the modeling process and with it the decision-making process during execution. In the following we will describe these components in detail to show the development process of the meta schema.

7.3 Levels of Evidence

The right choice of the main representation of LoEs is an important task, because it will affect all forthcoming development phases of the meta schema. By examining existing work we figured that so called evidence tables are used that consist of the representations of LoEs and their definitions. Evidence tables are means to check the consistency of data obtained in various kinds of studies or for subgrouping studies of similar study types, patient population, validity, or quality criteria [Oosterhuis *et al.*, 2004]. To live up to these expectations, evidence tables have to include, among others, the following information [(SIGN), 2001]:

- Study type
- Study size (e.g., number of patients included in the performed studies)
- Prevalence of the condition
- Outcomes measured (e.g., patient-related or laboratory-related outcomes)
- Effects measured, including measures of diagnostic accuracy with the level of uncertainty
- Comments on specific issues raised by the study (e.g., potential bias in the study)
- Classification and quality of levels of evidence in terms of the study

Typically, every guideline developing organization has its own method to identify the levels and the quality of the evidence. For example, SIGN uses search filters, which are pre-tested strategies identifying the higher quality evidence from the vast amount of the literature indexed in major medical databases [(SIGN), 2001].

However, developing evidence tables including the above mentioned information is the job of guideline developing organizations, because they are the domain-experts. This is a great advantage for us, because with these conscientiously developed and evaluated tables we are able to generate and derive a meta schema that covers existing information.

Usually, the process of abstraction or schematization is a fundamental principle of how a mass of information can be made manageable. Starting from this initial point of view, we focused on making the evidence information in CPGs manageable, so that it is clear and traceable. For this reason, we decided to develop a kind of hierarchy that assigns each evidence information represented in our CPGs a grade indicating the quality of the evidence. We think that a hierarchical structure is easy to comprehend and, if clearly defined, the user can separate each hierarchy level from one another. As a basis for our work, we used the grading system of SIGN, because they have a well structured and well defined grading system. We also used other grading systems to cover as much kinds of evidence information in our CPGs as possible.

We will present our meta schema for LoEs and its main attributes in the next subsection.

7.3.1 Concepts

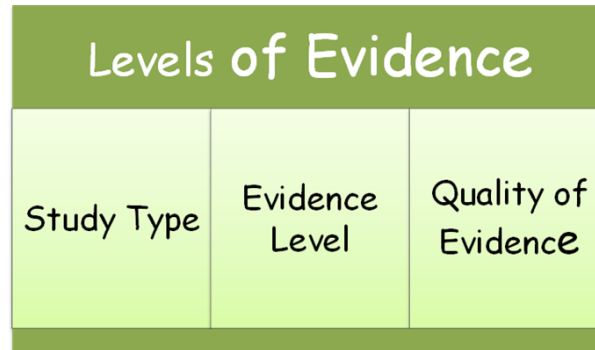


Figure 7.3: Attributes of Levels of Evidence (LoEs) of the meta schema

To get a hierarchical structure of evidence grades we classified the LoEs based on the weights of the study types. Therefore, we use on the definitions and descriptions of different types of study used in evidence-based medicine (compare Section 2.3). In our model the classification of LoEs (see Table 7.1) include the *study type* with various types of studies (e.g., meta-analysis, systematic reviews, randomized controlled trials (RCTs), cohort studies, case-control studies), the *classification system*, and a detailed definition of the *quality of the evidence*. These elements are described in the following (compare Figure 7.3):

Study type

There exist a great number of published studies for each and every single area of medicine. The type of a study is of particular importance to us, because it is going to be used to assign grades to ungraded evidence information. In our meta schema, the LoEs are based on seven different study types that occur in the CPGs we used and indicate the quality of evidence on which the study is based. In order to get a hierarchical structure, we defined the LoEs on the basis of these study types, where *meta-analysis* is on the top and *no study type* at the bottom of the hierarchy. We used Roman numbers to classify these study types as follows:

- I. Meta-Analysis
- II. Systematic Reviews
- III. Randomized Controlled Trials (RCTs)
- IV. Cohort Studies
- V. Case-Control Studies
- VI. Expert Opinion
- VII. No Study Type

Quality of Evidence

In addition to the type of study we also have to consider the quality of evidence on which a particular study type is based. This attribute is essential to present separately the evidence for each outcome measure and assure that the basis of the recommendations is transparent. Hereby, the quality of evidence is based on [(SIGN), 2001]:

- the study type
- the quality of each study
- the judgment on the consistency
- the clinical relevance

In Section 2.3 we discussed different types of studies, the quality of each study, and their importance in evidence-based medicine, to show how they are used in different kinds of situations. The judgment on the consistency is one of the concepts SIGN has introduced to ease the understandability and traceability of evidence information. It should help guideline users to relate how guideline developers were able to arrive at their recommendations, given the evidence they had [(SIGN), 2001].

We used the classification system of SIGN as a basis to define the quality of evidence in our meta schema, because their system comprehends a clear definition of the quality of evidence including the most important study types. In the following we will take a look at the method we used to develop our meta schema including the grading system of SIGN and integrating other grading systems existing in our CPGs.

Classification System to Classify the Evidence Levels

For each type of study, we defined a classification system to establish a relation between the study type and the quality of evidence the study is based on. As we mentioned before, LoEs are mostly explicitly represented in different CPGs but different labels are used to refer to them. Therefore, we introduce our own labels (e.g., I_3, II_6) that consists of a Roman numeral indicating the type of study and a Arabic numeral between one and seven indicating the quality of evidence the study is based on, separated by an underscore (see Figure 7.4).

To enable the smooth integration of evidence information into guideline representation languages, we decided to use a classification system, which gives us information about both the study type and the quality of the study. In this way, our classification system is appropriate for representing various kinds of information regarding the type and quality of a study in a simplified way. In Section 7.8 we describe how we extend the guideline representation languages Asbru and PROforma with this classification system.

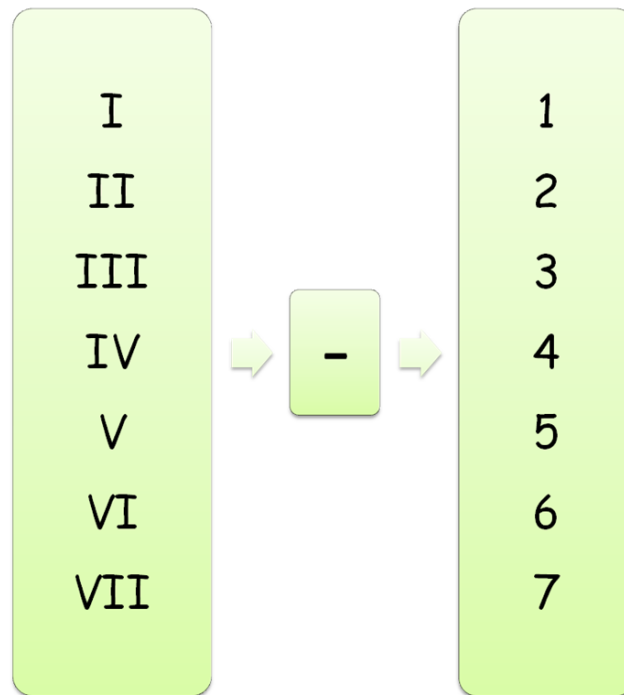


Figure 7.4: Labels of our classification system. The Roman numeral in the left box indicates the study type, whereas the Arabic numeral in the right box indicates the quality of the evidence. They are separated by an underscore showed in the middle box.

7.3.2 Definitions

We already mentioned that the attributes of LoEs are of particular importance, because all forthcoming tasks depend on the clear and transparent definitions of LoEs. Figure 7.5 shows the representation of the main attributes we will use in this section to represent the LoEs of our meta schema. In fact, we propose an incremental process in which each step corresponds to the representation of a specific type of study. We took the following steps in developing this hierarchical structure:

1. analyzing various kinds of study types to sort them according to their quality
2. analyzing various kinds of grading systems and their taxonomies to represent LoEs in our CPGs
3. analyzing various kinds of quality of evidence used by the guideline developing organizations
4. comparing descriptions of quality of evidence used by different guideline developing organizations
5. finding out parallel representations of the quality of evidence in different grading systems
6. developing a hierarchical structure with the study types in terms of their weight

7. grading the quality of evidence and assigning them with regard to the study types
8. developing a classification system including information about both the study type and the quality of evidence

In short, we considered various kinds of evidence information from existing grading systems to develop a meta system that covers the different classifications of the eight developing organizations and the one co-operation. We illustrate this procedure in Figure 7.6 by showing the definition of LoEs of SIGN with the parallels to our definitions. In this case, we see our representation of the study types *meta-analysis*, *systematic reviews*, and *cohort studies*. It is easy to see that not all descriptions of the quality of evidence have a parallel to the definitions of the grading system of SIGN. This is because our meta schema covers also grading systems of other organizations as can be seen in Figure 7.7 where other descriptions of the quality of evidence are used to define our meta representation for the particular study type *systematic reviews*.

As we mentioned before, we do not use only the SIGN approach as a basis for our meta schema. In addition to the grading system of SIGN we also used the other grading systems used in our CPGs. Our aim was to develop a meta schema, which represents a comparable means consisting all data needed to represent existing grading systems in a unique representation. As we are not medical experts and no guideline developers, we picked out the definitions of the quality of the evidence from the grading systems of other guideline developing organizations and included them into our meta schema. We discussed this approach with guideline developers and physicians to evaluate the correctness and usability of our meta schema (for more detail see Chapter 8). After incorporating the results from our evaluation we finalized our hierarchical representation of LoEs as shown in Table 7.1.

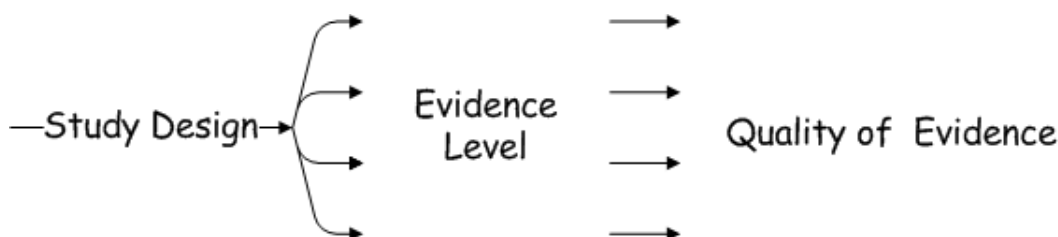


Figure 7.5: Our representation of the main attributes of the LoEs

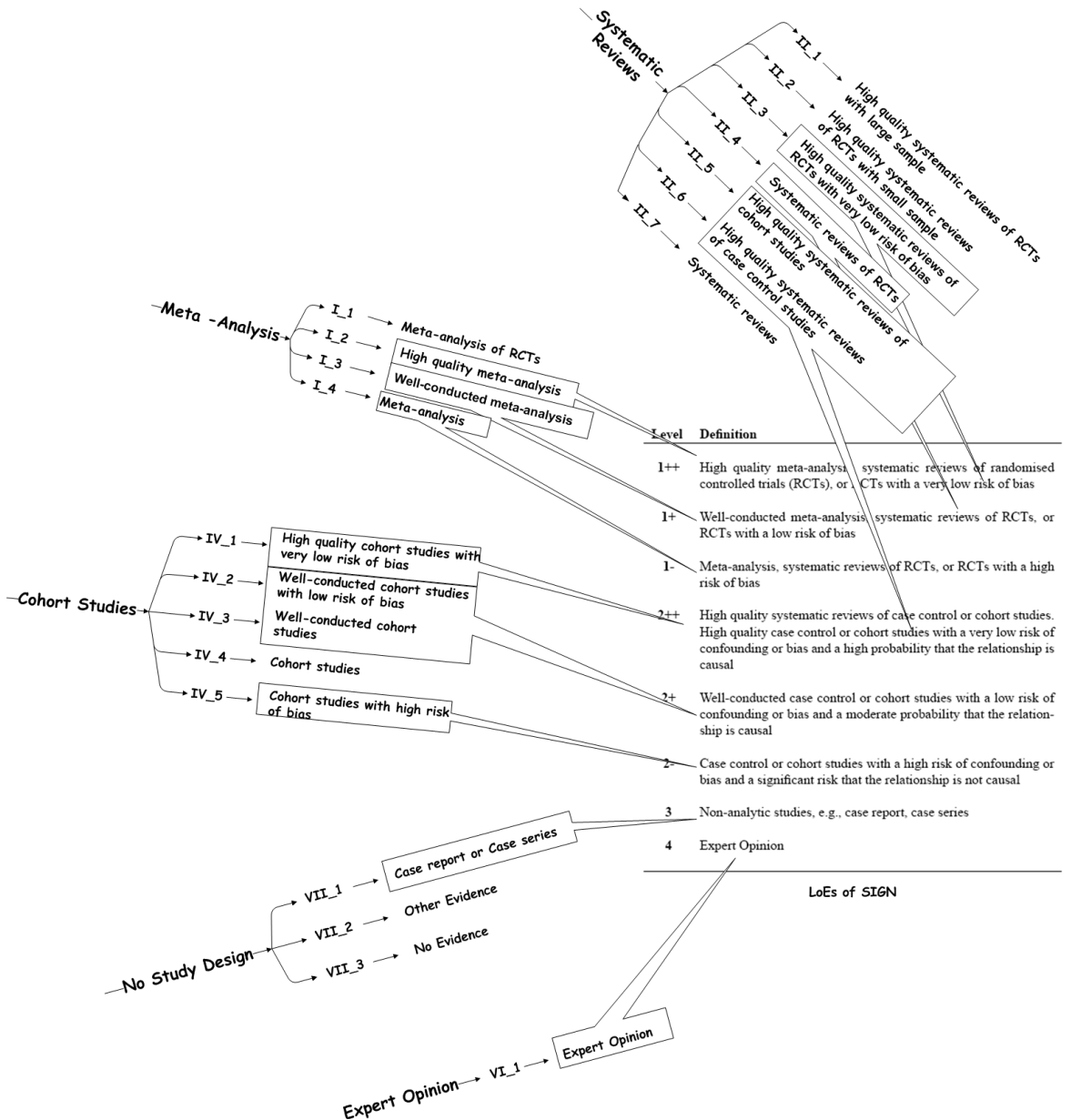


Figure 7.6: Parallels to the grading system of SIGN

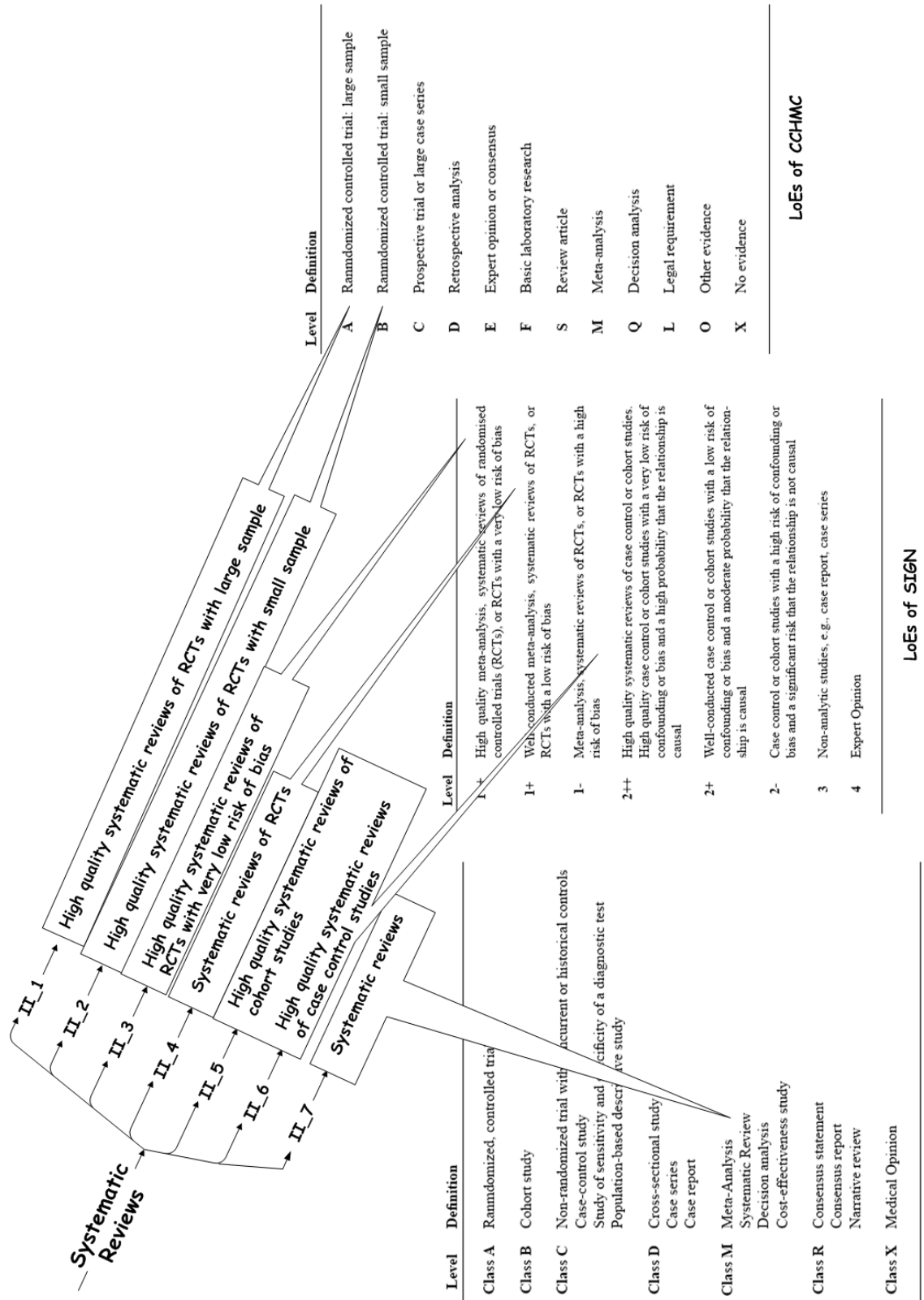


Figure 7.7: Development of the meta level of systematic reviews

Study Design	Evidence Level	Quality of Levels of Evidence
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
Randomized Controlled Trials (RCTs)	III_1	High quality RCTs with very low risk of bias
	III_2	High quality RCTs
	III_3	RCTs with large sample
	III_4	RCTs
	III_5	RCTs with small sample
	III_6	Well-conducted controlled trials without randomization
	III_7	Controlled trials without randomization
Cohort Studies	IV_1	High quality cohort studies with very low risk of bias
	IV_2	Well-conducted cohort studies with low risk of bias
	IV_3	Well-conducted cohort studies

	IV_4	Cohort studies
	IV_5	Cohort studies with high risk of bias
Case Control Studies	V_1	High quality case-control with very low risk of bias
	V_2	Well-conducted case-control with low risk of bias
	V_3	Well-conducted case-control
	V_4	Case-control
	V_5	Case-control with high risk of bias
Expert Opinion	VI_1	Expert Opinion
No Study Design	VII_1	Case reports or case series
	VII_2	Other Evidence
	VII_3	No Evidence

Table 7.1: Levels of Evidence in EviGuiDe

7.4 Strengths of Recommendation

Guideline developing organizations have to formulate their recommendations using the facts in the evidence tables. This is a challenging task, because several skills, methodological knowledge, and experience in decision analysis, is required. SoRs refer to the LoEs incorporating the components used to define the LoEs of a single study (e.g., study type and quality of the study type) and indicate to guideline users the likelihood that, if that recommendation is followed, the predicted outcome will be achieved [Oosterhuis *et al.*, 2004].

Several guidelines contain graded recommendations where the quality of the recommendation is explicitly given, whereas some guidelines do not contain any kind of grading of clinical recommendations. We decided not to neglect the second case by providing means to represent such ungraded recommendations.

7.4.1 Graded Recommendations

In Section 3.3 we stated several existing grading systems to represent the SoRs in our guidelines. A good grading system for SoRs has to offer the following elements to meet the requirements of evidence-based medicine [Oosterhuis *et al.*, 2004; West *et al.*, 2002]:

Quality of evidence. Here "quality" means the aggregate of quality ratings for individual studies, predicated on the extent to which bias was minimized. In

other words, quality of evidence is defined as the extent to which a study's type, conduct, and analysis have minimized selection, measurement, and confounding biases.

Quantity of evidence. Here "quantity" is described as the magnitude of effect, number of studies that have evaluated the given topic, or the overall sample size across all included studies.

Consistency of evidence. Consistency describes the extent to which similar findings are reported using similar or different study types for any given topic.

In addition to these elements the following considerations have to be taken into account when formulating a specific recommendation [Verkerk *et al.*, 2006]:

- Clinical relevance of the scientific evidence (e.g., size of effect, consistency of evidence, relative advantage, generalizability)
- Safety (e.g., harm or side effects)
- Patient perspectives (e.g., patient needs and expectations, therapy compliance, estimated satisfaction)
- Professional perspectives (e.g., professional advantages or risks, time needed for applying the intervention, change of attitudes, routines and habits)
- Availability of facilities (e.g., local, regional or national health services, knowledge and skills of professionals)
- Health care organization (e.g., do insurance companies compensate patients in case of (new) treatments, some recommendations require changes in hospital organizations)
- Cost and cost-effectiveness (e.g., potential cost implications, impact of health care budget)
- Legal aspects (e.g., legal consequences of applying or not applying the recommendations in the guideline)

SoRs are mostly based on how the guideline developers weigh the advantages and disadvantages associated with following the recommendation (e.g., benefits, harms, and costs). The real potential lies here, because if there are great advantages and little disadvantages, guideline developers can tell guideline users with authority that this recommendation can be accepted as a "strong recommendation".

Therefore, it has to be stated that medical decision-making is indeed a complex process requiring the understanding of a plenty of different outcomes, scores of factors, hundreds of relationships, and uncertainty regarding every component. It represents one of the most important factors affecting the cost and quality of health care [Eddv, 1986]. Therefore, recommendations stated in CPGs are one of the most important information sources to use during decision-making, as they provide physicians with various treatment options.

As we have seen, recommendations are graded to differentiate between those based on strong evidence and those based on weak evidence to facilitate an appropriate decision for an individual patient or a population [(SIGN), 2001]. Based on above mentioned factors guideline developing organizations developed their own SoRs to support guideline users during decision-making process. However, in most of the CPGs (16 guidelines out of 21) we considered, the SoRs are not explicitly represented. In the remaining five guidelines, different labels are used to refer to SoRs (compare Section 3.3). Therefore, we developed our own meta schema for SoRs that is compatible with the ones used in those five guidelines.

It was clear that our meta schema for SoRs has to be distinguishable, meaningful, clear, and unambiguous. We thought that more than four hierarchical levels would not satisfy these objectives and we defined the following SoRs:

Strong Recommendation. The quality of the supporting evidence is based on at least one of the following:

- meta-analysis
- systematic review of RCTs
- RCTs with very low risk of bias
- high quality RCTs
- high quality meta-analysis of observational studies
- high quality systematic reviews of observational studies

which are directly applicable to the target population.

Recommendation. The quality of the supporting evidence is based on:

- RCTs
- high quality observational studies with very low risk of bias
- high quality observational studies

which are directly applicable to the target population.

Weak Recommendation. The quality of the supporting evidence is based on :

- well-conducted observational studies with very low risk of bias
- well-conducted observational studies

which are directly applicable to the target population

or

The quality of the supporting evidence is based on:

- observational studies with a high risk of bias (directly applicable to the target population)
- observational studies

- high quality studies, which show little clear advantage to one approach over another
- expert opinion

No Recommendation. The quality of the supporting evidence is based on:

- no study design
- other evidence
- no evidence

7.4.2 Ungraded Recommendations

In addition to graded recommendations, there also exist ungraded recommendations where the guidelines do not contain any classification of the LoEs or SoRs. They usually appear in guidelines as ordinary text fragments. For instance:

”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” [Head and Neck Cancer Disease Site Group *et al.*, 2004]).

This circumstance makes the classification of ungraded evidence information a challenging task. However, such a classification is necessary, because evidence-based recommendations that are classified are better followed in practice than recommendations not based on any scientific evidence [Grol *et al.*, 1998].

In order to classify such ungraded evidence information, we analyzed these recommendations to get information about the above mentioned factors (e.g., study type, quality of study), which are indeed essential for assigning a grade to a particular recommendation. By examining these ungraded recommendations with the aim to get any data giving us information pertaining to the evidence information, we figured that these ungraded recommendations include information about the underlying study types. Consequently, this made our task easier to assign each recommendation a grade, because our meta schema is based on the study types, which are ordered in a hierarchical structure.

Now, with this data, we are able to assign these ungraded recommendations a grade. This classification can be achieved by following the below described steps (compare Figure 7.8):

Ungraded recommendation: analyze the ungraded recommendation in order to find information about the evidence

Evidence information: extract the evidence information including the study type from the ungraded recommendation

Relation to LoEs: establish a relationship between the LoEs of the meta schema and the ungraded recommendation

Study type: find the study type of the ungraded recommendation in the hierarchical structure of our LoEs

Relation to SoRs: establish a relationship between the study type of our LoEs and the SoRs of the meta schema

Classification of ungraded recommendation: assign a grade to the ungraded recommendation using the SoRs described in our meta schema

Figure 7.8 shows that the depicted clinical recommendation is supported by RCTs. This means, that we have information about the classification of the LoEs of RCTs, namely the classification system "III" and the quality of the evidence. But what needs to be considered here, is the recommendation strength. Therefore, the LoEs is only the basis for assigning a grade to the recommendations. We further have to take a look at our definitions of our SoRs. As we can see, the strengths *strong recommendation* and *recommendation* include the study type RCTs with different kinds of quality of evidence (e.g., RCTs with very low risk of bias, RCTs).

For example, we can say that RCTs can be assigned to the area of recommendation strengths *strong recommendation* and *recommendation*. Here, we again have to point out that 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.

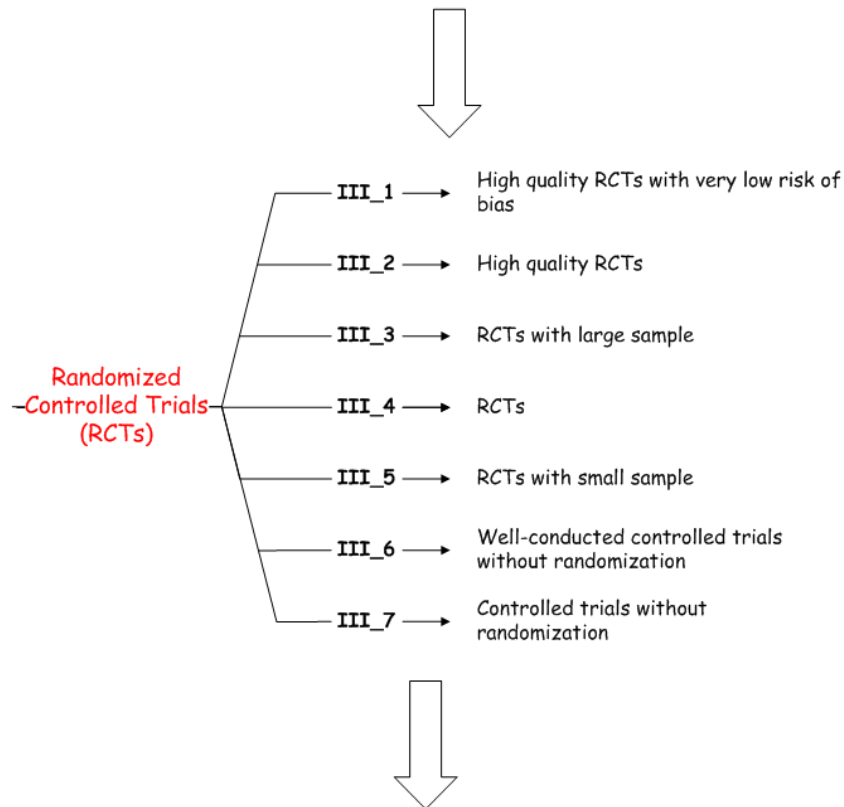
7.5 Trade-off between Benefits and Harms

Often CPGs contain texts about the benefits and possible harms of a particular treatment. The CPGs we used do not contain any explicit representation about the trade-off between the benefits and harms either, but only contain a common description of known benefits and harms. For instance in [(AAFP) *et al.*, 2004] the benefits of implementing the recommendations are described as follows:

"POTENTIAL BENEFITS

- Pneumatic Otoscopy: improved diagnostic accuracy; inexpensive equipment
- Tympanometry: increased diagnostic accuracy beyond pneumatic otoscopy; documentation
- Screening: potentially improved developmental outcomes, which have not been demonstrated in the best current evidence
- Documentation: defines severity, duration has prognostic value, facilitates future communication with other clinicians, supports appropriate timing of intervention, and, if consistently unilateral, may identify a problem with specific ear other than otitis media with effusion (OME) (e.g., retraction pocket or cholesteatoma).

”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”.



<p>Strong Recommendation</p>	<p>The quality of the supporting evidence is basing on at least one:</p> <ul style="list-style-type: none"> • meta-analysis • systematic review of RCTs • RCTs with very low risk of bias • High quality RCTs • high quality meta-analysis of observational studies • high quality systematic reviews of observational studies <p>which are directly applicable to the target population.</p>
<p>Recommendation</p>	<p>The quality of the supporting evidence is basing on:</p> <ul style="list-style-type: none"> • RCTs • high quality observational studies with very low risk of bias • high quality observational studies <p>which are directly applicable to the target population.</p>

Figure 7.8: Assigning a grade to ungraded recommendation

- Child at Risk: optimizing conditions for hearing, speech, and language; enabling children with special needs to reach their potential; avoiding limitations on the benefits of educational interventions because of hearing problems from OME.
- Watchful Waiting: avoid unnecessary interventions, take advantage of favorable natural history, and avoid unnecessary referrals and evaluations
- Medication: avoid side effects and reduce cost by not administering medications; avoid delays in definitive therapy caused by short-term improvement then relapse
- Hearing and Language: to detect hearing loss and language delay and identify strategies or interventions to improve developmental outcomes
- Surveillance: avoiding interventions that do not improve outcomes.
- Referrals: better communication and improved decision-making
- Surgery: improved hearing, reduced prevalence of OME, reduced incidence of acute otitis media, and less need for additional tube insertion (after adenoidectomy)
- Complementary and Alternative Medicine (CAM): not established
- Allergy Management: not established”

The harms of implementing the recommendations are described as follows [(AAFP) *et al.*, 2004]:

”POTENTIAL HARMS

- Pneumatic Otoscopy: cost of training clinicians in pneumatic otoscopy
- Tympanometry: acquisition cost, administrative burden, and recalibration
- Screening: inaccurate diagnosis (false-positive or false-negative), overtreating self-limited disease, parental anxiety, cost of screening, and/or unnecessary treatment
- Documentation: administrative burden
- Child at Risk: cost, time, and specific risks of medications or surgery
- Watchful Waiting: delays in therapy for otitis media with effusion (OME) that will not resolve with observation; prolongation of hearing loss

- Medication: adverse effects of specific medications: side effects of antihistamines and decongestants include insomnia, hyperactivity, drowsiness, behavioral change, and blood-pressure variability; side effects of antimicrobials may include rashes, vomiting, diarrhea, allergic reactions, alteration of the child's nasopharyngeal flora, societal impact of antimicrobial therapy on bacterial resistance and transmission of resistant pathogens, and cost; oral steroids can produce behavioral changes, increased appetite, weight gain, adrenal suppression, fatal varicella infection, and avascular necrosis of the femoral head
- Hearing and Language: parental anxiety, direct and indirect costs of assessment, and/or false-positive results
- Surveillance: allowing structural abnormalities to develop in the tympanic membrane, underestimating the impact of hearing loss on a child, and/or failing to detect significant signs or symptoms that require intervention
- Referrals: confidentiality concerns, administrative burden, and/or increased parent or caregiver anxiety
- Surgery: risks of anesthesia and specific surgical procedures; sequelae of tympanostomy tubes
- Complementary and Alternative Medicine (CAM): potentially significant depending on the intervention
- Allergy Management: adverse effects and cost of medication, physician evaluation, elimination diets, and desensitization.”

This example shows that guideline developing organizations need to be more explicit about the assessment of benefit and harms. We already mentioned that the GRADE working group, whose aim is to develop a common representation for grading LoEs and SoRs in CPGs, suggest to include explicit information about the balance between the main health benefits of a treatment while considering its costs, as they may play a significant role during the decision-making process [Atkins *et al.*, 2004a].

The GRADE working group recommends a sequential judgment about the following points to define the strengths of a recommendation indicating the extent to which one can be confident that adherence to the recommendation will do more good than harm [Atkins *et al.*, 2004a]:

- Quality of evidence across studies for each important outcome
- Which outcomes are essential to a decision
- Overall quality of evidence across these critical outcomes
- The balance between benefits and harms
- SoRs

We used the well defined categorization by the GRADE working group as a basis for defining our meta schema to represent the trade-off between benefits and harms (see [Atkins *et al.*, 2004a]). We want this attribute to be included in guideline representation languages, because this attribute is of particular importance for medical treatment planning even though the trade-off between benefits and harms are not explicitly represented in our CPGs. We think that in the near future this situation will change.

Our representation of the trade-off between benefits and harms has to be clear, traceable, unambiguous, and understandable. Therefore our representation consist of the following four classifications as defined in Table 7.2.

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.

Table 7.2: Trade-off between benefits and harms

As mentioned before, the balance between benefits and harms is of particular importance during decision-making, therefore this attribute has to be included into the grading systems and with it into the decision support systems to facilitate the decision-making process during execution.

7.6 Costs

CPGs have been developed to improve the quality of health care, while reducing avoidable costs of health care. Therefore, in addition to the balance between benefits and harms, the quality of evidence, and the applicability it is also important to consider costs during the judgment about the SoRs. For example, it can happen that the literature is of very high quality but that for instance the drug is very expensive, which leads to the result that it is not recommended to use the drug except in specified cases. In other words, a particular intervention recommended as a strong recommendation without considering its cost analysis may be inappropriate to follow in a situation with limited resources. So what to recommend depends on the study type, study quality, clinical relevance of the effect, and on "other considerations" such as costs, patients perspectives, and local circumstances. This information allows users to prioritize different options for treatment according to their value and

cost-effectiveness [Granata and Hillman, 1998]. Cost-effectiveness means that the enhancement in medical care should have acceptable costs [Thomas, 1999].

We propose this attribute to be included in our model and that, like the trade-off between benefits and harms, explicit information about the cost analysis has to be included into CPGs, as well.

7.7 Mapping Process

In order to establish a relationship between our meta schema and the existing grading systems in our selected set of CPGs, we used the following components (see Figure 7.9):

Guideline Developing Organization. Several classification systems to grade different kinds of recommendations have been developed over the years by different organizations. We propose a methodology that is based on several common grading systems from eight different organizations and one co-operation of three organizations (compare Table 6.1). The data about the guideline developing organization of a grading schema is essential to extract from and differentiate between the different grading systems. For example, the Finnish Medical Society Duodecim (FMSD) uses the letter *A* to grade a level of evidence, whereas the Allergic Rhinitis and its Impact on Asthma Workshop Group (ARIA) uses the letter *A* to classify the strength of a recommendation.

SoRs of the Organization. Having explored our CPGs and the grading systems, we received five different kinds of SoRs. The data about the SoRs is essential to assign each clinical recommendation a grade showing the weight of the recommended treatment option. We need this kind of information to map the existing SoRs to our meta schema.

LoEs of the Organization. As we have seen, not all organizations have their own SoRs to classify the recommendations in their CPGs. In this case the LoEs of these organizations are used to classify the major recommendations.

LoEs of EviGuiDe. This component presents the classification system of our meta schema including information about the study type and the quality of the study.

Scientific Conclusion. This attribute gives information about the proposed evidence area, where a graded and ungraded recommendation will be placed according to our meta schema.

The mapping process depicted in Figure 7.9 represents a fundamental principle to get a comparable means including definitions of existing grading systems. Such kind of information is essential to provide guideline users an overview of the relationship between the existing grading systems and our meta schema. Before describing the mapping process in detail, we will summarize the data we need to start with this process:



Figure 7.9: Main idea behind the mapping process

Various kinds of grading systems: Eight different kinds of LoEs and five different kinds of SoRs including the name of the guideline developing organizations and the definitions of the evidence information.

The meta schema: Our meta schema consisting definitions of the LoEs, SoRs, and trade-off between benefits and harms.

Based on these facts, the mapping process consists the following steps:

1. Search out a guideline developing organization
2. Compare definitions of LoEs of the organization and the LoEs of the meta schema (see Figure 7.10)
3. Compare definitions of SoRs (if available) of the organization and the SoRs of the meta schema (see Figure 7.11)
4. Compare definitions of trade-off between benefits and harms (if available) of the organization and the trade-off between benefits and harms of the meta schema (see Figure 7.12)
5. Extract the terminology of the guideline developing organizations used to classify the LoEs and the SoRs (e.g., A, B,... I, II,...Strong Recommendation, Option,...)
6. Map our terminology of the classification system (e.g., I_1, II_4,...Strong Recommendation, No Recommendation,...Clear Benefit, Benefit,...) with the existing terminology of the guideline developing organizations
7. Develop mapping tables for each guideline developing organization showing the relationship between their grading systems and our meta schema

In the following we will describe this mapping process regarding the grading system of the cooperation of AAFP, AAOHNS, and AAP as a representative example.

The co-operation of AAFP, AAOHNS, and AAP has both, a grading system to grade LoEs and SoRs (compare Section 3.3.1). Therefore, we have to consider the definitions of the LoEs as well as the definitions of the SoRs. The SoRs consist information about the trade-off between benefits and harms as well. This means,

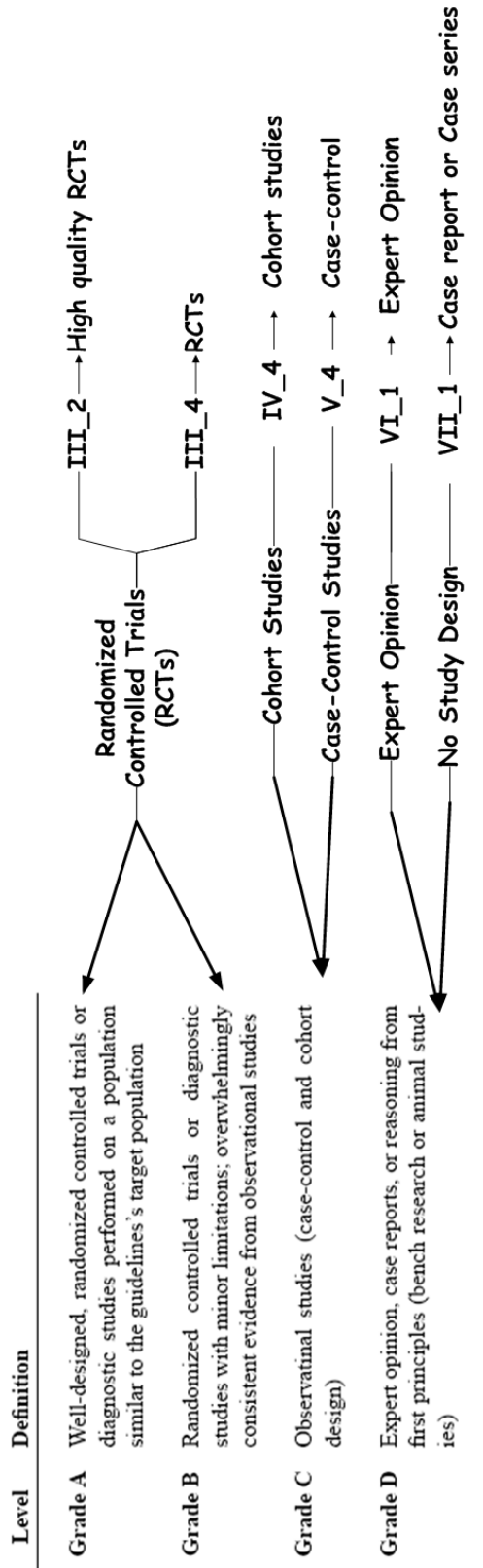


Figure 7.10: Mapping of the LoEs of the co-operation and our definitions of LoEs

Strong Recommendation	<p>The quality of the supporting evidence is basing on at least one:</p> <ul style="list-style-type: none"> • meta-analysis • systematic review of RCTs • RCTs with very low risk of bias • high quality RCTs • high quality meta-analysis of observational studies • high quality systematic reviews of observational studies <p>which are directly applicable to the target population.</p>
Recommendation	<p>The quality of the supporting evidence is basing on :</p> <ul style="list-style-type: none"> • RCTs • high quality observational studies with very low risk of bias • high quality observational studies <p>which are directly applicable to the target population.</p>
Weak Recommendation	<p>The quality of the supporting evidence is basing on :</p> <ul style="list-style-type: none"> • well-conducted observational studies with very low risk of bias • well-conducted observational studies <p>which are directly applicable to the target population</p> <p>or</p> <p>The quality of the supporting evidence is basing on:</p> <ul style="list-style-type: none"> • observational studies with a high risk of bias (directly applicable to the target population) • observational studies • high quality studies, which show little clear advantage to one approach over another • expert opinion
No Recommendation	<p>The quality of the supporting evidence is basing on:</p> <ul style="list-style-type: none"> • no study design • other evidence • no evidence

Figure 7.10

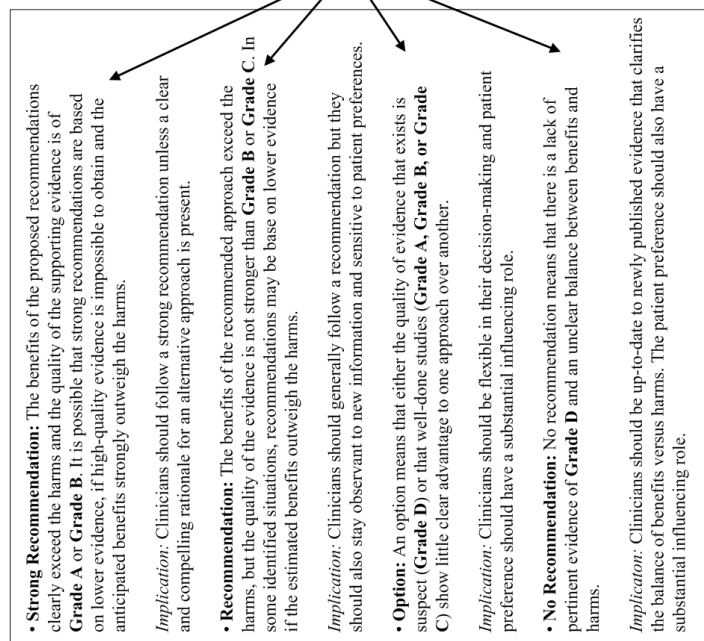


Figure 7.11: Parallels to our SoRs

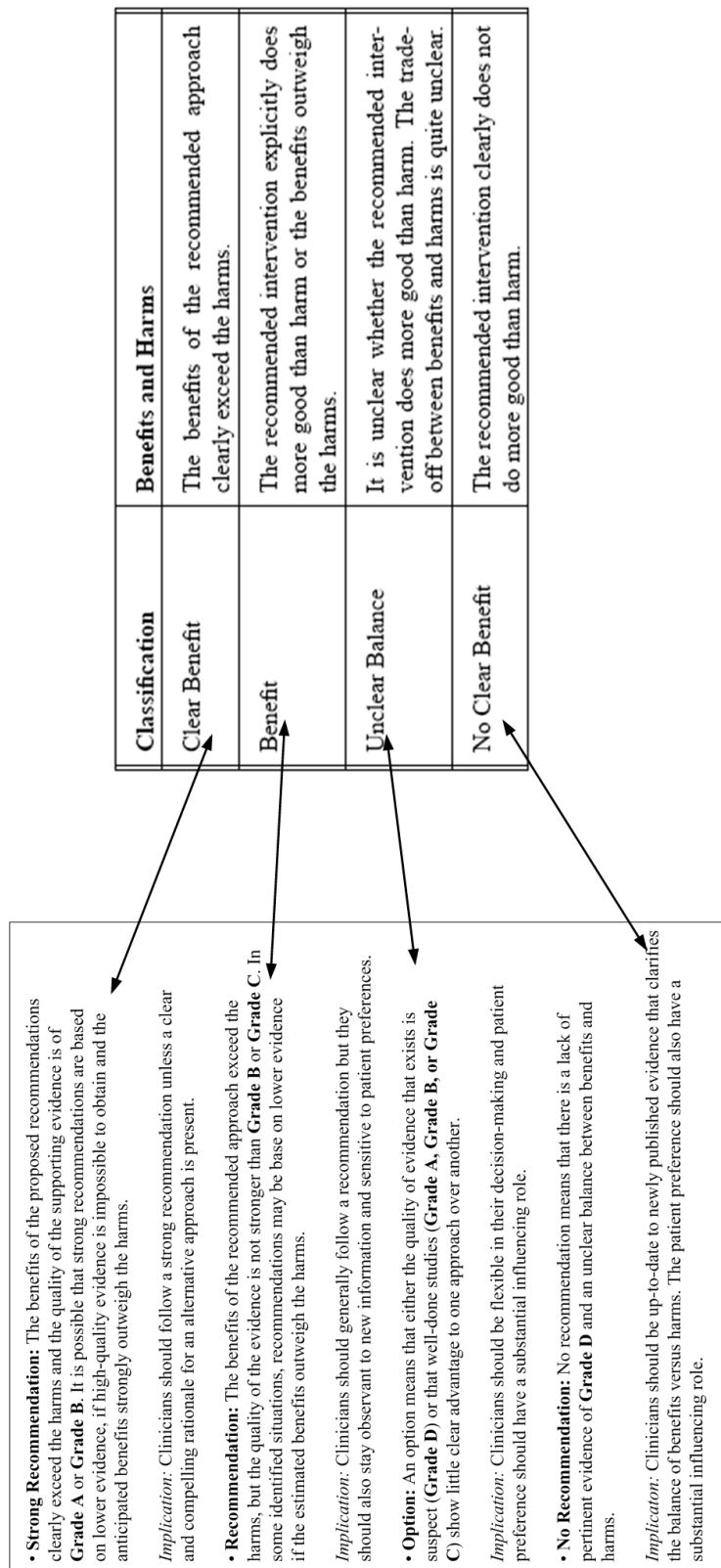


Figure 7.12: Extracting and mapping the trade-off between the benefits and harms represented in the SoRs of the cooperation with our definitions of the balance between benefits and harms

ARIA		EviGuiDe	
Strength of Recommendation	Levels of Evidence	Levels of Evidence	Proposed Recommendation Area
A	Ia	I_1	Strong Recommendation
	Ib	III_4	Recommendation
B	IIa	III_7	Recommendation
	IIb	IV_4	Weak Recommendation
C	III	V_4	Weak Recommendation
D	IV	VI_1	Weak Recommendation

Table 7.3: Mapping table of ARIA

we can include our classification of the trade-of between benefits and harms into this mapping table (compare Table 7.10). An important fact with this regard is that the grading system of this co-operation is the only one, which includes information about the trade-off between benefits and harms in their definitions of SoRs used in our CPGs.

Following the above mentioned steps we have to compare the LoEs used in the guideline and our LoEs of the meta schema. As our schema is based on the study types we extract the study types represented in the grading system of the cooperation. Figure 7.10 shows the results including the study types (RCTs, cohort studies, case-control studies, expert opinion, no study type), their classification of LoEs (III_2, III_4, IV_4, V_4, VI_1, VII_1), and the quality of evidence regarding to the LoEs used by the co-operation.

As mentioned before, SoRs are based on the definitions of LoEs. Therefore, we established the relationship between SoRs and LoEs of the co-operation as well as the relationship between our definitions of LoEs and SoRs (compare Figure 7.11). With this step we received the corresponding SoRs of the meta schema for the SoRs of the co-operation. Furthermore, we extracted the balance between the benefits and harms defined in the SoRs of the co-operation and mapped them to our definitions (see Figure 7.12).

Tables 7.3 to 7.10 show the mappings of the eight organizations and one co-operation.

AAP	EviGuiDe	
Levels of Evidence	Levels of Evidence	Proposed Recommendation Area
I	III_2	Strong Recommendation
II_1	III_6	Recommendation
II_2	IV_3, V_3	Weak Recommendation
II_3	IV_5	Weak Recommendation
III	VI_1	Weak Recommendation
	VII_1	No Recommendation

Table 7.4: Mapping table of AAP

CCHMC	EviGuiDe	
Levels of Evidence	Levels of Evidence	Proposed Recommendation Area
A	III_3	Strong Recommendation
B	III_5	Strong Recommendation
C	VII_1	No Recommendation
D	VII_2	No Recommendation
E	VI_1	Weak Recommendation
F	VII_2	No Recommendation
S	VII_2	No Recommendation
M	I_4	Strong Recommendation
Q	VII_2	No Recommendation
L	VII_2	No Recommendation
O	VII_2	No Recommendation
X	VII_3	No Recommendation

Table 7.5: Mapping table of CCHMC

FMSD	EviGuiDe	
Levels of Evidence	Levels of Evidence	Proposed Recommendation Area
A	I_1, I_2, I_3, I_4	Strong Recommendation
	II_1, II_2, II_3, II_4, II_5, II_6, II_7	
	III_1, III_2, III_3	
B	I_2, I_4	Strong Recommendation
	II_5, II_6	
	IV_1, V_1	Recommendation
C	IV_4, V_4	Weak Recommendation
D	VI_1	Weak Recommendation
	VII_1	No Recommendation

Table 7.6: Mapping table of FMSD

ICSI	EviGuiDe	
Levels of Evidence	Levels of Evidence	Proposed Recommendation Area
Class A	III_4	Recommendation
Class B	IV_4	Weak Recommendation
Class C	V_4	Weak Recommendation
Class D	VII_1	No Recommendation
Class M	I_4, II_7	Strong Recommendation
Class R	VII_2	No Recommendation
Class X	VI_1	Weak Recommendation

Table 7.7: Mapping table of ICSI

SIGN		EviGuiDe	
Strength of Recommendation	Levels of Evidence	Levels of Evidence	Proposed Recommendation Area
A	1++	I_2,II_4, III_1	Strong Recommendation
B			
A	1+	I_3,II_4, III_2	Strong Recommendation
B			
	1-	I_4,II_4	Strong Recommendation
B	2++	II_5, II_6	Strong Recommendation
C		IV_1, V_1	Recommendation
C	2+	IV_2, V_2	Weak Recommendation
	2-	IV_5, V_5	Weak Recommendation
D	3	VII_1	No Recommendation
D	4	VI_1	Weak Recommendation

Table 7.8: Mapping table of SIGN

UMHS	EviGuiDe	
Levels of Evidence	Levels of Evidence	Proposed Recommendation Area
A	III_4	Recommendation
B	III_7	Recommendation
C	IV_4, V_4	Weak Recommendation
D	VI_1	Weak Recommendation

Table 7.9: Mapping table of UMHS

AAFP, AAOHNS and AAP		EviGuiDe		
Strength of Recommendation	Levels of Evidence	Levels of Evidence	Scientific Conclusion	Benefit-Harm
Strong Recommendation	Grade A	III_2	Strong Recommendation	Clear Benefit
Strong Recommendation	Grade B	III_4	Recommendation	Clear Benefit
Recommendation				Benefit
Recommendation	Grade C	IV_4, V_4	Weak Recommendation	Benefit
Option				Unclear Balance
Option	Grade D	VI_1	Weak Recommendation	Unclear Balance
No Recommendation				No Recommendation

Table 7.10: Mapping grading system of the cooperation of AAFP, AAOHNS, and AAP with the meta schema

7.8 Extending Asbru and PROforma

Guideline representation languages with well-structured syntax and semantics, such as Asbru [Shahar *et al.*, 1998a] and PROforma [Fox *et al.*, 1998] are of particular importance, because they have been developed to handle various concepts that care formalization implies. Many guideline modeling tools (e.g., Asbru-View, Tallis) are based on these languages. Our proposed extensions can be embedded into the syntax and semantic of these languages so that such tools can provide a computer-interpretable representation of the evidence information in CPGs. Hence, we decided to use Asbru and PROforma to model the evidence information according to our meta schema.

In the following we will describe how the guideline representation languages Asbru (Section 7.8.1) and PROforma (Section 7.8.2) have to be extended to deal with the evidence information in CPGs a sufficient way.

7.8.1 Extending Asbru

Asbru is a time-oriented, intention-based representation language to represent CPGs and clinical protocols in XML as a set of hierarchical skeletal plans. Each plan consists of a name, a set of arguments, a time annotation, preferences, intentions, conditions, effects, and a plan body (see [Shahar *et al.*, 1998a; Seyfang *et al.*, 2002]). The *preferences* element is of particular importance to us, because it describes the resource constraints, the costs, and the responsible actor. To represent the attributes we proposed in our meta schema, we introduce the *scientific_conclusion* element. Table 7.11 shows the description of the *scientific_conclusion* element and its children, whereas Figure 7.13 illustrates its syntax diagram.

ELEMENT <i>scientific_conclusion</i>		
<i>Child Name</i>	<i>Occurrence</i>	
<i>any_comment</i>	zero or more	
<i>Attribute Name</i>	<i>Type</i>	<i>Default</i>
<i>organization</i>	NMTOKEN(String)	required
<i>levels_of_evidence</i>	NMTOKEN(String)	required
<i>study_type</i>	NMTOKEN(String)	required
<i>strength</i>	NMTOKEN(String)	required
<i>benefit_harm</i>	NMTOKEN(String)	optional
<i>costs</i>	NMTOKEN(String)	optional

Table 7.11: The element *scientific_conclusion* and the definitions of the attributes



Figure 7.13: Element `scientific_conclusion`

We describe this element also in form of a Document Type Definition (DTD) as Asbru is an XML-based language. This DTD describes all the data needed to formally represent evidence information in Asbru.

<!ELEMENT	scientific_conclusion	(#PCDATA)>
<!ATTLIST	scientific_conclusion	
organization	CDATA	#REQUIRED
study_type	(Meta Analysis Systematic Reviews Randomized Controlled Trials Cohort Studies Case Control Studies Expert Opinion No Study Type)	#REQUIRED
strength	(Strong Recommendation Recommendation Weak Recommendation No Recommendation)	#REQUIRED
level_of_evidence	(I_1 I_2 I_3 I_4 II_1 II_2 II_3 II_4 II_5 II_6 II_7 III_1 III_2 III_3 III_4 III_5 III_6 III_7 IV_1 IV_2 IV_3 IV_4 IV_5 V_1 V_2 V_3 V_4 V_5 VI_1 VII_1 VII_2 VII_3)	#REQUIRED
benefit_harm	(Clear Benefit Benefit Unclear Balance No Clear Benefit)	#IMPLIED
costs	CDATA	#IMPLIED
>		

We decided to embed the *scientific_conclusion* element into the definitions of the *preferences* element of Asbru , which is a *child* of the *plan element* containing various information used in the plan selection phase (compare [Seyfang *et al.*, 2002]). Table 7.12 shows the description of the extended *preferences* element and its children, whereas Figure 7.14 illustrates its syntax diagram.

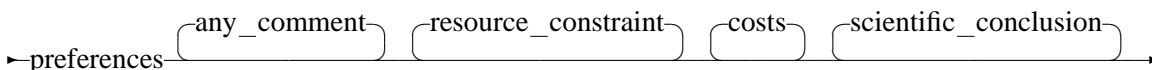


Figure 7.14: Element `preferences`

To illustrate the embedding of the *scientific_conclusion* element into Asbru we state the following example that shows an XML representation of a guideline modeled in Asbru using the above described attributes. This example states that, the execution of the plan requires a *device-A* for 1 to 2 hours and the strategy is considered *conservative*, whereas the *patient* is responsible for taking the action. Further the

ELEMENT preferences		
<i>Child Name</i>	<i>Occurrence</i>	
any-comment	zero or more	
resource-constraint	zero or more	
costs	zero or more	
scientific_conclusion	zero or more	
<i>Attribute Name</i>	<i>Type</i>	<i>Default</i>
strategy	NMTOKEN(String)	optional
responsible-actor	NMTOKEN(String)	optional

Table 7.12: Preferences and their children extended with the element `scientific_conclusion`

guideline is developed by *SIGN*, whereas the medical recommendation is classified with *III_4* referring to the study type *Randomized Controlled Trials* and indicating the strength *Recommendation*. Its *costs* are estimated to be 1000 Euro per month, whereas the patient's discomfort will be at a low level (see Listing 7.1).

Listing 7.1: Example of a XML representation in Asbru

```

<preferences responsible-actor = "patient" strategy="conservative">
  <resource-constraint name="device-A" type="obligatory">
    <time-annotation>
      <time-range>
        <duration>
          <minimum>
            <numerical-constant unit="h" value="1">
          </minimum>
          <maximum>
            <numerical-constant unit="h" value="2">
          </maximum>
        </duration>
      </time-range>
    </self/>
  </time-annotation>
</resource-constraint>
<scientific_conclusion organization="SIGN" level_of_evidence="III_4"
  study_type="Randomized_Controlled_Trials"
  strength="Recommendation">
  <costs name="monetary-costs">
    <numerical-constant unit="Euro" value="1000"/>
  </costs>
  <costs name="discomfort">
    <qualitative-constant value="low"/>
  </costs>
</scientific_conclusion>
</preferences>

```

7.8.2 Extending PROforma

PROforma is a guideline representation language to support the management of medical procedures as well as decision systems. In PROforma a guideline is modeled as a set of tasks and data items, where the main tasks are: *plan*, *decision*, *action*, and *enquiry*. The *decision* task (see Figure 7.15) is of particular importance to us, because it describes several decision options (*candidates*) and a set of rules to choose a candidate using pro and contra *arguments*.

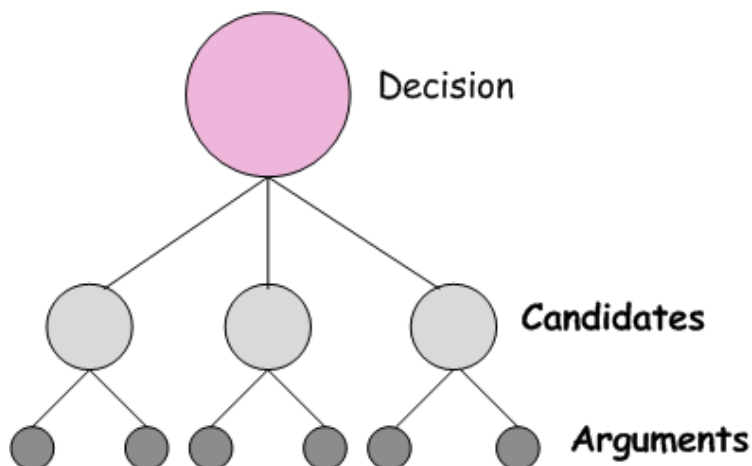


Figure 7.15: Structure of a decision in PROforma showing the relationship between the its candidates and arguments (adopted from [PROforma, 2007])

As we proposed our meta schema to support decision making by providing essential information about the evidence, we include our attributes in form of the *scientific_conclusion* element into the *arguments* of *candidates* in the decision task (see Table 7.13).

Property Name	Allowed Values	Intended Meaning
<i>value</i>	Any PROforma value	The value that has been assigned to this parameter
<i>expression</i>	Any PROforma expression	An expression that will be evaluated in order to assign a value to this parameter.
<i>scientific_conclusion</i>	Any sequence of <i>scientific_conclusion</i> identifiers	<i>Scientific_conclusion</i> associated with this argument.

Table 7.13: Properties of arguments extended with the attribute *scientific_conclusion* [Sutton and Fox, 2003a]

Table 7.14 describes the properties of the *scientific_conclusion* element:

Property Name	Allowed Values	Intended Meaning
<i>organization</i>	Any text string	The value that has been assigned to this parameter (e.g., SIGN, AAP)
<i>study_type</i>	Any text string	The value that has been assigned to this parameter (e.g., Meta-Analysis)
<i>strength</i>	Any text string	The value that has been assigned to this parameter (e.g., Recommendation)
<i>level_of_evidence</i>	Any text string	The value that has been assigned to this parameter (e.g., III_1, IV_1)
<i>benefit_harm</i>	Any text string	The value that has been assigned to this parameter (e.g., Unclear Balance)
<i>cost</i>	A sequence of integers	The value that has been assigned to this parameter

Table 7.14: Properties of *scientific_conclusion* [Sutton and Fox, 2003a]

Because the syntax of PROforma is described using the Backus Naur Form (BNF), we use the same syntax to present our extension that can be embedded into PROforma.

```

<argument>::="argument"::"<scientific_conclusion>","<expression>
<scientific_conclusion>::=<organization><study_type><strength> <level_of_evidence>
<scientific_conclusion>::=[<benefit_harm>]
<scientific_conclusion>::=[<cost>]
<organization>::=<atom>
<study_type>::=<atom>      (must be either "Meta Analysis", "Systematic Re-
                           views", "Randomized Controlled Trials", "Cohort
                           Studies", "Case Control Studies", "Expert Opin-
                           ion", "No Study Type")
<strength>::=<atom>      (must be either "Strong Recommendation", "Rec-
                           ommendation", "Weak Recommendation", "No
                           Recommendation")
<level_of_evidence>::=<atom> (must be either "I_1", "I_2", "I_3", "I_4",
                           "II_1", "II_2", "II_3", "II_4", "II_5", "II_6",
                           "II_7", "III_1", "III_2", "III_3", "III_4",
                           "III_5", "III_6", "III_7", "IV_1", "IV_2",
                           "IV_3", "IV_4", "IV_5", "V_1", "V_2",
                           "V_3", "V_4", "V_5", "VI_1", "VII_1",
                           "VII_2" or "VII_3")
<benefit_harm>::=<atom>   (must be either "Clear Benefit", "Benefit", "Un-
                           clear Balance", "No Clear Benefit")
<benefit_harm>::=<empty>
<cost>::=<atom>
<cost>::=<empty>

```

Our meta schema and thereby the extensions provide a possibility to handle the multitude of grading systems on an equal level and facilitates the flow of the underlying information to be quickly and traceable. We presented two extensions to the guideline representation languages Asbru and PROforma, to enable them to represent evidence information according to our proposed meta schema. We think that these extensions are essential to facilitate the decision-making process in computer-based medical treatment planning.

In the next section we present our evaluation process, which we performed to examine the correctness, feasibility, and understandability of our meta schema and the extensions.

Chapter 8

Evaluation

To evaluate the motivation to our meta schema, the grading schema itself, and the proposed extensions to Asbru and PROforma, we composed a questionnaire of ten questions, which were about the correctness, feasibility, and understandability of:

- the hierarchical structure of our meta schema
- the proposed LoEs and SoRs
- the mappings between existing grading schemes and our schema
- the attribute representing the trade-off between benefits and harms
- and the assignment of our LoEs and SoRs to ungraded evidence information.

We sent this questionnaire to 29 selected professionals consisting of physicians and persons who are familiar with guideline development. Eight of them replied to our call and provided us with different and useful insights regarding the mentioned attributes of our meta schema.

In the following we will first present our questionnaire and then will state selected comments we received from some of the professionals in order to explain the effects they had on the fine tuning of our meta schema.

8.1 Questionnaire

The questionnaire we formulated to perform a qualitative study contained the questions regarding the following points, which could be answered by choosing one of the five available options:

- the availability of needed information for the proposed attributes
- the expected extent of facilitation of the decision-making process when using our meta schema
- the representability of our schema with regard to existing grading schemes

The questionnaire included the following questions adapted from [Atkins *et al.*, 2005]:

1. The hierarchical structure of the levels of evidence is correct?
2. The quality of evidence is correctly defined?
3. The hierarchical structure of the strengths of recommendation is correct?
4. The balance between the benefits and harms is clearly defined?
5. The mapping correctly cover the levels of evidence and strengths of recommendation defined by different guideline developing organizations?
6. The needed information for the decision-making process is available?
7. This meta-schema facilitates the decision-making process during execution?
8. It is possible to model the evidence information by means of the meta schema so that it is representable for the existing grading schemes?
9. The meta-schema can assign the level of evidence to the ungraded evidence information in clinical practice guidelines?
10. The meta-schema can assign the strength of recommendations to the ungraded evidence information in clinical practice guidelines?

These questions could be answered with one of the following five answers:

1. strongly agree
2. agree
3. not sure
4. disagree
5. strongly disagree

8.2 Selecting Professionals

For the purpose of evaluating our methodology and get comments regarding the questionnaire, we sent it to 29 people. Eight of them independently judged the meta schema including the LoEs, SoRs, and trade-off between benefits and harms. They appraised the usability, understandability, and correctness of our approach and provided us their suggestions and comments. Six of the reviewers have had experience in using other grading systems to grade the evidence, classifying the recommendations, and making clinical decisions. Two of the reviewers have had experience in developing guideline representation languages and tools. Most of them have had experience in using and developing CPGs. They were also experienced in using,

modeling and assessing guidelines. Three reviewers had participated in developing guideline modeling tools and guideline assessing tools (e.g., AsbruView [Kosara *et al.*, 1998; Miksch *et al.*, 1998]). For selecting a group of professionals as reviewers for our qualitative study we looked for the following qualities:

- guideline developers
- physicians
- developers of guideline representation languages

8.3 Evaluating the Comments

As we have seen, we prepared questions for each part of our meta schema and sent a previous version of it to the above mentioned professionals. We used their comments and suggestions to enhance our meta schema, which resulted in the meta schema proposed in this thesis. Some of these comments that influenced our development of the meta schema are listed below (compare Figure 8.1):

1. The hierarchical structure of the levels of evidence is correct?

- "I strongly agree with the top level hierarchy when only study design is considered... not sure whether expert opinion should be placed above case series anyway...but it is bit more complicated when extra attributes like bias, sample size and quality are included." [Guideline representation language/system developer and physician]
- "Attributes like quality, biasness, and methodological weakness are difficult to model objectively. Possible objective attributes which could stand as surrogates for quality/bias etc. would be sample size, study power, confidence interval, etc." [Guideline representation language/system developer and physician]
- "...a study is of high quality if there is very low risk of bias...RCT is of a "higher" study design, but will not always be ethical to conduct..."[Guideline developer and physician]

2. The quality of evidence is correctly defined?

- "The recommendation = scientific conclusion + further considerations...so it can be that the literature is of very high quality but that for instance the drug is very expensive than we can recommend not to use the drug...so what the working group recommends depends on study quality (=level of evidence), the clinical relevance of the effect, and "other considerations" like costs, patient perspective, local circumstances,..." [Guideline developer and physician]

- "High quality RCTs with high risk of bias: this does not exist. A high quality RCT = a RCT with low risk of bias...studies with a high risk of bias = very low quality studies and are most of the time deleted (not cooperated in the guideline)..." [Guideline developer and physician]
 - "High quality studies, which show little clear advantage to one approach over another: this is a strong recommendation! You can strongly recommend that both option are equal." [Guideline developer and physician]
3. The hierarchical structure of the strengths of recommendation is correct?
- "...mostly agree...if we intend to make it machine interpretable clear definitions are required for terms like high quality, well conducted, large sample, low risk of bias, etc." [Guideline representation language/system developer and physician]
 - "I can not separate high quality meta-analysis from well-conducted meta-analysis, a high quality RCT from well-conducted RCT from a RCT with a large sample.. I believe these overlap" [Physician, guideline developer, developer of guideline assessing tools]
4. The balance between the benefits and harms is clearly defined?
- "You can make strong recommendation against when the harms strongly outweigh the benefits" [Physician, guideline developer, developer of guideline assessing tools]
5. The mapping correctly cover the levels of evidence and strengths of recommendation defined by different guideline developing organizations?
- "...why no recommendation? I would say weak, or even if more than one study maybe "recommendation" for D..." [Guideline developer and physician]
 - "Review article and expert opinion is the same kind of strength: I would say both weak." [Guideline developer and physician]
 - "...we normally do no make strong recommendation if we only have one RCT..." [Guideline developer and physician]
 - "The AAP explicitly mentions "at least one..." according to my opinion at least two...trials with consistent direction of effect and effect sizes are necessary in order to formulate a very strong recommendation" [Physician and epidemiologist]
 - "The main point is that just one good RCT ought not to be linked to a strong recommendation. At least two good RCTs with consistent effect size are needed for a strong recommendation. One of basic rules of good science is reproducibility of results.. in cases there is just one good RCT I would opt for a weak recommendation" [Physician and epidemiologist]

6. The needed information for the decision-making process is available?
 - "If it is about applying guideline recommendation to an individual decision-making (say for a patient) then some additional factors may be needed apart from LoEs and SoRs...the outcomes considered by guideline makers for particular recommendation (e.g., mortality, adverse events, costs, etc.)...for example, a particular intervention recommended as a strong recommendation without considering its cost analysis may be inappropriate to follow in a situation with limited resources." [Guideline representation language/system developer and physician]
 - "Decision making in the guideline development group is not only grading...it is evidence + the clinical expertise of the guideline working group..." [Guideline developer and physician]
 - "Developers need to be explicit about the benefit-harm assessment and often they are not." [Physician, guideline developer, developer of guideline assessing tools]

7. This meta-schema facilitates the decision-making process during execution?
 - " Usually a recommendation is based on number of different studies (some looking at efficiency some looking at adverse reactions, costs,...). Each study have its own levels of evidence...some explicit schema is required to allocate overall strength of recommendation" [Guideline representation language/system developer and physician]
 - "...certainly explicit statements about evidence quality and recommendation strength facilitate decision making" [Physician, guideline developer, developer of guideline assessing tools]

8. It is possible to model the evidence information by means of the meta schema so that it is representable for the existing grading schemes?
 - "The main challenges are the subjective nature of many of the concepts and incomplete (or implicit) information provided in guidelines" [Guideline representation language/system developer and physician]

9. The meta-schema can assign the level of evidence to the ungraded evidence information in clinical practice guidelines?
 - "Grading on the evidence consists of two steps: (1) what is the study design (RCT, case control,...), (2) how is the study performed (high or low)..." [Guideline developer and physician]

10. The meta-schema can assign the strength of recommendations to the ungraded evidence information in clinical practice guidelines?
 - "The main hurdle I guess will be incomplete information..." [Guideline representation language/system developer and physician]

In addition to the answers of the questions, we get general comments that were very useful for our purpose. Some of them are listed below:

- "Informing evidence quality and/or recommendation grade to decision-makers is very crucial in the specific context of medical/clinical guidelines. In a broader or general context, it is equivalent to the degree of belief that one has to assign to the arguments. In other words, when recommendation is based on stronger evidence like large RCTs, our degree of belief in the claim is high" [Guideline representation language/system developer and physician]
- "Evidence strength depends on multiple complex interacting factors (like study design, generalizability, population differences, study quality, directness,...) some of them are inherently subjective in nature. So reducing it to a single number could be over-simplistic and a bit artificial." [Guideline representation language/system developer and physician]
- "...several languages and applications currently does not support these capabilities..." [Guideline representation language/system developer and physician]
- "...does high quality also refer to the consistency in and effect sizes of the relevant outcomes of the included studies?...Underpowered studies are in my view by definition low quality studies, because evidence of absence of an effect could also be absence of evidence of an effect.." [Guideline developer an epidemiologist]
- "...what we want guideline developers to do is to tell us the strength that they assign to each recommendation...their assessment of the importance of adherence to a particular recommendation. A lot of attention has been paid to evidence quality, but it is not what is important to implementers. It serves only as an indicator of the confidence that developers may have in their statement of recommendation strength. Unfortunately, many guidelines still do not include statements of recommendation strength—only evidence quality, but I suspect this is changing." [Physician, guideline developer, developer of guideline assessing tools]
- "The evidence quality enters the picture in quantifying the advantages and disadvantages. One can be much more confident of the effectiveness of an intervention if it is supported by RCTs and meta-analyses than if it is merely supported by observational studies or expert opinion. One can likewise be more confident of risks, harms, and costs when they are supported by high-quality studies. But the evidence quality does not lead directly to a strength of recommendation. The assessment of the balance or imbalance between benefits and risks leads us to recommendation strength. For example, one can have a drug that has been demonstrated to be highly effective for treating heart disease in a large number of randomized controlled trials. However, there may be evidence of serious risks that are associated with taking the drug—also documented in large numbers of trials. Because effectiveness and risk

are balanced, one should not assign a strong recommendation, even though there is high quality evidence.” [Physician, guideline developer, developer of guideline assessing tools]

- ”The strength of recommendation does not only depend on the levels of evidence but also on the results. One may need to combine several different outcomes in one recommendation.” [Guideline developer and epidemiologist]
- ”...the strength of recommendation is a function of many different items...” [Guideline developer and epidemiologist]

Most of the respondents agreed or strongly agreed with the correctness, understandability, and traceability of:

- the hierarchical structure
- the quality of the LoEs and the SoRs
- the trade-off between benefits and harms
- the mapping tables for our schema
- the extensions of Asbru and PROforma

Some of them disagreed with the assignment of the SoRs to the ungraded evidence information. They believed that more information is needed than the LoEs and the SoRs to support a correct decision (compare Figure 8.1).

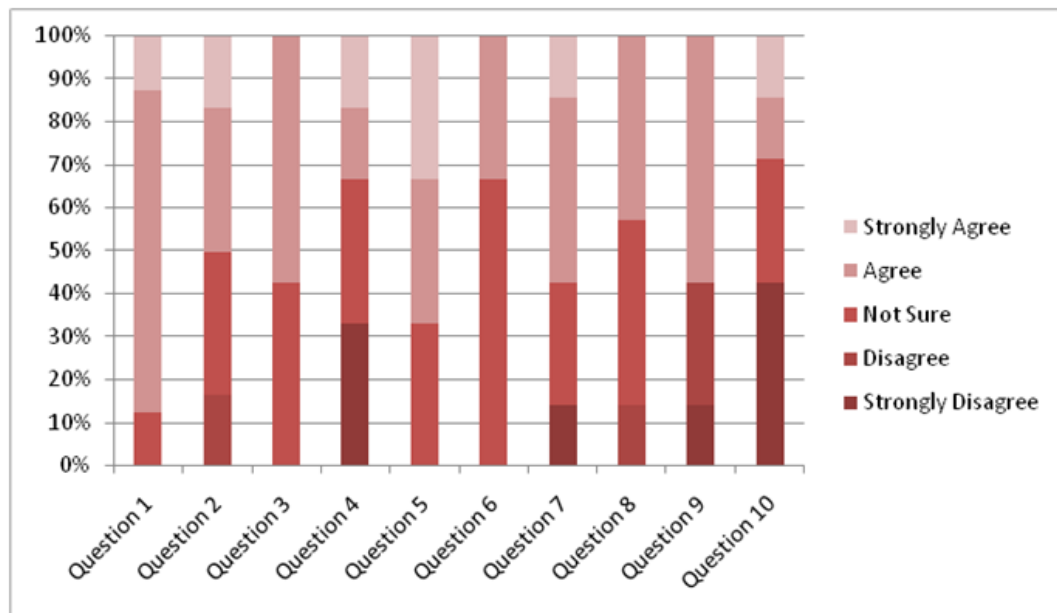


Figure 8.1: Evaluation

The previous version of our meta schema consisted of a large number of LoEs and SoRs including, among others, the following definitions, which we removed from our final meta schema:

- high quality RCTs with high risk of bias
- high quality RCTs with very high risk of bias
- RCTs with high risk of bias
- well-conducted RCTs
- high quality meta-analysis of RCTs with large sample
- high quality meta-analysis of RCTs with small sample

During the evaluation phase, we discussed various uncertainties to reach a consensus on our meta schema and updated our meta schema based on the above mentioned comments and answers to our questionnaire. For example, we replaced the expert opinion in our hierarchical structure above the case series, because we reached the consensus that an expert opinion has to be preferred over case series. Additionally, we updated our mapping tables with the correct assignments of the LoEs and SoRs and incorporated the explicit words "*at least one*" into the definitions of the SoRs of our meta schema to have the correct definition for a *Strong Recommendation* or *Recommendation*. We took all received answers into consideration to update our meta schema and representations, which yield to the enhanced version of our meta schema that we presented in this work.

Chapter 9

Supporting the Decision-Making Process

In this chapter we introduce the influence and benefits of our meta schema during the medical decision-making process. We already mentioned that evidence-based in terms of decision-making means to use the best available evidence in making decision about care. In order to provide the best treatment to a particular patient or population, physicians need various kinds of medical and non-medical information (e.g., patient data, type and stage of disease). For an optimal decision-making physicians have, among others, to [Friedland *et al.*, 1998]:

- identify each possible strategy
- accurately predict the probability of future events
- estimate the balance between the benefits and harms of each possible action.

In particular to facilitate the decision-making process, evidence-based CPGs provide physicians a means including recommendations based on scientific evidence. The main challenges with the existing grading systems are the subjective nature of many of the concepts and incomplete or implicit information provided within guidelines. During the development process of CPGs a lot of attention has been paid to evidence quality, but it is not what is important to guideline users. It serves only as an indicator of the confidence that developers may have in the statements of SoRs. Unfortunately, many CPGs still do not include statements of SoRs, but only LoEs.

An important factor with this regard play also CPGs without any classification systems. In cases where the guideline developing organizations do not use grading systems to represent the evidence information, recommendations in such CPGs are not classified. In either case, the decision-making process becomes more complicated, because evidence information about a particular recommendation provides guideline users the facility to choose between recommendations with high level and recommendations with low level. Certainly, recommendations based on scientific evidence are better followed in practice than recommendations not based on scientific evidence, and means the enhancement of the usability of CPGs.

Usually, physicians have to select and interpret guideline statements to provide an optimal treatment. They have to use more than one guideline from different guideline developing organizations to find the best available treatment for an individual patient or a population. In such cases they have to deal with various kinds of grading systems consisting different classifications and terminologies. However, the grading systems used for the evidence information differ widely, which makes the use hard, troublesome, and demanding. Therefore, guideline developers need to be explicit about the LoEs, SoRs, trade-offs between benefits and harms, and costs, because explicit statements about this components.

Based on these difficulties we can say that the EviGuiDe methodology offers a number of benefits. There are several reasons how our meta schema facilitates the decision-making process during execution. In summary, it:

- provides a unique representation of the evidence information instead of a set of different grading systems used in CPGs
- is an instrument to assign each recommendation a scientific conclusion, which provides guideline users information about the evidence area, where in a hierarchical structure such recommendations can be classified
- contains information about the trade-off between benefits and harms and costs (if available in CPGs)
- handles the multitude of grading systems on an equal level
- provides quick and traceable flow of the underlying information
- enhances the usability of CPGs by providing clear, comprehensive and unambiguous evidence information

In order to provide a computer-interpretable representation of the evidence information this approach facilitates the modeling process by providing the most relevant attributes needed to extend guideline representation languages.

Chapter 10

Summary and Future Work

In this work, we presented a meta schema, which is an instrument to connect different grading systems of evidence information in several CPGs. The meta schema is representable to eight different systems defining LoEs and three different systems defining SoRs and also incorporate the ideas and concepts of the GRADE Working Group to represent the trade-off between benefits and harms. Furthermore, using our meta schema it is possible to assign a LoE and a scientific conclusion to an ungraded evidence recommendation based on the study type and quality of evidence if this information is available. It covers also information about the trade-off between benefits and harms, which are mostly not included in the existing grading schemes. We proposed the used of the attributes LoEs, SoRs, the guideline developing organization, the trade-off between benefits and harms, and costs, because they proved to be significant during the development process. To enhance the traceability of our approach and establish a relation between the existing grading systems and our meta schema we mapped them and summarized them with the mapping tables in Section 7.7.

Furthermore, we presented two extensions to the guideline representation languages Asbru and PROforma, to enable them to represent evidence information according to our proposed meta schema. These extensions provide a possibility to handle the multitude of grading systems on an equal level and facilitate the quick and traceable flow of the underlying information. We think that these extensions are essential to facilitate the decision-making process in computer-based medical treatment planning.

10.1 Summary

In Chapter 1 we stated our research questions to be addressed in the course of our work. This section now summaries what we were able to discover in our effort to get satisfying answers.

- **Which influence has the evidence information in CPGs on the medical decision-making process?**

The evidence information in CPGs cover important factors for the decision-making process, because grading the LoEs and the SoRs enhances the practi-

cability and usefulness of CPGs. Therefore, professional societies and other guideline developing organizations developed different grading systems to classify the major recommendations in CPGs. For practical purposes, precise definitions of recommendations improve the use of CPGs in clinical practice [Grol *et al.*, 1998].

Our meta schema offers a number of benefits to facilitate the medical decision-making process. In particular, it provides a unique representation of the evidence information instead of a set of different grading systems. Our meta schema is a means to assign each recommendation a scientific conclusion, which provides guideline users information about the evidence area, where in a hierarchical structure such recommendations can be classified. We mentioned before that the trade-off between benefits and harms are not represented in CPGs sufficiently. Our meta schema contains information about the trade-off between the benefits and harms and cost if they are available in CPGs (compare Section 7.5). Additionally, our meta schema handles the multitude of grading systems on an equal level, facilitates the flow of the underlying information to be quick and traceable, and enhances the usability of CPGs by providing clear, comprehensive, and unambiguous evidence information.

- **Which components of CPGs are essential to represent the evidence information?**

CPGs are increasingly used to specify how a physician should behave in certain circumstances during the medical treatment and decision-making process [Bosse, 2001]. But making diagnostic or therapeutic decisions requires a sensitive interpretation of patient data of multiple types and the evidence information. Therefore, CPGs represent the best judgment of experienced clinicians and methodologists addressing the scientific evidence for a particular clinical topic [(SCQIM), 2004].

The main components which are of particular importance to represent the evidence information are described in Chapter 3 (e.g., LoEs, SoRs). Additionally, information about different types of studies and the guideline developing organizations are essential to be able to interpret different kinds of grading systems and assign, such different terminologies used to define the LoEs and SoRs, to the appropriate definitions. We selected and separated the different kinds of LoEs and SoRs according to the guideline developing organizations to get an overview about the existing grading systems used in our CPGs (compare Section 3.3).

- **Is it practicable to model both graded and ungraded evidence information using existing guideline representation languages, so that the evidence information influences the medical decision-making process in executing?**

Various guideline representation languages and tools have been developed to provide physicians a computer-interpretable representation of CPGs to enable automated decision-making support. However, evidence information

are inadequately treated in guideline representation languages, as they do not support the formalizing and modeling process of recommendations with regard to LoEs, SoRs, trade-off between benefits and harms, and costs sufficiently. Therefore, we decided to extend these guideline representation languages with the evidence information using our meta schema to facilitate the decision-making process.

- **Is it possible to assign a grade to ungraded evidence information using the meta schema?**

Given the well defined information about the major recommendations including data about the study type, it is indeed possible to assign a grade to ungraded evidence information according to our meta schema (compare Section 7.4.2 showing needed steps).

- **Is it applicable to map various kinds of graded and ungraded evidence information into the newly developed meta schema?**

In Section 7.7 we defined a mapping process to show that it is indeed possible to map existing grading systems into the meta schema consisting information about LoEs, SoRs, and the balance between benefits and harms if it is explicitly defined in the existing grading systems. In Chapter 8 we evaluated the correctness and understandability of this process to prove our method.

- **Which extensions are required to enable guideline representation languages to model the evidence information according to our meta schema?**

We decided to extend the guideline representation languages Asbru and PROforma to present our results. We used all attributes of the meta schema that are essential to provide a computer-interpretable representation of the evidence information and integrate them into the syntax of these languages (compare Section 7.8). We evaluated this extensions regarding their correctness with the aid of developers of guideline representation languages and tools.

10.2 Future Work

A lot of work has to be done in the field of bridging the gap between computer-interpretable and text-based CPGs. For this purpose, this section point out directions for future work which resulted from the research we have done in the course of this thesis.

Enhancing the meta schema

For the near future, our focus is on enhancing our meta schema in order to apply it to the following points:

- guidelines from other domains (e.g., cancer)
- other grading systems (e.g., GRADE)

- other guideline representation languages

Therefore, further refinement of the meta schema and the extensions is necessary. Additionally, some open questions of the meta schema are not solved so far, for instance how to deal with incomplete data and represent them with guideline representation languages.

Developing Information Extraction methods

Automatic extraction of the evidence information in CPGs according to our meta schema represents another important field for further research and will be addressed in the near future. We will develop Information Extraction methods that are able to (semi-)automatically extract evidence information in CPGs according to our proposed meta schema.

Appendix A

Evidence-based Clinical Practice Guidelines

Diagnosis and management of childhood otitis media in primary care. A national clinical guideline.

Scottish Intercollegiate Guideline Network (SIGN); year of publication: 2003

Disease / Condition(s)

- Acute otitis media
- Otitis media with effusion

Guideline Category

Diagnosis
Management
Treatment

Clinical Speciality

Family Practice
Otolaryngology
Pediatrics
Speech-Language Pathology

Intended Users

Advanced Practice Nurses
Nurses
Patients

Physician Assistants
Physicians
Public Health Departments
Social Workers
Speech-Language Pathologists

Guideline Objective(s)

- To provide recommendations based on current evidence for best practice in the management of acute otitis media and otitis media with effusion
- To provide evidence about detection, management, referral and follow-up of children with acute otitis media and otitis media with effusion

Note: This guideline excludes discussion of surgical management such as the insertion of grommets and does not address issues beyond childhood years. In addition, the needs of children with genetic or facial abnormalities are not considered.

Target Population

Children with acute otitis media or otitis media with effusion

Interventiones and Practices Considered

Diagnosis/Evaluation

1. History and clinical assessment, including evaluation of symptoms
2. Examination with otoscope
3. Audiometry
4. Tympanometry

Management/Treatment for Acute Otitis Media

1. Antibiotic treatment, particularly delayed antibiotic treatment

Note: Antibiotics should not routinely be prescribed as the initial treatment.

2. Analgesics, such as paracetamol

Note: Parents should be advised of the potential danger of overuse.

3. Follow up examination
4. Referral to otolaryngologist

Note: The following treatments should not be prescribed for children with acute otitis media: decongestants or antihistamines; oils (for pain).

Note: While homeopathy was considered, due to lack of evidence, no recommendation can be made at this time.

Management/Treatment for Otitis Media with Effusion

1. Autoinflation
2. Follow up evaluation
3. Referral to otolaryngologist

Note: The following treatments should not be used/are not recommended in the management of children with otitis media with effusion: antibiotics; decongestants; antihistamines or mucolytics; topical or systemic steroid therapy.

Note: While homeopathy was considered, due to lack of evidence, no recommendation can be made at this time.

Note: Several interventions intended for parents, teachers and caregivers were also considered, including advice on breastfeeding to reduce the incidence of otitis media; advice on smoking cessation; basic communication techniques; and advice on swimming and bathing following grommet insertion.

Major Outcomes Considered

- Symptom resolution
- Side effects of treatment
- Speech and language, development or behavioural problems

Methodology

Methods Used to Collect/Select the Evidence

Searches of Electronic Databases

Description of Methods Used to Collect/Select the Evidence

A thorough literature search was undertaken in Medline, Embase, and Healthstar to obtain material from 1985 to 1999 inclusive. Internet searches on key Web sites were also conducted and passed on to the group. Additional references were identified by group members and peer reviewers. All material was assessed and evidence synthesized in accordance with the Scottish Intercollegiate Guideline Network (SIGN) methodology.

Number of Source Documents

Not stated

Methods used to Assess the Quality and Strength of the Evidence

Weighting According to a Rating Scheme (Scheme Given)

Rating Schemes for the Strength of the Evidence

Levels of Evidence

1++ - High quality meta-analyses, systematic reviews of randomised controlled trials (RCTs), or RCTs with a very low risk of bias

1+ - Well-conducted meta-analyses, systematic reviews of RCTs, or RCTs with a low risk of bias

1- - Meta-analyses, systematic reviews of RCTs, or RCTs with a high risk of bias

2++ - High quality systematic reviews of case control or cohort studies. High quality case control or cohort studies with a very low risk of confounding or bias and a high probability that the relationship is causal

2+ - Well-conducted case control or cohort studies with a low risk of confounding or bias and a moderate probability that the relationship is causal

2- - Case control or cohort studies with a high risk of confounding or bias and a significant risk that the relationship is not causal

3 - Non-analytic studies, e.g., case reports, case series

4 - Expert opinion

Methods used to Analyse the Evidence

Systematic Review

Description of the Methods used to Analyze the Evidence

The Scottish Intercollegiate Guidelines Network (SIGN) carries out comprehensive systematic reviews of the literature using customized search strategies applied to a number of electronic databases and the Internet. This is often an iterative process whereby the guideline development group will carry out a search for existing guidelines and systematic reviews in the first instance and, after the results of this search have been evaluated, the questions driving the search may be redefined and focused before proceeding to identify lower levels of evidence.

Once papers have been selected as potential sources of evidence, the methodology used in each study is assessed to ensure its validity. SIGN has developed checklists to aid guideline developers to critically evaluate the methodology of different types of study design. The result of this assessment will affect the level of evidence allocated to the paper, which in turn will influence the grade of recommendation it supports.

Methods used to Formulate the Recommendations

Expert Consensus

Description of the Methods used to Formulate the Recommendations

The process for synthesizing the evidence base to form graded guideline recommendations is illustrated in the companion document titled "SIGN 50: A Guideline Developers' Handbook." (Edinburgh [UK]: Scottish Intercollegiate Guidelines Network. [SIGN publication; no. 50]), available from the SIGN Website.

Evidence tables should be compiled, summarizing all the validated studies identified from the systematic literature review relating to each key question. These evidence tables form an important part of the guideline development record and ensure that the basis of the guideline development group's recommendations is transparent.

In order to address how the guideline developer was able to arrive at their recommendations given the evidence they had to base them on, SIGN has introduced the concept of considered judgement.

Under the heading of considered judgement, guideline development groups are expected to summarise their view of the total body of evidence covered by each evidence table. This summary view is expected to cover the following aspects:

- Quantity, quality, and consistency of evidence
- Generalisability of study findings
- Applicability to the target population of the guideline
- Clinical impact (i.e., the extent of the impact on the target patient population, and the resources need to treat them.)

Guideline development groups are provided with a pro forma in which to record the main points from their considered judgement. Once they have considered these issues, the group are asked to summarise their view of the evidence and assign a level of evidence to it, before going on to derive a graded recommendation.

The assignment of a level of evidence should involve all those on a particular guideline development group or subgroup involved with reviewing the evidence in relation to each specific question. The allocation of the associated grade of recommendation should involve participation of all members of the guideline development group. Where the guideline development group is unable to agree a unanimous recommendation, the difference of opinion should be formally recorded and the reason for dissent noted.

The recommendation grading system is intended to place greater weight on the quality of the evidence supporting each recommendation, and to emphasise that the

body of evidence should be considered as a whole, and not rely on a single study to support each recommendation. It is also intended to allow more weight to be given to recommendations supported by good quality observational studies where randomised controlled trials (RCTs) are not available for practical or ethical reasons. Through the considered judgement process guideline developers are also able to downgrade a recommendation where they think the evidence is not generalisable, not directly applicable to the target population, or for other reasons is perceived as being weaker than a simple evaluation of the methodology would suggest.

On occasion, there is an important practical point that the guideline developer may wish to emphasise but for which there is not, nor is their likely to be, any research evidence. This will typically be where some aspect of treatment is regarded as such sound clinical practice that nobody is likely to question it. These are marked in the guideline as "good practice points." It must be emphasized that these are not an alternative to evidence-based recommendations, and should only be used where there is no alternative means of highlighting the issue.

Rating Scheme for the Strength of the Recommendations

The grade of recommendation relates to the strength of the evidence on which the recommendation is based. It does not reflect the clinical importance of the recommendation.

Grade A: At least one meta-analysis, systematic review of randomized controlled trials (RCTs), or randomized controlled trial rated as 1++ and directly applicable to the target population; or

A body of evidence consisting principally of studies rated as 1+, directly applicable to the target population, and demonstrating overall consistency of results

Grade 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+

Grade 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 rate as 2++

Grade D: Evidence level 3 or 4; or

Extrapolated evidence from studies rated as 2+

Good Practice Points: Recommended best practice based on the clinical experience of the guideline development group.

Cost Analysis

A formal cost analysis was not performed and published cost analyses were not reviewed.

Method of the Guideline Validation

External Peer Review

Internal Peer Review

Description of the Method of Guideline Validation

A national open meeting is the main consultative phase of the Scottish Intercollegiate Guidelines Network (SIGN) guideline development, at which the guideline development group presents their draft recommendations for comment. The national open meeting for this guideline was held in November 2001 and was attended by 80 representatives of all the key specialties relevant to the guideline. The draft guideline was also available on the SIGN web site for a limited period at this stage to allow those unable to attend the meeting to contribute to the development of the guideline.

The guideline was reviewed in draft form by a panel of independent expert referees, who were asked to comment primarily on the comprehensiveness and accuracy of interpretation of the evidence base supporting the recommendations in the guideline.

The guideline was then reviewed by an Editorial Group comprising relevant specialty representatives on SIGN Council, to ensure that the peer reviewers' comments had been addressed adequately and that any risk of bias in the guideline development process as a whole had been minimised.

Recommendations

Major Recommendations

Note from the Scottish Intercollegiate Guidelines Network (SIGN) and National Guideline Clearinghouse (NGC): In addition to these evidence-based recommendations, the guideline development group also identifies points of best clinical practice in the original guideline document.

The grades of recommendations (A-D) and levels of evidence (1++, 1+, 1-, 2++, 2+, 2-, 3, 4) are defined at the end of the "Major Recommendations" field.

Clinical Assessment

Diagnosis

B - Healthcare professionals should have an increased awareness of the possibility of the presence of otitis media with effusion in asymptomatic children. The following groups of children are at particular risk:

- Those in day care
- Those with older siblings
- Those with parents who smoke
- Those who present with hearing or behavioural problems

Medical Treatment

Acute Otitis Media

B - Children diagnosed with acute otitis media should not routinely be prescribed antibiotics as the initial treatment.

B - Delayed antibiotic treatment (antibiotic to be collected at parents' discretion after 72 hours if the child has not improved) is an alternative approach which can be applied in general practice.

B - If an antibiotic is to be prescribed, the conventional five day course is recommended at dosage levels indicated in the British National Formulary.

A - Children with acute otitis media should not be prescribed decongestants or antihistamines.

D - Parents should give paracetamol for analgesia but should be advised of the potential danger of overuse.

B - Insertion of oils should not be prescribed for reducing pain in children with acute otitis media.

Otitis Media with Effusion

D - Children with otitis media with effusion should not be treated with antibiotics.

B - Decongestants, antihistamines or mucolytics should not be used in the management of otitis media with effusion.

B - The use of either topical or systemic steroid therapy is not recommended in the management of children with otitis media with effusion.

D - Autoinflation may be of benefit in the management of some children with otitis media with effusion.

Follow up and Referral

Referral

D - Children with frequent episodes (more than four in six months) of acute otitis media, or complications, should be referred to an otolaryngologist.

A - Children under three years of age with persistent bilateral otitis media with effusion and hearing loss of ≥ 25 dB, but no speech and language, development or behavioural problems, can be safely managed with watchful waiting. If watchful waiting is being considered, the child should undergo audiometry to exclude a more serious degree of hearing loss.

B - Children with persistent bilateral otitis media with effusion who are over three years of age or who have speech language, developmental or behavioural problems should be referred to an otolaryngologist.

Patient Issues

Information for Parents, Teachers, and Carers

B - Parents of children with otitis media with effusion should be advised to refrain from smoking.

C - Parents should be advised that breastfeeding may reduce the risk of their child developing otitis media with effusion.

C - Grommet insertion is not a contraindication to swimming.

Definitions

Grades of Recommendations

Grade A: At least one meta-analysis, systematic review of randomized controlled trials (RCTs), or randomized controlled trial rated as 1++ and directly applicable to the target population; or

A body of evidence consisting principally of studies rated as 1+, directly applicable to the target population, and demonstrating overall consistency of results

Grade 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+

Grade 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 rate as 2++

Grade D: Evidence level 3 or 4; or

Extrapolated evidence from studies rated as 2+

Levels of Evidence

1++ - High quality meta-analyses, systematic reviews of randomised controlled trials (RCTs), or RCTs with a very low risk of bias

1+ - Well-conducted meta-analyses, systematic reviews of RCTs, or RCTs with a low risk of bias

1- - Meta-analyses, systematic reviews of RCTs, or RCTs with a high risk of bias

2++ - High quality systematic reviews of case control or cohort studies. High quality case control or cohort studies with a very low risk of confounding or bias and a high probability that the relationship is causal

2+ - Well-conducted case control or cohort studies with a low risk of confounding or bias and a moderate probability that the relationship is causal

2- - Case control or cohort studies with a high risk of confounding or bias and a significant risk that the relationship is not causal

3 - Non-analytic studies, e.g., case reports, case series

4 - Expert opinion

Evidence Supporting the Recommendations

Type of Evidence Supporting the Recommendations

The type of supporting evidence is identified and graded for each recommendation (see "Major Recommendations").

Benefits/Harms of Implementing the Guideline Recommendations

Potential Benefits

Antibiotics in comparison to placebo and observational treatment may have a modest benefit on symptom resolution and failure rates, as variously defined, in children over the age of two years with acute otitis media. The available evidence on natural

history of acute otitis media shows that in studies with close follow up, very few episodes of mastoiditis or other suppurative complications are reported in children with acute otitis media not initially treated with antibiotics.

Potential Harms

Acute Otitis Media

Although non-steroidal anti-inflammatory drugs (NSAIDs) are frequently used by parents, caution should be exercised because of the side effect profile.

List of Figures

2.1	General Procedure of EBM	10
2.2	Pillars of EBM	11
2.3	Representations of Meta-analyses (Blobbograms)	15
2.4	Funnel Plot to Identify Missing Studies	16
2.5	Process of Systematic Reviews	18
3.1	Guideline Development Process	22
4.1	Plan in Asbru	46
4.2	Time Annotation in Asbru	47
4.3	AsbruView: Topological and Temporal View	50
4.4	Document Exploration and Linking Tool / Addons (DELT/A).	51
4.5	Components of PROforma	53
4.6	Decision Task of PROforma	54
4.7	Components of Arezzo	56
4.8	Arezzo Composer Tool	57
4.9	Tallis Composer	58
7.1	The Meta Schema	67
7.2	Concept of EviGuiDe	69
7.3	Attributes of Levels of Evidence (LoEs) of the meta schema	71
7.4	Classification System	73
7.5	Our Representation of the Main Attributes of LoEs	74
7.6	SIGN and our Meta Schema	75
7.7	Developing the Meta Schema: an Example	76
7.8	Ungraded Evidence Information	83
7.9	Mapping Process	88
7.10	Mapping Process: Levels of Evidence	89
7.11	Mapping Process: Strengths of Recommendations	90
7.12	Mapping Process: Trade-Off between Benefits and Harms	91
7.13	Syntaxdiagram of the Element Scientific_Conclusion	98
7.14	Syntaxdiagram of the Element Preferences	98
7.15	Decision Task of PROforma	100
8.1	Results of the Qualitative Study	109

List of Tables

2.1	Hierarchy of the Levels of Evidence	17
2.2	Appraising Systematic Reviews	19
3.1	Organizations and Guidelines	30
3.2	Levels of Evidence (AAFP, AAOHNS, AAP)	32
3.3	Level of Evidence (AAP)	33
3.4	Category of Evidence (ARIA)	34
3.5	Strength of Recommendation (ARIA)	35
3.6	Classes of Evidence (CCHMC)	36
3.7	Evidence-Level (FMSD)	37
3.8	Classes of Research Reports (ICSI)	38
3.9	Levels of Evidence (SIGN)	40
3.10	Grades of Recommendations (SIGN)	41
3.11	Level of Evidence (UMHS)	42
4.1	The Element Preferences	48
4.2	The Element Costs	48
4.3	The Element Resource-Constraint	49
4.4	Properties of Decision	55
4.5	Properties of Candidates	55
4.6	Properties of Arguments	56
6.1	Organizations	64
7.1	Levels of Evidence in EviGuiDe	78
7.2	Trade-off Between Benefits and Harms	86
7.3	Mapping Table of ARIA	92
7.4	Mapping Table of AAP	93
7.5	Mapping Table of CCHMC	93
7.6	Mapping Table of FMSD	94
7.7	Mapping Table of ICSI	94
7.8	Mapping Table of SIGN	95
7.9	Mapping Table of UMHS	95
7.10	Mapping Table of the Co-operation	96
7.11	The Element Scientific_Conclusion	97
7.12	The Element Preferences	99
7.13	Extended Properties of Arguments	100
7.14	Properties of Scientific_conclusion	101

Bibliography

- [(AAFP) *et al.*, 2004] American Academy of Family Physicians (AAFP), American Academy of Otolaryngology Head and Neck Surgery, and American Academy of Pediatrics Subcommittee on Otitis Media with Effusion. Otitis Media with Effusion. *Pediatrics*, 113(5):1412–1429, May 2004.
- [(AAP), 2002] Committee on Infectious Diseases. American Academy of Pediatrics (AAP). Reduction of the Influenza Burden in Children. *Pediatrics*, 110(6):1246–1252, December 2002.
- [(AAP), 2004] American Academy of Pediatrics Committee on Infectious Diseases (AAP). Recommendations for Influenza Immunization of Children. *Pediatrics*, 113(5):1441–1447, May 2004.
- [Atkins *et al.*, 2004a] D. Atkins, D. Best, P. Briss, M. Eccles, Y. Falck-Ytter, S. Flottorp, G. Guyatt, R. Harbour, M. Haugh, D. Henry, S. Hill, R. Jaeschke, G. Leng, A. Liberati *et al.*, and the GRADE Working Group. Grading Quality of Evidence and Strength of Recommendations. *British Medical Journal*, 328(7454):1490–1498, June 2004.
- [Atkins *et al.*, 2004b] D. Atkins, M. Eccles, S. Flottorp, G. H. Guyatt, D. Henry, S. Hill, A. Liberati, D. O’Connell, A. D. Oxman, B. Phillips, H. Schünemann, T. Tan-Torres Edejer, G. Vist, J. W. Williams Jr., and The GRADE Working Group. Systems for Grading the Quality of Evidence and the Strength of Recommendations I: Critical Appraisal of Existing Approaches. *BMC Health Services Research*, 4(38), 2004.
- [Atkins *et al.*, 2005] D. Atkins, P. Briss, M. Eccles, S. Flottorp, G. Guyatt, R. Harbour, S. Hill, R. Jaeschke, A. Liberati, N. Magrini, J. Mason, D. O’Connell, A. Oxman, B. Phillips, H. Schünemann *et al.*, and the GRADE Working Group. Systems for Grading the Quality of Evidence and the Strength of Recommendations II: Pilot Study of a New System. *BMC Health Services Research*, 5(25), 2005.
- [Augustin and Fischer, 2002] M. Augustin and M. Fischer. Evidenzbasierte Medizin-Grundlagen und Anwendungen in der Phlebologie. *Phlebologie* 5, 31:100–107, 2002.
- [Belsey and Tony, 2001] J. Belsey and S. Tony. What is Evidence-Based Medicine? *Hayward Medical Communications Ltd*, 1(2), 2001.
- [Bosse, 2001] Tibor Bosse. An Interpreter for Clinical Guidelines in Asbru. Master’s thesis, Vrije Universiteit Amsterdam, Amsterdam, NL, August 2001.

- [Bousquet *et al.*, 2001] J. Bousquet, P. Van Cauwenberge, N. Khaltaev, Aria Workshop Group, and World Health Organization. Allergic Rhinitis and its Impact on Asthma. *J Allergy Clin Immunol*, 108(5):147–334, November 2001.
- [Bury *et al.*, 2000] Jonathan Bury, John Fox, and David R. Sutton. The PROforma Guideline Specification Language: Progress and Prospects. In Barbara Heller, Markus Löffler, Mark Musen, and Mario Stefanelli, editors, *Proceedings of the First European Workshop on Computer-Based Support for Clinical Guidelines and Protocols (EGWLP 2000)*, volume 83 of *Studies in Health Technology and Informatics*, Leipzig, Germany, Nov. 2000. IOS Press.
- [(CCHMC), 2004a] Cincinnati Children’s Hospital Medical Center (CCHMC). Evidence Based Clinical Practice Guideline for Medical Management of Acute Otitis Media in Children 2 Months to 13 Years of Age. Technical report, Children’s Hospital Medical Center (CHMC), Cincinnati, OH, October 2004.
- [(CCHMC), 2004b] Cincinnati Children’s Hospital Medical Center (CCHMC). Evidence Based Clinical Practice Guideline for Medical Management of Otitis Media with Effusion in Children 2 Months to 13 Years of Age. Technical report, Children’s Hospital Medical Center (CHMC), Cincinnati, OH, October 2004.
- [(CCHMC), 2006] Cincinnati Children’s Hospital Medical Center (CCHMC). Evidence Based Clinical Practice Guideline for Children with Acute Bacterial Sinusitis in Children 1 to 18 Years of Age. Technical report, Children’s Hospital Medical Center (CHMC), Cincinnati, OH, July 2006.
- [Cochrane, 1989] A. Cochrane. *Effectiveness and Efficiency – Random Reflections on Health Services*. Cambridge University Press, 1989.
- [Collaboration, 2001] The Cochrane Collaboration. The Cochrane Collaboration & Evidence-Based Mental Health. Information for Managers and Clinical Directors. newsletter, July 2001.
- [David, 2000] E. David. *Appraising Systematic Reviews: Changing Practice*. The Joanna Briggs Institute, 2000.
- [Davies and Crombie, 2001] T.O. Davies and I. Crombie. What is a Systematic Review? *Hayward Group*, 1(5), 2001.
- [Davies and Cromble, 2001] T.O. Davies and I. Cromble. What is Meta-Analysis? *Hayward Group*, 1(8), 2001.
- [de Clercq *et al.*, 2004] Paul A. de Clercq, Johannes A. Blom, Hendrikus H. M. Korsten, and Arie Hasman. Approaches for Creating Computer-interpretable Guidelines that Facilitate Decision Support. *Artificial Intelligence in Medicine*, 31(1):1–27, May 2004.
- [Ebell *et al.*, 2004] Mark H. Ebell, Jay Siwek, Barry D. Weiss, Steven H. Woole, Jeffrey Susman, Bernard Ewigman, and Marjorie Bowman. Strength of Recommendation Taxonomy (SORT): A Patient-Centered Approach to Grading Evidence in the Medical Literature. *American Family Physician*, 69(3):548–556, Feb. 2004.

- [Eddv, 1986] D.M. Eddv. Successes and Challenges of Medical Decision Making. *Health Affairs*, 5(2):108–115, 1986.
- [Eldredge, 2000] J. Eldredge. Evidence-Based Librarianship: an Overview. *Bull Med Libr Assoc*, 88(4):289–302, 2000.
- [Evans, 2001] D. Evans. *An Introduction to Systematic Review: Changing Practice*. The Joanna Briggs Institute, 2001.
- [Field and Lohr, 1990] Marilyn J. Field and Kathleen N. Lohr, editors. *Clinical Practice Guidelines: Directions for a New Program*. National Academies Press, Institute of Medicine, Washington DC, 1990.
- [(FMSD), 2005] Finnish Medical Society Duodecim (FMSD). *EBM Guidelines. Evidence-Based Medicine [CD-ROM]*, chapter Sore throat and tonsillitis. Duodecim Medical Publications Ltd., Helsinki, Finland, Dec 28 2005.
- [Fox *et al.*, 1998] John Fox, Nicky Johns, and Ali Rahmanzadeh. Disseminating Medical Knowledge – The PROforma Approach. *Artificial Intelligence in Medicine*, 14:157–181, Sept. 1998.
- [Friedland *et al.*, 1998] Daniel J. Friedland, J. Ben Davoren, Michael G. Shlipak, Stephen W. Bent, Leslee L. Subak, and Terrie Mendelson. *Evidence-Based Medicine: A Framework for Clinical Practice*. Appleton & Lange, 1998.
- [Granata and Hillman, 1998] Attilio V. Granata and Alan L. Hillman. Competing Practice Guidelines: Using Cost-Effectiveness Analysis To Make Optimal Decisions. *Annals of Internal Medicine*, 128(1):56–63, 1998.
- [Greenhalgh, 1997] T. Greenhalgh. How to Read a Paper: Getting your Bearings (Deciding What the Paper is About) [Education and Debate]. *BMJ*, 315(7102):243–246, 1997.
- [Grimshaw and Russel, 1993] J.M. Grimshaw and I.T. Russel. Effect on Clinical Guidelines on Medical Practice: A Systematic Review of Rigorous Evaluations. *Lancet*, 342:1317–1322, 1993.
- [Grol *et al.*, 1998] Richard Grol, Johannes Dalhuijsen, Siep Thomas, Cees in't Veld, Guy Rutten, and Henk Mokkink. Attributes of Clinical Guidelines that Influence use of Guidelines in General Practice: Observational Studies. *BMJ*, 317:858–861, August 1998.
- [Guyatt *et al.*, 1992] G. Guyatt, J. Cairns, and D. Churchill. Evidence-Based Medicine. A new Approach to Teaching the Practice of Medicine. *JAMA*, 268:2420–2425, November 1992.
- [Guyatt *et al.*, 2000] Gordon G. Guyatt, Maureen O. Meade, Roman Z. Jaeschke, Deborah J. Cook, and Brian R. Haynes. Practitioners of Evidence Based Care. *BMJ*, 320:954–955, April 2000.
- [Guyatt *et al.*, 2006] Gordon H. Guyatt, David Guttermann, Michael H. Baumann, Doreen Addrizzo-Harris, Elaine Hylek, Barbara Philips, Gary Raskob, Sandra Zelman Lewis, and Holger J. Schuenemann. 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.
- [Harbour and Miller, 2001] Robin Harbour and Juliet Miller. A New System for Grading Recommendations in Evidence Based Guidelines. *British Medical Journal*, 323:334–336, 2001.
- [Head and Neck Cancer Disease Site Group *et al.*, 2004] Head and Neck Cancer Disease Site Group, DI Hodson, T Haines, and M Berry. Symptomatic Treatment of Radiation-induced Xerostomia in Head and Neck Cancer Patients. Practice guideline report 5-5, Cancer Care Ontario (CCO), Toronto (ON), March 2004.
- [Hill and Spittlehouse, 2001] A. Hill and C. Spittlehouse. What is Critical Appraisal? *Hayward Group*, 3(2), 2001.
- [(ICSI), 2003] Institute for Clinical Systems Improvement (ICSI). Rhinitis. Technical report, Institute for Clinical Systems Improvement (ICSI), Bloomington, MN, May 2003.
- [(ICSI), 2004a] Institute for Clinical Systems Improvement (ICSI). Acute Sinusitis in Adults. Technical report, Institute for Clinical Systems Improvement (ICSI), Bloomington, MN, May 2004.
- [(ICSI), 2004b] Institute for Clinical Systems Improvement (ICSI). Diagnosis and Treatment of Otitis Media in Children. Technical report, Institute for Clinical Systems Improvement (ICSI), Bloomington, MN, May 2004.
- [(ICSI), 2004c] Institute for Clinical Systems Improvement (ICSI). Viral Upper Respiratory Infection (VURI) in Adults and Children. Technical report, Institute for Clinical Systems Improvement (ICSI), Bloomington, MN, May 2004.
- [(ICSI), 2005] Institute for Clinical Systems Improvement (ICSI). Acute Pharyngitis. Technical report, Institute for Clinical Systems Improvement (ICSI), Bloomington, MN, May 2005.
- [(ICSI), 2006] Institute for Clinical Systems Improvement (ICSI). Diagnosis and Treatment of Obstructive Sleep Apnea. Technical report, Institute for Clinical Systems Improvement, Bloomington, MN, March 2006.
- [InferMed, 2007] InferMed. AREZZO Technical White Paper, 2007. http://www.infermed.com/index.php/arezzo/case_studies. Retrieved at: Apr. 10, 2007.
- [Kaiser, 2005] K. Kaiser. *LASSIE: Modeling Treatment Processes Using Information Extraction*. PhD thesis, Technische Universität Wien, 2005.
- [Kosara and Miksch, 2001] Robert Kosara and Silvia Miksch. Metaphors of Movement: A Visualization and User Interface for Time-Oriented, Skeletal Plans. *Artificial Intelligence in Medicine, Special Issue: Information Visualization in Medicine*, 22(2):111–131, May 2001.
- [Kosara *et al.*, 1998] Robert Kosara, Silvia Miksch, Yuval Shahar, and Peter Johnson. Asbruv-View: Capturing Complex, Time-Oriented Plans — Beyond Flow-Charts. In *Thinking with Diagrams 98*, Aberystwyth, England, 1998.

- [Kosara *et al.*, 2002] Robert Kosara, Silvia Miksch, Andreas Seyfang, and Peter Votruba. Tools for Acquiring Clinical Guidelines in Asbru. In *Proc. of the Sixth World Conference on Integrate Design and Process Technology (IDPT'02)*, 2002.
- [Last, 1995] J. Last. *A Dictionary of Epidemiology*. Oxford University Press, 3rd edition, 1995.
- [Lichtenstein *et al.*, 1987] M. Lichtenstein, P. Elwood, and C. Mulrow. Guidelines for Reading Case-Control Studies. *J Chronic Dis*, 40(9):893–903, 1987.
- [Lobach and Hammond, 1997] D.F. Lobach and W.E. Hammond. Computerized Decision Support Based on Clinical Practice Guidelines Improves Compliance with Care Standards. *Am J. Med*, 102:89–98, 1997.
- [Miksch *et al.*, 1997] Silvia Miksch, Yuval Shahar, and Peter Johnson. Asbru: A Task-specific, Intention-based, and Time-oriented Language for Representing Skeletal Plans. In Enrico Motta, Frank van Harmelen, Christine Pierret-Golbreich, Ian Filby, and Niek Wijngaards, editors, *7th Workshop on Knowledge Engineering: Methods Languages (KEML-97)*, Milton Keynes, UK, 1997. Open University.
- [Miksch *et al.*, 1998] Silvia Miksch, Robert Kosara, Yuval Shahar, and Peter Johnson. AsbruView: Visualization of Time-Oriented, Skeletal Plans. In *Proceedings of the 4th International Conference on Artificial Intelligence Planning Systems 1998 (AIPS-98)*. Carnegie Mellon University, AAAI Press, 1998.
- [Miksch, 1999] Silvia Miksch. Plan Management in the Medical Domain. *AI Communications*, 12(4):209–235, 1999.
- [Molino, 1999] Gianpaolo Molino. From Clinical Guidelines to Decision Support. In Werner Horn, Yuval Shahar, Greger Lindberg, Steen Andreassen, and Jeremy Wyatt, editors, *Proceedings of the Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making (AIMDM'99)*, pages 3–12, Aalborg, Denmark, 1999. Springer Verlag Berlin/Heidelberg.
- [(NHS), 2006] National Institute for Health and Clinical Excellence (NHS). *The Guidelines Manual 2006*. NHS, 2006.
- [Oosterhuis *et al.*, 2004] Wytze P. Oosterhuis, David E. Bruns, Joseph Watine, Sverre Sandberg, and Andrea R. Horvath. Evidence-Based Guidelines in Laboratory Medicine: Principles and Methods. *Clinical Chemistry* 50, 5(5):806–816, 2004.
- [Pai *et al.*, 2004] Madhukar Pai, Michael McCulloch, Jennifer D. Gorman, Nitika Pai, Wayne Enanoria, Gail Kennedy, Prathap Tharyan, and John M. Jr. Colford. Systematic Reviews and Meta-Analyses: An Illustrated, Step-by-Step Guide. *The National Medical Journal of India*, 17(2):86–95, 2004.
- [Peleg *et al.*, 2000] Mor Peleg, Aziz A. Boxwala, Omolola Ogunyemi, Qing Zeng, Samson Tu, Ronilda Lacson, Elmer Bernstam, Nachman Ash, Peter Mork, Lucila Ohno-Machado, Edward H. Shortliffe, and Robert A. Greenes. GLIF3: The Evolution of a Guideline Representation Format. In *Proceedings of the AMIA Annual Symposium*, pages 645–649, Los Angeles, 2000.

- [Peleg *et al.*, 2003] M. Peleg, Samson W. T., J. Bury, P. Ciccarese, J. Fox, R. A. Greenes, R. Hall, P. D. Johnson, N. Jones, A. Kumar, S. Miksch, S. Quaglini, A. Seyfang, E. H. Shortliffe, and M. Stefanelli. Comparing Computer-Interpretable Guideline Models: A Case-Study Approach. *Journal of the American Medical Informatics Association (JAMIA)*, 10(1):52–68, Jan-Feb 2003.
- [PROforma, 2007] PROforma. Advanced Computation Laboratory. <http://www.acl.icnet.uk/lab/proformaModelling.html>, 2007. Last updated at: Sept. 13, 2006. Retrieved at: Apr. 10, 2007.
- [Raspe and Stange, 1999] H. Raspe and E.F. Stange. Evidence-Based Medicine: Kontext und Relevanz "Evidenzgestützter Medizin". *Z Gastroenterol*, 37:525–533, 1999.
- [Sackett *et al.*, 1996] David L. Sackett, William M.C. Rosenberg, J.A. Muir Gray, R. Brian Haynes, and W. Scott Richardson. Evidence Based Medicine: What it is and What it isn't. In *BMJ*, volume 312, pages 71–72, 1996.
- [Schneider and Tramèr, 1999] M. Schneider and M.R. Tramèr. Meta-Analyses – Some Explanations, Illustrated with an Example. *Praxis*, 49, 1999.
- [(SCQIM), 2004] Steering Committee on Quality Improvement and Management (SCQIM). Classifying Recommendations for Clinical Practice Guidelines. *Pediatrics*, 114(3):874–877, Sept. 2004.
- [Scuka, 2004] L. Scuka. A Meta-Analysis and Systematic Review of the Efficacy of Enrofloxacin-Infections with the Bacterium Escherichia Coli in Pigs. *Slov Vet Res*, 41 (3/4):121–127, 2004.
- [Senn, 2002] S. Senn. *Cross-Over Trials in Clinical Research*, page 3. John Wiley & Son Ltd., 2nd edition, September 2002.
- [Seyfang and Miksch, 2002] Andreas Seyfang and Silvia Miksch. Asgaard's Contribution to the Guideline Representation Comparison. Technical report, Institute for Software Technology and Interactive Systems, Vienna University of Technology, Austria, 2002.
- [Seyfang *et al.*, 2002] Andreas Seyfang, Robert Kosara, and Silvia Miksch. Asbru 7.3 Reference Manual. Technical Report Asgaard-TR-2002-1, Institute of Software Technology Interactive Systems, Vienna University of Technology, Vienna, Austria, Europe, 2002.
- [Shahar *et al.*, 1998a] Yuval Shahar, Silvia Miksch, and Peter Johnson. The Asgaard Project: A Task-Specific Framework for the Application and Critiquing of Time-Oriented Clinical Guidelines. *Artificial Intelligence in Medicine*, 14:29–51, Sept. 1998.
- [Shahar *et al.*, 1998b] Yuval Shahar, Silvia Miksch, and Peter Johnson. The Asgaard Project: A Task-Specific Framework for the Application and Critiquing of Time-Oriented Clinical Guidelines. *Artificial Intelligence in Medicine*, 14:29–51, Sept. 1998.
- [Shekelle *et al.*, 2001] P. Shekelle, MP. Eccles, JM. Grimshaw, and SH. Woolf. When Should Clinical Guidelines be Updated? *British Medical Journal*, 323(7305):155–157, July 2001.

- [Shiffman *et al.*, 2000] Richard N. Shiffman, Bryant T. Karras, Abha Agrawal, Roland Chen, Luis Marenco, and Sujai Nath. GEM: A Proposal for a More Comprehensive Guideline Document Model Using XML. *Journal of the American Medical Informatics Association (JAMIA)*, 7(5):488–498, 2000.
- [(SIGN), 2001] Scottish Intercollegiate Guidelines Network (SIGN). SIGN 50: A Guideline Developers’ Handbook. SIGN Publication 50, SIGN, Edinburgh, 2001.
- [(SIGN), 2003a] Scottish Intercollegiate Guidelines Network (SIGN). Diagnosis and Management of Childhood Otitis Media in Primary Care. A National Clinical Guideline. Technical Report 66, Scottish Intercollegiate Guidelines Network, Edinburgh, Scotland, Feb. 2003.
- [(SIGN), 2003b] Scottish Intercollegiate Guidelines Network (SIGN). Management of Obstructive Sleep Apnoea/Hypopnoea Syndrome in Adults. A National Clinical Guideline. Technical Report 73, Scottish Intercollegiate Guidelines Network, Edinburgh, Scotland, June 2003.
- [(SIGN), 2004] Scottish Intercollegiate Guidelines Network (SIGN). Management of Patients with Stroke: Identification and Management of Dysphagia. A National Clinical Guideline. Technical Report 78, Scottish Intercollegiate Guidelines Network (SIGN), Edinburgh, Scotland, September 2004.
- [Séroussi *et al.*, 1999] Brigitte Séroussi, Jaques Bouaud, and Éric-Charles Antoine. Enhancing Clinical Practice Guideline Compliance by Involving Physicians in the Decision Process. In Werner Horn, Yuval Shahar, Greger Lindberg, Steen Andreassen, and Jeremy Wyatt, editors, *Proceedings of the Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making (AIMDM’99)*, pages 3–12, Aalborg, Denmark, 1999. Springer Verlag.
- [Steele and Fox, 2002] Rory Steele and John Fox. Tallis PROforma Primer – Introduction to PROforma Language and Software with Worked Examples. Technical report, Advanced Computation Laboratory, Cancer Research, London, UK, 2002.
- [Sutton and Fox, 2003a] David Sutton and John Fox. Syntax and Semantics of PROforma. Technical Report 1.3.38, Cancer Research UK, March 2003.
- [Sutton and Fox, 2003b] David R. Sutton and John Fox. The Syntax and Semantics of the PROforma Guideline Modeling Language. *Journal of the American Medical Informatics Association (JAMIA)*, 10(5):433–443, Sep / Oct 2003.
- [Teranziani *et al.*, 2004] Paolo Teranziani, Stefania Montani, Alessio Bottrighi, Mauro Torchio, Gianpaolo Molino, and Gianluca Correndo. The GLARE Approach to Clinical Guidelines: Main Features. In Katharina Kaiser, Silvia Miksch, and Samson Tu, editors, *Computer-based Support for Clinical Guidelines and Protocols. Proceedings of the Symposium on Computerized Guidelines and Protocols (CGP 2004)*, volume 101, pages 162–166. IOS Press, 2004.
- [Thomas, 1999] Lois Thomas. Clinical Practice Guidelines. *BMJ*, 2(2):38–39, 1999.

- [Twaddle, 2005] Sara Twaddle. Clinical Practice Guidelines. *Singapore Medical Journal*, 46(12):681–687, Dec.] 2005.
- [(UMHS), 2002a] University of Michigan Health System (UMHS). Allergic Rhinitis. Technical report, University of Michigan Health System, Ann Arbor, MI, July 2002.
- [(UMHS), 2002b] University of Michigan Health System (UMHS). Management of Gastroesophageal Reflux Disease (GERD). Technical report, University of Michigan Health System, Ann Arbor, MI, March 2002.
- [(UMHS), 2002c] University of Michigan Health System (UMHS). Otitis Media. Technical report, University of Michigan Health System, Ann Arbor, MI, May 2002.
- [Verkerk *et al.*, 2006] Karin Verkerk, Haske Van Veenendaal, Johan L. Severens, Erik J. M. Hendriks, and Jako S. Burgers. Considered Judgment in Evidence-Based Guideline Development. *International Journal for Quality in Health Care*, 18(5):365–369, September 2006.
- [Vollebregt *et al.*, 1999] Arjen Vollebregt, Annette ten Teije, Frank van Harmelen, Johan van der Lei, and Mees Mosseveld. A Study of PROforma, a Development Methodology for Clinical Procedures. *Artificial Intelligence in Medicine*, 17(3):195–221, 1999.
- [Votruba *et al.*, 2003] P. Votruba, S. Miksch, and R. Kosara. Linking Clinical Guidelines with Formal Representations. In *Dojat, M. and Keravnou, E. and Barahona, P., editors, Artificial Intelligence in Medicine. Proceedings of the 9th Conference on Artificial Intelligence in Medicine in Europe (AIME 2003)*, pages 152–157, Berlin, 2003. Springer.
- [Votruba *et al.*, 2004] P. Votruba, S. Miksch, and R. Kosara. Tracing the Formalizing Steps of Textual Guidelines. In *Katharina Kaiser, Silvia Miksch, and Samson Tu, editors, Computer-based Support for Clinical Guidelines and Protocols. Proceedings of the Symposium on Computerized Guidelines and Protocols (CGP 2004)*, volume 101 Studies in Health Technology and Informatics., pages 172–176. IOS Press, 2004.
- [Weil, 1999] E. Weil. Evidence-Based Medicine Glossary. *University of Toronto Medical Journal*, 76(3):140–141, 1999.
- [West *et al.*, 2002] Suzanne West, Valerie King, Timothy S. Carey, Kathleen N. Lohr, Nikki McKoy, Sonya F. Suttton, and Linda Lux. Systems to Rate the Strength Of Scientific Evidence. Evidence Report/Technology Assessment 47, Agency for Healthcare Research and Quality U.S. Department of Health and Human Services, Rockville (MD), April 2002.
- [Zielinski, 2002] W. Zielinski. *Evidence-based Medicine: Einsatzmöglichkeiten in der Stationären Versorgung*. PhD thesis, Technische Universität Berlin, 2002.

Curriculum Vitae

Personal Information

Name	Alime Öztürk
Address	Arndtstrasse 31-33/1/21 1120 Wien
Phone	0699 123 78 372
E-Mail	e9925136@stud3.tuwien.ac.at
Nationality	Austria
Date of Birth	01.11.1977
Gender	Female
Family status	2 children (born 1999 and 2001)

Education

Sep.1984 - Jul.1988	Elementary school, Vienna
Sep.1988 - Jul.1992	Secondary school, Vienna
Sep.1992 - Jun.1996	Oberstufen Real Gymnasium (ORG), Vienna
June 1996	University entrance diploma Oberstufen Real Gymnasium (ORG), Vienna
Oct.1999 - Nov.2004	Computer Science, Vienna University of Technology Master's thesis: "Anforderungen an ein State-of-the-Art Case-Tool in Hinblick auf die durchgängige Verfolgbarkeit von Modellelementen", Supervisor: Prof. Thomas Grechenig
Nov.2004 - Jun.2007	PhD studies, Vienna University of Technology PhD-Thesis: "Embedding the Evidence Information in Computer-Supported Guidelines into the Decision-Making Process", Supervisor: Prof. Silvia Miksch

Job Experience

- Mar.2006 - Jul.2006 Student assistant for the courses "Computer-unterstuetzte Therapieplanung" at the
Institut of Software Technology & Iterative Systems
Information Engineering Group
- Nov.1996 - Sep.1998 Salesperson
Finanzdienstleistungsinstitut AG, Vienna
- Mar.1994 - Oct.1996 Telemarketing
Omnitel, Vienna

Membership

- OEGAI (Austrian Society for Artificial Intelligence)
VFA (Verband Freunde und Absolventen der Technischen Universitaet Wien)

Publications

- K. Kaiser, P. Martini, S. Miksch, A. Öztürk. *A Meta Schema for Evidence Information in Clinical Practice Guidelines as a Basis for Decision-Making*. In K. Kuhn, T. Y. Leong, J. Warren (eds.) 12th World Congress on Health (Medical) Informatics (Medinfo'2007), AMIA, Brisbane, Australia, 2007, forthcoming.
- A. Öztürk, K. Kaiser, P. Martini, S. Miksch. *Embedding the Evidence Information in Guideline Representation Languages*. In P. Kokol, T. Welzer-Družovec, D. Mičetić-Turk, M. Zorman (eds.) 20th IEEE International Symposium on Computer-Based Medical Systems (CBMS'07), IEEE Computer Society Press, Maribor, Slovenia, 2007, forthcoming.