From Clinical Practice Guidelines to Computer-interpretable Guidelines

A Literature Overview

A. Latoszek-Berendsen[1]; H. Tange[1]; H. J. van den Herik[2]; A. Hasman[3]
[1]Caphri School for Public Health and Primary Care, Maastricht University, Maastricht, The Netherlands;
[2]TiCC, Tilburg Centre for Cognition and Communication, Tilburg University, Tilburg, The Netherlands;
[3]KIK, University of Amsterdam, Amsterdam, The Netherlands

Keywords
Clinical practice guidelines, paper-based guidelines, computer-based guidelines, computerized (clinical) guidelines, computer-interpretable guidelines, guideline formalization, decision support system

Summary
Background: Guidelines are among us for over 30 years. Initially they were used as algorithmic protocols by nurses and other ancillary personnel. Many physicians regarded the use of guidelines as cookbook medicine. However, quality and patient safety issues have changed the attitude towards guidelines. Implementing formalized guidelines in a decision support system with an interface to an electronic patient record (EPR) makes the application of guidelines more personal and therefore acceptable at the moment of care.
Objective: To obtain, via a literature review, an insight into factors that influence the design and implementation of guidelines.
Methods: An extensive search of the scientific literature in PubMed was carried out with a focus on guideline characteristics, guideline development and implementation, and guideline dissemination.
Results: We present studies that enable us to explain the characteristics of high-quality guidelines, and new advanced methods for guideline formalization, computerization, and implementation. We show how the guidelines affect processes of care and the patient outcome. We discuss the reasons of low guideline adherence as presented in the literature and comment upon them.
Conclusions: Developing high-quality guidelines requires a skilled team of people and sufficient budget. The guidelines should give personalized advice. Computer-interpretable guidelines (CIGs) that have access to the patient's EPR are able to give personal advice. Because of the costs, sharing of CIGs is a critical requirement for guideline development, dissemination, and implementation. Until now this is hardly possible, because of the many models in use. However, some solutions have been proposed. For instance, a standardized terminology should be imposed so that the terms in guidelines can be matched with terms in an EPR. Also, a dissemination model for easy updating of guidelines should be established. The recommendations should be based on evidence instead of on consensus. To test the quality of the guideline, appraisal instruments should be used to assess the guideline as a whole, as well as checking the quality of the recommendations individually. Only in this way optimal guideline advice can be given on an individual basis at a reasonable cost.

Correspondence to:
Agnieszka Latoszek-Berendsen
Caphri School for Public Health and Primary Care
Maastricht University
P.O. Box 616
6200 MD Maastricht
The Netherlands
E-mail: aa.berendsen@gmail.com

Methods Inf Med 2010; 49: 550–570
doi: 10.3414/ME10-01-0056
received: August 4, 2010
accepted: October 8, 2010
prepublished: November 18, 2010

1. Introduction

Clinical practice guidelines (CPGs) are meant to improve the process and outcome of healthcare and to optimize resource utilization. CPGs are defined as systematically developed statements to assist physician and patient decisions about appropriate health care for specific circumstances [1].

The development of CPGs was a response to the discovery of large, unexplained variations in 1) the practice of the physician, 2) documentation of significant rates of inappropriate care, and 3) an interest in managing healthcare cost [2]. Important is the word assist: guidelines are guides rather than rules. In contrast, clinical protocols provide a comprehensive set of rigid criteria outlining the management steps for a single clinical condition or other aspects of the organization [3]. The text of CPGs usually contains a number of components such as: the intention(s) of the guideline, the medical background, patient eligibility criteria, procedural statements (e.g., drug recommendations), evidence for the advisories, treatment cost-benefit analyses, and references [4].

Guidelines are already among us for more than 30 years. Initially, the guidelines or the more algorithmic protocols were only used by nurses and other ancillary personnel. Many physicians did not use guidelines because individual patients differ from each other and guidelines usually cover a population. However, quality issues and patient safety issues somewhat changed their attitude towards guidelines. But, still, physicians do not use guidelines on a large scale. A variety of obstacles to the use of guidelines have been discovered. Some of the obstacles were directly related to the guideline: for example, the guidelines could contain ambiguous text

or even inconsistencies. Other reasons were, e.g., 1) paper-based guidelines are difficult to use in clinical practice, 2) physicians consider the use of guidelines as cookbook medicine, 3) physicians are not convinced that the use of guidelines would lead to better care, or 4) organizational barriers are present.

With the advent of decision support systems, including guideline implementation systems, the situation changed drastically. Formalized and computer-interpretable guidelines (CIGs) could be made (more) patient-specific. By coupling a guideline implementation system to an electronic patient record (EPR), alarms and reminders could be issued concerning the management of an individual patient. Computer-based guidelines can, to a certain extent, be automatically checked to discover inconsistencies or open ends. In this way the quality of the guidelines can be raised. Also, the compliance to guidelines can be checked by formal methods.

Early guidelines were often based on consensus among experts. Later guideline recommendations were increasingly more evidence-based. So, research findings were translated into actionable recommendations for clinical care. Since the scientific quality of the sources of the evidence on which the recommendations are based can vary, evidence is classified into a number of levels. For the purpose of guideline development usually three levels are distinguished: A, B, and C. The American College of Cardiology (ACC) and the American Heart Association (AHA) taskforce on practice guidelines also recognize the three levels of evidence [5].

Level A concerns data derived from multiple randomized clinical trials, level B concerns data derived from a single randomized trial or non-randomized studies, and level C concerns a consensus opinion of experts.

The taskforce emphasizes the importance of 1) mentioning the levels of evidence of the sources on which recommendations are based, and of 2) classifying the confidence in the recommendations itself. Three classes of recommendations are distinguished. Class I contains conditions for which there is evidence and/or general agreement that a given procedure or treatment (according to the recommendation) is useful and effective, Class II contains conditions for which there is conflicting evidence and/or a divergence of opinion about the usefulness/efficacy of a procedure or treatment, whereas Class III contains conditions for which there is evidence and/or general agreement that the procedure is not useful/effective, and in some cases may be harmful. By stating the level of evidence of the sources used for the recommendations, and the classes of recommendations, users of guidelines become aware of the strength of the research evidence supporting each recommendation mentioned in the guideline.

Below we provide three different groups of caveats (1.1) and an overview of the paper (1.2).

1.1 Caveats Regarding the Use of CPGs

CPGs could be of benefit for physicians because they receive a better overview of available treatments with recommendations for those who are uncertain about how to proceed. Obviously, guidelines may improve consistency of care and help to change the attitude of practitioners who became used to their, maybe outdated, medical practice. However, in situations where CPGs are flawed or misleading, professionals can be harmed by them. In the following, we list three different groups of concerns, expressed by the authors mentioned below, with respect to guidelines.

First, the type of advice in guidelines is influenced by the experience and knowledge of the guideline-developing group, since only a small subset of what is recommended in guidelines usually has been tested in well designed studies. The proportion of recommendations for which there is no conclusive evidence is also growing. Tricoci et al. [6] determined that recommendations issued in current ACC/AHA CPGs are largely developed from lower levels of evidence or expert opinion. The authors highlighted the need to improve the process of writing guidelines and to expand the evidence base from which CPGs are derived. Here we remark that Tricoci et al. were aware that their findings are reflective of a specialty that has a large pool of research to draw on for its care recommendations.

Second, guidelines in medical areas, in which large clinical trials are performed less frequently, may have an even weaker evidence-based foundation. Therefore, there is a risk for professionals that auditors or managers may unfairly judge the quality of their care based on criteria from invalid guidelines. Shaneyfelt and Centor placed this problem in the spotlight in an editorial in JAMA [7]. The authors were concerned that some of the consensus statements taken out of the guidelines will be turned into performance measures and other tools to critique the quality of physician care.

Third, Shaneyfelt and Centor [7] noted that guidelines often have become marketing tools for device and pharmaceutical manufacturers and that financial ties between guideline panel members and industry are common.

1.2 Goal and Overview

In this paper we present a literature survey of factors that influence the design and implementation of guidelines. In Section 1 we described our motivation for this literature search (improving quality of care and patient safety taking into account the three groups of serious caveats mentioned above). In Section 2 we present the method of literature search. Since only high-quality guidelines will probably be used in practice, in Section 3 we discuss how the quality of guideline development and implementation can be assessed. Then in Section 4 we describe formalization methods to translate paper guidelines into computer-based guidelines. Further on, in Section 5 we introduce models for computer-interpretable guidelines that facilitate their design. Additionally, we discuss problems that may occur during the formalization process. Further, in Section 6 we describe the effects of guidelines with respect to patient outcome and efficiency of care (including reasons why guidelines are not used). In Section 7 we present barriers to guidelines adherence. Finally, we discuss the current state of the art and present our ideas for further challenges in Section 8.

2. Methods

As stated in Section 1 we concentrate our search of the scientific literature on 1) guideline characteristics, 2) guideline development and implementation, and 3) guideline dissemination. After the following five steps we arrived at our defined set of articles.

First, an extensive search of the scientific literature in PubMed was carried out in October 2009 using the following key terms: clinical practice guidelines, paper(-)based guidelines, computer(-)based guidelines, computerized (clinical) guidelines, decision support system, computerized guidelines, computer interpretable guidelines, guideline formalization. By using these keywords incorporated in adequately formulated search queries 1727 articles were retrieved. We only took into account articles available with full text. Our query was as follows: (#1 AND #4) OR #2 OR (#3 AND #5) OR #6 where

#1 clinical practice guidelines (all fields);
#2 paper-based guidelines (all fields);
#3 computer-based guidelines (all fields);
#4 computerized guidelines (all fields);
#5 decision support system (all fields);
#6 computer interpretable guidelines (all fields).

Second, the titles of the articles were scanned and those articles were selected that matched at least one of the following five categories.

1. Characteristics of clinical practice guidelines
2. Paper-based guidelines
3. Development: guideline formalization
4. Implementation: computer-interpretable guidelines, feedback systems, decision support systems
5. Dissemination: approaches facilitating the design of computer-interpretable guidelines

Papers in the category 1 should provide information about attributes of clinical practice guidelines, or the definition of a "good" guideline, or mention standards to build clinical guidelines. Category 2 holds articles that contain details about paper-based clinical practice guidelines that focus on guideline acceptance, performance of the guidelines with respect to patients' outcome and with respect to the process of care. Category 3 contains articles that describe how to translate paper-based guidelines to computer-based guidelines. Category 4 captures papers that describe feedback systems and decision support systems, and/or includes effects of those systems. Category 5 includes papers that present methods for developing computer-based guidelines as well as decision support systems that incorporate these guidelines.

In total 696 articles matched the categories.

Third, the abstracts of those papers were independently studied by two evaluators (author 1 and 4 of this article, both being medical informaticians). This resulted in a set of 276 articles.

Fourth, the content of those 276 articles was examined. After reading the full papers, this set was reduced to 83 papers.

Fifth, by screening the references in these articles and reading new available sources published in 2010, 20 papers were added to the set that fulfilled the criteria mentioned above.

3. Quality of Guidelines

Research suggests that physicians do not extensively use CPGs at the point of care, because they are quite time-consuming to go through and unsuitable for just-in-time use. To overcome this obstacle many guidelines are therefore accompanied by algorithms which represent practice recommendations in diagrammatic form. Since the recommendations in paper-based guidelines do not always reflect the flow of patient encounters, the diagrams are portrayed as flowcharts to specify the recommended steps. With this addition recommendations can be extracted quickly. A disadvantage is that because of the condensed nature of the algorithms they are often rigid and cannot provide all the information present in the text-based guidelines which may be needed by non-expert physicians. Also, there is no room for explanation of counter-intuitive advice [8].

Guidelines based on the synthesis of the best, most recent evidence can help practitioners keep current with the literature and help them assimilate evidence into practice [9]. Given the resources required to identify all relevant primary studies many guidelines rely on systematic reviews that were either previously published or carried out by the guideline developers themselves. Systematic reviews can aid in guideline development because they involve searching for, selecting, critically appraising, and summarizing the results of primary research. However, sole reliance on systematic reviews will never adequately serve the development of guidelines. Guideline developers still have to appraise the original studies to understand the populations, interventions, and outcomes evaluated. When several systematic reviews exist on the same topic they may generate conflicting conclusions and then the developers have to deal with such information in a transparent way.

Yet, many guidelines represent expert consensus [7]. Since guideline committees begin with implicit biases and values, the recommendations they make may be affected. Bias may occur subconsciously and go unrecognized. Guideline users can only adjust for biases when guideline panels make their values and goals explicit, but usually they remain opaque.

Below we discuss characteristics of high-quality guidelines (3.1), guideline appraisal instruments (3.2), aspects of guidelines maintenance (3.3), and guideline dissemination and implementation (3.4).

3.1 Characteristics of High-quality Guidelines

Evidence-based guidelines can reduce the delivery of inappropriate care. Several medical organizations have published methodological standards for developing scientifically sound guidelines. Below we review four of such standards and opinions on what good quality may constitute. We emphasize the standards for computer-interpretable guidelines.

1) The American Institute of Medicine (IOM) defined eight desirable attributes of CPGs: validity, reliability and reproducibility, clinical applicability, clinical flexibility, clarity, documentation, development by a multi-disciplinary process, and plans for

review [1]. However, critical information that would attest to validity or would document fulfillment of the other IOM criteria is regularly absent from published guidelines [2]. For example, ambiguity and vagueness in the wording of the recommendations, incomplete decision logic pathways, and poor differentiation between evidence and opinion are common shortcomings of many guidelines [10].

2) Littlejohns and Cluzeau [11] define a 'good' guideline as a guideline that is valid, reproducible, and reliable. This means that there is evidence that i) a guideline, when implemented, will lead to outcomes that it intended to achieve; ii) given the same data, other guideline producers will offer the same recommendations, and iii) guideline users will interpret the recommendations in the same way.

3) Burgers et al. [12] define the quality of a guideline as the confidence that the potential biases inherent of guideline development have been addressed adequately and that the recommendations are both internally and externally valid, and are feasible for practice.

4) Shiffman et al. [13] observe that the quality of guidelines is related primarily to the scientific validity of guidelines and that, generally, the quality is assessed for the guideline as a whole. According to the authors the implementability of guidelines also should be assessed. Implementability is the aspect of guideline quality that indicates how easily individual recommendations within a guideline can be implemented in practice. Two dimensions of implementability are particularly important: executability (do the recommendations communicate what to do) and decidability (do the recommendations state when to do it). In addition, the aspect computability indicates the ease with which a recommendation might be operationalized in an electronic information system.

3.2 Guideline Appraisal Instruments

The process of guideline creation is time- and resource-consuming. To be sure that the guidelines will be used in clinical practice, they have to be of the highest quality. Therefore, appraisal instruments have been developed to guide the development process and to evaluate the quality of the developed guidelines.

Below, we present four guideline appraisal instruments that help to assess the quality of guidelines.

1) To assess the methodological quality of guidelines published in the peer-reviewed medical literature, Shaneyfelt et al. [2] developed and validated a 25-item guideline appraisal instrument. The instrument assessed the guidelines with respect to standards on format and development (10 items), identification and summary of evidence (10 items), and formulation of recommendations (5 items). They concluded that on average the 279 evaluated guidelines published in the peer-reviewed medical literature from 1985 to 1997 adhered to 43.1% of the 25 items. Guidelines show significant improvement over the years (from 36.9% in 1985 to (still only) 50.4% in 1997). Although guidelines are developed to improve health outcomes only 40% of the guidelines specified the outcomes of interest. Also, less than half of the guidelines described the population to which the guideline should be applied. The standards on the identification and summary of evidence were poorly followed (only one third of the guidelines adhered to these standards). Just few of the guidelines (14.3%) quantified the effects on healthcare costs. Most guidelines (82.1%) specified the preventive, diagnostic, or therapeutic options available.

2) The AGREE (Appraisal of Guidelines for Research and Evaluation) instrument [14] was developed to assess both the quality of the reporting and the quality of some aspects of the recommendations of the guideline. It provides an assessment of the validity of a guideline. In total six quality domains (scope and purpose, stakeholder involvement, rigor of development, clarity and presentation, applicability, and editorial independence) are considered by the instrument. A four-point Likert scale is used to score each of the 23 items.

Burgers et al. [12] sought to identify which external factors such as care level, scope, type of guideline, year of publication, type of developing agency and guideline program predict the quality of guidelines. The quality of a guideline was determined by the AGREE instrument and scores were obtained for the six quality domains that are considered in the instrument. The data collected for the validation of the AGREE instrument was analyzed. In total 86 guidelines developed by 62 different agencies from 11 countries were assessed using the instrument. Burgers et al. concluded that the type of developing agency was a major predictor. Guidelines produced within established guideline programs (e.g., structured and coordinated programs designed with the specific aim of producing several clinical practice guidelines) and by government-funded agencies had a higher quality than the guidelines produced by specialist societies. This is due to the fact that government-funded agencies have a guideline program that provides a systematic procedure with key elements such as a multidisciplinary guideline development group, systematic literature review, external peer review, and different products for dissemination. The influence of the other five investigated predictors of guideline quality was limited. The study also showed that although the type of developing agency was a good predictor, the agency scored lower on the AGREE quality domain "applicability" (applicability pertains to the likely organizational, behavioral and cost implications of applying the guideline) than on the other quality domains. Apparently agency policies and procedures are more concerned with the methodology of producing guidelines than with the effectiveness of guidelines in daily practice.

3) Shiffman et al. [15] convened the Conference on Guideline Standardization (COGS) to define a standard (the COGS statement) for guideline reporting that would promote guideline quality and facilitate implementation. In a modified Delphi study, participants had to rate their agreement with the statement that "(item name) is a necessary component of practice guidelines". This study resulted in the COGS statement, a checklist of 18 topics. Specifically one of the purposes was to address problems with implementing guidelines electronically. Guideline authors strive to make recommendations that accurately reflect the scientific evidence but

sometimes they intentionally introduce ambiguity into the recommendations to reflect their uncertainties. Often, developers use terms that are not clearly defined, thereby presenting difficulties when recommendations are integrated into a decision support system. The use of the checklist for a productive discussion among 1) guideline developers, 2) disseminators and implementers, and 3) knowledge managers about critical guideline items and clear statements of decidable and executable recommendations may overcome major impediments to guideline use.

Guidelines that have been reported according to the COGS statement are likely to be easier to assess for validity because they systematically report precise details that are critical for understanding a guideline's development, its recommendation statements, and potential issues in its application [16]. Shiffman et al. [15] note that the COGS checklist represents a common framework for guideline developers to help ensure that important information is included in the guideline documentation. For journal editors and other disseminators, the COGS checklist helps to determine that a guideline draft is ready for publication. For implementers, the checklist is to inform the selection process and to facilitate the representation and translation of recommendations into tools that influence clinical behavior.

4) Shiffman et al. [13] developed and validated a tool for appraisal of the implementability of clinical guidelines: GLIA (Guideline Implementability Appraisal). GLIA is intended to help anticipate barriers to implementation success. It can be used by the authors of the guideline and by those individuals who choose guidelines for applications within a health care delivery system. As a guideline is being developed GLIA can provide feedback to guideline authors about potentially remediable defects. Implementers can use GLIA to help select a guideline, to identify potential obstacles, and to target efforts toward addressing identified barriers.

Factors that indicate guideline quality but not implementability were not included in GLIA. Implementability was defined to refer to a set of characteristics that predict the relative ease of implementation of guideline recommendations. GLIA focuses on factors that are fundamental to the guideline. Extrinsic items relating to the effect of the recommendations on the process of care and items relating to the novelty or innovation of a guideline statement were retained because developers can anticipate these barriers and offer potential strategies for implementation success. The factors were grouped into categories and then specific questions were devised to characterize each of these categories. These questions became items of the instrument. GLIA consists of 31 items, arranged into 10 dimensions. Questions from 9 of the 10 dimensions are applied individually to each recommendation of the guideline. Decidability and executability are critical dimensions [13]. Application of GLIA to the measurement of decidability and executability requires that users translate guideline recommendations into statements comprising conditions and actions. Other dimensions of GLIA are: global (general characteristics of the guideline as a whole), presentation and formatting, measurable outcomes, apparent validity, and flexibility (degree to which a recommendation permits interpretation and allows for alternatives in its execution), effect on process of care, novelty/innovation, and computability.

Several GLIA items overlap with items in the AGREE instrument. However, the GLIA authors claim that GLIA is the only tool that emphasizes implementation concerns at the level of the individual recommendation. GLIA incorporates an optional category – computability – to indicate the ease with which a recommendation might be operationalized in electronic information systems.

From this overview we may conclude that there are sufficient instruments to evaluate the quality of existing guidelines. These instruments mainly evaluate the quality of the whole guideline. As we will show later, there is a trend now to evaluate the separate recommendations of the guideline rather than the guideline as whole. Some of the instruments not only assess the quality of the guideline as a whole but also 1) the separate recommendations, 2) whether guidelines are adequate for implementation in practice, and 3) whether guidelines can be used by decision support systems.

3.3 Maintenance of Guidelines

In this section we discuss three aspects of maintaining clinical practice guidelines, viz. maintenance, local adaptations, and augmented dissemination model.

Maintenance of guidelines is a notable problem to take into account, since guidelines have to be periodically updated. Tricoci et al. [6] noted with respect to the cardiovascular guidelines of the ACC/AHA that the mean time elapsing from a version to its update was 4.6 years for disease-based guidelines, 5.4 years for interventional procedure-based, and 8.2 years for diagnostic procedure-based guidelines. They also indicate that updated versions are not always more evidence-based than their predecessors. Tricoci et al. concluded that among ACC/AHA clinical practice guidelines with at least one revision or update, the number of recommendations increased with 48%, with the largest increase observed in the use of class-II recommendations. In the 16 current guidelines reporting levels of evidence, only 314 of the 2711 recommendations are classified as level of evidence A, whereas 1246 are level of evidence C. The level of evidence significantly varies across categories of guidelines (disease, intervention, or diagnostic) and across individual guidelines. Recommendations with level of evidence A are mostly concentrated in class I, but only 245 of 1305 class-I recommendations have level of evidence A.

Moreover, local adaptations to guidelines make maintenance more difficult. When revised guidelines are disseminated they have to be reintegrated into the local clinical settings. The revisions must be synchronized with the changes previously made to this guideline or a related guideline. Current guideline representations do not provide a sufficiently fine-grained representation for complete identification of revisions in a guideline. This problem again can be attacked better with the help of CIGs.

The Guideline Elements Model (GEM) (see Section 4.3) defines elements for identifying whether a guideline has been revised

and adapted, but not how. Boxwala et al. [17] suggested enhancing the GEM by adding components that 1) identify the specific elements that have been revised, 2) state the reason for the change, and 3) explain the revision. They proposed an augmented dissemination model. In this model the guideline is downloaded, its content is adapted to local practice preferences and then it is integrated with an institutional clinical information system. When revisions of the guideline are published, the revised guideline is downloaded, unmodified sections of the revised guideline are automatically synchronized with the previous local version of the guideline, and the modified sections are then manually adapted and integrated.

In summary, the update procedure is still in a state of flux, and needs to be improved.

3.4 Guideline Dissemination and Implementation

The use of guidelines is determined by 1) the quality of the guideline development (discussed above), 2) by the way the guidelines are disseminated, and 3) the way they are implemented. Points 2 and 3 are discussed below.

For busy clinicians, clinical practice guidelines can only be of help when they take notice of these guidelines. Insight into the best ways to spread the guidelines (dissemination) is therefore important. There are several dissemination and implementation strategies: educational strategies, educational outreach, audit/feedback/peer review, reminder and decision support systems, financial incentives, use of a local opinion leader, and combinations of these approaches (so-called multifaceted interventions). Beneath, we discuss the most prevailing ones.

The effectiveness of guideline dissemination and implementation strategies has been studied in systematic reviews. Below we discuss three important reviews. First, Grimshaw et al. [18] concluded that current guideline dissemination and implementation strategies can lead to improvements in care. However, there was considerable variation in the observed effects. Second, in their study Prior et al. [19] came to similar conclusions. Effective strategies must include multifaceted interventions, interactive education, and clinical reminder systems. Didactic education and passive dissemination strategies were ineffective. They concluded that successful strategies should actively engage clinicians throughout this process. Third, a study carried out among general practitioners in the Netherlands, to gain insight into processes of dissemination of guidelines, showed that the scientific journal was the most important, but certainly not the only source of information to become informed about them. Discussing the guidelines in the local family doctor group was also a means of dissemination. It was concluded that segmentation of the target group is necessary for effective dissemination of guidelines: for some doctors it is desirable to make evidence quickly available, for others spreading the guidelines through the local network may be effective, while for a third group a more active, personal approach may be necessary [20].

Below we focus on implementation details.

Guideline implementation involves the concrete activities and interventions undertaken to turn policies into desired results. Guidelines are not uniformly successful in improving care and several instances of implementation failure have been described [13]. Moreover, performance measurement and payment are increasingly linked to goals established by practice guidelines. Lin and Slawson [21] wrote an article on how to identify and use good practice guidelines. We conclude with them that the best guidelines are those based on systematic reviews and patient-oriented evidence, and are prospectively validated. The guidelines should have a transparent development process, identify potential conflicts of interest, and offer flexibility in various clinical situations.

Most CPGs are now easily accessible on specialized websites. For instance, the National Guideline Clearinghouse™ (NGC) is a public resource for evidence-based clinical practice guidelines. However, the wide diffusion of CPGs does not solve the problem of their effective use in daily practice (implementation).

Returning for a while to dissemination we may remark that disseminating CPGs in a textual format proved to be inefficient [22]. The practitioner has to read several pages in order to find the appropriate care recommendation for a specific clinical circumstance. Their practical aspect rather than their content is then at fault. Guidelines will be much more efficient if they are available in the healthcare setting, integrated in the health care information system, easily adaptable to given clinical situations/scenarios and able to avoid overloading physicians with non-essential information. Minimizing the time spent consulting CPGs is crucial when attempting to improve their usage in everyday practice. Computer-interpretable guidelines will therefore become a necessity.

4. Formalization of Guidelines

The effect of clinical practice guidelines (CPGs) and computer-interpretable guidelines (CIGs)) depends on the proper implementation of either type of guideline. It is unrealistic to expect that clinicians read and use all guidelines that are available; however they are still expected to possess all the knowledge. The increasing amount of CPGs and the encountered problems with CPGs was the reason that for over a decade researchers have looked for automated methods of delivering guidelines to clinicians when the guideline is most relevant to the care of patients: when the patient is seen by the clinician.

Below we discuss: standardization of decisions (4.1), knowledge-based approach (4.2), methodological issues (4.3), document-centric approach (4.4), the bridging approach (4.5), and implementation in decision support systems (4.6).

4.1 Standardization of Decisions

Morris [23], the proponent of the use of computer-based protocols, declared that standardization of clinical decisions is needed for clinical practice. Without explicitly defining the recommendations and the context in which they are appropriate,

the fundamental scientific requirement of replicability of results cannot be achieved. A guideline should not require judgments by a clinician. Any form of guideline or protocol can theoretically contain sufficient detail. In practice, however, paper-based versions of any but the simplest protocols cannot be made explicit and therefore remain dependent on clinician judgment. The need to standardize decisions provides a contrast to the equally compelling need to deliver individualized, patient-specific treatment. Computer-based protocols, which are explicit, detailed, and patient data–driven, can simultaneously achieve standardization of clinical decision making and individualization of patient care.

4.2 Knowledge-based Approach

The translation of paper-based into computer-based guidelines can be done in at least two different ways: 1) in a knowledge-based approach an expert extracts information from the guideline text, interprets it, and then encodes it using one of the guideline models described in Section 5; 2) in the document-centric approach mark-up methodologies are used to provide guideline text excerpts relevant to the patient context.

The development of clinical guidelines is quite expensive; they should be shared among clinical institutions. Patel et al. [24] described the process of shared guideline development. First, a paper-based guideline has to be created at authoring institutions. Second, paper-based guidelines have to be translated into a computer-based guideline representation language (such as GLIF); GLIF is a computer-interpretable language for modeling and executing clinical practice guidelines. GLIF has a formal representation. It defines an ontology for representing guidelines, as well as a medical ontology for representing medical data and concepts. GLIF was developed by the InterMed Collaboratory (Stanford Medical Informatics, Harvard University, McGill University, and Columbia University). Third, the representation has to be implemented within the clinical institution's application system. Fourth, the clinician has to interpret the guideline as it is represented in the guideline applications.

Following the steps of Patel, more groups have translated paper-based guidelines into a computer-based system. Gillois et al. [25] successfully implemented two cardiovascular CPGs and a hypercholesterolemia guideline into a computer system. First, they chose the national evidence-based guideline. Second, they translated it into a linear algorithm. Third, they represented it in GLIF, and fourth, they connected it with an EPR. The system outcome was the advice whether to start treatment or not. Gillois et al. concluded that to improve guideline computability and efficiency it is necessary to use a framework for guideline development.

Guideline developers tend to focus on specific tasks rather than on processes such as care plans and pathways which are extended in time. In contrast, research on business process modeling has demonstrated notions and tools which deal directly with process modeling, but has not been concerned with problems like data interpretation and decision making. Fox et al. [26] described these two approaches (guideline development vs. business process modeling) and compared some of their strengths and weaknesses.

4.3 Methodological Issues

Below, we show the impact of users (4.3.1), expert involvement (4.3.2), parallel guideline development (4.3.3), and used formalisms on guideline development (4.3.4).

4.3.1 User Involvement

Trivedi et al. [27] pointed out that user satisfaction leads to increased use of the system. Increased use of the system leads to end-users that are more efficient. The knowledge of end-users influences their perception of usefulness and it helps to give physicians as much information as possible before implementation. If end-users are left out of the decision making process they will have no personal investment in the use of the new system. User involvement in implementation can improve reliance and utilization.

4.3.2 Experts Involvement

CPGs need to be developed by experts from different fields. Already Patel et al. [24] in 1998 concluded that guidelines developed by teams that include both clinicians and experts in computer-based representations are preferable to those developed by individuals of either type working alone. At the same time, Woolf [28] (1999) warns that guidelines created only by medical or other groups not concerned with financial aspects may promote costly interventions that are unaffordable or that cut into resources needed for more effective services. There has to be a well defined balance between these matters while developing clinical guidelines. Still, in 2006, Biondich et al. [29] observed that even though clinical guideline authors, health information technology (HIT) developers, standards development organizations, and information system implementers all work to improve the processes of healthcare, they have long functioned independently towards realizing these goals. The independence has led to clinical standards of care that often poorly align with the functional and technical HIT standards developed to realize them. Biondich et al. describe the shortcomings and inefficiencies inherent in the current process and proposed two solutions. The first one is the creation of a technical expert panel. Among this group are representatives from medical informatics, quality improvement experts, disease content experts, and health service researchers. The second solution is the creation of a dictionary that clearly defines all decision variables with written definitions.

4.3.3 Parallel Guideline Development

Patel et al. [24] studied collaboration problems that arise when translating a clinical guideline from text into an encoded form. Guidelines should neither contain too much nor too little information. If they do, it may lead to erroneous and even dangerous interpretations. The translation has to take these considerations into account. Different experts with different backgrounds and expertise may construct different mental models of the guideline. Collabora-

tion may help to develop a shared mental model. However, building a shared model through collaboration may bring conflicts as every individual brings different knowledge and experience to the task. Yet, the negotiation of conflict may be beneficial as experts contribute relevant knowledge as they explain and justify their individual interpretations.

Based on the guideline development principles by Shekelle et al. [30], Goud et al. [31] designed a parallel guideline development strategy, in which a multidisciplinary group of experienced and knowledgeable professionals cooperate in creating both a textual and a computer-interpretable guideline. This strategy of parallel guideline development and formalization was applied during the development of the Dutch Cardiac Rehabilitation Guidelines of 2004 and appeared to be successful. In this approach guideline authors and a guideline formalization team closely collaborated. The guideline developers focused on relevant scientific evidence while the guideline formalization team looked for the right approach and tools, needed to formalize the guideline. Parallel guideline development gives an opportunity to eliminate, in an early stage, vague and incomplete concepts and recommendations.

From the above cited publications we may conclude, as Peleg et al. [32] also did, that the collaboration between guideline authors and (medical) informatics researchers in the process of developing guidelines improves the logical consistency and completeness of clinical algorithms or flowcharts. Here we remark that each collaborator brings in its own expertise. Any problems encountered or errors identified during guideline formalization should be directly discussed and, if needed, resolved.

4.3.4 Guideline Formalisms

Existing guideline formalisms try to create computer-interpretable models of clinical guidelines without the need to encode the abstractions and recommendations in a guideline by means of a generic rule or a procedural language. Nevertheless, these formalisms still aim for generality in being able to encode all kinds of guidelines. Therefore, this formalism also has to use abstractions, such as actions, decisions, and criteria, as their primitives. Clinicians and knowledge engineers still have to conceptualize a guideline in these abstract terms. Research in the 1980s has shown the power of domain-specific models. For example, OPAL, a domain-specific tool for acquiring cancer chemotherapy protocols, is adequate for being used by clinicians [33]. Peleg and Tu [34] suggest to support the translation of narrative guidelines by providing computer-interpretable templates for representing guideline knowledge using clinical abstractions that are appropriate for particular guideline sub-domains. These templates (also called design patterns) provide domain experts with abstractions that are defined at the same level as that of their guidelines. Screening and immunization templates have the potential to be applied to encode any guidelines belonging to these categories. This was demonstrated by their ability to encode all of the screening guidelines from the National Guideline Clearinghouse that contained a clinical algorithm and all of the CDC immunization guidelines. In addition, knowledge modeled in the immunization patterns 1) can be checked for completeness and consistency, 2) could be used by applications that execute workflow processes or that compute the due date of the next vaccination, and 3) could be automatically translated into a computer-interpretable guideline.

4.4 Document-centric Approach

Below, we discuss the document-based approach that uses the languages HTML and XML, and allows clinicians to browse through documents in an efficient manner. We provide a brief overview of this development and give some examples of the approach.

The first step is using relevant tags in the mark-up process. The result can be translated in one of the languages of the guideline models. Shiffman et al. [16] describe such a document-centric approach. They use the Guideline Elements Model (GEM) as a computable representation for clinical guidelines. GEM is an XML-based document model that uses a multilevel hierarchy to store the heterogeneous kinds of information contained in clinical practice guidelines. The hierarchy contains more than 100 tags by which guideline information can be classified (marked up) and modeled at varying levels of abstraction. In the mark-up process the user selects text from the guideline and places this text in the appropriate position in the GEM hierarchy. In this way an XML file is produced. The text is specified further by removing unnecessary words, changing verb phrases from passive to active voice, etc. Also, abstract recommendations are made clear and ambiguity is removed.

In the second development step, the text is translated into computable statements. The results have been used to obtain a description in the Arden syntax. Moreover, the possibility of translating the mark-up results into GLIF3 (a guideline-modeling language) by defining elements that represent GLIF constructs is investigated. The XML representation format used by GEM could limit its potential impact, as semantic web ontology languages, such as OWL, are becoming major knowledge representation frameworks in medical informatics. Nam Tran et al. [35] present a faithful translation of GEM from XML into OWL.

Next to GEM we can find more tools that support guideline implementation. We present here three of them.

Votruba et al. developed the Guideline Markup Tool (GMT) that not only supports the translation of clinical guidelines into a guideline-modeling language such as Asbru but also maintains the connection between the original guideline and its formal representation [36].

The Stepper tool was developed to assist a knowledge engineer in developing a computable version of narrative guidelines. The system is document-centric: it formalizes the initial text in multiple user-definable steps. They correspond to interactive XML transformations [37].

Shahar et al. [38] introduced a Web-based, modular, distributed architecture, the Digital Electronic Guideline Library (DeGeL), which facilitates gradual conversion of clinical guidelines from text to a formal representation in a chosen target guideline ontology. Their guiding principle is that expert physicians should be trans-

forming free-text guidelines into intermediate, semantically meaningful representations, while knowledge engineers should be converting these intermediate representations into a formal executable representation. Shahar et al. demonstrated the feasibility of the architecture and the tools for several guideline ontologies, including Asbru and GEM.

4.5 The Bridging Approach

Having developed many guidelines and facing even more recommendations, we may investigate the relation between both. Shankar et al. [4] report that in a survey users expressed the need for a presentation of the text of the guideline together with the recommendations. However, both the knowledge-based approach and the document-centric approach have difficulties in integrating guideline text with decision support systems. Therefore, Shankar et al. proposed a bridging approach for integrating structured knowledge bases with marked-up text. They introduced clinical queries that might be posed by a clinician to gain better insight into guideline recommendations. The elements in the knowledge base are then linked to the appropriate clinical queries. Subsequently, an information retrieval technology is used to satisfy the queries by retrieving relevant excerpts of the guideline document.

4.6 Implementation in Decision Support Systems

In the course of our research, it is important to examine to what extent the guidelines and the implementation in decision support systems influence the users.

To adapt the clinicians' behavior, CPGs should provide adequate patient-specific decision support during patient encounters, which is reliable for and understandable by the clinicians. Guidelines implemented in a computer-interpretable guideline (CIG) formalism providing support based on patient data may achieve this goal. Overtime, we observed an increase of guidelines implemented in decision support systems (DSSs) [39–41], where the system helps users to follow the guideline. In a DSS CIGs are implemented in a formalized way. The DSS retrieves all necessary patient data from an EPR and generates patient-specific advice based on these patient data and the guideline content.

DSSs cover more than only guidelines. Alerts and reminders, however, can be considered as simple guidelines. What can be said on the reasons for using or not using DSSs in general is, in our opinion, also valid for guideline implementation systems. In a systematic review, Kawamoto et al. [42] identified four features of DSSs that were critical for improving clinical practice: 1) decision support is provided automatically as part of clinical workflow, 2) decision support is delivered at the same time and location of decision making, 3) actionable recommendations are provided, and 4) DSSs are computer-based. A common theme of all four features is that they make it easier for clinicians to use such a system: the system must minimize the effort required by clinicians to receive and act on system recommendations. The success of a DSS may concern the process or patient outcome or both. In a recent systematic review Mollon et al. [43] concluded that the majority of 41 reviewed studies concerning DSSs for prescription reported changes in health care provider behavior (in 25 of the 37 successfully implemented systems), but that only five studies showed a successful impact on patient outcomes. DSSs may also provide a better overview of existing, implemented guidelines.

In 2008, Wright and Sittig [44] reviewed the history of clinical decision support and traced the development of architectures for clinical decision support systems. They arrived at four distinct architectural phases for decision support: 1) stand-alone decision support systems, 2) integrated systems, 3) standard-based systems, and 4) service models. Furthermore they developed a model for evaluating architectures for clinical decision support that focuses on 1) defining a set of desirable features for a decision support architecture; 2) building a proof-of-concept prototype; 3) demonstrating that the architecture is useful by showing that it can be integrated with existing decision support systems and comparing its coverage to that of other architectures [45].

5. Computer-interpretable Guidelines

In this section we discuss the knowledge-based approach to the creation of CIGs. In Subsection 5.1 we describe the general guideline model and distinguish the underlying model and the language in which the guidelines are specified. In Subsection 5.2 we discuss the representation in combination with its implementations. In Subsection 5.3 we provide an overview of the problems encountered while designing (or using) the CIGs.

5.1 General Models

It is possible to define the functionality of CIG approaches in terms of two main characteristics: a) the underlying model, i.e., the framework of the logical components (5.1.1), and b) the language in which guidelines are specified (5.1.2). We conclude by providing several approaches (5.1.3).

5.1.1 Logical Components

The underlying model is the core characteristic of every guideline approach. It must be able to represent various kinds of guidelines that differ considerably in complexity and in level of abstraction. Two telling examples are nesting and decomposition. The model must contain a set of building blocks (primitives) used to construct guidelines, such as tasks, rules, nodes, and frames. Most approaches model guidelines in terms of a task-network model, a (hierarchical) model of the guideline control-flow as a network of specific tasks (such as decisions, actions or hierarchically decomposed sub-guidelines). The control-flow part of the model should allow tasks to be executed sequentially or in parallel, allow cyclical or iterative tasks and permit entry points into the guidelines.

5.1.2 Formal Language and Execution Engine

The guideline model should be supported by a formal language which specifies the guidelines in terms of the above-men-

tioned model primitives. Usually, such a language consists of two parts: a control-flow language and an expression language. The control-flow language specifies the guideline structure in terms of primitives of the model and their (temporal) relations, whereas the expression language usually describes decision criteria. GLIF for example used an expression language called GEL, based on the Arden syntax's logic grammar. Later, GELLO, an object-oriented expression language, was developed. GELLO is better suited for GLIF's object-oriented data model, is extensible, and allows implementation of expressions that are not supported by the Arden syntax.

Patel et al. [46] compared the newest version of GLIF (GLIF3) with the previous one (GLIF2). GLIF2 (the first published version of GLIF, GLIF1 is only known as experimental version) supported guideline modeling as a flowchart of structured steps which represented clinical actions and decisions. However, the attributes of these constructs were defined as text strings that could not be parsed, preventing the resulting guidelines from being able to make inferences during computerized execution. Patel et al. analyzed the process of encoding guidelines in those two versions. They concluded that the use of GLIF3 appears to offer several improvements over GLIF2 in the ability to encode guidelines accurately and efficiently. The increased formality of the GLIF3 model and syntax leads to a guideline-encoding process that contains both a greater level of detail and less ambiguity than those in the previous version. The result of the improvements is an encoded guideline which is better equipped to aid practitioners as they make decisions in a variety of clinical settings.

A formal language must be interpretable by automatic parsers. Preferably each approach should include a guideline execution engine incorporating such a parser that is able to provide decision support based on the encoded guidelines and patient data.

Wang et al. [47] describe the design and implementation of the GLIF3 guideline execution engine (GLEE). The paper gives a good overview of the internal structure of the execution engine. It explains how GLEE executes the various primitives of GLIF and illustrates the role of GEL, the expression language used to encode decision criteria and patient states.

5.1.3 Modeling Approaches

Examples of guideline-modeling approaches are: PROforma, GLIF, GUIDE, PRODIGY, EON, GASTON, and Asbru. Overviews of computer-interpretable guideline formalisms were presented by de Clercq et al. [48], by Isern and Moreno [49], and by Sonnenberg and Hagerty [50]. Because of the available reviews on guideline-modeling approaches, we will not cover this topic in depth. A good overview of methods for formalizing clinical practice guidelines can be found in Peleg et al. [51]. They compared six generic computer-interpretable guideline models to determine their commonalities and differences.

Wang et al. [52] reviewed eleven guideline representation models and determined the primitives which these models use to represent guidelines. They discern action, decision, patient-state, and execution-state primitives. These primitives need to be organized to form a specific process model. Such a model also defines scheduling constraints on the primitives and nesting of guidelines during guideline application. Scheduling constraints specify the temporal order in which the primitives can be executed during guideline application. Nesting of guidelines defines the hierarchical relationship among guidelines. Patient states can be used to record a patient's clinical status in a specific context of a guideline. These patient states can then be used as entry points into a guideline. In this way multistep clinical processes that may take place over several encounters can be modeled. With execution states, steps scheduled, e.g., by GLEE, can be distinguished from those actually executed according to a user's decision.

Seroussi et al. [53] described the framework used by the ASTI system to represent therapeutic strategies in such a way that any patient's therapeutic history can be located within a guideline concerning hypertension. Because chronic diseases as well as the patient's response to treatments evolve over time, new therapeutic decisions depend on decisions made and actions taken at previous consultations, as well as the outcomes of those actions. This makes it difficult to locate the position in the guideline where to continue. To make such a position easily accessible, the guideline knowledge is represented as a two-level decision tree, a clinical level describing theoretical clinical situations and a therapeutic level formalizing the different steps of corresponding recommended therapeutic strategies. The therapeutic level of the knowledge base represents the sequence of treatments recommended for each clinical situation described at the clinical level. We remark and regret that the framework is only used in the guiding mode of ASTI (see Bouaud [54]).

5.2 Representation and Implementation

There is a large difference between the information contained in published guidelines and the knowledge and information that are necessary to implement them [51]. Implementers use poorly specified, mainly tacit knowledge-acquisition processes. Thus, such an implementation approach results in considerable inconsistencies in the encoding of guideline knowledge and in the functionality of the systems that are created [16]. For example, Patel et al. [24] found, that in some cases different recommendations would be given to the same patient when using computable representations of the same guidelines implemented by different individuals.

A second problem is the lack of standardization of EPRs. There are many vendors offering EPR systems. However, there are no standards yet that EPR software providers have to be compliant with. Guideline systems have to map the terms used in the guideline to the corresponding terms used in the EPR systems. Here we seem to face an analogous problem as above. Yet, in the last decade considerable progress has been made and standardized approaches for guideline representation and sharing are central to these efforts [55]. Among them are rule-based, logic-based, and task-based approaches. We discuss knowledge representation in 5.2.1, sharing the knowledge in 5.2.2, and data representation in 5.2.3

5.2.1 Knowledge Representation

Below we discuss the rule- and task-based approach.

The Arden syntax is a rule-based formalism developed for encoding individual clinical rules as medical logic modules (MLMs). MLMs do not provide full support for conceptualizing a multistep guideline that unfolds over time.

A number of other developments share a hierarchical decomposition of guidelines into networks of component tasks that unfold over time: the task-network models (TNMs). Unlike rule-based systems, TNMs can explicitly model alternative pathways or sequences of tasks, and they provide tools for visual representation of plans and the organization of tasks within them. All TNM languages decompose guidelines into networks of component tasks and express various arrangements of these components and interrelationships between them. The modeling formats use different terminology to refer to various types of task networks (called plans by Peleg et al. [51]) and their components. Peleg et al. note that authoring CIG models can be time-consuming and may require clinical knowledge as well as technical skill. Translating guidelines encoded in one format into systems using other formats would reduce duplication of effort. The GLIF project originally intended to devise an interchange format to facilitate this process. However, these goals appeared to be impractical at the time.

5.2.2 Sharing the Knowledge

Many parties are developing computer-based guidelines as well as decision support systems that incorporate these guidelines. The resulting products show much redundancy and overlap and there is little standardization to facilitate sharing of technical modules or to enable adaptation to local practice settings. This situation limits the sharing of guidelines. In a computer environment, implementation involves a number of steps to translate the knowledge contained in guideline text into a computable format. There is the problem that the terminology used in CIGs is different from the terminology used in EPRs. Several publications report about approaches that target the above-mentioned problems in different ways. Below we discuss eight of them.

1) Boxwala et al. [17] discuss the problem of sharable CIGs. They mention three ways to share CIGs: i) translation of the formats into the proprietary format of the receiving system, ii) sharing of CIGs as decision-support services provided through standard application programming interfaces adopted by all systems, and iii) adoption of a common format for CIGs. The first option has limitations because it is not clear that one format always can be translated into another format without problems. The second option has been successfully pursued. But also here, the problem is that relevant CIGs may be specified in different formats that may not be interpretable by the execution engine of the decision support service. Therefore, Boxwala et al. prefer the third option to provide requirements for a format for sharable CIGs.

2) A model to share guidelines encoded in different formats at the execution level was developed by Wang et al. [56]. They extracted a set of generalized execution tasks from existing guideline representation models. Then mappings were created between specific guideline representation models and the set of generalized execution tasks. They also developed a generic task-scheduling model to harmonize the existing approaches to guideline scheduling. This Guideline Execution by Semantic Decomposition of Representation (GESDOR) model is able to execute guidelines encoded either in GLIF or in PROforma.

3) Tu et al. [57, 58] started the SAGE (Standards-based sharable Active Guideline Environment) project and created a service model approach. They observed that more than one formalism, for medical logic, is needed to accomplish sharing of computable medical knowledge. Lack of standards in terminologies and in data models for patient information require recoding of parts of the guideline. Reuse of a guideline knowledge base is possible once an infrastructure is in place that includes a medical record query interface, terminology mediation, and an act interface. SAGE places a standard application programming interface in front of a clinical system. This solves the vocabulary problem – the SAGE virtual medical record specifies the vocabularies that will be used to access and process the medical data. The SAGE system will respond to opportunities for decision support in the care process. SAGE does not require a detailed workflow model but uses an event-driven approach. Here we remark that many workflow contexts need to be modeled to recognize appropriate events that should trigger decision-support services. These contexts identify i) opportunities for providing decision support, ii) the roles and information needs of care providers, iii) events that may activate the guideline system, and iv) the guideline knowledge relevant in these contexts. The major innovation of the SAGE guideline model was its demonstration that heterogeneous information sources, such as patient data, order sets, and external knowledge sources can be integrated and used within encoded guideline knowledge bases. The clinical information system (CIS) events encoded in the guideline are registered by the event manager in the CIS, thereby expressing the execution engine's interest in these CIS events. When a relevant event is detected, the engine starts the execution. The SAGE execution engine is able to execute the guideline by interpreting the encoded content, obtaining current patient data from the CIS, and invoking functionality within the CIS to implement an action specified in the guideline. The engine interprets the content of the context, action, and decision nodes in an encoded guideline, executes workflow and decision logic, and interacts appropriately with the CIS. The event listener is the mechanism by which the engine is notified of state changes in the CIS. As part of conforming to the SAGE engine, the CIS implements the module that forwards events of interest to the event listener.

4) Fox et al. [59] noted the possibility to capture guidelines in a computer-interpretable form, opening up the capability of using the internet to deliver patient-specific advice and other services. They described a development lifecycle and technology for publishing and delivering services at the point of care (called publets) and discussed the quality requirements.

5) Kawamoto and Lobach proposed a service-oriented architecture (SOA) for DSSs. DSS capabilities were implemented through the orchestration of independent software services of which the interfaces were standardized. Core services included: a decision support service, a common terminology service, and a retrieve, locate, and update service [60]. They also developed a decision support web service to address the difficulty of sharing medical knowledge in a machine-executable format based on this SOA approach: SEBASTIAN (System for Evidence-Based Advice through Simultaneous Translation with an Intelligent Agent across a Network) [61].

6) To make computerized CPGs potentially accessible to a large number of general practitioners (GPs), the PRESGUID project [22] proposed an online service enabling physicians to consult computer-based CPGs. PRESGUID provides a web-based guideline system that takes as input clinical data on a particular patient and returns, as output, the customized recommendation. If the recommendations require prescribing drugs, the system will query the drug database and will display detailed information about the relevant specific medications.

7) Dominguez et al. [62] reported about the creation of a traceable clinical guideline application based on model-driven development techniques, using UML statecharts. The models could be shared and implemented using a plug-in. Their approach focused on models conforming to domain application metamodels, allowing code programs to be automatically generated from them by means of a refinement process. They also described the way in which the structure of the execution traces was automatically generated, represented not as a linear sequence but as a hierarchy of states, providing one straightforward and reliable way of obtaining complete storage structures of guideline applications. The complete framework has been implemented as an Eclipse plug-in named GBDSSGenerator. Different types of clinical guidelines were generated including laboratory guidelines. However, until now, each guideline is translated into a separate program.

8) Skonetzki et al. [63] proposed a framework to represent clinical guidelines, called HELEN. It was not their intention to supersede the variety of proposed and implemented approaches for formal representation of clinical guidelines. Instead they focused on management and implementation of specific topics to bring CPGs into clinical practice. The complexity of the authoring process was considered to be the real bottleneck. To achieve the goal of developing a flexible, shareable, and computable description of the CPG, they used ontologies created with the help of PROTÉGÉ [64]. They showed the benefits of combining different knowledge representations such as narrative text, graphic illustrations, and algorithms. They introduced a possible approach for an explicit adaptation process of documented and auditable CPGs in order to make them institution-specific and support sharing with other organizations.

5.2.3 Data Representation

For implementing guidelines in systems, data representation is an important aspect, of course next to knowledge representation. For instance, item names used in the guidelines are usually different from the item names used in EPRs. The representation of data should therefore be seriously considered in order to facilitate that guideline implementation systems can be easily integrated with EPR systems. Standardization is of utmost importance here. This was already acknowledged in the SAGE project described above.

Johnson et al. [65] discussed the use of a virtual medical record to solve the problem of idiosyncratic data models in the deployment environment. In this case, guidelines are encoded, assuming 1) a uniform virtual medical record and 2) guideline-specific concept ontologies. For implementing a guideline-based decision support system in multiple deployment environments Johnson et al. created mapping knowledge bases to link terms in the concept ontology with the terminology used in the deployment system.

5.3 Validation and Verification of Computer-interpretable Guidelines

When translating CPGs to CIGs inconsistencies, ambiguities, and incompleteness can be detected in the paper-based guideline. Therefore, an adequate translation by a guideline-modeling language can increase the quality of guidelines. This was, among others, reported by Marcos et al. [66]. They tried to answer the question whether formalization could improve the quality of medical protocols. Their line of reasoning was as follows. Protocols show diversity in formats, e.g., texts, flow diagrams, and tables. Making these protocols more precise with the help of a formal language may reveal anomalies in the guideline description. In their work, they translated two protocols into a computer-based language, Asbru. During protocol formalization into the Asbru language, they found a number of anomalies in the guideline, such as ambiguity, incompleteness, inconsistency, and redundancy. The language sometimes demands explicit elements which are not explicit in the original protocol. Sometimes Asbru required information that otherwise would go unnoticed. An example is the minimal and maximal delay necessary for the specification of the retry time in a cyclical plan. The authors concluded that 1) removal of anomalies increased the internal consistency of the protocol, 2) formalization is a good foundation for detection of anomalies, and 3) formalization of the protocol is intensive and costly. In Section 8 we mention a number of formal methods that can be used for testing whether CIGs conform to their specifications.

Formalization can easily spot protocol aspects such as the specificity and certainty of recommendations that are stressed by the AGREE appraisal instrument. The strength of formalization when compared with other methods lies in the use of precise notions.

6. Effects of Guidelines

Many groups are working on the development and implementation of guidelines. It is interesting to know whether guidelines

produce any significant changes. CPGs may have potential benefits (see (a) below) but also limitations (see (b) below) for patients, physicians, and the health care system (see Woolf et al. [28]).

a) Patients may profit from CPGs that summarize the benefits and harms of available options. Patients can make more informed healthcare choices taking their personal preferences into account. Also, services which were not offered to the patient earlier may become available as a response to newly released guidelines.

Moreover, the healthcare system could profit from improved efficiency (standardized procedures) which can lower the costs of services, drug prescriptions [67], or laboratory test requests [68, 69].

b) In contrast to the above, patients may also be harmed by clinical guidelines. Cost of health care may influence the design of new guidelines and patients' individual needs may not be the only priority in making recommendations. Groups developing clinical practice guidelines need to be aware that what is best for patients overall, as recommended in the guideline, may be inappropriate for individuals.

Below we discuss the effect of using guidelines, both CPGs and CIGs, on the process of care (6.1) and on the patient outcome (6.2).

6.1 Effect on the Process of Care

Many groups focus their work on finding out whether CPGs or CIGs do make a difference, and whether they harm or benefit the process of care. In 6.1.1 we review three cases described in the literature that deal with the above-mentioned issue and in 6.1.2 nine cases that deal with CIG.

6.1.1 Clinical Practice Guidelines

In a recent study on the effects of evidence-based CPGs on quality of care in the Netherlands, Lugtenberg et al. [70] stated that there are positive indications for the effectiveness of Dutch evidence-based CPGs on the structure and process of care. The size of the effects varied largely across recommendations within the guidelines.

Butzlaff et al. [71] investigated in a randomized controlled trial whether access to four web-based and evidence-based guidelines (not computer-interpretable) increased the knowledge of the general practitioners (GPs) using the guidelines. No statistically significant knowledge increase was observed, but the authors did observe a low usage of the guidelines in the intervention phase. They assumed that the number of implemented guidelines (only 4) and the rather short time of intervention (2.5 months) may have played a role. Physicians explained that their hesitation to use the web was caused by the concern that patient data in their computer unintentionally and illegally might become available to hackers during their on-line time. The second main reason for not using web-based guidelines mentioned by the physicians was "lack of time".

Jousimaa et al. [72] observed no difference in guideline use or in impact on decision making between an intervention group of newly qualified primary care physicians that had access to an electronic narrative version of guidelines and a control group of similar physicians having access to the paper version. In this study, textbook- and computer-based guidelines were frequently used and were the most common source of information. The authors observed that the implementation of computer-based guidelines may need more training and investment in computer hardware. However, once computers are readily available and routinely used within consultations, the computer-based version offers many advantages, such as easy updating, low production costs, possibility to include other databases and audiovisual material, the possibility of linking computer-based guidelines to decision support systems, and the ability to monitor the guideline use.

6.1.2 Computer-interpretable Guidelines

In their systematic review, Shiffman et al. [73] looked at the functionality and effectiveness of computer-based guideline systems. They assessed functionality and effectiveness of DSSs and checked which of the eight services (see below) from their information management services model were implemented [74]. This services model was hypothesized to contain the factors that would influence the success of guideline implementations. The model facilitates workflow integration. It considers eight factors: recommendation, documentation, explanation, presentation, registration, communication, calculation, and aggregation. They found that 14 of the 18 described cases mentioned improvement in guideline adherence, although they could not prove that the presence of the mentioned factors was predictive for the success of the studied systems. To evaluate adequately the effect of those factors on the success or failure of a computer-based guideline implementation, more of the confounding variables (such as different types of guidelines, different settings, and different system implementations) need to be controlled.

Carton et al. [69] reported about a study where, concurrently to the care, the computer provided personalized advice for a patient. The authors found that only seven common clinical situations were responsible for the majority of unnecessary radiological requests (70%). When compared to the control situation in which no recommendations were provided, a decrease of 6% of requests not conforming to the guideline was observed. Yet, the overall effect of the intervention concerning radiological requests appeared to be weak. It became visible that in these situations the theoretical knowledge of junior practitioners was inadequate and additional education could provide the solution.

Hunt et al. [75] reviewed studies on the use of DSSs in clinical settings, a subject related to guideline implementation. A total of 68 controlled trials met their inclusion criteria. Sixty-five studies assessed physician performance (in 66% a benefit was found) while 14 studies assessed patient outcomes (in six cases a benefit was found).

In a follow-up study six years later Garg et al. [76] found that 64% of the studies on physician performance reported that the DSSs increased performance and 13% of the studies regarding patient outcomes established a benefit of DSSs. They concluded that DSSs can considerably enhance, for example, clinical performance for drug dosing and preventive care, but not so

much for diagnosis. They also concluded that the practitioners' performance in many cases did improve, but that more research is needed to explain the effects of such systems on patient health. Garg et al. raise an important question: to what extent should DSSs be proven beneficial before mass deployment?

Ambresin et al. [77] built a study website for the management of febrile patients returning from the tropics. The site was built to provide medical diagnostic assistance to primary care physicians. Physicians can interact with the computer in real-time. The authors called the study website a success. The interest of physicians in tropical diseases and evidence-based medicine increased, and therefore their knowledge increased as well. There are still problems which they need to address. For example, the observation that only 50% of the recommendations were followed, might be caused by approaches that are inappropriate or not feasible. Thus, the system needs revision and an update of the guideline before being adopted by the end-user. If that is achieved, an improvement of the quality of care is expected.

In their work Overhage et al. [78] point out that physicians often fail to order tests or treatments needed to monitor/ameliorate the effects of other tests or treatments (corollary orders). They wanted to see whether automated, guideline-based reminders to physicians, provided when they wrote orders, could reduce these errors of omission. To test whether guideline-based reminders could help, they carried out their study in the inpatient general medicine wards (six independent services) of a teaching hospital. They identified 87 target orders that could be paired with one or more corollary orders. The rule-based reminder program analyzed the data in the EPR for the omission of any corollary orders and presented these to the intervention group physicians (belonging to three of the six services). They concluded that computer suggestions about corollary orders had a large effect on the adherence to the guidelines, especially when measured in terms of immediate or 24-hour compliance. The interventions increased adherence to many guidelines that were being promoted by the Pharmacy and Therapeutic Committee. Pharmacists had to call physicians to ask about drug-related interventions one third less often for study patients than for control patients. They did not observe any effects on patient outcomes, such as length of stay, serum creatinine, or charges. However, since the guidelines handled only 9.6% of the orders written during this study, it was not expected to occur.

Bouaud et al. [54] noted that onscreen reminders and alerts automatically provided by a DSS do not guarantee a positive impact on a physician's compliance with recommendations. They hypothesized that two situations can be distinguished in primary care that have a different impact on the way decision support will be used. On the one side there are cases that the GPs know how to solve although they may make mistakes. These are usually relatively easy cases for which reminder-based interaction to detect these mistakes is mandatory. On the other side there are hard cases for which GPs may seek up-to-date evidence to support decision making. In such cases GPs would deliberately use guidance systems (the available knowledge has to be browsed by the GP). The authors used ASTI, a prototype guideline-based DSS, to test their hypothesis. ASTI has two modalities: a critiquing mode and a guiding mode. In the critiquing mode recommendations are automatically generated based on the patient data. In the guiding mode the knowledge base is not automatically processed but can be read by the GP using browsing tools. Bouaud et al. concluded that the critiquing mode was more used for easy cases and that the guiding mode was more used for medium and high complexity cases (the hard cases).

In their work Goud et al. [79] observed a positive impact on the professional's adherence to the cardiac rehabilitation guidelines when using a DSS. CARDSS (Cardiac Rehabilitation Decision Support System) did improve users' familiarity with the guidelines' recommendations and decision logic, and reduced guideline complexity by supporting calculations and interpretations of data. Goud et al. also observed that if the system recommendations were presented to patients, refusal to participate in therapies reduced. They concluded that a computerized decision support system can be effective in improving a multidisciplinary team's guideline concordance [80].

Chan et al. [81] warned that the measurement of provider adherence to determine the effect of CIGs may provide a too pessimistic view with respect to patient outcome. Because many factors next to decision support determine patient morbidity and mortality, it would be difficult to measure accurately the effect of a DSS on such outcomes. Evaluation of provider adherence to a particular recommendation made by the DSS is the most direct indicator of the ability of a DSS to affect clinical decision making. It was used to evaluate the effect of CIGs. In this approach Chan et al. distinguished between strict adherence and adherence to the spirit of the guideline. They analyzed provider adherence to the ATHENA DSS recommendations. From the results presented we can see that the providers strictly adhere to, e.g., a substitution (of drugs) advice only in 1.2% of the cases, whereas adherence to the spirit of the guideline was much higher: 37.5%. Therefore, Chan et al. concluded that a strict evaluation of actions and recommendations had failed to describe in full detail the providers' adherence to the guideline. Any valid measure of adherence must consider the extent to which the provider strictly follows the guideline and the provider's more clinically relevant higher-level intention.

6.2 Effect on the Patient Outcome

It is interesting to investigate whether 1) guidelines can cause changes in the physicians' behavior and 2) improve the process of care, and simultaneously whether there is any effect on the outcome such as patient morbidity and mortality. We discuss those questions for CPG (6.2.1) and CIG (6.2.2), and we conclude the section (6.2.3).

6.2.1 Clinical Practice Guidelines

In their work, Lugtenberg et al. [70] observed changes in patient outcomes. Yet, they concluded that they were generally modest and did hold only for some of the outcomes that were studied. Lugtenberg et al. suggested that guideline implementa-

tion should focus more on individual recommendations than on the guideline as a whole.

Worrall et al. [82] assessed the evidence for the effectiveness of clinical practice guidelines in improving patient outcomes in primary care. They identified 13 trials that examined patient outcomes. Four of the studies followed nationally developed guidelines and nine followed locally developed ones. Six of the studies involved computer-based or automated reminder systems. Only five studies reported statistically significant results. However, the lack of evidence of the effectiveness of the guidelines may well be due to the lack of methodologically sound studies among the conducted ones. Also, many of the studies that were evaluated, even those that were methodologically sound, may have examined guidelines that were based on out-of-date consensus statements or may have been poorly implemented.

Eagle et al. [83] described a project that applied guidelines in 10 hospitals in Michigan, USA, and concluded that the application of guidelines improved the delivery of evidence-based care (mortality reduced at one year by 5.1%).

6.2.2 Computer-interpretable Guidelines

Heselmans et al. [84] analyzed in a systematic review the effectiveness of computer-interpretable guidelines implemented in ambulatory care settings. They concluded that patient outcomes were not widely studied. No evidence was found of an effect on patient outcome. They also found that only seven of seventeen studies that investigated process outcomes showed improvements in process of care variables compared with the usual care. No incremental effect was found of the electronic implementation over the distribution of paper versions of the guideline, neither for the patient outcomes nor for the process outcomes. They concluded that there is little evidence for the effectiveness of electronic guidelines. The results by Heselmans et al. were less positive than those of the other studies mentioned. They explained the difference by noting that in their study the more effective straightforward reminder systems were excluded. Also, the definition of a successful intervention may have played a role.

Bryan et al. [85] evaluated the types and effectiveness of electronic DSSs in the primary care setting. They concluded that a DSS has the potential to produce statistically significant improvements in outcome, although there is much variability among the types and methods of DSS implementation and resulting effectiveness.

6.3 Section Conclusions

By reviewing studies that focus on the effects of guidelines on the process of care, we observe that guidelines can positively influence the work, improve guideline adherence, and increase efficiency (e.g., test ordering).

We notice that there are much less studies that focus on the patient outcomes. We need to bear in mind that since many factors, next to decision support, determine patient morbidity and mortality, it is difficult to measure accurately the effect of a DSS on such outcomes.

7. Barriers to Guideline Adherence

From studies and reviews we may conclude that the adherence by physicians to guidelines is low both for paper-based and computer-based guidelines. However, the problems causing this noncompliance are different in both cases. We have guideline-related barriers (7.1) and other barriers (7.2).

In case of narrative guidelines, physicians' low adherence is conditioned, for example, by lack of awareness that the guideline exists, lack of time to look up the answer especially in urgent situations, or vagueness of recommendations. In case of computer-based guidelines the physicians' adherence is related to workflow integration, general attitude towards computer systems, and agreement with the content of the guideline that underlies the knowledge base.

We discuss these barriers and in 7.3 we concentrate on improvement. Finally, in 7.4 we deal with the question: has guideline development gone astray?

7.1 Guideline-related Barriers

Already Grimshaw et al. [86] concluded that clinical practice guidelines increase quality and outcome of care. In their review they studied 59 published evaluations of clinical practice guidelines. They point out that the successful introduction of clinical practice guidelines depends heavily on the used methods of development, dissemination, and implementation. Different methods will be appropriate in different contexts. Studies that report large improvements in clinical care suggest the potential of guidelines when development, dissemination, and implementation are all appropriate. Studies that report small improvements or none may reflect failure at any stage during the introduction or evaluation of the guidelines. They may also reveal barriers. Grimshaw et al. presented a classification highlighting the more effective strategies and suggested that the strategies most likely to be effective were those that were internally developed, where the dissemination strategy was a specific educational intervention and where patient-specific advice was delivered concurrent with care. Having these characteristics may help considerably to overcome the guideline-related barriers.

Over time, guidelines were disseminated in many forms. They are published in magazines and journals, textbooks, CD-ROMs, and now also on the web. While electronic dissemination has broadened the availability of guidelines and enables guidelines to be retrieved in clinical settings, most guidelines have typically been specified in a non-computer-interpretable narrative text or in non-executable flowchart formats. These non-computable formats limit the usability of the guideline since the knowledge contained in the guideline may not be easily accessible during the patient encounter. This is a main limitation. Further, extracting recommendations from a non-computable document and determining their relevance for a specific patient require additional effort from the care provider. The usability issues mentioned have been identified as factors impeding compliance with the guideline [87].

Shaneyfelt and Centor [7] point to another limitation of guidelines: they are

often focused too narrowly on single diseases and are not patient-focused. Patients seldom have single diseases and few guidelines help clinicians in managing complexity.

7.2 Other Barriers to Physician Adherence

Other barriers exist too. Guideline adherence among Dutch general practitioners, for example, is not optimal [88]. An analysis of barriers to implementation was carried out by focusing on key recommendations rather than on guidelines as a whole. The barriers varied largely within guidelines (each key recommendation had its own barrier pattern). The most perceived barriers were lack of agreement with the recommendations due to lack of applicability or lack of evidence (68% of key recommendations), environmental factors such as organizational constraints (52%), lack of knowledge regarding the guideline recommendations (46%), and guideline factors such as unclear or ambiguous guideline recommendations (43%). To obtain more insight into these barriers we briefly discuss five studies.

Morris [23] in his work mentions the following eight barriers to protocol use: 1) lack of appreciation of the limitation of human decision making and the small number of variables on which decisions depend; 2) exclusion of practitioners from the protocol development process; 3) tendency to focus on infrequent but possible clinical scenarios that are not accommodated by protocol logic; 4) use of guidelines primarily for purposes other than improving quality of care and patient outcomes (e.g., reducing costs); 5) hubris among clinicians defending their autonomy; 6) concern that protocol logic is not correct; 7) concern about a reduced role of clinicians in medical practice; 8) insufficient technological infrastructure (no functional electronic patient records).

Cabana et al. [87] concentrate their work on finding reasons why physicians do not follow clinical practice guidelines. They investigated 76 articles and report seven general categories of barriers to physicians' guideline adherence. The barriers concern the physician's knowledge (lack of awareness, lack of familiarity), attitudes (lack of agreement, lack of self-efficacy) and behavior (external barriers, guideline factors, environmental factors). However, they do not point out which barriers are the most common and need to be taken into account in the first place while creating clinical practice guidelines and building decision support systems. Also, they concentrate on the guideline as a whole, whereas Lugtenberg et al. [88] suggest focusing on individual key recommendations.

Trivedi et al. [27] also looked at barriers in physician's acceptance of computerized decision support systems. They divided those barriers into three groups: 1) human issues, 2) organizational issues, and 3) technical issues. They extended the list of known barriers by a few, such as: physicians' knowledge of and experience with computers, privacy concerns, the security of text notes, depersonalization of care, and losing eye contact with patients. Physicians should receive additional time and support to facilitate a structured data entry that improves quality and enhances effective clinical operations. Software needs to be flexible, e.g., is the program easy to log-on to and capable of providing for physician episodic sessions, commonly interrupted due to patient care. In accordance with Shiffman et al. [15] Trivedi et al. concluded that the importance of integrating DSS into the clinical activities and workflow is a prerequisite to overcome the other barriers.

Espeland and Baerheim [89] identified and described factors that general practitioners considered. They may affect their decisions about ordering plain radiography for back pain. Here, barriers to guideline adherence suggested by such factors may play a part. In addition to those barriers, four other barriers reported by Cabana et al. [87] were found: one concerning attitude (lack of expectancy that guideline adherence will lead to the desired health care process), one feeling-related (emotional difficulty with adherence), and two external (improper access to actual/alternative health care services, and pressure from other health care providers/organizations). Cabana et al. concluded that their findings may provide a better understanding of how GPs try to achieve acceptable solutions in the face of conflicting pressure and uncertainty. Their findings may help to implement spine radiography guidelines by reducing the multiple barriers to change. Their review indicated that change can be achieved by addressing at least two of the following three types of factors: 1) predisposing (knowledge, attitudes), 2) enabling (skills, resources, reduction of external barriers), and 3) reinforcing (reward through feedback).

Tehrani et al. [90] wrote an overview of different methodologies used in various intelligent DSSs for mechanical ventilation. They tried to find the reasons for the infrequent use of DSSs in mechanical ventilation, especially because according to the authors mechanical ventilators will become more advanced and there is an increasing number of DSSs for mechanical ventilation. They compared different rule- and model-based techniques. Tehrani et al. concluded that the infrequent use of DSSs in mechanical ventilation is caused by: 1) lack of accessibility, 2) the system is not immune to noise and erroneous data, 3) inadequate training to use the system, 4) lack of implementation in commercial ventilators.

7.3 How to Improve Guideline Adherence?

Several instances of implementation failure have been described. In many cases factors extrinsic to the guideline itself, e.g., organizational and provider-specific obstacles inherent in a particular system of care interfered with the implementation success. In other cases factors intrinsic to the guideline have contributed to implementation failure, e.g., ambiguity, inconsistency, and incompleteness. These three factors can in many cases be ameliorated or fully remedied by guideline authors while the guideline is being developed.

During the translation of a narrative clinical practice guideline into a CIG, errors can occur. Peleg et al. [32] started a study investigating the translation process from a narrative guideline to a clinical algorithm. They studied this process by looking at intermediate versions produced during algorithm creation. They identified the types of

errors introduced during the development process. The classification scheme proposed by Knuth was used to classify modifications between narrative guideline text and the clinical algorithm produced from it. Knuth classified discrepancies between the requirement document for TeX and the resulting software. The first 12 of the 15 modification types are also applicable to the narrative guideline domain. They recorded modifications made during several translation steps and recorded if it concerned a positive or negative change; whether it was an important one or not; and the location of the change in the paper-based guideline. They found that much of the information that was relevant for creating an algorithm had its origin in the medical background component of the guideline and not in the recommendation component. They concluded that when a clinical expert works alone to create an algorithm, contrary to the theory of "learning by doing", he is likely going to make more errors than positive modifications. Collaborative work is likely to result in much broader improvements to the algorithm. Finally, team work seems to be the best solution for detecting errors. They notify that a physician who is presented with a clinical guideline is going to interpret it based on his[a] experience. Thus building a shared model may bring conflicts as every individual brings different knowledge and experience to the task. Depending on the group which is creating the guideline the final results may be different. That is why guidelines which are supposed to be broadly used have to be more universal, easy to interpret, and flexible.

Ambiguity and vagueness in clinical prac tice guidelines reduce the likelihood of clinician adherence according to Codish et al. [10]. Use of ambiguous and vague terms hampers communication and leads to uncertainty and to variable interpretation. The authors propose a model for ambiguity and vagueness in guideline recommendations. Their goal was 1) to provide guidance to guideline authors to enable them to reduce inadvertent use of ambiguous and vague language, 2) to improve transparency when vague language is used on purpose, 3) to create a framework to develop tools to apply the model during authoring and implementation of clinical practice guidelines. They propose a model containing three axes: the first axis differentiates vagueness from ambiguity and classifies each of them, the second axis indicates the intent of the author using vague or ambiguous terms, and the third axis defines what recommendation component of the guideline is affected by ambiguity or vagueness. The model can be used to develop software tools that are to be employed during the authoring and implementing process to identify and classify vague terms, inform authors of numerical values associated with the term, display an ordinal scale of similar terms, and suggest alternatives.

Grol et al. [91] determined attributes of clinical practice guidelines that influenced the use of guidelines in decision making in clinical practice. They concluded that evidence-based recommendations are better followed in practice than recommendations not based on scientific evidence. Also, precise definitions of recommended performance improved the use of guidelines. Testing the feasibility and acceptability of clinical practice guidelines among the target group is important for effective implementation.

LaBresh [92] described a program for quality improvement of acute stroke care. The program focuses on three domains: 1) diagnosing barriers to the delivery of care, 2) the clinical care system redesign process, and 3) a collaborative model. They mention knowledge, attitudinal and behavior barriers. Knowledge of clinical trial results and guideline recommendations are a necessary prerequisite for delivering evidence-based care. Traditional physician education has focused on presenting this evidence. Too often, however, the dissemination of guidelines was assumed to be sufficient to produce high levels of adherence. Yet, even with the knowledge and intention to deliver evidence-based care, the performance still might not be optimal because of organizational barriers, in particular primarily lack of well-designed systems to ensure reliable delivery of desired care. According to LaBresh the development of highly reliable systems requires a team approach. They presented a collaborative model (a group of participating ("collaborating") health care delivery organizations studying a specific health care quality problem, designing and implementing specific solutions, evaluating and refining these solutions, and disseminating findings to the other organizations) for quality improvement.

Fieschi et al. [93] took a look at the development of DSSs over time. According to Fieschi et al., the factors that influence the adoption of DSSs in today's medicine are: 1) 'evidence-based' practice is 'disease'- rather than 'patient'-oriented, 2) clinical practice guidelines while used in daily practice need to guide the decision making rather than force the decision, 3) collaborative medical practice is the means of improving the quality, continuity, and coordination of care, 4) the possibility of shared decision making, where DSS helps with, e.g., individual risk assessment or drug interactions.

Clinicians' adherence to DSSs is related to workflow integration [73], a general attitude towards DSSs and agreement with the content of the guideline that underlies the knowledge base [80]. This means that decision support needs to be provided at the right time and the right location, and the content needs to be trusted. This is precisely why guideline development by multidisciplinary teams, where medical professionals and system developers work together, is so important [29, 31, 32].

7.4 Has Guideline Development Gone Astray?

The quality of CIGs is clearly dependent on the quality of CPGs. As we have seen Tricoci et al. [6] concluded that updated guidelines contain more recommendations but many of these are based on lower-level evidence. In the British Medical Journal (BMJ) recently the question was raised whether guideline development has gone astray. Gibbons et al. [94] answered this question negatively. They pointed out that guideline development in cardiovascular diseases is a well developed process that has enhanced the delivery of proved treatments and improved patient outcomes. They illustrated

[a] For brevity, we use "he" and "his" whenever "he" or "she" and "his" or "her" is meant.

their opinion with evidence proving that the implementation of practice guidelines led to positive changes in clinical practice.

In contrast, Grol [95] answered the question positively. He stated that guidelines need to be integrated with other quality improvement initiatives, such as performance measurement and quality improvement programs. Guideline developers usually have quite different aims and interests from those in the quality improvement world. Guideline developers may have close ties with the industry. Many guidelines even today do not meet the AGREE criteria. Audits around the world show that guidelines are on average used in only 60–70% of day-to-day decisions. The cost effectiveness of guidelines compared with other methods for improving patient care is unknown. According to Grol procedures must be changed to speed up the development, minimize personal bias in recommendations and involve patients more actively in the process of both developing and using guidelines. Without such changes, guideline development may increasingly be seen as an expensive but unhelpful and ineffective toy for a happy few.

Also, in the medical informatics community it became clear that clinical decision support (CDS) implementation and use have been problematic. As a result, relevant medical knowledge that should be brought to bear is not always available or used for many health care decisions. On the request of the Office of the National Coordinator for Health Information Technology in the US, AMIA, the American Medical Informatics Association, established the CDS Roadmap Development Steering Committee to develop a tactical plan to guide federal and private sector activities to advance CDS. The goal of the Roadmap is to realize the vision of a U.S. health care system in which optimal, usable and effective clinical decision support is widely available to providers, patients, and individuals where and when they need it to make health care decisions.

The resulting roadmap [96] identified three pillars for fully realizing the promise of CDS: 1) make the best knowledge available when needed, 2) ensure high adoption and effective use of CDS, and 3) continually improve knowledge and CDS methods. Each pillar comprises two strategic objectives. The best knowledge can be made available when needed when 1a) clinical knowledge and CDS interventions are represented in standardized formats, and when 1b) clinical knowledge and CDS interventions are collected, organized, and distributed in one or more services from which users can readily find the specific material they need. High adoption and effective use can be reached when 2a) policy, legal, and financial barriers are addressed, and when 2b) clinical adoption and usage of CDS interventions are improved in such a way that they are easy to deploy and use by identifying and disseminating best practices for CDS deployment. Continuous improvement of knowledge and CDS methods can be obtained by 3a) assessing and refining the national experience with DSSs, and by 3b) advancing care-guiding knowledge by fully leveraging the data available in interoperable EHRs to enhance clinical knowledge and improve health management.

8. Discussion and Conclusions

Recent reviews show that decision support systems that issue alarms or reminders lead to a better care process. But still, it is not confirmed that the use of guideline systems that implement complex guidelines 1) improves the quality of healthcare, 2) lowers the medical costs of treatments, and 3) reduces practice variability. Below we provide our conclusions on: formalization of CIG (8.1), implementation (8.2), and effects (8.3), and add some discussion where appropriate.

8.1 Formalization of Computer-interpretable Guidelines

Formalization has many advantages but does not always work. We provide four conclusions based on our findings in the literature.

First, most guideline modeling approaches have facilities for testing that the model is unambiguous and syntactically correct [55].

Second, Marcos et al. [97] reversed the role of critiquing as used in some decision support systems. In such decision support systems the user solves a problem and the decision support system provides comments aimed at improving a suboptimal solution of the user. The critiquing approach assumes that the guideline is correctly implemented and that the contents of the guideline are error-free. The critiquing procedure can also be used to test the output of the system and thereby the implemented guideline by comparing the actions proposed by the system to those of an expert. Marcos et al. compared the solutions of an expert physician to the recommendations provided by Asbru and analyzed the differences to improve the guideline. Marcos et al. tried manually to match the actions of the expert physician with those produced by Asbru. However, the expert's actions did not always appear in the protocol as plans and therefore there is no 1:1 correspondence between an expert's actions on the one hand and protocol plans and corresponding intentions on the other. The direct use of intentions for critiquing, either for matching of actions or for studying their appropriateness, was also not possible. This was a problem for the vocabulary, and frequently led to differences in the degree of detail, abstraction level, etc. Marcos et al. therefore concluded that using the critiquing approach for guideline improvement is not promising.

Third, an advantage of software systems is that formal methods can be used during their development to test whether they conform to specifications. Formal methods are mathematics-based techniques for the specification and verification of software and hardware systems. As medical errors have far-reaching consequences, there is a good reason to investigate the usefulness of formal systems in medicine. Errors in guidelines may contribute to medical errors and mistakes. This may lead to unexpected harm to patients and to a low compliance of the use of guidelines. So, there is a growing interest to apply these methods carefully in the area of medical guidelines [98].

Fourth, given a specification that describes what the system should do, the formal verification is the act of proving or dis-

proving the correctness of an implementation with respect to the specification in a mathematically rigorous way. Like mathematical instruments, formal methods can be used in a number of ways for guidelines. We mention three of them. 1) The guideline can be considered as a system that is being developed. Verification then involves checking whether this guideline adheres to certain correctness or quality criteria [99]. 2) National guidelines could be considered as the golden standard when developing a local protocol [100]. 3) Verification can also be used to identify possibly undesired behaviors (non-compliance) as well as to analyze and comprehend features and characteristics of the real process that are not properly expressed in the modeled process [101]. An application of this type of verification (e.g., model checking) is the following. Groot et al. [102] employ model checking to investigate whether a part of the actual treatment is consistent with the guideline. They proposed a computational method for such critiquing, where the ideal actions were given by a formal model of a clinical guideline, and where the actual actions were derived from real world patient data. Moreover, they showed how critiquing can be cast in terms of temporal logic, and what can be achieved by using model checking. The method has been applied to a clinical guideline of breast cancer in conjunction with breast cancer patient data.

8.2 Implementation

Guideline development and implementation is time-consuming and expensive. Developing high-quality guidelines requires a sufficiently skilled team of people and sufficient budget. Guideline development panels should include members with different expertise, including medical informatics. Our main conclusion here is that sharing of CIGs is a critical requirement for guideline development, dissemination, and implementation. The use of a universal standard for guideline representation to encode all guidelines would be a solution. However, since no existing guideline representation model is dominant over the others this approach is impractical today. Since there is much overlap between the primitives of the various models, a universal representation may be possible in the future. Moreover, the GESDOR approach showed that guidelines encoded in different languages could be executed by a generalized execution engine. So this engine could be used in order to be able to use guidelines represented in different models. Also, service-oriented architectures may support the prompt execution of guidelines. But here standardization of the interfaces as well as of the data is at most important.

A paper-based and a semantically similar computer-based guideline should be designed simultaneously by groups of medical experts and knowledge engineers. This will make the step of the interpretation of a paper guideline superfluous. Because governmental agencies have larger resources than professional organizations and specialist societies and produce higher-quality guidelines [12], they should take the lead in developing high-quality CIGs.

Guideline implementation systems obtain their data from EPRs. A standardized terminology should be imposed so that the terminology used in the guideline matches the terminology used in EPR systems. Service-oriented approaches specifically take these problems into account.

8.3 Effects

Guideline adherence depends on a variety of factors, some related to guidelines themselves, some related to users, and some to the implementation context. Among the former are guideline quality, purpose, and implementation modality. Among the user-related factors are attitude to behavioral changes, authority interventions to foster adherence and eventually the type of users (general practitioners, hospital professionals, home caregivers, patients, etc.). Context is also crucial because organizational issues, such as lack of resources, can hamper guideline implementation and sometimes the original guideline intention is overridden by the guideline adaptation to a certain setting. Quaglini [103] discussed these factors and highlighted them by presenting a number of case studies. She stressed the importance of determining the motivation for non-compliance. She presented a classification system for non-compliance that is used in RoMa, an ancillary tool for non-compliance-management, coupled with a decision support system for stroke patients.

We end with two conclusions on effects.

1. By reviewing studies that focus on effects of guidelines on process of care, we may conclude that guidelines can positively influence the work, improve guideline adherence, and increase efficiency (e.g., test ordering).
2. We conclude that there are much less studies that focus on the effects on outcomes. We need to bear in mind that since many factors, next to decision support, determine patient morbidity and mortality it is difficult to measure accurately the effect of a DSS on such outcomes.

As it is clear from this paper, medical informatics researchers can play an important role in supporting the process of guideline development, dissemination, and implementation.

Acknowledgments

This work is financial sponsored by NWO, Token 2000 project MIA, grant number 634.000.021 and CAPHRI – School for Public Health and Primary Care.

References

1. Field M, Lohr K (eds). Guidelines for clinical practice: from development to use. Institute of Medicine, Washington, DC: National Academy Press; 1992.
2. Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. JAMA 1999; 281 (20): 1900–1905.
3. Clinical Practice Guidelines (cited 01.05.2009). Available from: http://www.openclinical.org/guidelines.html
4. Shankar RD, Tu SW, Martins SB, Fagan LM, Goldstein MK, Musen MA. Integration of textual guideline documents with formal guideline knowledge bases. Proc AMIA Symp 2001. pp 617–621.
5. American Heart Association (cited 15.10.2009). Available from: http://www.americanheart.org/downloadable/heart/1279224109481Methodology_Manual%206.2010.pdf

6. Tricoci P, Allen JM, Kramer JM, Califf RM, Smith SC, Jr. Scientific evidence underlying the ACC/AHA clinical practice guidelines. JAMA 2009; 301 (8): 831–841.
7. Shaneyfelt TM, Centor RM. Reassessment of clinical practice guidelines: go gently into that good night. JAMA 2009; 301 (8): 868–869.
8. Patel VL, Arocha JF, Diermeier M, How J, Mottur-Pilson C. Cognitive psychological studies of representation and use of clinical practice guidelines. Int J Med Inform 2001; 63 (3): 147–167.
9. Cook DJ, Greengold NL, Ellrodt AG, Weingarten SR. The Relation between Systematic Reviews and Practice Guidelines. Ann Intern Med 1997; 127: 210–216.
10. Codish S, Shiffman RN. A model of ambiguity and vagueness in clinical practice guideline recommendations. AMIA Annu Symp Proc 2005. pp 146–150.
11. Littlejohns P, Cluzeau F. Guidelines for evaluation. Fam Pract 2000; 17 (Suppl 1): S3–6.
12. Burgers JS, Cluzeau FA, Hanna SE, Hunt C, Grol R. Characteristics of high-quality guidelines: evaluation of 86 clinical guidelines developed in ten European countries and Canada. Int J Technol Assess Health Care 2003; 19 (1): 148–157.
13. Shiffman RN, Dixon J, Brandt C, Essaihi A, Hsiao A, Michel G, et al. The GuideLine Implementability Appraisal (GLIA): development of an instrument to identify obstacles to guideline implementation. BMC Med Inform Decis Mak 2005; 5: 23.
14. Appraisal of Guidelines for Research and Evaluation (cited 16.12.2009). Available from: www.agreecollaboration.org
15. Shiffman RN, Shekelle P, Overhage JM, Slutsky J, Grimshaw J, Deshpande AM. Standardized reporting of clinical practice guidelines: a proposal from the Conference on Guideline Standardization. Ann Intern Med 2003; 139 (6): 493–498.
16. Shiffman RN, Michel G, Essaihi A, Thornquist E. Bridging the guideline implementation gap: a systematic, document-centered approach to guideline implementation. J Am Med Inform Assoc 2004; 11 (5): 418–426.
17. Boxwala AA, Tu S, Peleg M, Zeng Q, Ogunyemi O, Greenes RA, et al. Toward a representation format for sharable clinical guidelines. J Biomed Inform 2001; 34 (3): 157–169.
18. Grimshaw J, Eccles M, Thomas R, MacLennan G, Ramsay C, Fraser C, et al. Toward evidence-based quality improvement. Evidence (and its limitations) of the effectiveness of guideline dissemination and implementation strategies 1966–1998.J Gen Intern Med 2006; 21 (Suppl 2): S14–20.
19. Prior M, Guerin M, Grimmer-Somers K. The effectiveness of clinical guideline implementation strategies – a synthesis of systematic review findings. J Eval Clin Pract 2008; 14 (5): 888–897.
20. Grol R, Zwaard A, Mokkink H, Dalhuijsen J, Casparie A. Dissemination of guidelines: which sources do physicians use in order to be informed? Int J Qual Health Care 1998; 10 (2): 135–140.
21. Lin KW, Slawson DC. Identifying and using good practice guidelines. Am Fam Physician 2009; 80 (1): 67–70.
22. Dufour JC, Fieschi D, Fieschi M. Coupling computer-interpretable guidelines with a drug-database through a web-based system – The PRESGUID project. BMC Med Inform Decis Mak 2004; 4: 2.
23. Morris AH. Developing and implementing computerized protocols for standardization of clinical decisions. Ann Intern Med 2000; 132 (5): 373–383.
24. Patel VL, Allen VG, Arocha JF, Shortliffe EH. Representing clinical guidelines in GLIF: individual and collaborative expertise. J Am Med Inform Assoc 1998; 5 (5): 467–483.
25. Gillois P, Chatellier G, Jaulent MC, Colombet I, Fieschi M, Degoulet P. From paper-based to electronic guidelines: application to French guidelines. Medinfo 2001; 10 (Pt 1): 196–200.
26. Fox J, Black E, Chronakis I, Dunlop R, Patkar V, South M, et al. From guidelines to careflows: modelling and supporting complex clinical processes. Stud Health Technol Inform 2008; 139: 44–62.
27. Trivedi MH, Kern JK, Marcee A, Grannemann B, Kleiber B, Bettinger T, et al. Development and implementation of computerized clinical guidelines: barriers and solutions. Methods Inf Med 2002; 41 (5): 435–442.
28. Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. BMJ 1999; 318 (7182): 527–530.
29. Biondich PG, Downs SM, Carroll AE, Shiffman RN, McDonald CJ. Collaboration between the medical informatics community and guideline authors: fostering HIT standard development that matters. AMIA Annu Symp Proc 2006. pp 36–40.
30. Shekelle PG, Woolf SH, Eccles M, Grimshaw J. Clinical guidelines: developing guidelines. BMJ 1999; 318 (7183): 593–596.
31. Goud R, Hasman A, Strijbis AM, Peek N. A parallel guideline development and formalization strategy to improve the quality of clinical practice guidelines. Int J Med Inform 2009; 78 (8): 513–520.
32. Peleg M, Gutnik LA, Snow V, Patel VL. Interpreting procedures from descriptive guidelines. J Biomed Inform 2006; 39 (2): 184–195.
33. Musen MA, Fagan LM, Combs DM, Shortliffe EH. Use of a domain model to drive an interactive knowledge-editing tool. Int J Man-Mach Stud 1987; 26 (1): 105–121.
34. Peleg M, Tu SW. Design patterns for clinical guidelines. Artif Intell Med 2009; 47 (1): 1–24.
35. Tran N, Michel G, Krauthammer M, Shiffman RN. Embedding the Guideline Elements Model in Web Ontology Language. AMIA Annu Symp Proc 2009. pp 645–649.
36. Votruba P, Miksch S, Seyfang A, Kosara R. Tracing the formalization steps of textual guidelines. Stud Health Technol Inform 2004; 101: 172–176.
37. Svatek V, Ruzicka M. Step-by-step mark-up of medical guideline documents. Int J Med Inform 2003; 70 (2–3): 329–335.
38. Shahar Y, Young O, Shalom E, Galperin M, Mayaffit A, Moskovitch R, et al. A framework for a distributed, hybrid, multiple-ontology clinical-guideline library, and automated guideline-support tools. J Biomed Inform 2004; 37 (5): 325–344.
39. Leslie SJ, Hartswood M, Meurig C, McKee SP, Slack R, Procter R, et al. Clinical decision support software for management of chronic heart failure: development and evaluation. Comput Biol Med 2006; 36 (5): 495–506.
40. Goldstein MK, Hoffman BB, Coleman RW, Tu SW, Shankar RD, O'Connor M, et al. Patient safety in guideline-based decision support for hypertension management: ATHENA DSS. Proc AMIA Symp 2001. pp 214–218.
41. Vogelzang M, Zijlstra F, Nijsten MW. Design and implementation of GRIP: a computerized glucose control system at a surgical intensive care unit. BMC Med Inform Decis Mak 2005; 5: 38.
42. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ 2005; 330 (7494): 765.
43. Mollon B, Chong J, Jr., Holbrook AM, Sung M, Thabane L, Foster G. Features predicting the success of computerized decision support for prescribing: a systematic review of randomized controlled trials. BMC Med Inform Decis Mak 2009; 9: 11.
44. Wright A, Sittig DF. A four-phase model of the evolution of clinical decision support architectures. Int J Med Inform 2008; 77 (10): 641–649.
45. Wright A, Sittig DF. A framework and model for evaluating clinical decision support architectures. J Biomed Inform 2008; 41 (6): 982–990.
46. Patel VL, Branch T, Wang D, Peleg M, Boxwala A. Analysis of the process of encoding guidelines: a comparison of GLIF2 and GLIF3. Methods Inf Med 2002; 41 (2): 105–113.
47. Wang D, Peleg M, Tu SW, Boxwala AA, Ogunyemi O, Zeng Q, et al. Design and implementation of the GLIF3 guideline execution engine. J Biomed Inform 2004; 37 (5): 305–318.
48. De Clercq P, Kaiser K, Hasman A. Computer-interpretable Guideline Formalisms. Stud Health Technol Inform 2008; 139: 22–43.
49. Isern D, Moreno A. Computer-based execution of clinical guidelines: a review. Int J Med Inform 2008; 77 (12): 787–808.
50. Sonnenberg FA, Hagerty CG. Computer-interpretable clinical practice guidelines. Where are we and where are we going? Yearb Med Inform 2006. pp 145–158.
51. Peleg M, Tu S, Bury J, Ciccarese P, Fox J, Greenes RA, et al. Comparing computer-interpretable guideline models: a case-study approach. J Am Med Inform Assoc 2003; 10 (1): 52–68.
52. Wang D, Peleg M, Tu SW, Boxwala AA, Greenes RA, Patel VL, et al. Representation primitives, process models and patient data in computer-interpretable clinical practice guidelines: a literature review of guideline representation models. Int J Med Inform 2002; 68 (1–3): 59–70.
53. Seroussi B, Bouaud J, Chatellier G. Guideline-based modeling of therapeutic strategies in the special case of chronic diseases. Int J Med Inform 2005; 74 (2–4): 89–99.
54. Bouaud J, Seroussi B, Falcoff H, Venot A. Design factors for success or failure of guideline-based decision support systems: an hypothesis involving case complexity. AMIA Annu Symp Proc 2006. pp 71–75.
55. de Clercq PA, Blom JA, Korsten HH, Hasman A. Approaches for creating computer-interpretable guidelines that facilitate decision support. Artif Intell Med 2004; 31 (1): 1–27.
56. Wang D, Peleg M, Bu D, Cantor M, Landesberg G, Lunenfeld E, et al. GESDOR – a generic execution model for sharing of computer-interpretable clinical practice guidelines. AMIA Annu Symp Proc 2003. pp 694–698.

57. Tu SW, Campbell J, Musen MA. The SAGE guideline modeling: motivation and methodology. Stud Health Technol Inform 2004; 101: 167–171.
58. Tu SW, Campbell JR, Glasgow J, Nyman MA, McClure R, McClay J, et al. The SAGE Guideline Model: achievements and overview. J Am Med Inform Assoc 2007; 14 (5): 589–598.
59. Fox JP, Bury J, Humber M, Rahmanzadeh A, Thomson R. Publets: clinical judgement on the web. Proc AMIA Symp 2001. pp 179–183.
60. Kawamoto K, Lobach DF. Proposal for fulfilling strategic objectives of the U.S. Roadmap for national action on clinical decision support through a service-oriented architecture leveraging HL7 services. J Am Med Inform Assoc 2007; 14 (2): 146–155.
61. Kawamoto K, Lobach DF. Design, implementation, use, and preliminary evaluation of SEBASTIAN, a standards-based Web service for clinical decision support. AMIA Annu Symp Proc 2005. pp 380–384.
62. Dominguez E, Perez B, Zapata M. Towards a Traceable Clinical Guidelines Application. Methods Inf Med 2010; 49: 571–580 (this issue).
63. Skonetzki S, Gausepohl HJ, van der Haak M, Knaebel S, Linderkamp O, Wetter T. HELEN, a modular framework for representing and implementing clinical practice guidelines. Methods Inf Med 2004; 43 (4): 413–426.
64. Musen MA, Eriksson H, Gennari JH, Tu SW, Puerta AR. PROTEGE-II: a suite of tools for development of intelligent systems from reusable components. Proc Annu Symp Comput Appl Med Care 1994. p 1065.
65. Johnson PD, Tu SW, Musen MA, Purves I. A virtual medical record for guideline-based decision support. Proc AMIA Symp 2001. pp 294–298.
66. Marcos M, Roomans H, ten Teije A, van Harmelen F. Improving medical protocols through formalization: a case study. In: Proc of the 6th Int Conf on Integrated Design and Process Technology (IDPT-02) 2002.
67. Martens JD, Werkhoven MJ, Severens JL, Winkens RA. Effects of a behaviour independent financial incentive on prescribing behaviour of general practitioners. J Eval Clin Pract 2007; 3: 369–373.
68. Bindels R, Hasman A, Kester AD, Talmon JL, De Clercq PA, Winkens RA. The efficacy of an automated feedback system for general practitioners. Inform Prim Care 2003; 11 (2): 69–74.
69. Carton M, Auvert B, Guerini H, Boulard JC, Heautot JF, Landre MF, et al. Assessment of radiological referral practice and effect of computer-based guidelines on radiological requests in two emergency departments. Clin Radiol 2002; 57 (2): 123–128.
70. Lugtenberg M, Burgers JS, Westert GP. Effects of evidence-based clinical practice guidelines on quality of care: a systematic review. Qual Saf Health Care 2009; 18 (5): 385–392.
71. Butzlaff M, Vollmar HC, Floer B, Koneczny N, Isfort J, Lange S. Learning with computerized guidelines in general practice? A randomized controlled trial. Fam Pract 2004; 21 (2): 183–188.
72. Jousimaa J, Makela M, Kunnamo I, MacLennan G, Grimshaw JM. Primary care guidelines on consultation practices: the effectiveness of computerized versus paper-based versions. A cluster randomized controlled trial among newly qualified primary care physicians. Int J Technol Assess Health Care 2002; 18 (3): 586–596.
73. Shiffman RN, Liaw Y, Brandt CA, Corb GJ. Computer-based guideline implementation systems: a systematic review of functionality and effectiveness. J Am Med Inform Assoc 1999; 6 (2): 104–114.
74. Shiffman RN, Brandt CA, Liaw Y, Corb GJ. A design model for computer-based guideline implementation based on information management services. J Am Med Inform Assoc 1999; 6 (2): 99–103.
75. Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. JAMA 1998; 280 (15): 1339–1346.
76. Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA 2005; 293 (10): 1223–1238.
77. Ambresin AE, D'Acremont V, Mueller Y, Martin O, Burnand B, Genton B. www.fevertravel.ch: an online study prototype to evaluate the safety and feasibility of computerized guidelines for fever in returning travellers and migrants. Comput Methods Programs Biomed 2007; 85 (1): 19–31.
78. Overhage JM, Tierney WM, Zhou XH, McDonald CJ. A randomized trial of "corollary orders" to prevent errors of omission. J Am Med Inform Assoc 1997; 4 (5): 364–375.
79. Goud R, Engen-Verheul Mv, Keizer Nd, Bal R, Hasman A, Hellemans IM, et al. The effects of computerized decision support on barriers to guideline implementation: A quality study in outpatient cardiac rehabilitation. Int JMed Inform 2010; 79 (6). 430–437.
80. Goud R, Jaspers MW, Hasman A, Peek N. Subjective usability of the CARDSS guideline-based decision support system. Stud Health Technol Inform 2008; 136: 193–198.
81. Chan AS, Coleman RW, Martins SB, Advani A, Musen MA, Bosworth HB, et al. Evaluating provider adherence in a trial of a guideline-based decision support system for hypertension. Medinfo 2004; 11 (Pt 1): 125–129.
82. Worrall G, Chaulk P, Freake D. The effects of clinical practice guidelines on patient outcomes in primary care: a systematic review. CMAJ 1997; 156 (12): 1705–1712.
83. Eagle KA, Montoye CK, Riba AL, DeFranco AC, Parrish R, Skorcz S, et al. Guideline-based standardized care is associated with substantially lower mortality in medicare patients with acute myocardial infarction: the American College of Cardiology's Guidelines Applied in Practice (GAP) Projects in Michigan. J Am Coll Cardiol 2005; 46 (7): 1242–1248.
84. Heselmans A, Van de Velde S, Donceel P, Aertgeerts B, Ramaekers D. Effectiveness of electronic guideline-based implementation systems in ambulatory care settings – a systematic review. Implement Sci 2009; 4 (1): 82.
85. Bryan C, Boren S. The use and effectiveness of electronic clinical decision support tools in the ambulatory/primary care setting: a systematic review of the literature. Inform Prim Care 2008; 16 (2): 79–91 (13).
86. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. Lancet 1993; 342 (8883): 1317–1322.
87. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA 1999; 282 (15): 1458–1465.
88. Lugtenberg M, Zegers-van Schaick JM, Westert GP, Burgers JS. Why don't physicians adhere to guideline recommendations in practice? An analysis of barriers among Dutch general practitioners. Implement Sci 2009; 4: 54.
89. Espeland A, Baerheim A. Factors affecting general practitioners' decisions about plain radiography for back pain: implications for classification of guideline barriers – a qualitative study. BMC Health Serv Res 2003; 3 (1): 8.
90. Tehrani FT, Roum JH. Intelligent decision support systems for mechanical ventilation. Artif Intell Med 2008; 44 (3): 171–182.
91. Grol R, Dalhuijsen J, Thomas S, Veld C, Rutten G, Mokkink H. Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. BMJ 1998; 317 (7162): 858–861.
92. LaBresh KA. Quality of acute stroke care improvement framework for the Paul Coverdell National Acute Stroke Registry: facilitating policy and system change at the hospital level. Am J Prev Med 2006; 31 (6 Suppl 2): S246–250.
93. Fieschi M, Dufour JC, Staccini P, Gouvernet J, Bouhaddou O. Medical decision support systems: old dilemmas and new paradigms? Methods Inf Med 2003; 42 (3): 190–198.
94. Gibbons RJ, Antman EM, Smith SC. Has guideline development gone astray? No. BMJ 340: c343.
95. Grol R. Has guideline development gone astray? Yes. BMJ 340: c306.
96. Osheroff JA, Teich JM, Middleton B, Steen EB, Wright A, Detmer DE. A roadmap for national action on clinical decision support. J Am Med Inform Assoc 2007; 14 (2): 141–145.
97. Marcos M, Berger G, van Harmelen F, ten Teije A, Roomans H, Miksch S. Using critiquing for improving medical protocols: harder than it seems. In: Proc of the 8th European Conference on Artificial Intelligence in Medicine (AIME-2001) 2001. pp 431–441.
98. Hommersom A, Groot P, Balser M, Lucas P. Formal methods for verification of clinical practice guidelines. Stud Health Technol Inform 2008; 139: 63–80.
99. Schmitt J, Balser M, Reif W. Verification of medical guidelines in KIV. Stud Health Technol Inform 2008; 139: 253–262.
100. Groot P, Hommersom A, Lucas P. Adaptation of clinical practice guidelines. Stud Health Technol Inform 2008; 139: 121–139.
101. Chesani F, Lamma E, Mello P, Montali M, Storari S, Baldazzi P, et al. Compliance checking of cancer-screening. Careflows: an approach based on Computational Logic. Stud Health Technol Inform 2008; 139: 183–192.
102. Groot P, Hommersom A, Lucas PJ, Merk RJ, Ten Teije A, van Harmelen F, et al. Using model checking for critiquing based on clinical guidelines. Artif Intell Med 2008.
103. Quaglini S. Compliance with clinical practice guidelines. Stud Health Technol Inform 2008; 139: 160–179.

