

QUALITY OF CLINICAL PRACTICAL GUIDELINES



Jako Burdaers

QUALITY OF CLINICAL PRACTICE GUIDELINES

The studies presented in this thesis were performed at the Centre for Quality of Care Research of the University of Nijmegen and the University of Maastricht, which participates in the Netherlands School of Primary Care Research (CaRe), acknowledged in 1995 by the Royal Dutch Academy of Science (KNAW).

The studies were performed in collaboration with the Appraisal Guidelines Research and Evaluation (AGREE) Collaboration, involving researchers from thirteen countries.

The studies presented in this thesis were supported by a grant of the European Union BIOMED2 Programme (BMH4-98-3669) and the Ministry of Health, Welfare and Sports (implementation project, nr. 545272).

Financial support for the publication of this thesis by the Centre for Quality of Care Research, the AGREE Collaboration and the Dutch Institute for Healthcare Improvement CBO is gratefully acknowledged.

Nijmegen, 2002

Copyright:	Oxford University Press (chapter 2) BMJ Publishing Group (chapter 3) Cambridge University Press (chapter 4) American Diabetes Association (chapter 5) The Royal College of General Practitioners, London (chapter 6) Bohn Stafleu Van Loghum bv, Houten (chapter 7)
Cover:	Jay Achterberg
Cover-photo:	Mont Ventoux, France, by Titus Ooteman, Landgraaf
Print:	Ponsen & Looijen BV, Wageningen
ISBN:	90-76316-16-3

QUALITY OF CLINICAL PRACTICE GUIDELINES

een wetenschappelijke proeve
op het gebied van de Medische Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen,
op gezag van de Rector Magnificus
prof. dr. C.W P.M. Blom,
volgens besluit van het College van Decanen
in het openbaar te verdedigen op
maandag 25 november 2002
des namiddags om 1 30 uur precies
door

Jacobus Simon Burgers

geboren op 3 september 1962 te Haarlem

Promotores: Prof. dr. R.P.T.M. Grol
Prof. dr. N.S. Klazinga, Universiteit van Amsterdam

Co-promotor: Dr. J.O.M. Zaat, Nederlands Huisartsen Genootschap

Manuscriptcommissie: Prof. dr. P. de Vries Robbé
Prof. dr. P.M.M. Bossuyt, Universiteit van Amsterdam
Dr. J. Kremer

Contents

Preface	7
Introduction	9
1 Twenty-five years of clinical guideline development: strengths, limitations and questions	15
Section I - Clinical guideline programmes	43
2 Towards evidence-based clinical practice: An international survey of 18 clinical guideline programmes <i>International Journal for Quality in Health Care; in press</i>	45
Section II - Clinical guidelines	67
3 Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project <i>Quality and Safety in Health Care; in press</i>	69
4 Characteristics of high quality guidelines: evaluation of 86 clinical guidelines developed in ten European countries and Canada <i>International Journal of Technology Assessment in Health Care 2002; 18(4): 955-964</i>	85
Section III - Recommendations in clinical guidelines	97
5 Inside guidelines: comparative analysis of recommendations and evidence in diabetes guidelines from 13 countries <i>Diabetes Care; in press</i>	99
6 Characteristics of effective clinical guidelines for general practice <i>British Journal of General Practice; in press</i>	115
7 The quality of Dutch clinical guidelines for general practice: evaluation of 130 key recommendations from 28 guidelines. <i>Huisarts en Wetenschap 2002; 45(7): 349-353</i>	125
8 Discussion: conclusions and recommendations	135
Summary	159
Samenvatting	169
Appendix A. The AGREE Collaboration	181
Appendix B. Questionnaire for description of clinical guideline programmes	185
Appendix C. 1. The AGREE Instrument (original English version)	191
2. The AGREE Instrument (translated Dutch version)	213
Appendix D. User guide for appraisal of recommendations in clinical guidelines	235
Dankwoord/Acknowledgement	241
Curriculum vitae	245

In this thesis the terms 'clinical practice guidelines', 'clinical guidelines' and 'guidelines' are considered as synonyms.

Preface

It is unusual to write a foreword for a thesis. A foreword is more appropriate to the publication of a book, and is not usually written by the author of the book himself. Nevertheless, I avail myself of the opportunity to make a few remarks about my motives in writing this thesis.

During my medical studies I had a growing interest in the epistemological background of medical science. I was especially interested in the lectures in medical philosophy, in which the empirical roots of medical theory and practice were sharply criticised. It was stated that medical facts are *constructed* within a presupposed model, which is necessarily based on value judgments.

As a staff member at the Dutch College of General Practitioners, involved in guideline development, I was confronted with the gaps in empirical knowledge and the lack of evidence for procedures in general practice. Nevertheless, it was a real challenge to develop guidelines on the basis of good arguments. Logical and consistent reasoning and the exposure of myths, supposed truths, or misconceptions were rewarded with a product that was approved by the profession and frequently used in practice.

Although guidelines often meet resistance in practice, as a general practitioner I only experienced the benefits of guidelines. However, I was an insider and I knew how the guidelines were developed and that it is never intended that they should be applied rigidly. Almost every week, during or after my surgeries, I consulted a number of guidelines or their summaries. Although I did not follow the recommendations in every case, they always helped me to substantiate and explain my arguments to the patient. Often I let the patient choose among the resulting options for treatment or management.

Hence both the development and the flexible application of guidelines have held great interest for me. I had already collected some literature on this subject with the idea that I would someday 'do something' with it. And so I was extremely pleased when in 1998 I was given the opportunity to undertake scientific research on the quality of clinical guidelines. Now I was able to test my views in a comprehensive framework and to learn from the experience in other countries. Thus this thesis is not only the traditional 'test of competence' but also evidence of a personal involvement in the subject that began much earlier.

I hope that I succeed in passing some of my enthusiasm for guidelines to the reader of this thesis and that any natural resistance to guidelines will thereby be overcome.

Jako Burgers

Introduction

People appreciate good health. Good health is an important part of the quality of life. We do not often realise that. When we become ill, we painfully discover that being healthy is not a matter-of-course. In many cases, Mother Nature is benignant, in that time takes care of our complaints and troubles. However, when pain and fear take over and we no longer feel in control of our bodies, we gratefully use health services. In general, we have great confidence in these services, which is partly due to the respect we have for the technical achievements of physicians. However, the basis of good health care originates from a relationship of trust. We rely on the sincerity, benevolence, and truthfulness of others, and usually do not question the quality of care.

Quality of health care

When we observe unexpected variations in medical care, the quality of care becomes a subject of discussion. Small variations are acceptable, and even desirable when taking into account differences in patient populations. However, when variations exceed certain limits, at least some of the variation might reflect overuse, inadequate use, or underuse of services by physicians in certain areas. The lack of practice standards or policies could explain some of the variation. Health care professionals primarily act on individual or local opinions, habits, and traditions and Mother Nature often offers them a helping hand. Common infections tend to heal with or without antibiotics, and most babies are born healthy, with or without a caesarean section.

Overall appreciation of medical care and general respect for the medical profession might explain why it took so long—until the late 1970s in the United States and the early 1980s in the Netherlands—before quality of care became of common interest. An important trigger was the increasing strain on government budgets and a pressing need for cost-effective health care. Moreover, professional organisations were also concerned with quality of care. They aimed at presenting the state of the art to colleagues, patients, and policy makers and preferred to develop guidelines themselves in order to minimise outside regulation.

Quality of clinical guidelines

In many countries, professional organisations took the lead in developing clinical practice guidelines. Initially, the experience and opinions of a number of clinical experts formed the basis of guidelines ('Good Old Boys Sat Around the Table'). However, in the 1990s, scientific requirements for clinical research and guidelines

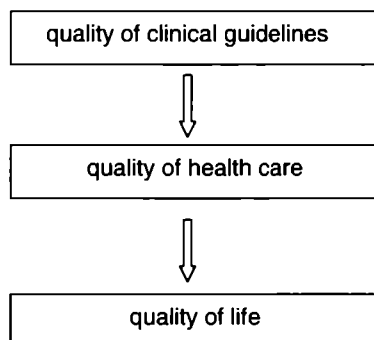
became more strict, partly due to the evidence-based medicine movement. Quality criteria were formulated for selection and assessment of scientific literature. A uniform approach became possible thanks to new electronic database services such as MEDLINE that made research literature easy and readily available. Also, large-scale systematic reviews were conducted, like those by the Cochrane Collaboration. These efforts may be considered to be precursors to the development of 'evidence-based guidelines', covering the most important medical fields. Health care professionals and patients worldwide were to benefit from clear and specific recommendations based on scientific evidence.

During the last decade, a 'guideline industry' emerged in many Western countries. Numerous guidelines have been developed by different agencies for the same clinical condition, even within a particular country (e.g., in the United States or Canada or the United Kingdom). However, the content of these guidelines can vary, with potentially great consequences for patient care. Methods for implementing guidelines also vary and it is often unclear to what extent the guidelines are being used in practice. Finally, the development of guidelines needs extensive resources, but it is uncertain whether all investments actually lead to better and more efficient patient care.

This thesis

This thesis concerns the quality of clinical guidelines, a rather abstract subject. The above shows that when we are concerned about the quality of life—which is not abstract at all and will be felt every day—we need good medical care. Good clinical guidelines might help ensuring such care (Figure 1).

Figure 1. The goal of clinical guidelines



Quality of guidelines is a subject that has attracted much international attention. In 1998, a group of researchers from thirteen countries—the AGREE collaboration (Appraisal Guidelines Research and Evaluation)—started a research project on the quality of guidelines and guideline development (Appendix A). The project was funded by the European Union. The aim of the project was to investigate variation between guidelines and guideline development models in order to advise the European Commission on guideline development, dissemination, and implementation. The central question was: *what are the criteria for good clinical guidelines and guideline development?* This is also the main research question of this thesis. A number of studies conducted within the AGREE project are presented in this thesis (chapters 2, 3, 4, and 5). In addition, we conducted a study on the quality and effectiveness of Dutch guidelines for general practice (chapters 6 and 7).

Outline of the thesis

Chapter 1 defines the clinical guideline research at three analytic levels:

1. the level of the clinical guideline programme
2. the level of the clinical guideline document
3. the level of specific recommendations in the guideline

The terms 'guideline programme', 'clinical guideline', and 'recommendation' are defined and the historical background of guideline development is discussed from an international perspective. The strengths and limitations of guidelines are discussed, leading to a number of research questions that were addressed in the studies presented in the following chapters (Table 1).

Table 1. Outline of this thesis

Subject	Method	Chapter
Definition of the problem	Literature review and theoretical considerations	1
Quality of guideline programme	International survey of guideline programmes	2
Quality of guidelines	International validation study of quality criteria of guidelines	3, 4
Quality of recommendations	Comparative analysis of guidelines from different countries	5
	Evaluation of Dutch recommendations for general practice	6, 7
Discussion	Synthesis of results	8

Chapter 2 presents a survey of eighteen guideline programmes from thirteen countries. A questionnaire was sent to key informants of these programmes. Questions included the aims and scope of the programme, methods of guideline development, dissemination and implementation strategies, and future plans. The results are discussed in light of the need for international collaboration in the field of guideline development.

Chapter 3 discusses the development and validation of the AGREE Instrument, the main product of the AGREE collaboration. The validation study included 100 guidelines from eleven countries assessed by 264 appraisers in two rounds. Data on reliability, validity, and usefulness of the instrument are presented.

Chapter 4 describes a study in which data from the validation study of the AGREE Instrument were further analysed using multi-level techniques. The characteristics of the guidelines with the highest scores are presented.

Chapter 5 presents a study in which fifteen guidelines on type 2 diabetes mellitus from thirteen countries were compared and analysed. Both the specific recommendations and their underlying evidence were analysed. Two areas (metformin use in obese patients and self-monitoring of blood glucose) were studied in more detail with regard to the similarities and differences between recommendations and the supporting evidence. The often presupposed 'one-to-one-relation' between recommendations and evidence is challenged.

Chapter 6 describes a study on characteristics of effective recommendations. Data were used from the 'Toetsen-aan-Standaarden-project' (Evaluating Guidelines project) in which the adherence to Dutch guidelines for general practitioners (NHG-standaarden) was measured. Characteristics of recommendations with high compliance were compared to those of recommendations with low compliance.

Chapter 7 describes a study in which 130 recommendations from 28 NHG guidelines were assessed by a panel of twelve general practitioners, using ten quality criteria.

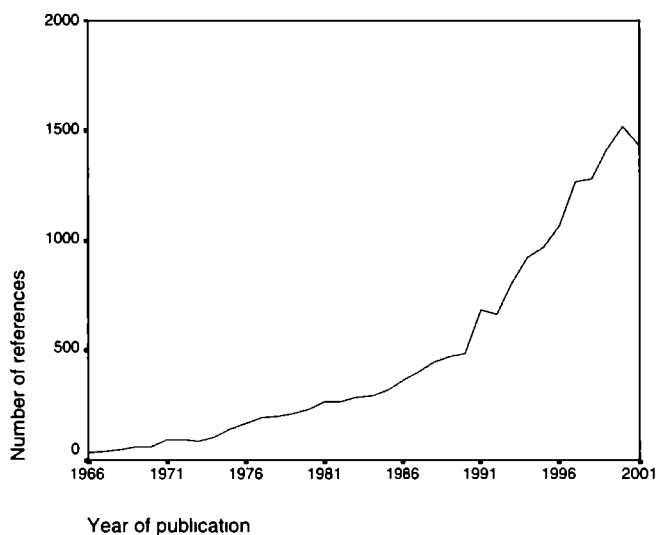
Chapter 8 discusses and synthesises the main findings and conclusions of the separate studies. The consequences for guideline developers, for clinical practice and for policy makers are considered. Finally, suggestions for future research and international collaboration are presented.

Chapter 1

Twenty-five years of clinical guideline development: strengths, limitations and questions

Clinical practice guidelines are a rather new phenomenon in medical science. They present not only medical facts but also concrete recommendations for clinicians in practice. In this regard clinical guidelines differ from research reports and literature reviews. Although clinical guidelines were already being developed in the 1930s, for instance, by some American specialist associations, systematic guideline development has made progress since 1977, when the National Institutes of Health (NIH) in the United States started with developing 'consensus statements' within a well-defined programme [1]. The NIH guideline programme has been a model for various other countries in the development of 'health technology assessment' and guidelines in the 1980s [2]. In the 1990s the production of guidelines increased exponentially, especially under the impetus of the evidence-based medicine movement. Both governmental agencies and numerous professional organisations took the initiative in the systematic development of clinical guidelines. This has led to an explosion of guidelines [3]. In the United States more than 1000 guidelines are included in the database of the 'National Guideline Clearinghouse' [4]. Guidelines and papers on guidelines are also increasingly included in databases of medical research evidence such as MEDLINE (Figure 1).

Figure 1. Number of references containing the word 'guideline(s)' in the title in MEDLINE by year of publication



Note: The decrease in 2001 could be partly due to a lag in entering data in MEDLINE

For a proper evaluation we should consider not only the quantity but also the quality of guidelines. This chapter explores the strengths and limitations of clinical guidelines. Before this, I will provide a definition of the concepts 'clinical guideline' and 'quality of guidelines' and will describe some historical details of guideline development from an international perspective. This chapter ends with the formulation of research questions, which are answered in the next chapters of this thesis.

Concepts and definitions

Clinical guidelines

The following definition from the Institutes of Medicine is generally accepted and often cited: *'clinical practice guidelines are systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances'* [5]. This definition emphasises that guidelines are an aid in daily practice, not only for health care professionals but also for patients. A prerequisite is that the guideline has been developed according to a systematic procedure. Similarly, the National Library of Medicine defines practice guidelines as *'directions or principles presenting current or future rules of policy for the health care practitioner to assist him in patient care decisions regarding diagnosis, therapy, or related clinical circumstances'* [6]. To that is added: *'The guidelines may be developed by government agencies at any level, institutions, professional societies, governing boards, or by the convening of expert panels. The guidelines form a basis for the evaluation of all aspects of health care and delivery'*. This definition does not include systematic development, but on the other hand, it mentions potential organisations and groups that could develop guidelines.

Remarkably, both definitions only concern the concrete recommendations and not the document in which the methodological background and supporting evidence are described. One often speaks of 'a guideline' (for example, for diagnosis of asthma), while implying the complete document, including a set of recommendations, methods, rationale, and discussion of evidence. Hence there is a need for defining this document. The Dutch Institute for Healthcare Improvement CBO uses the following definition: *'a 'guideline' is a document with recommendations, guidance and instructions to support daily practice in health care, based on the results of scientific research and the consequent discussion and formation of opinion, aimed at the explicit statement of good medical practice'*

[7]. This definition applies more to how the term 'guideline' is used in practice [8]. A systematic development is not required to fulfil the definition, but it is assumed that scientific research must be discussed.

In order to describe the development procedure, it is also necessary to define the programme in which guidelines are developed. Most guideline development organisations have such a programme in which guidelines are developed according to a formalised procedure. However, our search of the literature revealed no definition of the term 'guideline programme'. Our proposal is '*A guideline programme is a structured and coordinated programme designed with the specific aim of producing several clinical practice guidelines*'. The guideline development is usually coordinated by organisations with well-defined goals and missions, such as governmental or quasi-governmental agencies, professional organisations, or independent quality institutes.

In summary, for a proper analysis of the quality of guidelines, three levels should be distinguished (Table 1).

Table 1. Levels of analysis of clinical guidelines

Level	Definition
Guideline programme	a structured and coordinated programme designed with the specific aim of producing several clinical practice guidelines
Clinical guideline	a document that includes a set of statements about appropriate health care to support daily practice, based on evidence and critical appraisal, aimed at the explicit statement of good medical practice
Recommendation	a systematically developed statement to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances

Sometimes other terms are used. Eddy defines a practice policy as a '*standard*', 'if the health and economic consequences of an intervention are sufficiently well known to permit decisions and if there is virtual unanimity among patients about the desirability (or undesirability) of the intervention and about the proper use (or non-use) of the intervention' [9]. According to Eddy, standards *must* be followed in all cases, while guidelines in most cases *should* be followed, but depending on the individual patient, the setting, and other factors. The term 'standard' is also applied to the guidelines of the Dutch College of General Practitioners (NHG), which defines its guidelines as 'NHG Practice Guidelines'. According to Eddy's definition, however, these are guidelines and not standards, because a substantial part of the guideline is not based on evidence but on consensus [10]. Moreover, the

consequences of most of the recommended interventions for health and costs are not well known.

Finally, the term '*protocol*' is often used when there is a programmed and detailed description of a practice policy with clear, well-defined decisions. Usually protocols are regionally or locally developed, using a national guideline as a starting point, and then they formulate more specific recommendations that should be applied in certain local health care settings. In the development of a protocol, 'integrated care pathways' (or 'critical pathways') are often used. These are structured multidisciplinary care plans which detail essential steps in the care of patients with a specific clinical problem [11].

Quality of clinical guidelines

Clinical guidelines aim for maintaining or improving the quality of health care by providing recommendations about appropriate health care. In addition, guidelines can shorten the delay in implementation of research findings and uptake of innovations by summarising research evidence and current scientific insights [12]. However, guidelines also need to be implemented. Therefore, the concept of quality of clinical guidelines includes the methodological rigour as well as the usefulness of a guideline in practice, which are important factors in changing provider behaviour [13-15]. Hence, the AGREE Collaboration used the following definition: '*by quality of clinical practice guidelines we mean the confidence that the potential biases of guideline development have been addressed adequately and that the recommendations are both internally and externally valid, and are feasible for practice*' [16]. And further: 'This process involves taking into account the benefits, harms, and costs of the recommendations, as well as the practical issues attached to them. Therefore, the assessment includes judgements about the methods used for developing the guidelines, the content of the final recommendations, and the factors linked to their uptake'. In other words, the 'implementability' of guidelines also determines their quality [17].

From a wider perspective, clinical guidelines are part of a quality cycle. If guidelines are introduced to improve the quality of care, an implementation plan is required [18]. For successful use of guidelines in practice, a combination of different dissemination and implementation strategies is often needed [19,20]. For measuring the actual effect of guidelines on care, guidelines should be translated

into 'review criteria' and 'indicators' [21]. Review criteria are '*systematically developed statements that can be used to assess the appropriateness of specific health care decisions, services, and outcomes*' [1]. These are preferably based on the key recommendations in a guideline. Indicators are '*measurable elements of practice performance for which there is evidence or consensus that it can be used to assess the quality, and hence change of the quality of care provided*' [22]. Thus the 'adherence' or 'compliance'—the extent to which the recommendations are actually followed—is measured and compared with previously established 'standards of quality'. These can indicate the minimal as well as the optimal level of care [21].

Table 2 presents the concepts and definitions that apply to quality of care.

Table 2. Concepts and definitions concerning quality of health care

Concept	Definition
Quality of clinical guidelines	the confidence that the potential biases of guideline development have been addressed adequately and that the recommendations are both internally and externally valid, and are feasible in practice
Review criteria	systematically developed statements that can be used to assess the appropriateness of specific health care decisions, services, and outcomes
Indicator	a measurable element of practice performance for which there is evidence or consensus that it can be used to assess, and hence change, the quality of care provided
Standards of quality	authoritative statements of (1) minimum levels of acceptable performance or results, (2) excellent levels of performance or results, or (3) the range of acceptable performance or results

Clinical guideline development in some western countries

Almost all western countries are currently active in the field of clinical guideline development. The background and goals nevertheless differ and depend on the political context and the health care system. I will describe the historical background of guideline development in some countries, not aiming for completeness but outlining the most salient developments.

United States and Canada

One of the reasons for systematic development of clinical guidelines was that, in the 1970s, studies revealed a striking regional variation in health care [23,24]. It also appeared that the costs of health care had increased so much that they endangered the financing of the federal Medicare programme (a national

insurance plan especially for the elderly and the handicapped) [25]. Rapidly advancing technology held promise but made the care much more expensive. There was also an impression that unnecessary care was often given, for instance, carrying out procedures such as gastroscopy and coronary bypass operations for the wrong reasons [26]

The *National Institutes of Health* (NIH) took the lead in clinical guideline development and selected especially those topics about which there were controversies and which had public health importance [27]. The 'consensus development conference', in which the final decisions were taken, played a central role in the development of its guidelines. Prior to the conference, a systematic literature study was conducted. The recommendations were finally formulated on the basis of discussions and consensus. In 1977 the NIH produced its first consensus guideline, on screening for breast cancer.

At about the same time, the clinical guideline movement became active in Canada. In 1976 the *Canadian Task Force on the Periodic Health Examination*, now the *Canadian Task Force on Preventive Health Care* (CTFPHC) was established. Its purpose was to evaluate the effectiveness of preventive manoeuvres and to issue recommendations on the periodic health examination [28]. They first developed a system of 'grades of evidence', in which the design of the study determined the strength of the evidence [29]. Here can be seen the roots of the development of evidence-based guidelines. Yet the clinical guideline movement in the 1980s was dominated by the consensus-development method of the NIH, which gradually was adopted by more and more European countries (Sweden 1981, the Netherlands 1982, Denmark 1983, Great Britain 1984, Finland 1985, Norway 1986) [2].

In 1989 in the United States, the *Agency for Health Care Policy and Research* (AHCPR), now the *Agency for Healthcare Research and Quality* (AHRQ), was established [30]. This organisation was of great importance for the national and international approach of clinical guideline development. It embraced the principles of clinical epidemiology and formulated strict criteria for clinical guidelines [31,32]. Clinical guidelines should be based as much as possible on the results of well-designed studies, of which the randomised clinical trial was considered to be the best. An extensive and systematic study of the literature was essential to guideline development. Moreover, this was considerably simplified by the availability of electronic databases such as MEDLINE [33]. Between 1992 and 1996, the AHCPR produced 19 guidelines (such as benign prostatic hyperplasia,

depression, and low back pain), published in various formats: a thick volume of hundreds of pages with background information on the evidence, a 'quick reference guide' for the practitioner, and a version for the patient [34]. The recommendations were supported with levels of evidence, based on the original system of the CTFPHC and adapted in 1985 by the *US Preventive Services Task Force* [28].

The clinical guideline programme of the AHCPR ended in 1996, for political reasons [35]. The problem was that central, national guidelines were not accepted by all stakeholders and sometimes were seen as threatening. There was, for example, great resistance among chiropractors towards the guideline on low back pain, which advocated a policy of watchful waiting. Consensus over the policy was lacking, which severely endangered implementation of the guideline. Hence, it was decided that the AHCPR itself would no longer make clinical guidelines but would leave that to other authorities. Its task is now primarily to support the process of guideline development and implementation. In spite of the failed guideline programme, the AHCPR has greatly encouraged the development of evidence-based clinical guidelines in the United States and in other countries.

In 1999 AHCPR was renamed the Agency for Healthcare Research and Quality (AHRQ), the focus being shifted from policy to research on the quality of care. At the moment, one of the most important facilities of the AHRQ is the *National Guideline Clearinghouse* (NGC), which is administered and supported by cooperation between the American Medical Association (AMA) and the American Association of Health Plans (AAHP) [4]. This is a database with more than 1000 evidence-based clinical guidelines. Not only American but also foreign guidelines can be registered if they fulfil the following four criteria: (1) the guideline meets the IOM definition of clinical practice guidelines, (2) the guideline is produced under the auspices of 'medical specialty associations; relevant professional societies, public or private organisations, government agencies at the Federal, State, or local level; or health care organisations or plans', (3) the guideline is developed on the basis of a systematic literature search and review of existing scientific evidence published in peer-reviewed journals and for which sufficient background information is provided, and (4) the guideline is in English and not older than five years. The NGC also offers the possibility of comparing guidelines on the basis of numerous variables.

Initiatives have also been taken in Canada to improve the methodology of guideline development, such as the Cancer Care Ontario Practice Guideline

Initiative that developed and implemented evidence-based guidelines concerning cancer [36,37]. Attention was also given to how the guideline is received in practice, by including practitioners' opinions about draft guidelines using a survey [38]. In this way a broad participation of target users is obtained, which should minimise the chance of a debacle such as occurred with the AHCPR guidelines.

United Kingdom

In United Kingdom the boom in clinical guideline development began a few years later than in the United States [35]. The initiative came mainly from the professional societies. In 1986 *The Royal College of General Practitioners* developed a series of guidelines for good practice management. *The Royal College of Physicians* began guideline development in 1990, followed a few years later by the *Royal College of Surgeons*, among others. The first years of guideline development were rather uncoordinated and there were many guidelines of only moderate quality [39].

In the 1990s the British government, which is responsible for health care by means of the National Health Service (NHS), began to acquire an interest in clinical guidelines. In 1993 they published a document in which a strategy for promoting the effectiveness of medical procedures and interventions were set out [40]. On the basis of this report financial support was granted to such centres as the *Cochrane Collaboration* and the *NHS Centre for Reviews and Dissemination*.

A remarkable initiative came from the *Centre of Health Services Research* of the University of Newcastle upon Tyne, which set up a guideline programme, the *North of England Evidence Based Guideline Development Project*, started in 1995 [41]. The ultimate goal was to formulate methodological principles for clinical guideline development. Seven guidelines developed according to these principles have been published. The systematic review of research evidence is fundamental in this methodology, just as in the programme of the AHCRP.

In 1996 the NHS published a pamphlet in which an important role for guidelines was established [42]. Clinical guidelines should improve patient care as well the cost-effectiveness of health care. It was emphasised that the development of guidelines would be the responsibility of professional societies, to prevent the medical profession from turning against the plan. To aid the implementation of guidelines, the existing local audit systems could be useful by providing feedback about the actual care provided [43].

To coordinate and monitor clinical guideline development, implementation, and evaluation, the *National Institute for Clinical Excellence* (NICE) was established in 1999 to serve England and Wales [44]. It did not develop guidelines itself but commissioned guideline development to well-known and credible organisations.

Since 1993, the *Scottish Intercollegiate Guidelines Network* (SIGN) has functioned in a similar way for Scotland [45]. SIGN is a professional association that coordinates clinical guideline development and implementation. The SIGN guideline programme pays much attention to the quality of guidelines, as well as to the multidisciplinary acknowledgement of the guidelines. SIGN has developed more than 40 clinical guidelines thus far. Some of these are updates of earlier versions. All guidelines are available via Internet [46]. SIGN is also active in the area of methodology of guideline development. It has recently developed a handbook for guideline developers, in which a new system of levels of evidence and grades of recommendations is presented [47].

France

Since 1992 national clinical guidelines in France have been developed in the guideline programme of the *Agence Nationale d'Accréditation et d'Évaluation en Santé* (ANAES, known as ANDEM before 1997). ANAES is an organisation set up on the initiative of the French government for the purpose of promoting quality of health care and cost containment. Because physicians are paid on the basis of the volume of their service, the cost of health care in France has increased steadily. Clinical guidelines should reduce the volume and costs [48]. Based on the guidelines, so-called *Références Médicales Opposables* (RMOs) were formulated. These specify the conditions under which a given procedure may *not* be performed [49]. This should prevent superfluous diagnostic procedures and ineffective treatments. The RMOs were not optional. Physicians who have a contract with insurance companies would be sanctioned if they ignore the RMOs. A control apparatus that takes random samples among physicians should reveal irregularities [50]. The entire system is an example of a top-down approach of dubious effectiveness [51]. Consequently, no additional RMOs have been developed in recent years. The ANAES guidelines are developed by multidisciplinary work groups, consisting of 12 to 15 experts, on the basis of systematic analyses of the literature. The method is comparable to that of the AHCPR. There is an external review procedure in which 25 to 40 experts are requested to give comments on the guideline under development. Over 100

guidelines and 140 RMOs have been published and disseminated to all practising physicians. In addition to the guidelines based on the AHCPR method, ANAES also provides its label to consensus based guidelines developed by scientific societies when specific methodological requirements are met.

Another large-scale guideline initiative began in 1993 as the 'Standards, Options, and Recommendations' (SOR) project, in which guidelines for cancer are developed through the cooperative effort of the *Fédération Nationale des Centres de Lutte Contre le Cancer* (FNCLCC), with 20 regional cancer centres, various French hospitals, and medical specialist associations [52]. The terms 'standard' and 'options' express the degree of certainty of the recommended intervention and are derived from Eddy [9]. A literature review and external review are essential to the guideline development, as in the ANAES guidelines.

Germany

National clinical guideline development in Germany was begun in 1995 by the *Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften* (AWMF), the coordinating association of medical specialists [53]. Most specialist associations then became active and developed many monodisciplinary guidelines. By the end of 1999 the total number of clinical guidelines had risen to 556 and there were 241 in preparation. Most guidelines were developed by consensus procedures, in which clinical experience and intuition played an important role. In the past few years more attention has been given to the principles of evidence-based medicine [54]. Evaluation studies showed that clinical guidelines were often of low quality [55]. Literature reviews were not carried out systematically and the formulation of recommendations was rather arbitrary. This led to the establishment of a national clearinghouse for clinical guidelines ('Leitlinien Clearingverfahren'), following the example of the National Guideline Clearinghouse in the United States. The clearinghouse only includes guidelines that fulfil certain quality criteria [56,57]. The clearinghouse is administered by the *Ärztliche Zentralstelle Qualitätssicherung* (ÄZQ), established in 1995 with the task of advising and supporting the national association of physicians and the national association of health care insurers [58].

Netherlands

Compared with other European countries, the Netherlands was a forerunner in the area of clinical guideline development. In 1982, the *Dutch Institute for Healthcare Improvement CBO* began developing national guidelines for medical specialists. The objective was to make good medical practice explicit in controversial multidisciplinary subjects [59]. The method for guideline development was derived from the National Institutes of Health (NIH). Subsequently, in the 1990s, the development of consensus statements was gradually abandoned and guidelines were developed in accordance with the principles of evidence-based medicine. Until now, the CBO has produced more than 70 multidisciplinary, evidence-based guidelines.

Following the example of the CBO, in 1989 the *Nederlands Huisartsen Genootschap* (NHG) (*Dutch College of General Practitioners*) began developing 'NHG Practice Guidelines' [60]. In contrast to the CBO guidelines, these are monodisciplinary guidelines developed 'by general practitioners for general practitioners' [61]. The goal of the guidelines is to help the general practitioner in daily practice. In addition, the guidelines contribute to the professionalisation of family practice [62]. The NHG guideline programme gives much attention to the implementation of their guidelines, using a multifaceted approach with written materials (publication in scientific journal, educational packages) and personal approaches (contact with colleagues, outreach visits) [63]. Similar to the CBO, the NHG has produced more than 70 guidelines thus far.

Finland

In Finland, *Duodecim*, the umbrella association of physicians, has been active in guideline development since 1989. More than 1000 clinical guidelines have been developed and they have been made available in an international version on CD-ROM in 2000 and via the Internet in 2002 [64]. The collection of clinical guidelines can be seen as a handbook, which also includes rare conditions. The target group is general practitioners, but specialists can also benefit from the guidelines [65].

In 1996 Duodecim started a guideline programme—the Current Care project—in cooperation with the medical specialists association, with the aim of producing multidisciplinary, evidence-based guidelines on topics with a large impact on public health [66]. More than 30 guidelines have been developed within this project thus far.

Australia and New Zealand

In Australia, clinical guidelines have been produced since 1980. The most influential organisation is the *National Health and Medical Research Council* (NHMRC), an independent advisory body for health care in Australia. Since 1995 it has systematically produced multidisciplinary guidelines [67]. They mainly concern 'large' subjects such as breast cancer, stroke, and benign prostatic hyperplasia. The guidelines appear in different versions, including patient versions, and have an impact similar to that of the guidelines of the AHCPR a few years ago. Clinical guideline development by the NHMRC serves as an example for other Australian organisations. To support the development of guidelines, the NHMRC produced a guide [68], supplemented by a series of handbooks that explain the different steps in clinical guideline development, such as searching and reviewing evidence and translating evidence into recommendations for practice [69].

From 1991 to 1995, clinical guideline development in New Zealand was supported and stimulated by the government [70]. Because the quality of the guidelines varied, the *New Zealand Guidelines Group* (NZGG) was established in 1996, with the purpose of promoting the principles of evidence-based medicine. About 25 guidelines have been developed according to these principles. The guidelines on hypertension and cholesterol included tables with absolute cardiovascular risk figures [71,72], which were taken over by several guideline organisations in other countries.

To summarise, in the past 25 years there has been an impressive increase in clinical guideline activities. Guideline development has gradually been formalised and institutionalised. Table 3 presents a number of organisations that currently play a prominent role in guideline development. Due to the rise of evidence-based medicine, more attention is being given to the methodology of guideline development. Most national guidelines are widely disseminated. Not only the health care professionals but also policy makers and government organisations are interested in clinical guidelines and are increasingly supporting guideline development. However, it remains unclear whether the high expectations of guidelines can be fulfilled. For a proper evaluation it is necessary to examine how guidelines can contribute to improving quality of care and the limitations that can occur in this process.

Table 3. Prominent organisations involved in guideline development in some western countries

Country	Name of organisation	Website
Australia	- National Health and Medical Research Council (NHMRC)	www.health.gov.au/hfs/nhmrc
Canada	- Canadian Task Force on Preventive Health Care (CTFPHC)	www.ctpfhc.org
	- Canadian Medical Association (CMA)	www.cma.ca/cpgs
	- Cancer Care Ontario Practice Guidelines Initiative (CCOPGI)	www.cancercare.on.ca/ccopgi
New Zealand	- New Zealand Guidelines Group (NZGG)	www.nzgg.org.nz
Finland	- Finnish Medical Society Duodecim	www.duodecim.fi
France	- Agence Nationale d'Accreditation et d'Évaluation en Santé (ANAES)	www.anaes.fr
	- Fédération Nationale des Centres de Lutte Contre le Cancer (FNCLCC)	www.fnclcc.fr
Germany	- Ärztliche Zentralstelle Qualitätssicherung (AZQ)	www.leitlinien.de
	- Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF)	www.awmf.de
Netherlands	- Dutch Institute for Healthcare Improvement CBO	www.cbo.nl
	- Dutch College of General Practitioners (NHG)	www.artsen.net/nhg
UK	- National Institute for Clinical Excellence (NICE)	www.nice.org.uk
	- Centre for Health Services Research Unit of University of Newcastle upon Tyne (North of England)	www.ncl.ac.uk/chsr
	- Scottish Intercollegiate Guidelines Network (SIGN)	www.sign.ac.uk
USA	- National Guideline Clearinghouse (NGC)	www.guideline.gov
	- National Institutes of Health (NIH)	consensus.nih.gov
	- US Preventive Services Task Force (USPSTF)	www.ahrq.gov/clinic/uspstfix.htm
	- Centers for Disease Control and Prevention (CDC)	www.cdc.gov

Strengths of clinical guidelines

Clinical guidelines are documents that define how optimal health care should be provided. The strengths of guidelines are related to their intention. Roughly speaking, we can distinguish two intentions: (a) the guideline as source of information and (b) the guideline as policy document.

a. Source of information

Clinical guidelines are texts containing information. The information can be considered and used in different ways. However, the use of guidelines does not necessarily change provider behaviour.

SOURCE OF KNOWLEDGE. Clinical guidelines can be regarded as a reflection of the current state of knowledge, provided that there has been a systematic review of the literature and that the literature has been formally assessed [73]. They summarise, analyse, and synthesise the research evidence, as do systematic reviews [74]. In contrast to reviews, guidelines also give answers to a range of questions about specific clinical problems, while a review usually answers a single question (for example, 'what is the most effective drug treatment for dyspepsia?'). Another difference concerns how to proceed in the absence of clinical evidence. A review usually concludes with the statement that there is insufficient evidence available, while a guideline can still formulate recommendations or options on the basis of arguments or clinical experience. The SIGN guidelines label these statements as 'good practice points' [47].

BASIS FOR EDUCATION. Clinical guidelines can be used for academic courses and continuing medical education (CME) because they provide an integrated view on how to manage a condition [75]. The guideline bridges the gap between research and practice and presents the 'state of the art'. Thus without much effort, one becomes aware of recent developments concerning effective and efficient care.

TOOLS IN PRACTICE. In practice, summaries of clinical guidelines, such as quick-reference guides or flow charts, can serve as reminders and can even be referred to during a consultation. They help the health care professional to explain clinical decisions and to involve the patient in decision making [76]. Other tools, such as computer applications or tables showing absolute risks for cardiovascular diseases, can also be used during consultations.

SOURCE OF INFORMATION FOR PATIENTS. Many clinical guidelines are also accessible to the public via the Internet and can be consulted by patients. Some guidelines are accompanied by patient versions that explain things in simple language [77]. They provide comprehensive information about optimal care or consider different options for the management of a condition [78]. This may give the patient the opportunity to weigh the advantages and disadvantages and to share in decision making [79]. Finally, guidelines explain what patients may expect and can protect them from overuse as well as underuse of services.

b. Policy document

As policy documents, clinical guidelines are intended to influence or change provider behaviour. In this case they may be part of a coherent aggregate of policy measures.

BASIS FOR QUALITY ASSESSMENT. Recommendations in clinical guidelines can be translated into 'review criteria' and 'indicators' [80,81]. With indicators, performance can be tested in practice and compared with 'standards of care' [22,82]. Thus, health care professionals can receive feedback on their performance. This information can be used to select specific measures for quality improvement.

BASIS FOR INTERDISCIPLINARY COLLABORATION. Many clinical guidelines concern topics involving different disciplines, such as low back pain (general practitioners, neurologists, orthopaedic surgeons, rheumatologists, rehabilitation physicians, and paramedics such as physiotherapists, ergotherapists, and psychologists). A guideline can serve as a basis for interdisciplinary agreements about the management of a condition. Examples are the agreements between general practitioners and medical specialists established by the Dutch College of General Practitioners in collaboration with specialist associations [83]. National guidelines can also be used at regional and local levels as the starting point for making local agreements [84].

CONTRIBUTION TO APPROPRIATE CARE. Some clinical guidelines primarily aim at preventing unnecessary care and unnecessary costs [85,86]. Examples are the AHCPR and the ANAES guidelines [30,48]. In these cases, the financiers of care, such as the insurers and the government, also have an interest in the guidelines. For insurers, guidelines can be used to negotiate with health care providers about the contracts or budgets [87,88]. For the government, guidelines can be used to reduce the costs of health care [35]. An example is a guideline on influenza vaccination, which helps the general practitioner to establish the indications for flu vaccination, but also helps to decrease costs by reducing hospital admissions due to pneumonia influenzae [89].

TOOLS FOR RATIONING. Clinical guidelines can be used to allocate medical technologies and as rationing policies [90,91]. The guidelines should then be developed through a fair process involving patient and public views [92]. In the

Netherlands, however, the project *Passende medische zorg* (Appropriate health care) concluded that it is not desirable to use a national guideline—top-down—to restrict health care services on the basis of costs [93,94]. Clinical practice can only be rationalised to a limited level, partly due to the lack of research evidence [95]. On a local level, choices could be made, if all considerations and arguments are explicitly discussed.

Limitations of clinical guidelines

Clinical guidelines are often represented as a 'magic bullet' for problems in health care [96]. However, they only are one of the various methods that could be used to improve quality of care [14]. They are especially useful when practitioners are uncertain about the appropriate care, for which scientific evidence can provide an answer. In other situations, other methods such as establishing a multidisciplinary care plan or the change of tasks or care processes may be indicated [97]. Even if guidelines are applied adequately, in practice they can come up against a number of limitations. We can distinguish limitations that concern (a) the content of guidelines and (b) the effects of guidelines.

a. Content

LACK OF EVIDENCE. In spite of the overflow of medical literature, guideline developers often encounter a lack of evidence. Many questions remain unanswered because of the lack of well-designed studies. The lack of evidence is often supplemented by statements based on consensus [98]. The power of such statements is, however, unclear. Guidelines that use 'levels of evidence' and 'grades of recommendations' suggest that 'level A' is the best [99]. However, since most recommendations in clinical guidelines are not based on level A evidence, the scientific power of guidelines could be questioned.

SUBJECTIVE INTERPRETATION OF EVIDENCE. Even if there is sufficient evidence, the critical appraisal of the evidence is in a certain sense subjective [100]. If formal evaluation criteria are used, the weighing of individual criteria is often unclear and a matter of taste. Prior beliefs can dominate the interpretation, in particular when there are conflicts of interest [101]. These could include professional as well as commercial interests. This might explain why guidelines of different organisations sometimes contain conflicting recommendations [102,103].

'COOKBOOK MEDICINE'. A clinical guideline assumes a 'standard patient', while in practice one deals with individual patients with their own views and preferences. Although different patient groups could be described in guidelines, it is impossible to cover the entire spectrum of individual conditions. There is a danger that guidelines simplify clinical practice and disregard the individual needs of patients. [104,105]. Rigid and uncritical application of guidelines can even endanger quality of care, because in some specific situations it is desirable to diverge from guidelines [96,106].

b. Effect

UNREALISTIC EXPECTATIONS. Clinical guidelines formulate the way in which optimal health care should be provided. Adequate application of guidelines suggests a certain health gain [107]. However, it is not clear to what extent this can be achieved in practice [108]. Many guidelines are produced on the basis of results of clinical studies in selected populations in standard settings. Clinical practice is different and results achieved in well-designed studies are often not achieved in daily practice [109]. Therefore, the effect of interventions recommended in the guideline cannot be predicted with certainty.

PROFESSIONAL RESISTANCE. In general, health care professionals strive for professional autonomy. The need to follow guidelines could threaten this autonomy [110]. Similarly, some professionals fear that guidelines will increase their medico-legal exposure [111,106]. In court, guidelines could overrule clinical judgement [104,105]. Clinical guidelines can also be misused by governmental and other authorities. Introducing guidelines together with sanctions could harm the image of guidelines and increase professional resistance [51].

LACK OF IMPLEMENTATION PLANS. Guidelines do not implement themselves [63]. Even if health care professionals are willing to apply guidelines, changing behaviour and routines demands much effort [15]. Effective guideline implementation requires a multifaceted approach [19,112]. Guideline developers do not always feel responsible for implementation and sometimes leave it to regional or local authorities. The danger is that even evidence-based guidelines could disappear into the bookcase of health care professionals and that clinical practice will not be affected in any way.

DOUBTS ABOUT COST-EFFECTIVENESS. The development of clinical guidelines makes large demands on resources. The cost of developing a national guideline varies from 50,000 to more than 250,000 EURO [113,114]. There are also costs for dissemination and implementation of the guidelines. Whether guidelines improve the cost-effectiveness of health care has not yet been demonstrated [115].

Table 4 summarises the strengths and limitations of clinical guidelines.

Table 4. Strengths and limitations of clinical guidelines

Strengths	Limitations
<ul style="list-style-type: none"> - source of knowledge and summary of research evidence - basis for academic courses and continuing medical education - tools in practice to explain decisions - source of information for patients to share in decision making - basis for audit and quality assessment - basis for interdisciplinary collaboration and agreements - contribution to appropriate care - tools for rationing and policy decisions 	<ul style="list-style-type: none"> - lack of evidence - subjective interpretation of evidence and undesirable influence of interests - cookbook medicine - unrealistic expectations about effects in practice - professional resistance through threatening professional autonomy and legal implications - lack of implementation plans - uncertain cost-effectiveness

Research questions

Twenty-five years of guideline development resulted in many ideas about clinical guidelines. In the past few years, a growing need has been felt to synchronise guideline development, nationally as well as internationally. This was the main reason for starting the AGREE (Appraisal Guidelines Research and Evaluation) project in 1998, aiming for collaborative research on quality of clinical guidelines and the methodology of guideline development. Researchers from thirteen countries, including the Netherlands, participated in this project. We formulated a number of specific problems and research questions, which formed the basis for the studies presented in this thesis (Table 5).

1. Pluralism

At present there is an industry of guideline activities, but it lacks sufficient coordination. A survey of guideline activities in various European countries showed that there were great differences between countries [116]. Some countries did not have well-structured programmes to develop and implement guidelines.

We hypothesise that such programmes are needed to ensure the high quality of clinical guidelines and to prevent uncontrolled proliferation of guidelines. The question is which requirements a guideline programme ought to fulfil. This was the motivation for one of the AGREE studies described in chapter 2 of this thesis.

2. Lack of quality

In spite of the mass of clinical guidelines, several studies suggested that many guidelines are of poor quality [117-120]. However, there are no validated and generally-accepted criteria for assessing guideline quality [121]. This is in contrast to the numerous instruments that have been developed for the assessment of randomised clinical trials (RCTs) [122] and the worldwide consensus about the reporting of RCTs [123]. Consequently, there is a need for clarity and agreement about the criteria for good clinical guidelines. This was the motivation for the development and validation of the AGREE Instrument, which is described in chapter 3. In applying the instrument to guidelines we raised the question of what the characteristics of good clinical guidelines are. This is described in chapter 4.

3. Differences in recommendations

Different guidelines about the same clinical condition often contain different and sometimes conflicting recommendations, while the 'body of evidence' (e.g., electronic access to MEDLINE) is shared [102,124-127]. These differences cannot always be explained by differences in quality or method of guideline development. A systematic comparison of recommendations and their supporting evidence can give insight into the way in which evidence is collected, interpreted, and synthesised, and which other factors (e.g., professional, social, and cultural factors) play a part in the formulation of recommendations. This was the motivation for the study described in chapter 5, which was also carried out as part of the AGREE project.

The AGREE Instrument assesses especially the quality of clinical guidelines (i.e., guideline documents), but there is also a need for statements about the quality of the individual recommendations in guidelines. The question is whether the criteria of the AGREE Instrument are suitable for appraisal of recommendations. Chapter 7 describes a study in which a modified instrument was used to assess recommendations in different clinical guidelines developed by the same guideline organisation.

4. Limited effect in practice

While guidelines are intended to bridge the gap between theory and practice, they create a new gap because they do not implement themselves. In this sense, problems that arise with the introduction of clinical guidelines are comparable to those that accompany the introduction of new technologies [128,129]. Different studies have shown that the effects of clinical guidelines vary and depend partly on the implementation strategy [15]. Nevertheless, that does not explain all effects or lack of effects. There appear to be successful and less successful guidelines [17]. The question is what are the characteristics of effective recommendations. This formed the motivation for the study described in chapter 6 of this thesis.

Table 5. Research questions and studies of this thesis

Research question	Study	Chapter
What are the basic requirements for a guideline programme?	Towards evidence-based clinical practice: an international survey of 18 clinical guideline programmes	2
What are the criteria for good clinical guidelines?	Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project	3
What are the characteristics of good clinical guidelines?	Characteristics of high-quality guidelines: evaluation of 86 clinical guidelines developed in ten European countries and Canada	4
How can differences between recommendations in clinical guidelines be explained?	Inside guidelines: comparative analysis of recommendations and evidence in diabetes guidelines from 13 countries	5
What are the characteristics of effective clinical guidelines?	Characteristics of effective clinical guidelines for general practice	6
How can the quality of recommendations in clinical guidelines be assessed?	The quality of the Dutch guidelines for general practice: evaluation of 130 key recommendations from 28 standards	7

References

1. Institute of Medicine. Field MJ, Lohr KN (eds.). Guidelines for Clinical Practice: From Development to Use. Washington DC: National Academy Press, 1992.
2. Goodman C, Baratz SR (eds.). Council on Health Care Technology & Institute of Medicine. Improving Consensus Development for Health Technology Assessment: An International Perspective. Washington DC: National Academy Press, 1990.
3. Hibble A, Kanka D, Pencheon D, Pooles F. Guidelines in general practice: the new Tower of Babel? *BMJ* 1998; 317: 862-863.
4. National Guideline Clearinghouse. www.guideline.gov/index.asp.
5. Institute of Medicine. Field MJ, Lohr KN (eds.). Clinical Practice Guidelines: Directions for a New Program. Washington DC: National Academy Press, 1990.
6. National Library of Medicine. PubMed. www.ncbi.nlm.nih.gov/entrez/query.fcgi.
7. Kwaliteitsinstituut voor de Gezondheidszorg CBO. Richtlijnontwikkeling binnen het Kwaliteitsinstituut voor de Gezondheidszorg CBO. Handleiding voor werkgroepleden (Guideline development at the Dutch Institute for Healthcare improvement. User guide for working group members). Utrecht: CBO, 2000.
8. Mäkelä M, Thorsen T (eds.). Changing Professional Practice: Theory and Practice of Clinical Guidelines Implementation. Copenhagen: Danish Institute for Health Services, 1999.
9. Eddy DM. A Manual for Assessing Health Practices and Designing Practice Policies. Guidelines for the Development and Implementation of Clinical Practice Guidelines: the Explicit Approach. Philadelphia: American College of Physicians, 1992.
10. Burgers JS. De wetenschappelijke onderbouwing van de NHG-Standaarden (The scientific base of the guidelines of the Dutch College of General Practitioners). *Huisarts Wet* 1994; 37: 188-193.
11. Campbell H, Hotchkiss R, Bradshaw N, Porteous M. Integrated care pathways. *BMJ* 1998; 316: 133-137.
12. Haines A, Donald A. Getting research findings into practice. Making better use of research findings. *BMJ* 1998; 317: 72-75.
13. Grimshaw JM, Russell IT. Achieving health gain through clinical guidelines II: Ensuring guidelines change medical practice. *Qual Health Care* 1994; 3: 45-52.
14. Grol R. Improving the quality of medical care. Building bridges among professional pride, payer profit, and patient satisfaction. *JAMA* 2001; 286: 2578-2585.
15. Grimshaw JM, Shirran L, Thomas R, Mowatt G, Fraser C, Bero L, et al. Changing provider behaviour. An overview of systematic reviews interventions. *Med Care* 2001; 39(Suppl 2): 2-45.
16. The AGREE Collaboration. Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument. www.agreecollaboration.org.
17. Grol R, Dalhuijsen J, Thomas S, in 't Veld C, Rutten G, Mekkink H. Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. *BMJ* 1998; 317: 858-861.
18. Grol R. Beliefs and evidence in changing clinical practice. *BMJ* 1997; 315: 418-421.
19. Wensing M, van der Weijden T, Grol R. Implementing guidelines and innovations in general practice: which interventions are effective? *Br J Gen Pract* 1998; 48: 991-997.
20. Solberg LI, Brekke ML, Fazio CJ, Fowles J, Jacobsen DN, Kottke TE, et al. Lessons from experienced guideline implementers: attend to many factors and use multiple strategies. *Jt Comm J Qual Improv* 2000; 26:171-188.
21. Agency for Health Care Policy and Research. Using Clinical Practice Guidelines to Evaluate Quality of Care. Rockville, MD: Agency for Health Care Policy and Research, 1995 (AHCPR Publication no. 95-0045).
22. Lawrence M, Olesen F, for the Equip working party on indicators. Indicators of quality in health care. *Eur J Gen Pract* 1997; 3: 103-108.
23. Lewis CE. Variations in the incidence of surgery. *N Engl J Med* 1969; 281: 880-884.
24. Wennberg JE, Gittelsohn A. Small-area variation in health care delivery. *Science* 1973; 142: 1102-1108.

25. Greenberg B, Derzon RA. Determining health insurance coverage of technology: problems and options. *Med Care* 1981; 19: 967-978.
26. Chassin MR, Koseoff J, Park RE, et al. Does inappropriate use explain geographic variations in the use of health care services? A study of three procedures. *JAMA* 1987; 258: 2533-2537.
27. Jacoby I. The consensus development program of the National Institutes of Health: current practices and historical perspectives. *Int J Technol Assess Health Care* 1985; 1: 420-432.
28. Woolf SH, Battista RN, Anderson GM, Logan AG, Wang E. Assessing the clinical effectiveness of preventive maneuvers: analytic principles and systematic methods in reviewing evidence and developing clinical practice recommendations. A report by the Canadian Task Force on the Periodic Health Examination. *J Clin Epidemiol* 1990; 43: 891-905.
29. Fletcher SW, Spitzer WO. Approach of the Canadian Task Force to the Periodic Health Examination. *Ann Intern Med* 1980; 92: 253-254.
30. Woolf SH. Practice guidelines: a new reality in medicine. I. Recent developments. *Arch Intern Med* 1990; 150: 1811-1818.
31. Woolf SH. Interim Manual for Clinical Practice Guideline Development. Rockville, MD: Agency for Health Care Policy and Research, 1991 (AHCPR Publication no. 91-0018).
32. Sackett DL, Haynes RB, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. Boston: Little Brown, 1985.
33. Dalrymple PW. CD-ROM MEDLINE use and users: information transfer in the clinical setting. *Bull Med Libr Assoc* 1990; 78: 224-232.
34. Agency for Health Care Policy and Research. Clinical Practice Guideline Development. Rockville: Agency for Health Care Policy and Research, 1994 (AHCPR Publication no. 93-0023).
35. Day P, Klein R, Miller F. Hurdles and Levers. A comparative US-UK study of guidelines. Nuffield Trust Series No.4. London: Nuffield Trust, 1998.
36. Browman GP, Levine MN, Mohide EA, Hayward RS, Pritchard KI, Gafni A, et al. The practice guidelines development cycle: a conceptual tool for practice guidelines development and implementation. *J Clin Oncol* 1995; 13: 502-512.
37. Levine M, Browman G, Newman T, Cowan DH. The Ontario Cancer Treatment Practice Guidelines Initiative. *Oncology* 1996; 10(Suppl): 19-22.
38. Browman GP, Newman TE, Mohide EA, Graham ID, Levine MN, Pritchard KI, et al. Progress of clinical oncology guidelines development using the Practice Guidelines Development Cycle: the role of practitioner feedback. *J Clin Oncol* 1998; 16: 1226-1231.
39. Cluzeau F, Littlejohns P, Grimshaw J, Hopkins A. Appraising clinical guidelines and the development of criteria: a pilot study. *J Interprofes Care* 1995; 9: 227-235.
40. NHS Management Executive Improving Clinical Effectiveness Executive Letter EL(93)115 21. Leeds: NHS Management Executive, 1993.
41. Eccles M, Clapp Z, Grimshaw J, Adams PC, Higgins B, Purves I, et al. North of England evidence based guidelines development project: methods of guideline development. *BMJ* 1996; 312: 760-762.
42. NHS Executive. Clinical guidelines: using clinical guidelines to improve patient care within the NHS. London: NHS Executive, 1996.
43. Miles A, Bentley P, Polychronis A, Price N, Grey J. Clinical audit in the National Health Service: fact or fiction? *J Eval Clin Pract* 1996; 2: 29-35.
44. Rawlins M. In pursuit of quality: the National Institute for Clinical Excellence. *Lancet* 1999; 353: 1079-1082.
45. Petrie JC, Grimshaw JM, Bryson A. The Scottish Intercollegiate Guidelines Network initiative: getting validated guidelines into local practice. *Health Bull* 1995; 53: 345-348.
46. Scottish Intercollegiate Guidelines Network. www.sign.ac.uk.
47. Scottish Intercollegiate Guidelines Network. SIGN 50: A guideline developers' handbook. SIGN publication No. 50, 2001.
48. Durand-Zaleski I, Colin C, Blum-Boisgard C. An attempt to save money by using mandatory practice guidelines in France. *BMJ* 1997; 315: 943-946.

49. Maisonneuve H, Cordier H, Durocher A, Matillon Y. The French clinical guidelines and medical references programme: development of 48 guidelines for private practice over a period of 18 months *J Eval Clin Pract* 1997; 3: 3-13.
50. Matillon Y, Goldberg J. Clinical practice guidelines *Lancet* 1997; 349: 795-796.
51. Durieux P, Chaix-Couturier C, Durand-Zaleski I, Ravaud P. From clinical recommendations to mandatory practice. The introduction of regulatory guidelines in the French Healthcare system. *Int J Technol Assess Health Care* 2000; 16: 969-975.
52. Fervers B, Hardy J, Blanc-Vincent MP, Theobald S, Bataillard A, Farsi F, et al. SOR: project methodology. *Br J Cancer* 2001, 84(Suppl 2). 8-16.
53. Ollenschläger G, Oesingmann U, Thomeczek C, Kolkmann FW. Medical guidelines in Germany-current status and future developments. *Z Arztl Fortbild Qualitatssich* 1998; 92: 273-280.
54. Helou A, Lorenz W, Ollenschläger G, Reinauer H, Schwartz FW. Methodological standards of the evidence-based approach of clinical guidelines development in Germany. Consensus between the scientific community, self-governed bodies and practice. *Z Arztl Fortbild Qualitatssich* 2000; 94: 330-339.
55. Helou A, Perleth M, Bitzer EM, Dörning H, Schwartz FW. Methodological quality of clinical practice guidelines in Germany: results of a systemic assessment of guidelines presented on the Internet. *Z Arztl Fortbild Qualitatssich* 1998; 92: 421-428
56. Ollenschläger G, Helou A, Kostovic-Cilic L, Perleth M, Raspe HH, Rienhoff O, et al. Checklist for methodological quality of guidelines. A contribution to quality promotion of medical guidelines. *Z Arztl Fortbild Qualitatssich* 1998; 92: 191-194
57. Lauterbach KW, Lubecki P, Oesingmann U, Ollenschläger G, Richard S, Straub C. Concepts of a clearing procedure for guidelines in Germany. *Z Arztl Fortbild Qualitatssich* 1997; 91: 283-288.
58. Leitlinien-In-Fo. www.leitlinien.de.
59. Klazinga NS, Casparie AF, Van Everdingen JJE. Profile of the consensus development programme in the Netherlands. National Organisation for Quality Assurance in Hospitals (CBO). In Goodman C, Baratz SR (eds.) Council on Health Care Technology & Institute of Medicine. Improving Consensus Development for Health Technology Assessment: An International Perspective. Washington, DC National Academy Press, 1990: 110-118.
60. Grol R. Development of guidelines for general practice care. *Br J Gen Pract* 1993; 43: 146-151.
61. Thomas S. Standard setting in The Netherlands. impact of the human factor on guideline development *Br J Gen Pract* 1994; 44: 242-243.
62. Grol R, Thomas S, Roberts R. Development and implementation of guidelines for family practice: lessons from The Netherlands. *J Fam Pract* 1995; 40: 435-439.
63. Grol R. Successes and failures in the implementation of evidence-based guidelines for clinical practice. *Med Care* 2001; 39(Suppl 2): 46-54.
64. EBM-guidelines www.ebm-guidelines.com.
65. Jousimaa J, Kunnamo I, Makelä M. An implementation study of the PDRD primary care computerized guidelines *Scand J Prim Health Care* 1998, 16: 149-153.
66. Karma P, Makela M. Current clinical guideline project on the go. *Duodecim* 1997; 113: 9-10
67. Smallwood RA, Lapsley HM. Clinical practice guidelines: to what end? *Med J Aust* 1997; 166: 592-595.
68. National Health and Medical Research Council. A Guide to the Development, Implementation and Evaluation of Clinical Practice Guidelines. Canberra. National Health and Medical Research Council, 1999
69. National Health and Medical Research Council. Handbook Series on Preparing Clinical Practice Guidelines. www.health.gov.au/hfs/nhmrc/publicat/synopses/cp65syn.htm.
70. Scott CD. Reform of the New Zealand health care system. *Health Policy* 1994, 29: 25-40
71. Jackson R, Beaglehole R. Evidence-based management of dyslipidaemia *Lancet* 1995; 346: 1440-1442.

- 72 Jackson R, Sackett D Guidelines for managing raised blood pressure evidence based or evidence burdened? *BMJ* 1996, 313 64-65
- 73 Woolf SH Evidence-based medicine and practice guidelines an overview *Cancer Control* 2000, 4 362-367
- 74 Cook DJ, Greengold NL, Ellrodt AG, Weingarten SR The relation between systematic reviews and practice guidelines *Ann Intern Med* 1997, 127 210-216
- 75 Feder G, Eccles M, Grol R, Griffiths C, Grimshaw J Clinical guidelines using clinical guidelines *BMJ* 1999, 318 728-730
- 76 Eddy DM Clinical decision making from theory to practice Practice policies – what are they? *JAMA* 1990, 263 877-878
- 77 Hagen NA, Whyllie B Putting clinical practice guidelines into the hands of cancer patients *CMAJ* 1998, 158 347-348
- 78 Entwistle VA, Watt IS, Davis H, Dickson R, Pickard D, Rosser J Developing information materials to present the findings of technology assessments to consumers The experience of the NHS Centre for Reviews and Dissemination *Int J Technol Assess Health Care* 1998, 14 47-70
- 79 Woolf SH Shared decision-making the case for letting patients decide which choice is best *J Fam Pract* 1997, 45 205-208
- 80 Baker R, Fraser RC Development of review criteria linking clinical guidelines and assessment of quality *BMJ* 1995, 311 370-373
- 81 Hadorn DC, Baker DW, Kamberg CJ, Brook RH Phase II of the AHCPR-Sponsored heart failure guideline translating practice recommendations into review criteria *Jt Comm J Qual Improv* 1997, 22 265-276
- 82 Kotsanos JG, for the Health Outcomes Work Group of the Pharmaceutical Research and Manufacturers of America Development of performance measures for seven chronic diseases *Jt Comm J Qual Improv* 1997, 23 150-161
- 83 Hulshof NA, Geijer RMM, van Suijlekom LWA, van Essen EEM, Flikweert S, et al Twee kapiteins op een schip? De landelijke transmurale afspraak 'Asthma bij kinderen' [Two captains on board The national interdisciplinary agreement 'Asthma with children'] *Med Contact* 1999, 54 340-343
- 84 Capdenat Saint-Martin E, Michel PM, Raymond JM, Iskandar H, Chevalier C, Petitpierre MN, et al Description of local adaptation of national guidelines and of active feedback for rationalising preoperative screening in patients at low risk from anaesthetics in a French university hospital *Qual Health Care* 1998, 7 5-11
- 85 O'Brien JA, Jacobs LM, Pierce D Clinical practice guidelines and the cost of care *Int J Technol Assess Health Care* 2000, 16 1077-1091
- 86 Eccles M, Mason J, Freemantle N Developing valid cost effectiveness guidelines a methodological report from the North of England evidence based guideline development project *Qual Health Care* 2000, 9 127-132
- 87 O'Connor PJ, Solberg LI, Christianson J, Amundson G, Mosser G Mechanism of action and impact of a cystitis clinical practice guideline on outcomes and costs of care in an HMO *Jt Comm J Qual Improv* 1996, 22 673-682
- 88 Richman R, Lancaster DR The clinical guideline proces within a managed care organization *Int J Technol Asses Health Care* 2000, 16 1061-1076
- 89 Van Essen GA, Kuyvenhoven MM, Melker RA Implementing the Dutch College of General Practitioners' guidelines for influenza vaccination an intervention study *Br J Gen Pract* 1997, 47 25-29
- 90 Giacomini MK, Streiner DL, Anand SS Using practice guidelines to allocate medical technologies *Int J Technol Assess Health Care* 2000, 16 987-1002
- 91 Feek CM Rationing healthcare in New Zealand the use of clinical guidelines *Med J Aust* 2000, 173 423-426
- 92 Norheim OF Healthcare rationing - are additional criteria needed for assessing evidence based clinical practice guidelines? *BMJ* 1999, 319 1426-1429

93. De Neeling JND, namens de projectgroep Passende Zorg. Eindrapport 'Passende medische zorg. Over keuzen en richtlijnen' (Final report 'Appropriate healthcare. On priority setting and guidelines'). Utrecht: KNMG, 2000.
94. Mulder JH. Healthcare rationing in the Netherlands: the need for specific guidelines. *Med J Aust* 2000; 172: 329-331.
95. Naylor CD. Grey zones of clinical practice: some limits to evidence-based medicine. *Lancet* 1995; 345: 840-842.
96. Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Potential benefits, limitations, and harms of clinical guidelines. *BMJ* 1999; 318: 527-530.
97. Grol R, Wensing M, eds. Implementatie: Effectieve Verandering in de Patiëntenzorg (Implementation: Effective Change in Patient Care). Maarssen: Elsevier, 2001.
98. Connis RT, Nickinovich DG, Caplan RA, Arens JF. The development of evidence-based clinical practice guidelines. Integrating medical science and practice. *Int J Technol Assess Health Care* 2000; 16: 1003-1012.
99. Harbour R, Miller J, for the Scottish Intercollegiate Guidelines Network Grading Review Group. A new system for grading recommendations in evidence-based guidelines. *BMJ* 2001; 323: 334-336.
100. Sniderman AD. Clinical trials, consensus conferences, and clinical practice. *Lancet* 1999; 354: 327-330.
101. Choudhry NK, Stelfox HT, Detsky AS. Relationships between authors of clinical practice guidelines and the pharmaceutical industry. *JAMA* 2002; 287: 612-617.
102. Thomson R, McElroy H, Sudlow M. Guidelines on anticoagulant treatment in atrial fibrillation in Great Britain: variation in content and implications for treatment. *BMJ* 1998; 316: 509-513.
103. Lewis S. Paradox, process and perception: the role of organizations in clinical practice guidelines development. *CMAJ* 1995; 153: 1073-1077.
104. Tonelli MR. The philosophical limits of evidence-based medicine. *Acad Med* 1998; 73: 1234-1240.
105. Garfield FB, Garfield JM. Clinical judgement and clinical practice guidelines. *Int J Technol Assess Health Care* 2000; 16: 1050-1061.
106. Hurwitz B. Legal and political considerations of clinical practice guidelines. *BMJ* 1999; 318: 661-664.
107. Woolf SH. Practice guidelines: a new reality in medicine III. Impact on patient care. *Arch Intern Med* 1993; 153: 2646-2655.
108. Worrall G, Chaulk P, Freake D. The effects of clinical practice guidelines on patient outcomes in primary care: a systematic review. *CMAJ* 1997; 156: 1705-1712.
109. Knottnerus JA, Dinant GJ. Medicine based evidence, a prerequisite for evidence based medicine. *BMJ* 1997; 315: 1109-1110.
110. Tunis SR, Hayward RSA, Wilson MC, Rubin HR, Bass EB, Johnston M, et al. Internists' attitudes about clinical practice guidelines. *Ann Intern Med* 1994; 120: 957-963.
111. Kapp MB. 'Cookbook' medicine. A legal perspective. *Arch Intern Med* 1990; 150: 496-500.
112. Cretin S, Farley DO, Dolter KJ, Nicholas W. Evaluating an integrated approach to clinical improvement. *Med Care* 2001; 39: 70-83.
113. Thomas S, Grol R. Richtlijnen voor huisartsen, een vergelijking tussen Nederlands en de Verenigde Staten (Guidelines for family physicians, a comparison between the Netherlands and the United States). *Kwaliteit en Zorg* 1995; 4: 166-175.
114. Grol R, Eccles M, Maisonneuve H, Woolf S. Developing clinical practice guidelines: the European experience. *Dis Manage Health Outcomes* 1998; 4: 255-266.
115. Haycox A, Bagust A, Walley T. Clinical guidelines the hidden costs. *BMJ* 1999; 318: 391-393.
116. The Appraisal of Guidelines, Research and Evaluation in Europe Collaborative Group (AGREE). Guideline development in Europe. An international comparison. *Int J Technol Assess Health Care* 2000; 16: 1039-1049.
117. Ward JE, Grieco V. Why we need guidelines for guidelines: a study of the quality of clinical practice guidelines in Australia. *Med J Aust* 1996; 165: 574-576.

118. Cluzeau FA, Littlejohns P, Grimshaw JM, Feder G, Moran SE. Development and application of a generic methodology to assess the quality of clinical guidelines. *Int J Qual Health Care* 1999; 11: 21-28.
119. 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: 1900-1905
120. Grilli R, Magrini N, Penna A, Mura G, Liberati A. Practice guidelines developed by specialty societies: the need for a critical appraisal. *Lancet* 2000; 355: 103-105.
121. Graham ID, Calder LA, Hebert PC, Carter AO, Tetroe JM. A comparison of clinical practice guideline appraisal instruments. *Int J Technol Assess Health Care* 2000; 16: 1024-1038.
122. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials* 1995; 16: 62-73
123. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001; 357: 1191-1194.
124. Fahey TP, Peters TJ. What constitutes controlled hypertension? Patient based comparison of hypertension guidelines. *BMJ* 1996; 313: 93-96.
125. Eisinger F, Geller G, Burke W, Holtzman NA. Cultural basis for differences between US and French clinical recommendations for women at increased risk of breast and ovarian cancer. *Lancet* 1999; 353: 919-920.
126. Psaty BM, Furberg CD. British guidelines on managing hypertension. Provide evidence, progress, and an occasional missed opportunity. *BMJ* 1999; 319: 589-590.
127. Vogel N, Burnand B, Vial Y, Ruiz J, Paccaud F, Hohlfeld P. Screening for gestational diabetes: variation in guidelines. *Eur J Obstet Gynecol Reprod Biol* 2000; 91: 29-36
128. Rogers EM. *Diffusion of Innovations*, 3rd Edition. New York: Free Press, 1983
129. Berg M. *Rationalizing Medical Work. Decision Support Techniques and Medical Practices* [Thesis] Maastricht: Maastricht University, 1995.

Chapter 2

Towards evidence-based clinical practice: an international survey of 18 clinical guideline programmes

Jako S Burgers

Richard PTM Grol

Niek S Klazinga

Marjukka Mäkelä

Joost OM Zaat,

for The AGREE Collaboration

International Journal for Quality in Health Care; in press.

Abstract

Objective. To systematically describe the structures and working methods of guideline programmes.

Design. Descriptive survey using a questionnaire with 32 items based on a framework derived from the literature. Answers were tabulated and checked by participants.

Study participants. Key informants of 18 prominent guideline organisations in the United States, Canada, Australia, New Zealand, and nine European countries.

Main outcome measures. History, aims, methodology, products and deliveries, implementation, evaluation, procedure for updating guidelines, and future plans.

Results. Most guideline programmes were established to improve the quality and effectiveness of health care. Most use electronic databases to collect evidence and systematic reviews to analyse the evidence. Consensus procedures are used when evidence is lacking. All guidelines are reviewed before publication. Authorisation is commonly used to endorse guidelines. All guidelines are furnished with tools for application and the Internet is widely used for dissemination. Implementation strategies vary among different organisations, larger organisations leaving this to local organisations. Almost all have a quality assurance system for their programmes. Half of the programmes do not have formal update procedures.

Conclusions. Principles of evidence-based medicine dominate current guideline programmes. Recent programmes are benefiting from the methodology created by longstanding programmes. Differences are found in the emphasis on dissemination and implementation, probably due to differences in health care systems and political and cultural factors. International collaboration should be encouraged to improve guideline methodology and to globalise the collection and analysis of evidence needed for guideline development.

Introduction

Clinical practice guidelines are developed throughout the world to improve the quality of health care. The methods used to develop guidelines vary among organisations [1,2] and the quality of the methods has long given cause for concern [3,4]. With the growth of evidence-based medicine in the 1990s, there has been a shift from professional consensus to scientific rigor, employing systematic reviews and meta-analyses as the basis for developing valid guidelines [5]. In addition, Eddy introduced the 'explicit approach', in which the recommendations are linked to the supporting scientific evidence and the benefits, harms, and costs of the interventions are transparently presented (e.g., by using balance sheets) [6]. Ideally, the recommendations should be accompanied by a statement of the strength of the underlying evidence and expert judgment, as well as by projections of the relevant health consequences of alternative courses of care [7]. The effect of clinical guidelines on medical practice and their impact on patient care is, however, often limited [8,9]. Hence, guideline development needs to be complemented by evidence-based implementation [10]. All efforts should ultimately lead to 'evidence-based clinical practice' in which the clinician 'uses the best evidence available in consultation with the patient, to decide upon the option which suits that patient best' [11,12].

Guides for the development, implementation, and evaluation of clinical guidelines have been developed in different countries, such as Australia [13] and the United Kingdom [14], but it is not known whether the recommended approaches are actually used in current guideline programmes. Recent studies of international guideline activities were not conducted systematically or did not describe the content of existing guideline programmes [2,15-17].

The aim of this survey was to systematically describe the structures and working methods of current guideline programmes in different countries throughout the world, covering the entire scope of guideline development, dissemination, implementation, and evaluation. Our study was conducted within the context of an international research project—the *AGREE (Appraisal Guideline Research and Evaluation) project*—which aimed at harmonisation of guideline development methods in order to reduce duplication of efforts and to ensure efficient use of resources [18].

Methods

For this study we adopted the definition of the Institutes of Medicine (IOM) of *clinical practice guidelines* as 'systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances' [19]. Furthermore, we defined a *guideline programme* as 'a structured and coordinated programme designed with the specific aim of producing several clinical practice guidelines'.

Selection of guideline programmes

We aimed at studying a wide range of programmes from different countries. We only included national programmes or programmes with a high impact on a national level. The sample consisted of programmes from countries involved in the AGREE project, with a maximum of two programmes per country. To widen the scope, we also included the well-known technology-assessment programme from Sweden and the national guideline programme used in Australia as an example for other guideline organisations in that country. In all, eighteen guidelines programmes were selected.

Design of the questionnaire

We produced a conceptual framework covering relevant aspects of guideline programmes using criteria for guideline programmes from different authors. Starting points were the 'criteria for good guideline programmes' formulated by Lohr and the framework for guideline implementation studies produced by Mäkelä and Thorsen [20,21]. We also used the IOM provisional instrument and the Cluzeau instrument, both of which provide criteria for assessing clinical guidelines, among which some criteria could also be applied to guideline programmes [7,22]. The first version of the framework was tested by describing a few programmes. Valid information was difficult to obtain for some criteria, such as 'credibility of agency responsible for guideline development' and 'process for selection of panel members for the guideline development group'. These criteria were therefore discarded. Based on the final framework (Box 1) we designed a questionnaire that covered all of the items (Appendix B). The answer categories for some of the items were derived from the National Guideline Classification Scheme of the US National Guideline Clearinghouse [23].

Box 1. Framework for description of clinical guideline programmes**Basic characteristics of guideline organisation**

- name, country, website
- type of organisation¹
- historical details (year of first guideline, reason for guideline development)
- funding
- estimated budget for guideline development and dissemination

Purpose and topics

- objectives
- care level
- target users¹
- scope (screening/prevention/diagnosis/treatment)
- topic selection (who selects topics)

People involved in guideline development

- size of guideline development group
- number of disciplines in guideline development group
- involvement of experts (e.g., epidemiologists, statisticians, health economists)
- involvement of patients
- editorial support

Methodology of guideline development

- methodological training for group members
- methods used to collect evidence¹
- methods used to analyse evidence¹
- methods used to formulate recommendations¹
- methods of review¹
- authorisation

Products and deliveries

- total number of guidelines produced
- average size of guideline (number of pages)
- guideline products (e.g., extensive/short/patient versions)
- tools for application (e.g., algorithms/flow charts, balance sheets, risk tables)
- media used (paper/CD-ROM/Internet)

Implementation, evaluation and update procedure

- implementation strategies (e.g., educational materials, conferences, audit and feedback)
- use of monitoring and documentation
- quality system for guideline programme (e.g., use of quality criteria, guideline clearinghouse)
- update procedure

Future plans

- plans for further development of guideline programme in the near future

¹ For these items answer categories were derived from the US National Guideline Clearinghouse Classification Scheme²³

Data collection and analysis

The questionnaire was sent to key informants of the guideline programmes. These were persons in a leading role in the guideline development organisation or having long experience in the guideline programme. Their answers were tabulated in

simple linear classes (as shown in the Tables below), except answers to open questions, which were summarised in short statements. When responses were not clear, we sent 4-8 additional specific questions to the key informant. For validation, we sent the first draft of the results back to the informants, asking them to check our interpretations. They did so, enabling them at the same time to compare their responses with those of others, and all gave their approval of our interpretations.

Results

All key informants responded to the original questionnaire and the validation procedures.

Basic characteristics of guideline organisations

Nine organisations were professional societies, six were governmental agencies, two were national (or central) but not governmental, and one was an academic institution (Table 1). In the late 1970s, the National Institutes of Health Consensus Development Program led the development of consensus statements. The Dutch and the Swedish organisations started guideline development in the 1980s, the others in the 1990s or in 2000. The common reasons given for establishing a guideline programme were to improve the quality of health care, to support evidence-based care, to improve cost-effectiveness of care, and to contribute to more effective care. Some programmes were intended to increase equity or to strengthen the medical profession, or were part of a research effort.

All guideline programmes except the Swiss programme receive governmental support. Some agencies are funded exclusively by the government, but usually professional organisations also fund guideline development. The average budget for developing a single guideline varies from US \$10,000–25,000 in New Zealand to \$200,000 in the United States. The differences in the budget for dissemination are even larger, varying from nothing to \$200,000 per guideline.

Purpose and topics

All of these guideline programmes aim at appropriate clinical care and six of the programmes also attempt to contain health care costs (Table 2). Most programmes target primary as well as secondary care and their guidelines have a broad scope that covers prevention, diagnosis, and management of a wide range of clinical topics. Two of the programmes focus on prevention and two are limited to cancer care. Four programmes target primary care exclusively and two are for

hospital specialists. Eleven programmes also consider patients and policymakers as target users of their guidelines.

In most programmes the organisation responsible for guideline development coordinates the selection of topics for guidelines. People outside the organisation—in some cases policymakers or health authorities—can propose topics for guideline development. In Italy the topics are identified by the national health plan.

People involved in guideline development

Guideline development groups are typically fairly large, consisting of 10 to 20 members (Table 3). In four programmes, smaller groups are preferred, and two programmes have groups of more than 20 persons. The number of disciplines per group is most often three to five. Most programmes invite methodological experts to participate, typically epidemiologists (15 programmes), library scientists (12), statisticians (4), communication experts (4), health economists (3), and clinical or social psychologists (2). The remaining three programmes include experts if they are necessary. Patient representatives participate in guideline groups in eleven programmes and are involved during the prerelease review in two. Editorial support is given by permanent guideline staff in 14 programmes (four of which also employ temporary staff per guideline) and by temporary committees only in four programmes, while one programme has no systematic arrangement for editorial support.

Methodology of guideline development

Training in the methodology of guideline development is offered to the members of the guideline development group in almost all programmes. In seven programmes the training is obligatory for all group members. All guideline programmes use electronic database searches to collect evidence and most also use searches by hand (Table 4). Most evidence is analysed by systematic reviews, supported in two programmes by decision analyses. All but one programme link recommendations to evidence, and seven programmes use formal consensus methods to formulate recommendations. External review is used in all but one programme, and the majority ask for formal authorisation from outside. Guideline comparison is used in seven programmes, pilot testing before release in two. Authorisation by the professional organisation of target users is generally employed to endorse the guidelines.

Table 1. Basic characteristics of guideline organisations

Country	Name of organisation (acronym/short name)	Type of organisation	Website	Year of first guideline	Reason for guideline development
Australia	National Health and Medical Research Council (NHMRC)	National government	www.health.gov.au/hfs/nhmrc	1995	Pilot test development of national standards for others to follow
Canada	Cancer Care Ontario Practice Guidelines Initiative (CCOPGI)	Provincial government	www.cancercare.on.ca/ccopgi	1994	Facilitate evidence based-decision making in cancer care
Denmark	Danish College of General Practitioners (DSAM)	Professional	www.dsam.dk	1998	Quality improvement
England	Centre for Health Services Research, University of Newcastle-upon-Tyne (North of England)	Academic institution	www.ncl.ac.uk/chsr	1995	Develop guidelines for research purposes
	Royal College of Physicians London (RCP London)	Professional umbrella	www.rcplondon.ac.uk	1990	Provide guidance for multidisciplinary management of patient problems after stroke
Finland	Finnish Medical Society Duodecim (Duodecim)	Professional umbrella	www.duodecim.fi	1997	Provide necessary instruments for evidence-based, equitable cost-effective health care
France	Agence Nationale d'Accréditation et d'Évaluation en Santé (ANAES)	National government	www.anaes.fr	1993	Improvement of quality of care
	Fédération Nationale des Centres de Lutte Contre le Cancer (FNCLCC)	Professional umbrella	www.fnclcc.fr/sor.htm	1993	Professional initiative to improve quality of cancer care following evidence for practice variation at local, regional and national level
Germany	Association of the Scientific Medical Societies in Germany (AWMF)	Professional umbrella	www.awmf.de	1992	Quality improvement and improvement of clinical research, now recommended by the High Advisory Board of the Federal Ministry of Health
Italy	Agency for Regional Health Services (ASSR)	Central but not governmental	www.assr.it	2000	Provide tools to be used at regional level in the promotion of effective and appropriate use of health services

Country	Name of organisation (acronym/short name)	Type of organisation	Website	Year of first guideline	Reason for guideline development
Netherlands	Dutch Institute for Healthcare Improvement CBO	Professional	www.cbo.nl	1980	Support medical audit in hospitals
	Dutch College of General Practitioners (NHG)	Professional	www.nhg.artsennet.nl	1989	Professionalisation of GP's with formulating state of the art in order to give other parties (specialists, government) a clear view of their competence
New Zealand	New Zealand Guidelines Group (NZGG)	National but not governmental	www.nzgg.org.nz	1998	Reduce gap between current and appropriate care
Scotland	Scottish Intercollegiate Guidelines Network (SIGN)	Professional	www.sign.ac.uk	1995	Support health care quality improvement by promoting effective clinical care to reduce variation in clinical practice
Sweden	Swedish Council on Technology Assessment in Health Care (SBU)	National Government	www.sbu.se	1989	Not applicable
Switzerland	Swiss Medical Association (FMH)	Professional	www.fmh.ch	2000	One component of a global programme to promote quality of care in Switzerland
United States	US Preventive Services Task Force (USPSTF)	National Government	www.ahrq.gov/clinic/uspstfix.htm	1989	Confusion over preventive care, reluctance of insurers to cover preventive care, and reluctance of providers to provide preventive care
	National Institutes of Health Consensus Development Program (NIHCDP)	National Government	consensus.nih.gov	1977	Facilitate translating medical scientific findings into practice

Table 2. Purpose and topics

Organisation	Objectives	Level of care	Target users	Scope of guidelines	Who selects topics?
NHMRC (Australia)	Appropriate care	Primary, secondary, and tertiary care	Physicians, nurses, patients, health care organisations, hospitals, policymakers	Diagnosis, treatment/management	Initially NHMRC, more recently specialist colleges
CCOPGI (Canada)	Appropriate care, cost containment, equal access to healthcare	Primary, secondary, and tertiary care	Physicians, patients, policy makers, regional cancer systems	Screening, prevention, diagnosis, treatment/management	Guideline development committees, policy advisory committee may request for guideline on new, expensive drug
DSAM (Denmark)	Appropriate care, cost containment	Primary care	Family physicians	Prevention	Danish College of GP's
North of England	Appropriate care, cost containment	Primary care	Physicians, nurses	Treatment, management	Program leader
RCP London (England)	Appropriate care	Secondary and tertiary care	Physicians, paramedics, nurses, patients, health care organisations, hospitals, policymakers	Prevention, diagnosis, treatment/management	Department of Health and National Institute for Clinical Excellence (NICE)
Duodecim (Finland)	Appropriate care, equitability	Public health, primary, secondary, and tertiary care	Physicians, paramedics, nurses, patients, health care organisations, hospitals, policymakers	Screening, prevention, diagnosis, treatment/management	Current Care board (representatives of major stakeholders)
ANAES (France)	Appropriate care, cost containment	Primary, secondary, and tertiary care	Physicians, paramedics (in private and public settings)	Screening, prevention, diagnosis, treatment/management	ANAES, medical speciality societies, health insurance
FNCLCC (France)	Appropriate care, cost containment	Public health, secondary care	Physicians, paramedics, nurses, patients, health care organisations, hospitals, policymakers	Diagnosis, treatment/management	Professionals, scientific committee, commissioned by national government agency
AWMF (Germany)	Appropriate care	Primary, secondary, and tertiary care	Health care providers, self-governmental bodies in health care	Screening, prevention, diagnosis, treatment/management	Medical speciality organisations
ASSR (Italy)	Appropriate care and organisation of health services	Public health, primary, secondary, and tertiary care	Physicians, nurses, patients, health care organisations, hospitals, policymakers	Screening, prevention, diagnosis, treatment/management	National health plan issued by Ministry of Health

Organisation	Objectives	Level of care	Target users	Scope of guidelines	Who selects topics?
CBO (Netherlands)	Appropriate care, effective health care	Secondary care	Physicians, paramedics, nurses	Screening, prevention, diagnosis, treatment/management	Committee of independent medical specialists and hospitals
NHG (Netherlands)	Appropriate care	Primary care	Family physicians	Screening, prevention, diagnosis, treatment/management	Independent advisory board of family physicians
NZGG (New Zealand)	Appropriate care, cost-effectiveness	Primary, secondary, and tertiary care	Physicians, paramedics, nurses, patients, health care organisations, hospitals, policymakers	Screening, prevention, diagnosis, treatment/management	Practitioners, using suitability screen developed by NZGG
SIGN (Scotland)	Appropriate care, reduce variation in clinical practice	Public health, primary, secondary, and tertiary care	Physicians, paramedics, nurses, patients, health care organisations, hospitals, policymakers	Screening, prevention, diagnosis, treatment/management	SIGN council (representatives from medical specialist societies, medical colleges, and funding organisation)
SBU (Sweden)	Appropriate care, cost-effectiveness	Public health, primary, secondary, and tertiary care	Physicians, paramedics, nurses, patients, health care organisations, hospitals, policymakers	Screening, prevention, diagnosis, treatment/management	SBU board
FMH (Switzerland)	Appropriate care	Public health, primary, secondary, and tertiary care	Physicians	Screening, prevention, diagnosis, treatment/management	Swiss Medical Association, specialist societies
USPSTF (USA)	Appropriate care	Primary care	Physicians, nurses, health care organisations, hospitals, policymakers	Screening, prevention, diagnosis	USPSTF members, with input from outside groups including primary care professional societies, prevention experts, government health experts
NIHCDP (USA)	Appropriate care	Public health, primary, secondary and tertiary care	Physicians, paramedics, nurses, patients, health care organisations, hospitals, policymakers	Screening, prevention, diagnosis, treatment/management	The NIH Office of Medical Applications of Research with input from NIH institute directors

Table 3. People involved in guideline development

Organisation	Composition of guideline development group				Editorial support
	Average number of members	Average number of disciplines	Experts always Involved (beyond clinical experts)	Involvement of patients	
NHMRC (Australia)	10 - 15	3 - 5	Epidemiologists, health economists	Yes	Standing staff
CCOPGI (Canada)	15 - 20	3 - 5	Library scientists, epidemiologists, staticians, communication experts	Yes	Standing staff
DSAM (Denmark)	5 - 10	3	Only if necessary	No	Standing staff and hearing staff
North of England	10 - 15	3 - 5	Epidemiologists, health economists	Yes	Standing staff
RCP London (England)	> 20	> 5	Library scientists, epidemiologists, clinical psychologists	Yes	Committee that varies for different guidelines
Duodecim (Finland)	5 - 10	3 - 5	Library scientists, epidemiologists	No ¹	Standing staff and appointed 'group editor'
ANAES (France)	> 20	3 - 5	Library scientists, epidemiologists	No	Standing staff and committee that varies for different guidelines
FNCLCC (France)	10 - 15	3 - 5	Library scientists, epidemiologists, staticians	Yes	Standing staff
AWMF (Germany)	5 - 10	1 - 20	Library scientists, epidemiologists, social psychologists	No ¹	Committee that varies for different guidelines
ASSR (Italy)	10 - 15	0 - 3	Library scientists, epidemiologists, staticians, communication experts	No	Committee that varies for different guidelines
CBO (Netherlands)	15 - 20	> 5	Library scientists, epidemiologists	Yes	Standing staff
NHG (Netherlands)	5 - 10	0 - 3	Only if necessary	No	Standing staff
NZGG (New Zealand)	10 - 15	3 - 5	Epidemiologists	Yes ²	Standing staff and budgeted for each guideline
SIGN (Scotland)	15 - 20	> 5	Library scientists, epidemiologists	Yes	By standing staff and committee that varies for different guidelines
SBU (Sweden)	10 - 15	3 - 5	Library scientists, epidemiologists, communication experts, health economists	Yes	Chairman and project co-ordinator
FMH (Switzerland)	10 - 15	3 - > 5	Only if necessary	No	No usual support, varies for different guidelines
USPSTF (USA)	10 - 15	> 5	Library scientists, epidemiologists	No	Standing staff
NIHCDD (USA)	15 - 20	5	Library scientists, epidemiologists, staticians, communication experts	Yes	Standing staff

¹ patients are not member of guideline development group but are involved by reviewing representatives for patient organisations

² usually consumer groups representatives rather than patients

Table 4. Methodology of guideline development

Organisation	Methods used to collect the evidence ¹	Methods used to analyse the evidence ²	Methods used to formulate recommendations ³	Method of review ⁴
NHMRC (Australia)	By hand, electronic	Meta, systematic	Evidence, informal	Comparison, external
CCOPGI (Canada)	By hand, electronic, unpublished data	Meta, systematic	Evidence, formal	External, internal
DSAM (Denmark)	By hand, electronic	Systematic	Informal	Comparison, internal
North of England	Electronic	Meta, systematic	Evidence, informal	External
RCP London (England)	Electronic	Meta, systematic, non-systematic, experience	Evidence, formal, informal	Comparison, external
Duodecim (Finland)	Electronic	Meta, systematic	Evidence, informal	External, internal
ANAES (France)	By hand, electronic, unpublished data	Systematic, experience	Evidence, formal, informal	Pilot, external, internal
FNCLCC (France)	By hand, electronic	Systematic, non-systematic, experience	Evidence, informal	External, internal
AWMF (Germany)	By hand, electronic	Meta, systematic, non-systematic, experience	Evidence, formal, informal, subjective	Pilot, comparison, external, internal
ASSR (Italy)	Electronic	Meta, systematic	Evidence, formal	Pilot, comparison, external, internal
CBO (Netherlands)	By hand, electronic	Systematic, non-systematic	Evidence, formal, informal	External
NHG (Netherlands)	By hand, electronic	Non-systematic, experience	Evidence, informal	External, internal
NZGG (New Zealand)	By hand, electronic, patient data, unpublished data	Decision, meta, systematic, non-systematic, experience	Evidence	Comparison, external, internal
SIGN (Scotland)	By hand, electronic	Systematic, experience	Evidence, informal	External, internal
SBU (Sweden)	By hand, electronic, patient data, unpublished data	Decision, meta, systematic	Evidence	External
FMH (Switzerland)	By hand, electronic	Systematic, non-systematic, experience	Evidence, informal	External, internal
USPSTF (USA)	Electronic	Meta, systematic	Evidence	Comparison, external, internal
NIHCDP (USA)	By hand, electronic, unpublished data	Meta, systematic	Evidence, formal	External

¹ By hand = hand searches of published literature, electronic = searches of electronic databases, patient data = searches of patient registry data, unpublished data = searches of unpublished data

² Decision = decision-analysis, experience = experience based, meta = meta-analysis, non-systematic = non-systematic review, systematic = systematic review

³ Evidence = evidence-linked, formal = formal expert consensus, informal = informal expert consensus, subjective = subjective review

⁴ Comparison = comparison with guidelines from other groups, external = external peer review, internal = internal peer review, pilot = pilot testing

Products and deliveries

Longstanding programmes have produced more guidelines than those started recently (Table 5). Average guideline length varies among programmes, but guidelines tend to consist of more than 15 pages. Guidelines are usually presented in both a summary or short version and an extended version with notes or references or both. Eleven programmes also produce patient versions. Almost all programmes develop tools for application, such as flow charts or algorithms. Balance sheets are produced in three programmes and risk tables in four. All but one programme provide their guidelines on the Internet.

Implementation, evaluation, and update procedure

A wide range of strategies is used to implement guidelines and the strategies vary according to guideline topics (Table 6). Most often used are educational materials and conferences. A specific group of ANAES guidelines is implemented by health insurers using financial disincentives. Some agencies do not take responsibility for implementing their guidelines but leave this to regional or local organisations. More than half of the programmes monitor or evaluate the effects of at least some guidelines. Almost all programmes use some type of quality system for good guideline development. Five organisations submit their guidelines to a guideline clearinghouse. All programmes report that they update their guidelines at least occasionally. Half of the programmes do not have formal update procedures.

Table 5. Products and deliveries

Organisation	Total number of guidelines	Average number of pages	Products ¹	Media used
NHMRC (Australia)	10 - 20	> 50	Extensive, short, summary, patient, flowcharts	Paper, Internet
CCOPGI (Canada)	30 - 50	15 - 25	Extensive, summary, patient, flowcharts	Paper, CD-ROM, Internet
DSAM (Denmark)	0 - 10	15 - 25	Extensive, summary, flowcharts, risk tables	Paper, CD-ROM, Internet
North of England	0 - 10	> 50	Extensive, summary, balance sheets	Paper, Internet
RCP London (England)	0 - 10	> 50	Extensive, summary, patient	Paper, Internet
Duodecim (Finland)	20 - 30	15 - 50	Extensive, short, patient, flowcharts	Paper, CD-ROM, Internet
ANAES (France)	> 50	> 50	Extensive, short, summary ² , patient ² , flowcharts	Paper, Internet
FNCLCC (France)	> 50	25 - 50	Extensive, short, patient, flowcharts	Paper, CD-ROM, Internet
AWMF (Germany)	> 50	15 - 25	Extensive, short, summary, patient, flowchart	Paper, Internet
ASSR (Italy)	0 - 10	25 - 50	Extensive, summary, patient ² , flowcharts ²	Paper, Internet
CBO (Netherlands)	> 50	25 - 50	Extensive, short, summary, risk tables	Paper, Internet
NHG (Netherlands)	> 50	10 - 15	Extensive, summary, patient, risk tables	Paper
NZGG (New Zealand)	5 - 10	25 - 50	Extensive, short, summary, patient, flowcharts, balance sheets, risk tables	Paper, Internet
SIGN (Scotland)	30 - 50	25 - 50	Extensive, summary, flowcharts	Paper, CD-ROM, Internet
SBU (Sweden)	30 - 50	> 50	Extensive, short, patient	Paper, Internet
FMH (Switzerland)	0 - 10	10 - 15	Extensive, flowcharts	Paper, Internet
USPSTF (USA)	> 50	10 - 15	Extensive, short, summary, balance sheets	Paper, Internet
NIHCDP (USA)	> 50	15 - 25	Extensive, short	Paper, Internet

¹ Extensive = extensive version with notes/references, flow charts = flow charts /algorithms, patient = patient version, short = short version, summary = one or two pages summary

² planned products, thus not available yet

Table 6. Implementation, evaluation, and update procedure

Organisation	Implementation strategies ¹	Use of monitoring	Quality system ²	Update procedure
NHMRC (Australia)	Educational, conferences, leaders, visits, audit, organisational	Yes	Criteria, comments, appraisal	Not formal
CCOPGI (Canada)	Educational, conferences, leaders	Yes	Comments, clearing house	Formal, regular
DSAM (Denmark)	Educational, conferences, leaders, visits, audit, organisational, financial ³	Yes	Criteria, comments	Formal, every 2-3 year
North of England	Educational, conferences, visits ⁴ , reminders ⁴	No	Criteria, comments	Not formal, irregular
RCP London (England)	Educational, conferences, leaders, visits, audit, patient, organisational	Yes	Comments, clearing house	Formal, regular
Duodecim (Finland)	Educational, conferences, visits, audit, organisational	Yes, for some guidelines	Criteria, comments	Formal, every 2 year
ANAES (France)	Educational, leaders, audit, organisational, financial ⁵	No	Criteria	Formal, irregular
FNCLCC (France)	Educational, conferences, leaders, audit, reminders, organisational	No	Criteria, comments	Formal, irregular
AWMF (Germany)	Educational, conferences, leaders, audit, patient, organisational, financial ⁶	Yes, for some guidelines	Criteria, comments, appraisal, clearing house	Not formal, regular
ASSR (Italy)	Educational, audit, reminders	Yes	Criteria, appraisal	Not yet but planned
CBO (Netherlands)	conferences, audit	No	Criteria	Formal, every 5 year
NHG (Netherlands)	Educational, conferences, visits, reminders, organisational, financial ³	Yes	Comments	Formal, every 3 year
NZGG (New Zealand)	Educational, conferences, leaders, audit, organisational	Yes	Criteria, comments, appraisal	Formal, irregular
SIGN (Scotland)	Conferences, leaders, organisational	Yes	Criteria, comments, clearing house	Formal, every 2 year
SBU (Sweden)	Educational, conferences, leaders, visits, organisational ³	Yes	Not applicable	Formal, every 2-3 year
FMH (Switzerland)	Conferences	Yes	Criteria	Formal, regular
USPSTF (USA)	Conferences, reminders ⁷	No	Criteria, comments, clearing house	Not formal, regular
NIHCDP (USA)	Educational, conferences	Yes	Not available	Not formal, irregular

¹ Audit = audit and feedback, educational = educational materials, financial = financial incentives, leaders = local opinion leaders, organisational = organisational interventions, patient = patient mediated interventions, reminders = (computer) reminders

² Appraisal = appraising existing guidelines, comments = revising guidelines based on comments from the professional community, criteria = developing and publishing criteria for good guideline development ('guidelines for guidelines'), clearing house = submitting guidelines to guideline clearing house

³ for one guideline ⁵ used by health insurers

⁴ used in implementation trials ⁶ strategies vary between different medical societies

⁷ computerised systems are developed by others

Future plans

The future plans of guideline programmes reflect active development. Nine programmes consider themselves to be in a transitional phase, so their plans are evolving very rapidly. Plans for better management of the quality of the guideline process were mentioned by half of the respondents. Many of them plan to increase the amount of training for guideline development groups. Four programmes intend to create a strategy for more active implementation or dissemination. The remaining plans are divided evenly among creation of a better evidence base, more patient involvement, better updating procedures, increased attention to cost-effectiveness or economic issues in guidelines, and more international collaboration. The only programme that does not yet present its guidelines on the Internet has plans to do so, and one programme plans to translate its native-language guidelines into English.

Discussion

All of the guideline programmes included in this study intend to develop clinical guidelines rigorously. While their integration in health care systems varies, there appears to be a trend in guideline development methodology toward the increasing use of similar procedures. Various organisations seem to be using the available information on good guideline development methods, and newcomers are modeling their programmes on existing programmes. In particular, the evidence-based approach (i.e., using electronic database searches, systematic review, and evidence linkage) is being adopted with greater consistency by all organisations. Longstanding programmes do not necessarily have stricter procedures for development than more recent programmes, but the governmental agencies in our sample tend to utilise more quality assurance measures than do professional societies.

Most guideline programmes combine an evidence-based approach with formal or informal consensus procedures. In particular, when evidence is contradictory, controversial, or lacking, consensus procedures are needed to solve problems in health care. Consensus could be considered to be an additional source of evidence when it is obtained from formal surveys of experts and the broader population of practitioners, or from feasibility studies [24]. Exploring and comparing existing guidelines could provide additional insight into how evidence and consensus could be combined.

While the programmes share basic principles, we found some important differences in the details. Patients are not involved in all programmes and pilot testing and guideline comparison are only used in a few. National agencies take less responsibility for implementation of guidelines than do professional organisations. Larger organisations seem to prefer leaving implementation to regional and local organisations, while guideline development organisations in smaller countries are more involved in implementing their guidelines. Finally, professional organisations use more formal update procedures than do other organisations.

Differences among guideline programmes could be partly due to differences in resources. For instance, governmental agencies have larger budgets for guideline development, which could explain why their guideline development groups include more members and more disciplines than those of the professional organisations. Differences in scope and purpose due to different health care systems and political and cultural factors could explain differences in dissemination and implementation strategies.

Even with small budgets, professional organisations can develop high-quality guidelines if they work within a structured programme, adopting quality criteria of other programmes and using evidence collected elsewhere. In some countries guideline development is facilitated by large government-funded organisations, such as the Agency for Healthcare Research and Quality in the United States, the National Health and Medical Research Council in Australia, and the National Institute for Clinical Excellence (NICE) in England and Wales.

The future plans show that guideline organisations are aiming at active international collaboration. There is a growing awareness that cooperative partnerships such as the AGREE Collaboration (Appendix A) may contribute to improving methods of guideline development, implementation, and evaluation and to avoiding duplication of efforts. International databases such as the Cochrane Library are useful sources of evidence but only provide part of the evidence needed for guidelines. Pragmatic approaches are needed for subjects not covered by existing reviews [25]. Browman suggested establishing a registry of clinical guidelines under development [26]. Guideline organisations could thereby benefit from the evidence collected and work done by others. Exchanging guidelines that fulfill agreed quality criteria [18], as in using guideline clearinghouses [23,27], and sharing the monitoring of emerging literature in order to keep guidelines up-to-date will prevent duplication of effort [28]. In this way the collection and analysis of

evidence can become a worldwide effort. However, the formulation of recommendations will still depend on country-specific or local decisions, influenced by professional and cultural values and considerations of the cost of applying the evidence. Therefore, aiming for international guidelines will probably be 'a step too far' [29].

Our study is the most recent survey of clinical guideline programmes throughout the world. McGlynn et al. [3] did similar work in 1990 but only included consensus development conference programmes. The studies of Audet et al. and the Institutes of Medicine [4,7] only covered American programmes. In contrast, we collected structured information on 18 organisations responsible for guideline development programmes from the United States, Canada, Australia, New Zealand, and nine European countries. We did not aim at conducting a comprehensive review of guideline programmes. Many programmes, in particular those of professional organisations in Canada and the United States, were not included in our sample. Nevertheless, by providing models of good guideline development in each country, our sample can be considered to be representative of large national programmes with a high impact.

Conclusions

Principles of evidence-based medicine have largely affected the methodology of guideline development. Consensus on the essential features of guideline programmes is growing. Recent new programmes are benefiting from the more advanced methodology created by experienced, longstanding programmes. However, there are still differences between programmes with respect to the ownership (i.e., governmental agencies versus professional organisations) and the emphasis on dissemination and implementation. International collaboration should be encouraged, to improve guideline methodology and to promote worldwide collection and analysis of evidence needed for guideline development. Patient involvement could be improved, to enhance the use of guidelines in practice. Thus, we may anticipate that, ultimately, evidence-based guidelines will lead to evidence-based clinical practice.

References

1. Woolf SH. Practice guidelines: a new reality in medicine. II. Methods of developing guidelines. *Arch Intern Med* 1992; 152: 946-952.
2. Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Potential benefits, limitations, and harms of clinical guidelines. *BMJ* 1999; 318: 527-530.
3. McGlynn EA, Koseoff J, Brook RH. Format and conduct of consensus development conferences. *Int J Technol Assess Health Care* 1990; 6: 450-469.
4. Audet A, Greenfield S, Field M. Medical practice guidelines: current activities and future directions. *Ann Intern Med* 1990; 113: 709-714.
5. Grimshaw J, Russell I. Achieving health gain through clinical guidelines. I: Developing scientifically valid guidelines. *Qual Health Care* 1993; 2: 243-248.
6. Eddy DM. A Manual for Assessing Health Practices and Designing Practice Policies. Guidelines for the Development and Implementation of Clinical Practice Guidelines: the Explicit Approach. Philadelphia: American College of Physicians, 1992.
7. Institute of Medicine. Field MJ, Lohr K (eds). Guidelines for Clinical Practice: from Development to Use. Washington DC: National Academy Press, 1992.
8. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 1993; 342: 1317-1322.
9. Bero LA, Grilli R, Grimshaw JM, Harvey E, Oxman AD, Thomson MA. Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. *BMJ* 1998; 317: 465-468.
10. Grol R, Grimshaw J. Evidence-based implementation of evidence-based medicine. *Jt Comm J Qual Improv* 1999; 25: 503-510.
11. Mur Gray JA. Evidence-based Healthcare. How to make Health Policy and Management Decisions. London: Churchill Livingstone, 1997.
12. Lohr KN, Eleazer K, Mauskopf J. Health policy issues and applications for evidence-based medicine and clinical practice guidelines. *Health Policy* 1998; 46: 1-19.
13. National Health and Medical Research Council. A Guide to the Development, Implementation and Evaluation of Clinical Practice Guidelines. Canberra: National Health and Medical Research Council, 1998.
14. Scottish Intercollegiate Guidelines Network (SIGN). A guideline developers' handbook. SIGN Publication No. 50: <http://www.sign.ac.uk>.
15. Grol R, Eccles M, Maisonneuve H, Woolf S. Developing clinical practice guidelines: The European experience. *Dis Manage Health Outcomes* 1998; 4: 255-266.
16. The AGREE Collaborative Group. Guideline development in Europe: an international comparison. *Int J Technol Assess Health Care* 2000; 16: 1039-1049.
17. Engelbrecht R, Courté-Wienecke S, eds. A Survey on The Current State of Development, Dissemination and Implementation of Guidelines of Clinical Practice in European Countries. Neuherberg: GSF – National Research Center for Environment and Health, 1999. Available at: <http://www-mi.gsf.de/proguide/survey.htm>.
18. The AGREE Collaboration: <http://www.agreecollaboration.org>.
19. Institute of Medicine. Field MJ, Lohr K (eds). Clinical Practice Guidelines: Directions for a New Program. Washington DC: National Academy Press, 1990.
20. Lohr KN. The quality of practice guidelines and the quality of health care. In: Selbmann HK. Guidelines in Health Care: Report of a WHO-Conference. Baden-Baden: Nomos Verlagsgesellschaft, 1998: 42-52.
21. Makela M, Thorsen T. A framework for guidelines implementation studies. In: Changing Professional Practice: Theory and Practice of Clinical Guidelines Implementation. Copenhagen: Danish Institute for Health Services, 1999: 23-53.
22. Cluzeau F, Littlejohns P, Grimshaw J, Feder G, Moran SE. Development and application of a generic methodology to assess the quality of clinical guidelines. *Int J Qual Health Care* 1999; 11: 21-28.
23. National Guideline Clearinghouse, United States: <http://www.guideline.gov/index.asp>.
24. Connis RT, Nickinovich DG, Arens JF. The development of evidence-based clinical practice guidelines. *Int J Technol Assess Health Care* 2000; 16: 1003-1012.
25. Browman GP. Improving clinical practice guidelines for the 21st century: attitudinal barriers and not technology are the main challenges. *Int J Technol Assess Health Care* 2000; 16: 959-968.

26. Browman GP. Development and aftercare of clinical guidelines: the balance between rigor and pragmatism. JAMA 2001; 286: 1509-1511.
27. German Guideline Information Service: <http://www.leitlinien.de>.
28. Baker R, Feder G. Clinical guidelines: where next? Int J Qual Health Care 1997; 9: 399-404.
29. De Maeseneer J, Derese A. European general practice guidelines: a step too far? Eur J Gen Pract 1999; 5: 86-87.

Chapter 3

Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project

The AGREE Collaboration (see page 192 for contributions)

Quality and Safety in Health Care; in press.

Abstract

Background. International interest in clinical practice guidelines has never been greater. However many published guidelines do not meet even the basic quality requirements. There have been renewed calls for validated criteria that can be used to assess the quality of guidelines.

Objective. To develop and validate an international instrument for assessing the quality of the process and reporting of clinical practice guideline development.

Methods. The instrument was developed through a multi-staged process of item generation, selection and scaling, field testing and refinement procedures. 100 guidelines selected from 11 participating countries were evaluated independently by 194 appraisers with the instrument. Following refinement the instrument was further field tested on 3 guidelines per country by a new set of 70 appraisers.

Results. The final version of the instrument contains 23 items grouped into six quality domains with a 4-point Likert scale to score each item (scope and purpose, stakeholder involvement, rigour of development, clarity and presentation, applicability, editorial independence). 95% of appraisers found the instrument useful to assess guidelines. Reliability was acceptable for most domains (Cronbach's Alpha ranged from 0.64 to 0.88). Guidelines produced as part of an established guideline programme had significantly higher scores on editorial Independence and after the publication of a national policy had significantly higher quality scores on rigour of Development ($p < 0.05$). Guidelines with technical documentation had higher scores on that domain ($p < 0.01$).

Conclusions. This is the first time an appraisal instrument for clinical practice guidelines has been developed and tested internationally. The instrument is sensitive to differences in important aspects of guidelines, and can be used consistently and easily by a wide range of professionals from different backgrounds. The adoption of common standards should improve the consistency and quality of the reporting of guideline development worldwide and provide a framework to encourage international comparison of clinical practice guidelines.

Introduction

Clinical practice guidelines are now a common feature of clinical practice and are of interest worldwide. They are expected to facilitate more consistent, effective and efficient medical practice, and improve health outcomes [1]. Governments, professional associations and healthcare organisations are increasingly sponsoring the development and dissemination of clinical guidelines [2]. There is also a growing number of guidelines developed by European or international groups.

Although the principles for the development of sound guidelines are well established [3-5], many published guidelines fall short of the basic quality criteria identified in two recent studies [6,7]. Defining the quality of guidelines is not straightforward. In principle a 'good' guideline is one that eventually leads to improved patients outcome. It needs to be scientifically valid, usable and reliable. However, this evidence is rarely available. Often, the best that can be expected is some information on whether the guideline producers have attempted to minimise all the biases that can occur in the complex process of creating a guideline and how well this is reported.

As the number of published guidelines proliferates there have been calls for the establishment of internationally recognised standards to improve the development and reporting of clinical guidelines [6]. Moreover there is a pressing need for internationally recognised criteria that are valid, reliable and useful for various assessment purposes in different countries, both for guideline developers and clearinghouses as well as individual users of guidelines.

In response, an international group of researchers from thirteen countries, the Appraisal of Guidelines, REsearch and Evaluation (AGREE) Collaboration, has developed and validated a generic instrument that can be used to appraise the quality of clinical guidelines. The AGREE Instrument is designed to assess the process of guideline development and how well this process is reported. It does not assess the clinical content of the guideline nor the quality of evidence that underpins the recommendations. In this paper we report the development and validation of the AGREE instrument (Appendix C)¹.

¹ The AGREE Instrument is also available on the AGREE website: www.agreecollaboration.org

Methods

A multi-staged approach was used that included an item generation, selection and scaling process and field-testing and refinement procedures.

Item generation, selection, and scaling

To develop the framework for the instrument, quality was defined as the confidence that the biases linked to the rigour of development, presentation and applicability of a clinical practice guideline have been minimised and that each step of the development process is clearly reported. We considered five theoretical quality domains: scope and purpose, stakeholder involvement, rigour of development, clarity and presentation, and applicability. A small working group (Françoise Cluzeau, Jako Burgers, Richard Grol, Peter Littlejohns) generated an initial list of 82 items from validated appraisal instruments and relevant literature [6,8-12] that addressed these domains. The working group examined the list for coverage, overlap and content validity, and reduced it to 34 items. The list and a user guide describing the items were pretested on two Dutch and two English guidelines and refinements were made in response to the comments received.

The refined list and user guide were then circulated to all the AGREE partners and to 15 international experts for their views on the clarity, comprehensiveness, relevance and ease of use. In addition, the AGREE partners were asked to apply the instrument to two guidelines each. The feedback from this process led to reformulation of ambiguous items and removal of overlapping and value laden items. The result was the first draft instrument comprising of 24 items, grouped into the five domains identified in the development phase. We also modified the user guide to reflect changes made to the items. A four-point Likert scale was used to score each item (1=strongly disagree, 2=disagree, 3=agree, 4=strongly agree). A three-point scale (1=not recommend, 2=recommend with provisos or modifications, and 3=strongly recommend) was used to score an overall judgement on whether the guideline ought to be recommended for use.

Field testing and refinement

The AGREE collaborators field tested the instrument following a research protocol that covered selection criteria for the guidelines, methods for recruiting appraisers, and time scales (Box 1). Each country coordinated the appraisal of at least seven guidelines. Each guideline was assessed independently by four appraisers and,

where possible, each appraiser assessed two guidelines. The appraisers received a standard letter with instructions on how to complete the instrument. Most used an English version of the draft AGREE instrument. If necessary, the materials or the user guide only were translated to ensure appraisers' understanding of the items. Feedback on the instrument, user guide, and the appraisal process was solicited with a standard letter, translated into a national language where necessary.

Box 1. Participating countries, and selection criteria for guidelines and appraisers

Countries ¹	Canada, Denmark, England, Finland, France, Germany, Italy, Netherlands, Scotland, Spain, Switzerland
Guidelines	<ul style="list-style-type: none"> - guidelines published between 1992 and 1999 - preferred disease areas: asthma, breast cancer, and diabetes - documents that contain specific recommendations for clinical practice (excluding systematic reviews or service documents)
Appraisers	<ul style="list-style-type: none"> - broad range of professions including clinical experts, nurses, researchers and policy makers - different health care settings including primary care, secondary care, teaching hospitals - excluding members from guideline development group

¹ England and Scotland were considered separately because they have independent guideline programmes.

The field test was conducted in winter 1999-2000 with the 24-item draft instrument. For this phase, 100 guidelines from 11 countries (mode = 8, range 7-22) were evaluated by 194 appraisers. The results of this field test were reviewed at an AGREE workshop in spring 2000 and the instrument and user guide were refined in response to the results. The final version of the instrument underwent further field-testing in autumn 2000. In this phase, a random sample of 3 guidelines per country from the original 100 were assessed by 70 newly recruited appraisers.

Analyses

Mean item scores for each guideline were calculated by averaging the scores across the four appraisers. Standardised domain scores for each guideline were calculated by summing scores across the 4 appraisers and standardising them as a percentage of the possible maximum score a guideline could achieve. Mean item and standardised domain scores were used in the analyses unless otherwise noted below.

To guide the refinement of the instrument from the draft version to the final version, a principal components analysis was undertaken with data from the first field test. The mean item scores for each of the one hundred guidelines were included in the analysis, with the eigen value limit set at one and the criteria for the minimum loading score set at 0.52 [13,14].

Final Instrument Properties

Reliability: two measures of reliability were conducted:

- a) Using mean item scores, the Cronbach Alpha coefficient was calculated to measure internal consistency of each domain of the final instrument [15].
- b) Intraclass correlations (ICC) were calculated to assess the reliability within each domain. ICCs based on single appraisers' ratings and the means of 2, 3 and 4 appraisers were calculated [16].

Validity: several measures of validity were considered.

- a) FACE VALIDITY: appraisers' attitudes about the instrument and user guide were collected by questionnaire and used to assess face validity.
- b) CONSTRUCT VALIDITY: three hypotheses were considered for tests of construct validity:
 - 1. Established guideline programmes have opportunities to compose and refine guideline development methodologies, create efficiencies of process, and access committed funds. Thus, it was hypothesised that guidelines originating from established programmes would have higher domain scores than those produced outside of an established system. To test this hypothesis, a series of one-way ANOVAS on quality scores was undertaken for each domain with type of guideline programme (established/not established) as the between-subjects factor.
 - 2. It can be argued that guidelines supported by well-documented technical information, either within the guideline itself, or as part of supporting reports or publications, will have domain scores higher than those without this documentation. To test this notion, Kendall's Tau B rank correlation tests on quality scores for each domain were undertaken.
 - 3. Guidelines developed as national policies should be particularly robust because of the authority conferred to them. Thus, it was predicted that guidelines created on a national level should be of higher quality than regional or local ones. To test this notion, a series of one-way ANOVAS on

quality scores was undertaken for each domain with level status (national/other guidelines) as the between-subjects factor.

c) CRITERION VALIDITY: as there is no gold standard in this area, participants' overall assessment scores were used as a proxy measure. Assessments of criterion validity were assessed by calculating the Kendall's Tau B rank correlation coefficients between the appraisers' domains scores and the overall assessment scores.

Results

The median time for appraising a guideline was 1.5 hour in both field studies. This included reading a guideline and completing the instrument. All appraisals were completed and returned.

Refinement of Instrument

The Principal Components Analysis of the draft instrument items yielded a five-factor solution that generally supported the domains of quality identified in the development phase. Table 1 shows the list of items and their loading (correlation) coefficients on each of the five domains from the rotated factor matrix.

Editorial independence appeared to load on several domains. In response, it was shifted to a sixth domain in the final version of the instrument, and a new item addressing conflicts of interest was included. Two items: 'The guideline is clearly structured' and 'The potential problems with changes of attitude or behaviour of health care professionals in applying the guidelines have been considered' were removed from the final version of the instrument due to failure to establish adequate reliability in the first field test. Finally, 10 items were reworded slightly in the final version of the instrument in response to feedback received from the appraisers (see Face validity below). The refined instrument, the final version, contained 23-items grouped into six domains with the 4-point Likert scale to score each item (Table 1).

Table 1. Domain structure for guideline quality obtained from Principal Components Analysis, the mean and standard deviations of domain scores, and percentage of variance explained by each domain. Item numbers represent the order in the instrument.

		Coefficient*
Domain 1. Scope and purpose		
<i>Mean percentage domain score = 69.3, sd = 21.3, range (16.7-97.2), % variance = 4.6</i>		
1	The overall objective(s) of the guideline is(are) specifically described	0.594
2	The clinical question(s) covered by the guideline is(are) specifically described	0.768
3	The patients to whom the guideline is meant to apply are specifically described	0.702
Domain 2. Stakeholder involvement		
<i>Mean percentage domain score = 36.1, sd = 18.9, range (4.2-68.7), % variance = 6.6</i>		
4	The guideline development group includes individuals from all the relevant professional groups	0.643
5	The patients' views and preferences have been sought	0.580
6	The target users of the guideline are clearly defined	0.683
7	The guideline has been piloted among end users	0.471
Domain 3. Rigour of development		
<i>Mean percentage domain score = 40.7, sd = 25.0, range (0-89.3), % variance = 42.3</i>		
8	Systematic methods were used to search for evidence	0.794
9	The criteria for selecting the evidence are clearly described	0.763
10	The methods used for formulating the recommendations are clearly described	0.750
11	The health benefits, side effects and risks have been considered in formulating the recommendations	0.689
12	There is an explicit link between the recommendations and the supporting evidence	0.753
13	The guideline has been externally reviewed by experts prior to its publication	0.589
14	A procedure for updating the guideline is provided	0.619
Domain 4. Clarity and presentation		
<i>Mean percentage domain score = 65.8, sd = 14.1, range (37.5-91.7), % variance = 8.6</i>		
15	The recommendations are specific and unambiguous	0.716
16	The different options for management of the condition are clearly presented	0.589
17	Key recommendations are easily identifiable	0.739
18	The guideline is supported with tools for application	0.640
Domain 5. Applicability		
<i>Mean percentage domain score = 36.9, sd = 23.2, range (0-91.7), % variance = 6.1</i>		
19	The potential organisational barriers in applying the recommendations have been discussed	0.804
20	The potential cost implications of applying the recommendations have been considered	0.697
21	The guideline presents key review criteria for monitoring and/or audit purposes	0.684
Domain 6. Editorial independence		
<i>Mean percentage domain score = 30.3, sd = 22.4, range (0-72.2),</i>		
22	The guideline is editorially independent from the funding body	
23	Conflicts of interest of guideline development members have been recorded	New item

* Coefficients from varimax rotated factor matrix

Final Instrument Properties

Reliability: internal consistency ranged between 0.64 to 0.88 and was acceptable for most domains (Table 2). The lower alpha coefficient found for domain 6, Editorial Independence, was not surprising as this domain was composed of only two items. Table 2 also displays the intraclass correlations for each domain as a function of number of raters. As would be expected, the number of appraisers evaluating a guideline affected reliability; increasing the number of raters resulted in substantially higher ICCs.

Table 2. Interrater reliability and internal consistency for each quality domain (n=33)

Domains	Intraclass Correlation ¹				Cronbach α
	1 appraiser	2 appraisers	3 appraisers	4 appraisers	
1 Scope and purpose	0.44	0.61	0.70	0.76	0.88
2 Stakeholder involvement	0.47	0.64	0.72	0.78	0.72
3 Rigour of development	0.71	0.83	0.88	0.91	0.88
4 Clarity and presentation	0.25	0.39	0.49	0.57	0.69
5 Applicability	0.50	0.67	0.75	0.80	0.79
6 Editorial independence	0.34	0.51	0.61	0.67	0.64

¹ The Spearman-Brown formula to obtain the ICC for the average of k ratings from the ICC of 1 rating is

$$ICC_k = \frac{k(ICC_1)}{1 + (k-1)ICC_1}$$

Validity

a) **FACE VALIDITY:** results from the first field test indicated that the appraisers found the instrument useful to assess guidelines (95%) and the user guide helpful (98%). However, almost half of the participants reported having difficulties with at least one item of the instrument (49%). The most commonly reported problem was that guidelines lacked the detailed information necessary to assign a score. After refinement of the instrument results from the second field test showed that the percentage of appraisers reporting difficulties with at least one item in the instrument decreased to 29%.

b) **CONSTRUCT VALIDITY:** tests of the first hypothesis showed that guidelines produced as part of a guideline program had significantly higher scores on domain 6, Editorial Independence, than those published outside a programme ($p < 0.05$). Tests of the second hypothesis showed that guidelines with technical documentation had higher scores on domain 3, Rigour of Development, than those published without documentation ($p < 0.01$). Finally, tests of the third hypothesis

revealed that guidelines produced after the publication of a national policy had significantly higher quality scores on domain 3, Rigour of Development, than did their counterparts ($p < 0.05$). No other significant differences emerged on any of the other domains for any of the contrasts (see table 3 for details).

c) CRITERION VALIDITY: Kendall's Tau B rank correlation coefficients between the appraisers' domain scores and their overall assessments were all highly significant ($p < 0.001$), providing some evidence of criterion validity using this proxy measure. Table 4 shows the correlation matrix of the six quality domains. With one exception, the domains tended to be more highly correlated with overall judgement than with each other.

Table 4. Correlation between each domain and overall judgement

	Domain 1	Domain 2	Domain 3	Domain 4	Domain 5	Domain 6	Overall
Domain 1	1 00						
Domain 2	0 81	1 00					
Domain 3	0 56	0 71	1 00				
Domain 4	0 56	0 60	0 56	1 00			
Domain 5	0 49	0 55	0 38	0 57	1 00		
Domain 6	0 48	0 56	0 56	0 59	0 49	1 00	
Overall	0 79	0 88	0 87	0 77	0 67	0 74	1 00

Domain 1 Scope and purpose

Domain 2 Stakeholder involvement

Domain 3 Rigour of development

Domain 4 Clarity and presentation

Domain 5 Applicability

Domain 6 Editorial independence

Table 3. Standardised guidelines scores and their confidence intervals for each domain according to guideline programme, level of background information and national policy

	Domain 1. Scope and purpose	Domain 2. Stakeholder involvement	Domain 3. Rigour of development	Domain 4. Clarity and presentation	Domain 5. Applicability	Domain 6. Editorial independence
<i>All guidelines (n=33)</i>	69.3 (61.7-76.9)	36.1 (29.4-42.8)	40.7 (31.9-49.6)	65.8 (60.8-70.8)	36.9 (28.7-45.1)	30.3 (22.3-38.2)
<i>Guideline programme</i>						
Developed within a guideline programme (n=20)	68.2 (58.5-78.0)	35.6 (27.5-43.8)	44.2 (33.1-55.3)	66.6 (59.8-73.4)	34.9 (24.9-44.9)	36.7* (26.5-47.0)
Developed outside a guideline programme (n=13)	70.9 (57.1-84.7)	36.9 (23.8-50.0)	35.3 (19.1-51.5)	64.4 (56.1-72.7)	39.8 (24.0-55.7)	20.3 (8.1-32.5)
<i>Level of background information</i>						
No information (n=7)	63.5 (42.2-84.8)	29.5 (13.8-45.1)	23.8 (6.9-40.8)	58.6 (43.2-74.1)	38.1 (12.2-64.0)	30.4 (12.1-48.7)
Some information / references (n=10)	67.2 (47.8-86.7)	31.1 (15.1-47.3)	29.4 (16.5-42.4)	64.2 (51.2-77.1)	29.4 (17.1-41.6)	26.1 (7.0-45.3)
Detailed documentation (n=16)	73.1 (64.1-82.0)	42.1 (33.4-50.8)	55.1** (42.5-67.8)	69.8 (65.3-74.3)	41.0 (27.9-54.0)	32.8 (21.3-44.3)
<i>National policy</i>						
Guidelines developed before (n=13)	71.2 (58.1-84.2)	34.2 (22.9-45.5)	29.0 (16.3-41.6)	67.8 (59.9-75.7)	41.8 (25.8-57.8)	25.9 (10.6-41.1)
Guidelines developed after (n=20)	68.1 (57.9-78.2)	37.3 (28.2-46.4)	48.3* (36.7-60.0)	64.4 (57.4-71.3)	33.6 (23.9-43.3)	33.1 (23.6-42.7)

* p<0.05

** p<0.01

Discussion

This is the first time an appraisal instrument for clinical practice guidelines has been developed and tested at international level. Created through a rigorous and iterative process by a collaboration of international experts in clinical guidelines, the instrument was applied to 100 guidelines by over 260 appraisers from 11 countries. Previous studies on similar instruments have been limited to appraisers working in the same institution and from the same country [3,7]. This study resulted in a rigorously developed set of criteria for appraising guidelines that can be helpful for clinical practice in two ways: firstly to help clinicians to differentiate between guidelines from different sources; secondly, as a support to the development of high quality guidelines for medical practice.

Our results show that the instrument is sensitive to differences in important aspects of clinical practice guidelines, and it can be used consistently by a wide range of professionals from different cultural backgrounds. Health professionals, policy makers, and consumers were all able to appraise guidelines with the AGREE questions and user guide. The appraisers found the instrument easy to apply and perceived it to be useful for judging the quality of guidelines.

When interpreting the results, several considerations must be kept in mind. First, the factor analysis confirmed our conceptual framework, lending support to the assumption that the quality of clinical guidelines is composed of distinct domains, each assessing key quality attributes. However, the concept of guideline quality is still grounded in assumptions that need testing empirically, and we do not know the relative contribution of each domain to the overall quality of a guideline. Construct validity, based on three a priori hypotheses, was not strong. It was somewhat surprising to observe that national (as opposed to local) development and established (as opposed to more recent) programmes supporting production did not predict quality more strongly. The high correlations found between the domain scores and the overall assessment corroborated the modest criterion validity, although the effect may be attenuated by the fact that the appraisers made their global ratings after assessing the guidelines.

Second, the reliability of the domains is directly affected by the number of appraisers assessing one guideline. Thus using four appraisers will yield a more reliable assessment than using a single appraiser [17]. In this study, average ratings of four raters provided the most reliable assessment and we recommend that at least four raters should be used when using the instrument.

Finally, we were not able to demonstrate conclusively the validity of our instrument. The instrument assesses the methodological quality of a guideline and this relies heavily on how well documented the guideline development process is [18]. However explicit reporting does not guarantee optimal recommendations. A well-reported guideline may contain flawed recommendations, and conversely an unsystematically constructed one may provide sound evidence [19]. Nevertheless, the criteria we used are accepted as key determinants of valid and effective guidelines amongst methodologists, and the domains are quite clear. Validation of the instrument is a challenging task. We are currently undertaking detailed content analysis of the appraised guidelines as part of our research programme. This will provide a separate measure of construct validity.

AGREE has considerable implications for research and policy. These standards for the development and reporting of clinical practice guidelines can be used by guideline producers worldwide. The adoption of such standards can improve the consistency and quality of the reporting process [20]. The sharing of standards across countries will facilitate international comparison of guidelines and can provide a framework for studies aimed at understanding why guidelines for the same condition may produce differing recommendations [21,22].

As the number of clinical practice guidelines submitted for publication increases there is a need to ensure that they satisfy certain minimum requirements. AGREE can be adopted by editors of peer reviewed journals as a framework to assess the quality of clinical guidelines in the same way that CONSORT is used to judge the quality of randomised controlled trials and meta-analyses [23,24].

Given the expansion of national guideline programmes, governments and other agencies must ensure the guidelines of the highest quality before they endorse them or promote their use in practice. Furthermore, as international co-operation between countries grows there is a strong incentive for policy makers to develop a concerted approach to quality management initiatives, including clinical practice guidelines. The AGREE instrument can enhance this process. This is already taking place, as several agencies, such as National Institute for Clinical Excellence (NICE) in the UK, the National Federation of Cancer Centres (FNCLCC) in France, The Agency for Quality in Medicine in Germany (ÄZQ) and the Scottish Intercollegiate Guidelines Network (SIGN), are using AGREE in the context of their guidelines programme. The World Health Organisation has adopted the AGREE instrument to assess its guidelines.

Conclusions

The AGREE collaboration has developed an instrument for guideline appraisal using a rigorous methodology. The instrument has been applied to different clinical practice guidelines in 11 countries by a large number of appraisers from a variety of backgrounds. We recommend that guideline producers use this instrument while planning their programmes, and potential guideline users use it to evaluate the quality of guidelines before adopting them.

References

1. Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Potential benefits, limitations, and harms of clinical guidelines. *BMJ* 1999; 318: 527-530.
2. Woolf SH, Grol R, Hutchinson A, Eccles MP, Grimshaw JM. An international overview. In: Eccles MP, Grimshaw JM (eds). *Clinical Practice Guidelines*. Oxford: Radcliffe Medical Press, 2000.
3. Institute of Medicine. Field MJ, Lohr KN (eds). *Guidelines for clinical practice. From Development to Use*. Washington DC: National Academy Press, 1992.
4. Shekelle PG, Woolf SH, Eccles M, Grimshaw J. Clinical guidelines: Developing guidelines. *BMJ* 1999; 318: 593-596.
5. Grimshaw JM, Russell IT. Achieving health gain through clinical guidelines. I: Developing scientifically valid guidelines. *Qual Health Care* 1993; 2: 243-248.
6. Grilli R, Magrini N, Penna A, Mura G, Liberati A. Practice guidelines developed by specialty societies. The need for a critical appraisal. *Lancet* 2000; 355: 103-106.
7. 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: 1900-1905.
8. Lohr KN, Field MJ. A Provisional Instrument for Assessing Clinical Practice Guidelines. In: Institute of Medicine. Field MJ, Lohr KN (eds). *Guidelines for Clinical Practice. From development to use*. Washington DC: National Academy Press, 1992.
9. Cluzeau F, Littlejohns P, Grimshaw J, Feder G, Moran S. Development and application of a generic methodology to assess the quality of clinical guidelines. *Int J Qual Health Care* 1999; 11: 23-28.
10. Grol R, Dalhuijsen J, Thomas S, in 't Veld C, Rutten G, Mookink H. Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. *BMJ* 1998; 317: 858-861.
11. Thomson R, Lavender M, Madhok R. How to ensure that guidelines are effective. *BMJ* 1995; 311: 237-242.
12. Lohr KN. The quality of practice guidelines and the quality of health care. In: *Guidelines in health care. Report of a WHO Conference*. January 1997. Baden-Baden: Nomos Verlagsgesellschaft, 1998.
13. MacCallum RC, Widaman Keith F, Zhang S, Hong S. Sample size in factor analysis. *Psychol Meth* 1999; 4: 84-99.
14. Norman GR, Streiner DL. *Biostatistics. The Bare Essentials*. Second Edition, St Louis, Mosby, 2000.
15. Bland JM, Altman DG. Cronbach's alpha. *Statistics notes*. *BMJ* 1997; 314: 572.
16. Fleiss J.L. The Measurement of Interrater Agreement. In: *Statistical Methods for Rates and Proportions*. New York: John Wiley & Sons, 1981.
17. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; 86: 420-428.
18. Hayward RS, Wilson MC, Tunis SR, Bass EB, Guyatt G. Users' guides to the medical literature. VIII. How to use clinical practice guidelines. A. Are the recommendations valid? The Evidence-Based Medicine Working Group. *JAMA* 1995; 274: 570-574.
19. Moher D, Ba P, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; 352: 609-613.
20. Cluzeau F, Littlejohns P. Appraising clinical guidelines in England and Wales. The development of a methodological framework and its application to policy. *Jt Comm J Qual Improv* 1999; 25: 514-521.
21. The AGREE Collaborative Group. Guideline development in Europe: an international comparison. *Int J Technol Assess Health Care* 2000; 16: 1039-1049.
22. Grol R, Eccles M, Maisonneuve H, et al. Developing clinical practice guidelines: the European experience. *Dis Manage Health Outcomes* 1998; 4: 255-266.
23. Moher D, Schulz KF, Altman D for the CONSORT Group: The CONSORT Statement: Revised Recommendations for Improving the Quality of Reports of Parallel-Group Randomized Trials. *JAMA* 2001; 285: 1987-1991.
24. Moher D, Jones A, Lepage L for the CONSORT Group: Use of the CONSORT Statement and Quality of Reports of Randomized Trials. A Comparative Before-and-After Evaluation. *JAMA* 2001; 285: 1992-1995.

Chapter 4

Characteristics of high-quality guidelines: evaluation of 86 clinical guidelines developed in ten European countries and Canada

Jako S Burgers

Françoise A Cluzeau

Steven E Hanna

Claire Hunt

Richard PTM Grol,

and the AGREE Collaboration

International Journal for Technology Assessment in Health Care 2002; 18(4): 955-964.

Abstract

Objectives. To identify predictors of high-quality clinical practice guidelines.

Methods. A total of 86 guidelines from 11 countries were assessed by four independent appraisers per guideline using the AGREE Instrument (23 items). Six aspects of guideline development were considered to explain the variation in quality scores: care level (primary/secondary care), scope (diagnosis/treatment), type of guideline (new/update), year of publication, type of agency (governmental/professional), and whether the guideline was produced within a structured and coordinated programme.

Results. Guidelines produced within a guideline programme and by governmental agencies had higher scores than their counterparts. Differences in the applicability of the guidelines could not be explained by the variables studied.

Conclusion. To ensure high quality, clinical guidelines should be produced within a structured and coordinated programme. Professional organisations or specialist societies that aim to develop guidelines may adopt quality criteria from leading guideline agencies.

Introduction

Within the last decade the body of available clinical guidelines has expanded enormously. Guidelines are increasingly used in health care systems throughout the world to improve the quality of patient care [1]. To ensure good quality of care, the guidelines used should meet specific criteria for quality. Quality of guidelines can be defined 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' [2]. However, recent studies have reported that the methodologic quality of guidelines is often modest and varies among different guidelines and different agencies [3-7]. Whereas variation in health care is a common reason for developing guidelines, variation in the quality of guidelines will be counterproductive. To address this issue, we should learn more about the characteristics of high-quality guidelines aiming at ensuring improvement of clinical practice and patient care. This knowledge could help policy makers and healthcare providers in selecting the best guidelines and guideline developers in setting or refining their guideline development programme.

There is little research regarding the characteristics of guidelines or guideline agencies predicting guideline quality. Studies conducted in the United Kingdom [3] and in Finland [6] concluded that national guidelines had higher quality scores than local guidelines. In addition, Grilli et al. suggested that guidelines produced by major technology assessment agencies are probably better than those developed by specialty societies [4]. Other predictors of guidelines quality are not known yet.

In this study we sought to identify predictors of guideline quality by analysing data collected for validation of the AGREE (Appraisal of Guidelines for Research & Evaluation) instrument (Appendix C) [2]. This instrument was developed by an international group of researchers from thirteen countries (The AGREE Collaboration) with the aim to create a common, valid and transparent approach to the appraisal of clinical guidelines [8]. The instrument was the result of a multi-staged process of item generation, selection and scaling, field testing, and refinement procedures. As part of the validation of the instrument, a study was conducted to assess the quality of a sample of clinical guidelines developed in ten European countries and Canada. As part of this project, information about several possible predictors was collected. We examined which of these guideline and

agency characteristics were predictive of scores on the quality domains of the AGREE instrument.

Methods

Instrument development

To set the framework of the instrument, six theoretical quality domains were considered: Scope and Purpose, Stakeholder Involvement, Rigour of Development, Clarity and Presentation, Applicability, and Editorial Independence. An initial list of 82 items from existing instruments and checklists and relevant literature addressed these domains [3,5,9-12]. This list was examined for coverage, overlap and content validity and reduced to 34 items. The refined list was then circulated for external review, including all AGREE partners and 15 international experts. The feedback from the reviewers led to reformulation of ambiguous items and removal of overlapping and value-laden items. The final instrument included 23 items (Appendix C). A four-point Likert scale was used to score each item (4=strongly agree, 3=agree, 2=disagree, 1=strongly disagree).

Selection of guidelines

We defined a *guideline* as 'a set of systematically developed statements to assist practitioner and patient decisions about appropriate health care for one specific clinical condition or disease area' [13]. Documents that did not contain recommendations for clinical practice (e.g., systematic reviews, service documents) were excluded. All country coordinators were asked to select 7 to 10 guidelines, published between 1992 and 1999. Coordinators were instructed to provide guidelines that they regarded as both high and low in quality, in order to test the discriminative value of the instrument. In all, 86 guidelines developed by 62 different agencies and organisations from 11 countries were selected.

Selection of appraisers

In each country four independent appraisers per guideline were recruited. Where possible, each appraiser assessed two guidelines. The appraisers included medical practitioners, clinical experts, clinical researchers, and methodologists. Members of the guideline development group, members of the secretariat that produced the guidelines, and external referees were excluded.

Variables

To explain the variation in the quality of the guidelines, the following six characteristics of guidelines were considered:

1. care level (primary, secondary/tertiary care, all levels);
2. scope (prevention/diagnosis, treatment, combination);
3. type of guideline (new, update);
4. year of publication (1992-1994, 1995-1997, 1998-1999);
5. type of agency (professional/specialist societies, government funded agencies, other);
6. guideline programme (part of guideline programme, not part of guideline programme).

A *guideline programme* was defined as 'a structured and coordinated programme designed with the specific aim of producing several clinical practice guidelines' [14].

The country coordinators were asked to include information about these variables for each guideline on a standardised form.

Analysis

We analysed the scores according to the six quality domains of the instrument. Standardised guideline domain scores were calculated by summing the scores across the four appraisers and standardising them as a percentage of the maximum possible score. Each guideline variable was entered into a multilevel model in order to consider the clustering effect of the agency responsible for the guideline [15]. The significance of differences in standardised domain scores between guidelines with different characteristics was studied using one-way analysis of variance (ANOVA) as part of the multilevel model. We identified the proportion of variance in scores between guidelines *between* agencies and guidelines *within* agencies. Multilevel modeling also provides tests to measure the extent to which each variable could explain the variance. Analyses were performed using SPSS 9.0 and NLME 3.2 library for S-PLUS 2000 [16].

Results

The standardised guideline domain scores ranged from 31.3 ('Applicability') to 66.1 ('Scope and Purpose') (Table 1). The range of scores was broad within all six domains.

Table 1. Domain scores of guidelines clustered according to six variables of guidelines

	Scope and Purpose	Stakeholder Involvement	Rigour of Development	Clarity and Presentation	Applicability	Editorial Independence
<i>Care level</i>						
Primary care (n=21)	65.7	34.2	22.4 ^a	57.1	29.6	48.0
Secondary/tertiary care (n=32)	64.5	37.3	45.5	60.2	27.5	49.7
All levels (n=33)	68.0	29.7	37.8	54.9	36.0	45.7
<i>Scope^b</i>						
Prevention/diagnosis (n=9)	73.8	32.6	38.1 ^a	61.3	35.2	56.5
Treatment (n=27)	64.2	37.9	45.2 ^a	61.0	26.5	49.1
Combination (n=47)	64.4	30.3	32.5	55.2	31.6	44.1
<i>Type of guideline^b</i>						
New (n=60)	66.2	33.9	38.7	57.3	31.9	48.9
Update (n=25)	65.9	32.8	32.7	58.3	29.4	45.0
<i>Year of publication^b</i>						
1992 – 1994 (n=7)	60.7	30.1	19.4	54.5	32.1	38.1
1995 – 1997 (n=25)	61.0	32.8	34.4	53.8	32.3	47.0
1998 – 1999 (n=52)	70.2	34.9	41.2	60.4 ^a	31.2	50.2
<i>Authors</i>						
Professional/specialist societies (n=39)	64.2	29.9	26.5	51.3	28.2	35.3
Government funded organisations (n=35)	71.2	39.6	48.8	64.6	36.1	59.8 ^a
Other (n=12)	57.6	28.3	36.0	55.9	27.3	53.5
<i>Guideline programme</i>						
Part of guideline programme (n=55)	67.7	35.6	43.7 ^a	63.2 ^a	32.2	49.8
Not part of guideline programme (n=31)	63.5	30.2	25.3	47.5	29.8	44.3
All guidelines (n=86)	66.1	33.6	36.9	57.4	31.3	47.8

^a p<0.05^b total number is not 86 due to missing values

One-way ANOVA results from the multilevel models indicated that most significant differences were found for 'Rigour of Development'. Three variables accounted for these differences (level of care, scope and guideline programme). Overall, guidelines developed by government funded agencies had the highest scores on all domains. However, the scoring differences between these agencies and professional or specialist societies were only significant on the domain 'Editorial Independence'. Guidelines developed within a guideline programme had higher scores than their counterparts on all domains, but these were only significant for

'Rigour of Development' and 'Clarity and Presentation'. For domains 'Scope and Purpose', 'Stakeholder Involvement', and 'Applicability' significant differences were absent for all variables.

Multilevel modeling provides separate estimates of the variance in quality scores among guideline agencies and among guidelines within agencies. These estimates are reported in table 2 as percentages of total variance. There is more between-agency than within-agency variation in quality scores for 'Stakeholder Involvement', 'Clarity and Presentation', and especially, 'Rigour of Development'. Thus, variations in these aspects of quality of guidelines are primarily associated with characteristics of guideline agencies. By contrast, variation in 'Applicability' scores was more associated with differences among guidelines than differences among agencies.

Table 2. Standard deviation (95% confidence limits) and proportion of variance of domain scores occurring between agencies (agency level) and within agencies (guideline level)

	Scope and Purpose	Stakeholder Involvement	Rigour of Development	Clarity and Presentation	Applicability	Editorial Independence
<i>Standard deviation</i>						
Agency (n=62)	14.5 (9.3-22.4)	14.5 (10.9-19.3)	22.7 (18.3-28.2)	15.5 (11.9-20.3)	9.7 (5.5-17.2)	19.0 (12.7-28.6)
Guideline (n=86)	14.7 (10.9-19.8)	10.1 (7.5-13.5)	10.9 (8.3-14.3)	10.5 (7.9-13.8)	15.5 (12.4-19.4)	21.1 (16.3-27.3)
<i>% variance</i>						
Agency (n=62)	49.1	67.4	81.3	68.6	28.1	44.8
Guideline (n=86)	51.0	32.6	18.7	31.4	71.9	55.2

For 'Rigour of Development' and 'Clarity and Presentation', the variance of scores could be partly explained by certain characteristics of guidelines (Table 3). The level of care and scope of the guideline significantly explained variance *within* agencies, whereas the author and guideline programme particularly explained variance *between* agencies. For 'Clarity and Presentation', the guideline programme and year of publication accounted for most of the variance.

Table 3. Relative reduction of variance by different predictors for domains 'Rigour and Development' and 'Clarity and Presentation'

	Rigour and Development		Clarity and Presentation	
	<i>between agency</i>	<i>within agency</i>	<i>between agency</i>	<i>within agency</i>
Care level	ns	11.1	ns	ns
Scope	ns	10.2	ns	ns
Type of guideline	ns	ns	ns	ns
Year of publication	ns	ns	-16.4	29.5
Author	7.4	ns	ns	ns
Guideline Programme	7.6	ns	19.5	ns

ns = $p > 0.05$

Discussion

The main finding of this study is that high-quality clinical guidelines were particularly produced within established guideline programmes and by government funded agencies. This is consistent with the study of Grilli et al. [4] which showed that guidelines produced by specialist societies were lower in quality than guidelines produced by major agencies such as the Scottish Intercollegiate Guidelines Network (SIGN) and the Agence Nationale d'Accréditation et d'Évaluation en Santé (ANAES) in France. These agencies have a structured guideline programme providing a systematic procedure with key elements such as a multidisciplinary guideline development group, a systematic literature review, external peer review and different products for dissemination [14]. These elements ensure high scores on several domains, in particular on 'Rigour of Development'. On the other hand, our study also showed that the agency responsible for guideline development had less influence on 'Applicability' than on other domains (Table 2). This suggests that agency policies and procedures are more concerned with the methodology of producing guidelines than with the effectiveness of guidelines in daily practice.

Developing high-quality guidelines requires a sufficiently skilled team of people and sufficient budget. In general, governmental agencies have greater resources than professional organisations and specialist societies, which might explain why their guidelines have higher quality scores. Nevertheless, we still believe that professional organisations can develop high-quality guidelines, provided they develop their guidelines within a structured programme and adopt quality criteria of other programmes.

The influence of other characteristics on the quality scores was limited. Guidelines with a narrow scope, (i.e., exclusively focusing on prevention/diagnosis

or treatment) had higher scores on 'Rigour of Development' than guidelines that covered both prevention/diagnosis and treatment. The quality of a guideline might be improved by providing recommendations on a few well-defined issues instead of covering the whole clinical area of the condition selected for guideline development. As a consequence, guidelines produced for primary care had lower scores on 'Rigour of Development', because these were broader in scope than guidelines in secondary care that focus on an already established diagnosis.

Surprisingly, the year of publication and the type of guideline (new versus updated) had little influence on the scores. However, there was a small trend of overall improvement over time.

Estimates of the variance *between* agencies are difficult due to the low number of guidelines per agency on average. This could explain the odd increase (i.e., the reduction of variance is negative) in the estimate when year is added to the domain 'Clarity and Presentation' analysis (Table 3). In contrast, *within* agencies the clarity and presentation of their guidelines obviously improves over time.

The strength of our study is that we assessed the guidelines with a rigorously developed instrument created by a collaboration of international experts in guideline development. There is insufficient evidence for adopting any other existing guideline appraisal instrument [17]. In contrast to other studies [4,5] our sample of guidelines was not restricted to guidelines included in MEDLINE, thus representing a broad range of guidelines that are not necessarily representative of the quality of guidelines produced by the agencies selected. Moreover, we did not aim to provide a general statement about 'the quality of clinical guidelines'. We aimed to explain the variance in quality by characteristics of the guidelines. Therefore, we collected additional information about the background and context of the guidelines (e.g., guideline programme) that enabled us to explain differences in quality scores. So far, this is the first study to achieve this. However, it is uncertain whether the selection process is related to other variables that have not been studied.

Our study was limited by the lack of information on the ultimate adherence to the guidelines. Evidence based guidelines do not guarantee that they will be followed [18]. Other factors, such as attitudinal and organisational barriers, should be overcome to ensure any effect of the guideline in daily practice [19,20]. It would be interesting for future research to study the relationship between the 'quality' of guidelines and the effectiveness of guidelines.

Policy implications

Clinical guidelines should be produced within a structured and coordinated programme to ensure that they are of high quality. Professional organisations or specialist societies that aim to develop guidelines may adopt quality criteria from leading guideline agencies. International collaboration is needed to set standards for guideline quality. As an example, the AGREE instrument for assessing the quality of clinical guidelines [2] is a recent product of international collaboration that can be used by policy makers to help them decide which guidelines could be recommended for use in practice and by guideline developers to follow a structured and rigorous development methodology. A collaborative network of guideline organisations will contribute to further improvement of guideline methodology and implementation and to avoiding duplication of efforts. Guideline clearinghouses (e.g., the U.S. National Clearinghouse [21]) can contribute to this process by disseminating high quality guidelines internationally that can be used by different organisations for local adaptation. The overall cost of developing guidelines could be reduced considerably, if guideline developers used high-quality guidelines as a basis for producing their own guidelines.

References

- 1 Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J Potential benefits, limitations, and harms of clinical guidelines *BMJ* 1999, 318 527-530
- 2 The AGREE Collaboration Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument www.agreecollaboration.org
- 3 Cluzeau F, Littlejohns P, Grimshaw J, Feder G, Moran S Development and application of a generic methodology to assess the quality of clinical guidelines *Int J Qual Health Care* 1999, 11 21-28
- 4 Grilli R, Magrini N, Penna A, Mura G, Liberati A Practice guidelines developed by specialty societies the need for a critical appraisal *Lancet* 2000, 355 103-105
- 5 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 1900-1905
- 6 Varonen H, Makela M Practice guidelines in Finland availability and quality *Qual Health Care* 1996, 6 75-79
- 7 Ward JE, Grieco V Why we need guidelines for guidelines a study of the quality of clinical practice guidelines in Australia *Med J Aust* 1996, 165 574-576
- 8 The Appraisal of Guidelines, Research, and Evaluation in Europe (AGREE) Collaborative Group Guideline development in Europe an international comparison *Int J Technol Assess Health Care* 2000, 16 1029-1049
- 9 Grol R, Dalhuijsen J, Thomas S, in 't Veld C, Rutten G, Mookink H Attributes of clinical guidelines that influence use of guidelines in general practice observational study *BMJ* 1998, 317 858-861
- 10 Lohr KN, Field MJ A provisional instrument for assessing clinical practice guidelines In Institute of Medicine Field MJ, Lohr KN (eds) *Guidelines for clinical practice* Washington DC National Academy Press, 1992
- 11 Lohr KN The quality of practice guidelines and the quality of health care In Selbmann HK *Guidelines in Health Care Report of a WHO-Conference* Baden-Baden Nomos Verlagsgesellschaft, 1998 42-52
- 12 Thomson R, Lavender M, Madhok R How to ensure that guidelines are effective *BMJ* 1995, 311 237-242
- 13 Institute of Medicine Field MJ, Lohr KN (eds) *Clinical practice guidelines directions for a new program* Washington DC National Academy Press, 1990
- 14 Burgers JS, Grol R, Klazinga NS, Makela M, Zaat J, for the AGREE collaboration Towards evidence-based clinical practice an international survey of 18 clinical guideline programs *Int J Qual Health Care*, in press
- 15 Rice N, Leyland Multilevel models applications to health data *J Health Serv Res Policy* 1996, 1 154-164
- 16 Pinheiro JC, Bates DM *Mixed-effects models in S and S-PLUS* Heidelberg Springer Verlag, 2001
- 17 Graham ID, Calder LA, Hébert PC, Carter AO, Tetroe JM A comparison of clinical practice guideline appraisal instruments *Int J Technol Assess Health Care* 2000, 16 1024-1038
- 18 Grol R Beliefs and evidence in changing clinical practice *BMJ* 1997, 315 418-421
- 19 Browman GP Improving clinical practice guidelines for the 21st century *Int J Technol Assess Health Care* 2000, 16 959-968
- 20 Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al Why don't physicians follow clinical practice guidelines? *JAMA* 1999, 282 1458-1465
- 21 U S National Guideline Clearinghouse <http://www.guideline.gov/index.asp>

Section III Recommendations in clinical guidelines

Chapter 5

Inside guidelines: comparative analysis of recommendations and evidence in diabetes guidelines from 13 countries

Jako S Burgers

Julia V Bailey

Niek S Klazinga

Akke K van der Bij

Richard PTM Grol

Gene Feder,

for the AGREE Collaboration

Diabetes Care; in press.

Abstract

Objective. To compare guidelines on diabetes mellitus from different countries in order to examine whether differences in recommendations could be explained by use of different research evidence.

Research design and methods. We analysed 15 clinical guidelines on type 2 diabetes mellitus from 13 countries, using qualitative methods to compare the recommendations, and bibliometric methods to measure the extent of overlap in citations used by different guidelines. A further qualitative analysis of recommendations and cited evidence for two specific issues in diabetes care explored the apparent discrepancy between recommendations and evidence.

Results. The recommendations made in the guidelines were in agreement about the general management of type 2 diabetes mellitus, with some important differences in treatment details. There was little overlap in evidence cited by the guidelines, with 18% (185/1033) of citations shared with any other guideline, and only 10 studies (1%) appearing in six or more guidelines. The measurable overlap in evidence between guidelines increases if multiple publications from the same study and the use of reviews were taken into account. Research originating from the United States predominated (40% of citations), however, nearly all (11/12) guidelines were significantly more likely to cite evidence originating from their own countries.

Conclusion. Despite the variation in cited evidence and preferential citation of evidence from a guideline's country of origin, we found a high degree of international consensus in recommendations made for the clinical care of type 2 diabetes. The influence of professional bodies such as the American Diabetes Association may be an important factor in explaining international consensus. Globalisation of recommended management of diabetes is not a simple consequence of the globalisation of research evidence.

Introduction

Over the past twenty years clinical guidelines have been developed to bridge the gap between research and practice [1]. There has been a concerted effort to base clinical decisions on research evidence [2] and, particularly through the Cochrane collaboration, to make this evidence available globally [3]. Guideline development groups aim to use the totality of relevant research evidence to formulate recommendations [4]. Since bibliographic databases (for example MEDLINE and EMBASE) are easily available, one might expect that this would lead to international consensus on the evidence chosen to underpin recommendations for clinical care, and a consequent convergence of recommendations made in guidelines.

Nevertheless, recommendations often differ in guidelines on the same topic, particularly when evidence for treatment decisions is weak. For example, Eisinger and colleagues found substantial differences between recommendations from the United States and France about prophylactic mastectomy or oophorectomy in high-risk women [5]. Differences were attributed to cultural variation in ideas about patient autonomy and involvement in health care, differing national views on aesthetics of the breast and about fertility. Even where there is good trial evidence, recommendations vary. For instance, analysis of hypertension guidelines from New Zealand, United States, Canada, United Kingdom and the World Health Organisation showed wide variation in the criteria for blood pressure treatment decisions [6]. Differences persisted between more recent editions of national hypertension guidelines, even with more systematic and transparent methods of guideline development [7].

It is evident that there are disparities in recommendations in guidelines for a range of different clinical conditions. Investigators hypothesise that differences are due to insufficient evidence [6,8,9], differing interpretations of evidence [10], unsystematic guideline development methods [11,12], the influence of professional bodies [13], cultural factors such as differing expectations of apparent risks and benefits [5,6], socio-economic factors or characteristics of health care systems [14].

In this study we compared recommendations between a range of guidelines on the management of type 2 diabetes and analysed to what extent the variation (or concordance) between recommendations was explained by the evidence cited in the guidelines.

Methods

Selection of guidelines

We applied the Institute of Medicine's definition of clinical guidelines: 'systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances' [15]. Systematic reviews and evidence reports that did not contain specific recommendations were not included in this study. Because of the large number of clinical issues related to diabetes mellitus, the selection of guidelines was limited to two areas: (i) ambulatory or outpatient care, excluding guidelines exclusively covering type 1 diabetes mellitus, complications of diabetes that need specialist care (retinopathy, diabetic foot, nephropathy, neuropathy) and gestational diabetes; (ii) treatment of diabetes, excluding guidelines on prevention and diagnosis.

The sample consisted of a total of 15 guidelines for the clinical care of type 2 diabetes (Table 1) representing the national guidelines of the AGREE (Appraisal Guidelines Research & Evaluation) collaboration. This international group of researchers has investigated variation between guidelines and guideline development models with the aim of advising the European Commission on guideline development, dissemination and implementation. The east London guideline was chosen because there were no national English guidelines available. Two French guidelines were complementary and were analysed as one guideline. The guidelines from Australia, New Zealand, Canada, and the United States were identified through a web-based search and consultation with colleagues. Four guidelines (CA, NL2, US1, US2) were updated versions of earlier guidelines. Two guidelines (CA, DK) were funded by pharmaceutical companies; the others were funded by national or regional government agencies (EN, FR, NZ, SC, SP) or state health care systems (AU, US2), by national professional organisations (FI, IT, NL2, US1), or by hospitals (NL1, SW).

Selection of comparable sections

Because the guidelines varied in their scope, we selected sections that covered the treatment and monitoring of hyperglycemia and cardiovascular risk.

The guidelines were in seven different languages. Members of the study team translated those guidelines written in French or Dutch; relevant sections of guidelines in Finnish, Danish, and Spanish were translated by guideline

developers in their respective countries. The Italian guideline was excluded from the analysis of recommendations because of its length and lack of structure.

Table 1. Description of selected guidelines

Country (ID code)	Organisation responsible for guideline development	Title in English	Year of publication
1 Australia (AU)	NSW (New South Wales) Health Department	Improving diabetes care and outcomes Principles of care and guidelines for the clinical management of diabetes mellitus	1996
2 Canada (CA)	Canadian Medical Association	Clinical practice guidelines for the management of diabetes in Canada	1998
3 Denmark (DK)	Danish College of General Practitioners	Non insulin demanding diabetes – NIDDM A practical guidance for therapists	1998
4 England (EN)	East London Clinical Guidelines Project Department of General Practice and Primary Care	Clinical guidelines for the management of diabetes in East London	1996
5 Finland (FI)	Finnish Diabetes league	Type II diabetes clinical guideline	1994
6 France (FR)	Agence Nationale d'Accréditation et d'Evaluation en Santé (ANAES)	a Strategy for monitoring of type 2 diabetics, excluding monitoring of complications	1999
		b Strategy for management of type 2 diabetics, excluding management of complications	2000
7 Italy (IT)	Italian Society for Diabetology	Diabetes mellitus Practical guide for diagnosis and treatment	1997
8 Netherlands (NL1)	Dutch Institute for Healthcare Improvement CBO	Guidelines diabetic nephropathy and cardiovascular diseases with diabetes mellitus	1998
9 Netherlands (NL2)	Dutch College of General Practitioners (NHG)	NHG Practice Guideline diabetes mellitus	1999
10 New Zealand (NZ)	New Zealand Guidelines Group	Guidelines for the management of core aspects of diabetes care	1999
11 Scotland (SC)	Scottish Intercollegiate Guidelines Network (SIGN)	Management of diabetic cardiovascular disease	1997
12 Spain (SP)	Catalan Society of Primary Care	Guideline on treatment of diabetes mellitus type 2 in primary care	1996
13 Switzerland (SW)	University Hospital of Geneva	Detection of diabetes mellitus Guidelines for the outpatient's clinic	1996
14 USA (US1)	American Diabetes Association (ADA)	Standards of medical care for patients with diabetes mellitus	2000
15 USA (US2)	Institute for Clinical System Improvements	Management of Type 2 diabetes mellitus	2000

Extraction and comparison of recommendations

We defined recommendations as any statements that promote or advocate a particular course of action in clinical care. Two investigators with medical training, working independently, extracted the recommendations. We resolved discrepancies through discussion within the study team. A panel of four investigators (JSB, JVB, GF, NK) judged the extent of accordance or discordance of recommendations across guidelines.

Extraction and measurement of overlap of citations

One member of our team selected all references linked to the relevant sections chosen for study and another crosschecked this selection. Each citation was entered onto a Reference Manager database (Version 8.5), adding a unique identifier code for each guideline. We excluded the Danish, Finnish and Swiss guidelines from this part of the study because they cited fewer than three references each. We used the Reference Manager search facility to quantify the numbers of citations in common with other guidelines, the type of citation (e.g. meta-analysis, review, or guideline), and the address of the first author as a proxy for the country of origin of the cited study. The proportion of shared references between guidelines was expressed as a percentage of the maximum possible score according to the publication dates both of the guideline and its linked references.

Examination of link between recommendations and citations

To explore the discrepancy between disparate citations and largely concordant recommendations, we purposively selected [16] two areas for further analysis: use of metformin in obese patients and self-monitoring of blood glucose. We selected citations that were explicitly linked to the recommendations or listed at the end of relevant sections and compared citations between guidelines. For each citation we tabulated the type of study, country of origin, study subjects, conclusions and any recommendations made by the authors. Where secondary citations were used (i.e. meta-analyses, systematic reviews or other guidelines), we included the evidence cited by these documents. We compared the publication dates of citations, and the dates of the latest evidence cited by guidelines (censoring dates). We did not appraise the quality of the studies, but examined the consistency between the study conclusions and recommendations made in the guidelines.

Results

Guidelines varied considerably in length (range 3-350 pages), format and number of references (Table 2). Nine guidelines linked their recommendations to citations; four of these (FR, SC, CA, US2) also used grading systems to appraise the evidence.

Table 2. Length of guidelines, number of references and shared references

ID Code	Number of references	Number of pages	Number of references linked to relevant sections	% (Number) of shared references	Weighted shared score (ranking) ^a
FR	590	312	422	20.4 (86)	16.0 (9)
CA	302	29	158	46.2 (73)	19.6 (6)
NL1	246	164	127	42.5 (54)	18.8 (7)
US1	233 ^b	93 ^b	171	42.7 (73)	18.1 (8)
IT	218	350	83	31.3 (26)	15.2 (11)
NL2	190	18	132	44.7 (59)	20.6 (5)
SP	95	85	73	39.7 (29)	15.5 (10)
SC	77	21	56	42.9 (24)	18.1 (8)
US2	67	52	57	63.2 (36)	35.3 (1)
AU	65	92	12	66.6 (8)	24.3 (3)
NZ	44	19	25	56.0 (14)	33.5 (2)
EN	40	36	30	53.3 (16)	21.8 (4)
SW	2	3	excluded from further analysis		
FI	1	55	excluded from further analysis		
Total	2170	1329	1346	37.0 (498)	

^a weighted shared score = number of shared references x 100, divided by maximum possible number of shared references according to publication dates of the guideline and its linked references.

^b for the 11/42 selected ADA position or consensus statements only

Guidelines varied in their coverage; for example, the Danish and Spanish guidelines allocated more than 10% of the text to detailed dietary recommendations, whereas the English and New Zealand guidelines only made a few general statements. Guidelines also varied in their scope; for example the Scottish, the Australian and the Dutch guidelines (NL1) did not cover drug treatment of hyperglycemia.

Comparison of recommendations

The guidelines largely agreed on general management of patients with type 2 diabetes, which was covered by the following recommendations:

- All patients should be offered dietary advice and overweight/obese patients should be offered weight management advice.

- The diet should be low in sugar, fat content and overall calories, and should be combined with exercise.
- All patients should stop smoking to reduce cardiovascular risk.
- Patient education is necessary to promote good diabetic control.
- Poor glycaemic control should be tackled initially with diet alone, followed by oral medication, and insulin if necessary, unless the patient is acutely unwell.
- Sulphonylureas or biguanides are recommended in patients with normal Body Mass Index (BMI); metformin is recommended in obese patients.
- A second oral agent should be added to maximum doses of an initial agent in case of poor glycaemic control.
- HbA1c is suitable for long-term monitoring and should be lower than 8%.
- If on insulin, self-monitoring of blood glucose is recommended.
- Screening and treatment of raised blood pressure, microalbuminuria, and hyperlipidemia is recommended.
- ACE inhibitors are recommended in patients with hypertension and renal disease.
- Aspirin is recommended for secondary prevention of cardiovascular disease.

Differences between the recommendations were found in the following areas:

- Length of trial of diet and exercise before oral treatment ranged from 2 to 9 months; some guidelines recommended a longer period in obese patients compared to non-obese patients.
- BMI used to define obesity ranged from 25 to 30.
- Widely varying indications were suggested for the use of alphaglucohydrolase inhibitors.
- No consensus on the value or indications of combination therapy with oral hypoglycaemics and insulin.
- Target HbA1c ranged from 6.5 to 7.5%; target blood pressure ranged from <130/80 to <160/90.
- Frequency of monitoring HbA1c and blood pressure ranged from once to four times a year and one to six times a year.
- There was no consensus on self-monitoring of blood glucose in patients on diet alone or on oral medication.
- There was no consensus on the first-line drug for raised blood pressure

- Widely differing opinions were given on the value of aspirin use as primary prevention in 'high risk' patients
- Widely differing targets were given for lipid control (e.g. total cholesterol 4.5-6.5 mmol/l); there was no consensus on the use of absolute cardiovascular risk or isolated lipid levels for treatment decisions
- Routine annual ECG was recommended by half of the guidelines; others recommended ECG for specific indications or did not mention it.

Comparison of linked citations

We selected a total of 1346 references from 12 guidelines (Table 2); 1033 of these were different citations. Only 18% (185/1033) of the unique citations were shared with any of the other 11 guidelines. Considering all of the references made in the guidelines, on average 37% (498/1346) of these were shared with any other guideline (range 20-67%). The Diabetes Control and Complications Trial (DCCT) [17] was most frequently cited (in 11 guidelines). A randomised controlled trials addressing intensive insulin therapy with patients with type 2 diabetes was cited by eight guidelines [18]. If all 45 publications of the American Diabetes Association (ADA) were analysed as one document, it would be shared with eight guidelines. Two studies (one randomised controlled trial and one cohort study) were shared with seven guidelines; six trials were shared with six guidelines. Six guidelines referred to the WHO St. Vincent Declaration. Four of the twelve most frequent citations were from the USA, three from the UK (all three United Kingdom Prospective Diabetes Study publications), two from Israel, one each from Finland and Japan, and one was a WHO document.

The largest proportion of lead authors of papers cited in the guidelines (40%) originated from the United States (Table 3). All guidelines, except the Australian, cited a significantly higher proportion of studies from authors of their own countries compared to the origin overall of citations in the database ($p < 0.02$). Citations in the English, Scottish and New Zealand guidelines were predominantly from the United Kingdom; citations in all other guidelines, except the Dutch general practice guideline (NL2), were predominantly from the USA.

Sixteen of the total 1033 citations (2%) were meta-analyses, 89 (9%) were reviews or overviews (including four systematic reviews) and 55 (5%) were existing guidelines (including practical guides and clinical practice recommendations) or consensus statements. Twenty of these 160 secondary citations (13%) were American Diabetes Association publications.

Table 4. Countries of authors of citations (%)

ID Code	AU/ NZ	CA	UK/ IR	FR	IT	NL	SP	US	Scan- dinavian	Other	Multi- national	Un- known
AU	8	-	17	8	-	-	-	67	-	-	-	0
NZ	16	-	40	-	-	-	-	16	12	16	-	0
CA	3	6	13	1	3	1	-	45	17	7	1	5
EN	10	-	37	-	-	-	-	27	23	-	-	3
SC	2	-	39	4	-	2	-	23	23	5	-	2
FR	2	4	11	11	4	3	-	38	11	11	1	4
IT	-	1	6	1	30	-	-	41	4	2	4	10
NL1	-	1	13	1	3	18	1	32	21	6	2	1
NL2	3	1	13	1	1	36	-	24	12	10	-	0
SP	3	-	11	1	-	1	11	51	11	7	1	3
US1	1	1	9	1	3	1	-	59	18	5	1	2
US2	2	2	14	2	-	-	2	63	7	5	2	2
Total	2.6	2.7	11.8	5.2	4.7	7.1	1.0	40.4	11.8	7.8	1.3	3.8

Bold italics contain % of citation of studies from authors of the own country.

Examination of link between recommendations and citations (case studies)

USE OF METFORMIN IN OBESE PATIENTS

Eleven guidelines covered the use of oral medication. Nine explicitly recommended metformin as a first choice oral treatment for hyperglycemia in the obese; the Canadian and US1 guidelines recommended tailoring treatment for the individual. We compared the citations from six guidelines; the others had no citations linked to their recommendations on use of metformin. There was little overlap in the 20 citations given: one (UKPDS 34) was shared by four out of five guidelines with a censoring date that would allow use of this paper [19]. The UKPDS 13 paper was shared by three out of six guidelines [20], and three other citations were shared by two. Over half of the linked citations [11/20] were randomised controlled trials, one a meta-analysis, and the remainder were non-systematic reviews. All studies concluded that metformin was useful in obese patients. Whilst the choice of citations varied, publications from one trial (UKPDS) predominated and each guideline cited at least one publication that explicitly supported the recommendation.

SELF-MONITORING OF BLOOD GLUCOSE

Nine guidelines covered self-monitoring, and were unanimous in recommending the self-monitoring of blood glucose in type 2 diabetes treated with insulin. We

compared the citations from seven guidelines. Only two citations were present in more than one guideline: the DCCT trial [17] was cited in two and the ADA consensus statement [21] was shared with four. However, when we considered the primary studies in systematic reviews, meta-analyses or guideline and consensus documents, the overlap between citations increased substantially: 17 out of 33 references were then shared with at least two guidelines. For example, the Dutch and French guideline had seven citations in common by virtue of a systematic review conducted by Faas et al. [22]. Of the seven citations that specifically addressed self-monitoring in type 2 diabetes, five (two randomised controlled trials, one cross-sectional study, one review and one comment) concluded that there was no evidence to support its use. The two supportive citations were guidelines (an ADA consensus statement, and a Canadian guideline).

Discussion

This is the first study comparing both guideline recommendations and cited evidence across national guidelines. Our bibliometric analysis included more than 1000 citations. We minimised selection and observer bias by prospective choice of inclusion criteria for recommendations and citations and independent extraction by two researchers.

We found a high degree of international consensus on the clinical care of people with type 2 diabetes, despite differences in detailed recommendations. This was in contrast to what we expected, considering the range of influences on the guideline development process and the variation in organisation of care and health care system between countries [23]. Yet the citations linked to and, presumably, justifying, these guideline recommendations were widely disparate. The influence of large pragmatic treatment trials (e.g., DCCT [17] and UKPDS studies [19,24,25]) was nevertheless visible in most of the guidelines and apparent even in guidelines without references.

Little use was made of systematic reviews (for example Cochrane reviews), which is consistent with the findings of Silagy et al. [26]. National guidelines were significantly more likely to cite research from investigators from the same country, explaining some of the variation in citations between guidelines. Others have found that local sources of evidence are over-represented in guidelines [27], and that the results of trials conducted in the same country may be given more prominence [28].

We used the case studies to generate hypotheses to explain the small degree of overlap in citations between guidelines. Recommendations for the use of metformin in obese patients drew on supportive trial and review evidence. The different studies linked to these concordant recommendations often had similar conclusions. We also observed a consensus in recommendations for the use of self-monitoring of blood glucose, despite citation of evidence which did not support this position. The overlap in evidence would have been larger if we had aggregated citations from the same study (e.g. UKPDS), and if we had included the primary citations made within reviews and meta-analyses. Even taking this into account, the evidence cited in Type 2 diabetes guidelines largely does not overlap. Therefore we hypothesise that there are other potential influences on guideline developers. For example, the recommendations of the American Diabetes Association, strongly influenced the other guidelines on diabetes, with the exception of the English and Scottish. Similarly, Littlejohns and others found that professional opinion expressed in a consensus statement from the Royal College of General Practitioners and the Royal College of Physicians influenced the recommendations made in nine United Kingdom guidelines for the treatment of depression in primary care [29].

Guideline development is a social as well as technical process, which is affected by access to and choice of research evidence and decisions about the interpretation of evidence and formulation of recommendations [30-32]. Our study suggests that research evidence is not necessarily the most powerful influence on the content of recommendations in the current generation of guidelines on the management of type 2 diabetes mellitus. Guideline developers might first aim to achieve consensus about recommendations and then switch to the evidence as a rhetorical device to support decisions 'post hoc'. Thus, the relationship between choice and interpretation of research evidence and the formulation of guideline recommendations is neither necessarily linear nor uni-directional. However, we are not suggesting a complete epistemological divide between evidence as represented by research papers and guidelines recommendations. As Greenhalgh and McCormack have argued with regards to the UKPDS study, the interpretation of results within primary research studies is also debatable, influenced by prior beliefs, and open to challenge [33].

There are several sources of imprecision in our analysis. First, the guidelines were partly selected by researchers participating in the AGREE Collaboration. Therefore the sample might be biased towards guidelines developed with more explicit and robust methods, such as systematic searching and the use of evidence grading systems. Nevertheless, the extent and format of the guidelines differed widely. Six guidelines did not link their recommendations to evidence, which complicated the data extraction.

Secondly, we did not record the extent of initial agreement on choice of recommendations, judgement on their concordance or discordance, nor on linkage between citations and recommendations. However, there were few disagreements and these were easily resolved by panel consensus.

Thirdly, some of the variation in the content of the guidelines might be explained by the different publication dates of the guidelines and the rapid shift of information during the period studied. For instance, nine of the guidelines included in our study could not consider the UKPDS data that were published in 1998. In our analysis of the citations, we dealt with this confounding by correcting for publication dates of the guidelines and the cited evidence.

Finally, analysis of shared references is a blunt instrument for exploring the relationship between guideline recommendations and evidence. High-quality and large trials should be given more weight in the analysis. That is why we included two case studies exploring in more detail the relationship between recommendations and evidence in diabetes guidelines. Other clinical issues will need this kind of analysis to test the generalisability of our findings.

The process of formulating guideline recommendations and the social determinants of guidelines require further investigation. Decisions about choice of evidence and the role of international conferences, pharmaceutical companies and opinion forming bodies, such as the American Diabetes Association, on national guidelines is not well understood. The growing availability of high quality systematic reviews may support more uniformity in the use of research evidence in guidelines [34]. Nevertheless, guidelines go beyond simple reviews of available evidence, and necessarily reflect value judgements in considering all the issues relevant to clinical decision-making. Transparency by guideline developers about how their judgements have been made would allow clinicians to evaluate the applicability of guideline recommendations to their own health care context, and to individual patients.

References

1. Bero LA, Grilli R, Grimshaw JM, Harvey E, Oxman AD, Thomson MA. Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. The Cochrane Effective Practice and Organization of Care Review Group. *BMJ* 1998; 317: 465-468.
2. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence-based medicine: what it is and what it isn't. *BMJ* 1996; 312: 71-72.
3. Cochrane Collaboration. The Cochrane Library [database on disk and CD-ROM] Update Software, Oxford 2002. Updated quarterly.
4. National Health and Medical Research Council. A Guide to the Development, Implementation and Evaluation of Clinical Practice Guidelines. National Health and Medical Research Council, Canberra 1998.
5. Eisinger F, Geller G, Burke W, and Holtzman NA. Cultural basis of differences between US and French clinical recommendations for women at increased risk of breast and ovarian cancer. *Lancet* 1999; 353: 919-920.
6. Fahey TP, Peters TJ. What constitutes controlled hypertension? Patient based comparison of hypertension guidelines. *BMJ* 1996; 313: 93-96.
7. Psaty BM, Furberg CD. British guidelines on managing hypertension. Provide evidence, progress, and an occasional missed opportunity. *BMJ* 1999; 319: 589-590.
8. Dinkevitch E, Hupert J, Moyer VA. Evidence based paediatrics. Evidence based well child care. *BMJ* 2001; 323: 846-849.
9. Vogel N, Burnard B, Vial Y, Ruiz J, Paccaud F, Hohlfeld. Screening for gestational diabetes: variation in guidelines. *Eur J Obstet and Gynecol Reprod Biol* 2000; 91: 29-36.
10. Ravago TS, Mosniam J, Alem F. Evaluation of Community Acquired Pneumonia Guidelines. *J Med Sys* 2000; 24: 289-296.
11. Thomson R, McElroy H, Sudlow M. Guidelines on anticoagulant treatment in atrial fibrillation in Great Britain. variation in content and implications for treatment. *BMJ* 1998; 316: 509-513.
12. Beck C, Cody M, Souder E, Zhang M, Small GW. Dementia Diagnostic Guidelines: Methodologies, Results and Implementation Costs. *J Am Geriatr Soc* 2000; 48: 1195-1203.
13. Littlejohns P, Cluzeau F, Bale R, Grimshaw J, Feder G, Moran S. The quality and quantity of clinical practice guidelines for the management of depression in primary care in the UK. *Br J Gen Pract* 1999; 49: 205-210.
14. De Maeseneer J, Derese A. European general practice guidelines: a step too far? *Eur J Gen Pract* 1999; 5: 86-87.
15. Institute of Medicine. Field MJ, KN Lohr (eds). Clinical practice guidelines: directions for a new agency. Washington DC. National Academy Press, 1990.
16. Patton MQ. Qualitative Evaluation and Research Methods. In. Patton MQ. Designing Qualitative Studies. Newbury Park, CA: Sage, 1990: 145-198.
17. Diabetes Control and Complications Trial (DCCT) Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med* 1993; 329: 977-986.
18. Ohkubo Y, Kishikawa H, Araki E, Miyata T, Isami S, Motoyoshi S et al. Intensive insulin therapy prevents the progression of diabetic microvascular complications in Japanese patients with non-insulin-dependent diabetes mellitus: a randomized prospective 6-year study. *Diabetes Res Clin Pract* 1995; 28: 103-117.
19. UK Prospective Diabetes Study (UKPDS) Group. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). *Lancet* 1998; 352: 854-865.
20. UK Prospective Diabetes Study (UKPDS) Relative efficacy of randomly allocated diet, sulphonylurea, insulin, or metformin in patients with newly diagnosed non-insulin dependent diabetes followed for three years (UKPDS 13). *BMJ* 1995; 310: 83-88.
21. American Diabetes Association. Self-monitoring of blood glucose (Consensus Statement). *Diabetes Care* 1994; 17: 81-86.
22. Faas A, Schellevis FG, van Eijk JT. The efficacy of self-monitoring of blood glucose in NIDDM subjects. A criteria-based literature review. *Diabetes Care* 1997; 20: 1482-1486.

23. Payer L. *Medicine and culture: varieties of treatment in the United States, England, West Germany, and France*. New York: Penguin Books, 1989.
24. UK Prospective Diabetes Study (UKPDS) Group. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes (UKPDS 38). *BMJ* 1998; 317: 703-713.
25. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* 1998; 352: 837-853.
26. Silagy CA, Stead LF, Lancaster T. Use of systematic review in clinical practice guidelines: case study of smoking cessation. *BMJ* 2001; 323: 833-836.
27. Grant J, Cottrell R, Cluzeau F, Fawcett G. Evaluating "payback" on biomedical research from papers cited in clinical guidelines: applied bibliometric study. *BMJ* 2000; 320: 1107-1111.
28. Ramsay LE, Wallis EJ, Yeo WW, Jackson PR. The rationale for differing national recommendations for the treatment of hypertension. *Am J Hypertens* 1998; 11: 79S-88S.
29. Littlejohns P, Cluzeau F, Bale R, Grimshaw J, Feder G, Moran S. The quality and quantity of clinical practice guidelines for the management of depression in primary care in the UK. *Br J Gen Pract* 1999; 49: 205-210.
30. McIntosh A. Guidance: what role should consensus play in guideline development? In: Hutchinson A, Baker, eds. *Making Use of Guidelines in Clinical Practice*. Abingdon, Oxon: Radcliffe Medical Press, 1999: 63-81.
31. Berg M, Meulen RT, van den Burg M. Guidelines for appropriate care: the importance of empirical normative analysis. *Health Care Anal* 2001; 9: 77-99.
32. Hayward RS, Wilson MC, Tunis SR, Bass EB, Guyatt G. Users' guides to the medical literature. VIII. How to use clinical guidelines. A. Are the Recommendations valid? *JAMA* 1995; 274: 570-574.
33. McCormack J, Greenhalgh T. Seeing what you want to see in randomised controlled trials: versions and perversions of UKPDS data. *BMJ* 2000; 320: 1720-1723.
34. Cook DJ, Greengold NL, Gray Ellrodt A, Weingarten SR. The relation between systematic reviews and practice guidelines. *Ann Intern Med* 1993; 127: 210-216.

Chapter 6

Characteristics of effective clinical guidelines for general practice

Jako S Burgers

Richard PTM Grol

Joost OM Zaat

Teun H Spies

Akke K van der Bij

Henk GA Mookink

British Journal of General Practice; in press.

Abstract

Background. The use of clinical guidelines in general practice is often limited. Research on barriers to guideline adherence usually focuses on attitudinal factors. Factors linked to the guideline itself are much less studied.

Aim. To identify characteristics of effective clinical guidelines for general practice, and to explore whether these differ between therapeutic and diagnostic recommendations.

Methods. A set of 12 attributes, including 6 potential facilitators and 6 potential barriers to guideline use, was formulated. A panel of 12 general practitioners assessed the presence of these attributes in 96 guideline recommendations formulated by the Dutch College of General Practitioners. Compliance rates were derived from an audit study of 200 general practitioners. The attributes of recommendations with high compliance rates (70-100%) were compared to those with low compliance rates (0-60%).

Results. High compliant recommendations were to a lesser extent requiring new skills (7% compared to 22% in low compliant recommendations), were less often part of a complex decision tree (12% versus 25%), were more compatible with existing norms and values in practice (87% versus 76%) and more often supported with evidence (47% versus 31%). For diagnostic recommendations the ease of applying them and the potential (negative) reactions of patients were more relevant than for therapeutic recommendations.

Conclusions. To bridge the gap between research and practice, the evidence as well as the applicability should be considered in formulating recommendations. If the recommendations are not compatible with existing norms and values, not easy to follow, or require new knowledge and skills, appropriate implementation strategies should be designed to ensure change in daily practice.

Introduction

Within the past decade considerable time and energy have gone into the development of evidence-based guidelines for improving clinical practice. Unfortunately, not all guidelines actually improve the quality of care [1]. Why are some guidelines successful in changing care and others not? Research on barriers to guideline adherence is often qualitative in nature and focuses on guideline users and their behaviour [2-5]. In contrast, factors linked to the guideline itself are much less studied. Literature on characteristics of effective guidelines is very limited. Rogers suggested that attributes as relative advantage, compatibility, complexity, triability and observability may influence the adoption of an innovation [6]. Grilli and Lomas confirmed that the complexity and triability of recommendations could partly predict the level of compliance with a guideline [7]. Based on literature, Grol et al. extended the number of attributes that might influence the use of guidelines in practice [8]. Their study showed that controversial recommendations, vague and non-specific recommendations and recommendations that demand changing existing routines and habits were less likely to be followed than their counterparts. However, their study was limited in the number of recommendations studied. Further research in this area is necessary to ensure that guidelines are developed in a way that they are optimally effective in improving patient care [9]. A better understanding of those aspects of a guideline that make a difference in daily clinical practice may guide the setting of guidelines and recommendations for practice in a positive way.

In this study we aimed to identify characteristics of effective guidelines using a large sample of concrete recommendations with contrasting compliance rates. We examined to what extent the attributes of recommendations with high compliance rates differed from those of recommendations with low compliance rates. In addition, we explored differences between diagnostic and therapeutic recommendations.

Methods

Clinical guidelines are documents that contain a set of individual recommendations covering one specific disease area. For this study we included recommendations defined as 'systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances' [10]. Effective guidelines are guidelines whose recommendations are actually followed in practice.

Formulation of attributes of recommendations

Based on previous work [8] and supplemented with items derived from other instruments assessing the quality of guidelines [11 12], we started with a set of 16 attributes of recommendations. In a pilot, using 28 recommendations, it appeared that two attributes (i.e., 'concerned with a relevant aspect of daily care' and 'concretely and specifically formulated') were present in more than 25 of the recommendations and two attributes (i.e., 'supported with a discussion of costs' and 'demands extra resources') were present in less than two recommendations. These four attributes did not sufficiently discriminate and were therefore excluded. Thus, the final instrument consisted of 12 attributes (Box 1). Half of these are expected to facilitate adherence, the other half can be seen as potential barriers to physician guideline use.

Box 1. Attributes of recommendations

Potential facilitators

1. The recommendation is largely supported with scientific evidence (e.g. clinical trials, cohort or case-control studies).
2. The recommendation is supported with a discussion of the benefits (e.g. health gain).
3. The recommendation is supported with a discussion of the harms and risks (e.g. drug side effects).
4. The recommendation is easily to follow.
5. The effects of the recommendation can be seen quickly.
6. The recommendation is compatible with existing norms and values in practice.

Potential barriers

7. The recommendation is part of a complex decision tree
8. The application of the recommendation requires new knowledge.
9. The application of the recommendation requires new skills.
10. The application of the recommendation demands changes in the organisation.
11. The application of the recommendation requires changes in existing routines and habits.
12. The recommendation can evoke negative reactions in patients.

Selection of recommendations

We used performance data derived from an audit study of 200 general practitioners in the Netherlands conducted in 1997 [13]. The general practitioners were randomly selected from proportional samples in different regions of the country. Data were collected from 7,614 consultations using validated self-recording forms that were filled in immediately or shortly after the consultation. With these data compliance was determined with key recommendations in 29 guidelines selected from a total of 51 available guidelines developed by the Dutch College of General Practitioners ('NHG Practice Guidelines'). The selected guidelines were equally distributed over different disease areas and covered acute

as well as chronic diseases. They were rigorously developed according to principles of evidence-based medicine and were published in *Huisarts en Wetenschap*, the Dutch scientific journal for general practitioners, which reaches about 85% of the practitioners [14]. The guidelines are widely accepted and play a prominent role in programmes of continuing medical education in the Netherlands [15-17]. The mean compliance rate in the audit study was 71%.

For our study, we excluded recommendations with compliance rates between 60 and 70% because these can be considered as neither 'effective' nor 'non-effective' and thus not of interest to this study. We selected recommendations proportionally distributed over the categories diagnosis, education, treatment, follow up and referral, with a maximum of six per guideline. Thus, 63 recommendations with high compliance rates (70-100%) and 33 recommendations with low compliance rates (0-60%) were selected (Table 1).

Table 1. Compliance rates of selected recommendations

Compliance rate	Number of recommendations (%)
0 - 10%	2 (2.1)
11 - 20%	4 (4.2)
21 - 30%	11 (11.5)
31 - 40%	9 (9.4)
41 - 50%	5 (5.2)
51 - 60%	2 (2.1)
61 - 70%	0
71 - 80%	10 (10.4)
81 - 90%	24 (25.0)
91 - 100%	29 (30.2)
Total	96 (100)

Formal assessment of recommendations

We composed a panel of twelve experienced general practitioners who were familiar with guideline methodology but not directly involved in the formulation of the recommendations included in this study. The recommendations were divided into six clusters and were independently assessed by two panel members. Each pair assessed one cluster. The panel members were asked to determine whether the 12 attributes (Box 1) were present or not present in the recommendations. They were blind to the actual compliance rates of the recommendations. We provided the panel members with a user guide to help them with the assessment.

The results of the assessment were returned to each pair. Disagreement was resolved by discussion. Consensus was achieved in 99% of the assessments.

Analysis

We compared the presence of attributes of recommendations with high compliance rates (71-100%) with that of recommendations with low compliance rates (0-60%). Crosstable statistics were used to calculate odds ratios (OR) that can be considered as a measure of association between individual attributes and compliance rate. We ranked the attributes using the reciprocal values of odds ratios between 0 and 1. We also analysed differences in odds ratios between diagnostic and therapeutic recommendations. The remaining categories (i.e., education, follow up and referral) included too few recommendations (19 and 9 respectively) to calculate odds ratios. All analyses were performed using SPSS 9.0.

Results

Four attributes were positively associated with high compliance rate; eight attributes had a negative effect on the compliance rate (Table 2). The effects of 'supported discussion of harms' and 'effects can be seen quickly' on the compliance were negative in contrast to what was expected. All six potential barriers (attributes 7-12) had indeed a negative effect on the compliance rate. The strength of the association varied among different attributes and was the highest for 'requires new skills' (OR = 0.25, 1/OR = 4.00), followed by 'part of complex decision tree' (OR = 0.40, 1/OR = 2.50) and 'compatible with norms and values' (OR = 2.20).

The influence of the different attributes on the compliance rate was not similar for diagnostic and therapeutic recommendations (Table 3). The support of the recommendation with a discussion of benefits and harms (attribute 3 and 4) was only positively associated with high compliance rates for therapeutic recommendations. For diagnostic recommendations the influence of 'part of complex decision tree' and 'easy to follow' was more relevant than for therapeutic recommendations. 'Evoke negative reactions in patients' was negatively associated with high compliance rates for diagnostic recommendations but positively associated with high compliance rates for therapeutic recommendations in contrast to what was expected.

Table 2. Attributes of recommendations with high or low compliance rates. Figures are % of recommendations (number of recommendations)

	Attribute present in recommen- dations with high compliance rates (71-100%) (n = 63)	Attribute present in recommen- dations with low compliance rates (0-60%) (n = 33)	odds ratio 95% CI	ranking ^a
<i>Potential facilitators</i>				
1 supported with scientific evidence	46.8 (29)	31.3 (10)	1.93 (0.87-4.75)	4
2 supported with discussion of benefits	40.3 (25)	36.4 (12)	1.18 (0.49-2.83)	12
3 supported with discussion of harms	9.5 (6)	15.2 (5)	0.59 ^b (0.17-2.10)	6
4 easy to follow	88.9 (56)	81.3 (26)	1.85 (0.56-6.04)	5
5 effects can be seen quickly	59.7 (37)	71.0 (22)	0.60 ^b (0.24-1.53)	7
6 compatible with norms and values	87.3 (55)	75.8 (25)	2.20 (0.74-6.53)	3
<i>Potential barriers</i>				
7 part of complex decision tree	11.7 (7)	25.0 (8)	0.40 (0.13-1.22)	2
8 requires new knowledge	19.0 (12)	28.1 (9)	0.60 (0.22-1.63)	8
9 requires new skills	6.6 (4)	21.9 (7)	0.25 (0.07-0.93)	1
10 requires changes in organisation	8.1 (5)	12.1 (4)	0.64 (0.16-2.55)	10
11 requires changes in routines	36.5 (23)	48.5 (16)	0.61 (0.26-1.43)	9
12 can evoke negative reactions in patients	38.1 (24)	42.4 (14)	0.83 (0.35-1.97)	11

^a Ranking of attributes was determined by using the reciprocal values of odds ratios between 0 and 1^b Reverse to what was expected**Table 3. Attributes of diagnostic and therapeutic recommendations**

	Diagnostic recommendations (n = 37)		Therapeutic recommendations (n = 31)	
	odds ratio (95% CI)	ranking ^a	odds ratio (95% CI)	ranking ^a
<i>Potential facilitators</i>				
1 supported with scientific evidence	2.81 (0.50-15.7)	4	3.71 (0.70-19.6)	2
2 supported with discussion of benefits	0.54 ^b (0.13-2.30)	7	3.20 (0.72-14.1)	4
3 supported with discussion of harms	NA ^c		1.13 (0.21-6.17)	11
4 easy to follow	3.83 (0.55-26.9)	3	0.62 ^b (0.05-7.75)	7
5 effects can be seen quickly	0.60 ^b (0.10-3.55)	9	0.30 ^b (0.06-1.49)	3
6 compatible with norms and values	NA ^c		1.07 (0.41-2.79)	12
<i>Potential barriers</i>				
7 part of complex decision tree	0.19 (0.04-0.85)	1	0.80 (0.04-14.2)	9
8 requires new knowledge	0.55 (0.12-2.56)	8	0.54 (0.10-2.94)	6
9 requires new skills	0.19 (0.01-2.33)	2	0.24 (0.02-2.68)	1
10 requires changes in organisation	0.43 (0.05-3.54)	6	0.87 (0.05-15.3)	10
11 requires changes in routines	0.67 (0.16-2.73)	10	0.73 (0.17-3.11)	8
12 can evoke negative reactions in patients	0.40 (0.10-1.64)	5	2.22 ^b (0.50-9.96)	5

^a Ranking of attributes was determined by using the reciprocal values of odds ratios between 0 and 1^b Reverse to what was expected ^c NA = not assessable due to low numbers

Discussion

Our study shows that the applicability of recommendations is at least as relevant as their support with evidence to guarantee adherence to guidelines. The most important barriers to the application of recommendations are concerned with the need for new skills and the complexity of the recommendations. When the recommendations are easy to follow and compatible with norms and values, then the application will be facilitated.

For diagnostic recommendations, the ease of applying them in practice seems to be more important than for therapeutic recommendations. Complex diagnoses (e.g. syndromes with more than four criteria) or inconvenient procedures (e.g. gastroscopy) may hinder physicians to follow guidelines, even if there is sufficient evidence for them. In contrast, for ensuring use of therapeutic recommendations, the strength of the evidence seems relatively more important than factors as complexity and patient expectations.

The findings of our study help to understand why guidelines may not be used and why certain recommendations are more likely to be followed than others. This may be useful to guideline development organisations as well as primary care groups responsible for implementing clinical governance in primary care [18].

Some limitations of this study should be mentioned. Despite the large number of recommendations, the confidence intervals for the odds ratios were quite wide. Only for one attribute ('requires new skills') the confidence intervals did not include one. Nevertheless, the influence of 10 (out of 12) attributes confirmed our hypotheses, which is probably not being due to random effects. Conducting a study on a larger scale would be difficult, because each recommendation requires as much decisions as the number of attributes. This will progressively increase the task of the panel assessing the recommendations.

We could not test the influence of attributes related to the clarity of recommendations, because almost all selected recommendations were concrete and specific. This is not surprising because monitoring and audit studies use review criteria that are primarily based on concrete and specific recommendations [19]. We are still convinced that recommendations should be concrete and specific in order to change behaviour or practice. In a randomised controlled trial Shekelle et al. confirmed that the clarity of a guideline significantly contributed to its effect [20].

Although the guidelines were discussed in the context of continuing medical education, one might argue whether the general practitioners were aware of all

recommendations included in our study. However, their decisions could still comply with the recommendations if these reflect current practice.

Our study builds on previous work of Grol et al. and its findings are largely consistent with this study [8]. However, both studies are limited by a cross-sectional design and could therefore not determine the ability of guidelines to change practice. In contrast, Foy et al. examined attributes of recommendations in a retrospective study using compliance rates before and after audit and feedback [21]. The results confirmed that recommendations compatible with existing norms and values, and not requiring changes to fixed routines, were associated with greater compliance. However, significant changes in compliance were only measured for recommendations seen as incompatible. In other words, the more compatible the recommendation the smaller the behavioural change.

Guidelines are developed to close the gap between research and practice, but the appearance of guidelines creates a new gap between their development and use in practice. Whereas guidelines essentially aim to influence or change practice, they would be of little value if they are not used. Therefore guideline developers should consider the evidence as well as the applicability in formulating recommendations. For each recommendation they should ask themselves whether the recommendation is compatible with existing norms and values in practice and easy to follow, or complex and requires new knowledge and skills. Pilot testing of the guidelines among target users may provide additional information on barriers of implementation. If the application of the recommendation is expected to be difficult but the supportive evidence is strong, appropriate implementation strategies should be designed to ensure change in daily practice. For instance, tools for application such as algorithms or balance sheets might facilitate the adherence to complex recommendations. If the recommendations require new skills, workshops should be organised shortly after dissemination of the guideline. If negative reactions of patients can be expected, specific mass media information for the general public may be helpful. Thus, anticipating on the specific barriers of implementations will increase the effectiveness of the guidelines.

Future research should provide more information on the ability of guidelines to change practice. A prospective study with baseline and follow-up compliance data may be set up to study this aspect in more detail. The next step might be to measure the effect of clinical guidelines on changing patient outcomes [22]. Finally, qualitative studies concerning the reasons why physicians follow or do not follow guidelines could complement the knowledge about effective guidelines.

References

1. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 1993; 342: 1317-1322.
2. Watkins C, Harvey I, Langley C, Gray S, Faulkner A. General practitioners' use of guidelines in the consultation and their attitudes to them. *Br J Gen Pract* 1999; 49: 11-15.
3. 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: 1458-1465.
4. Freeman AC, Sweeney K. Why general practitioners do not implement evidence: qualitative study. *BMJ* 2001; 323: 1-5.
5. Grimshaw JM, Shirran L, Thomas R, Mowatt G, Fraser C, Bero L, Grilli R, Harvey E, Oxman A, O'Brien MA. Changing provider behaviour: an overview of systematic reviews of interventions. *Med Care* 2001; 39(Suppl): I12-45.
6. Rogers EM. *Diffusion of Innovations*. 4th Ed. New York: Free Press, 1995.
7. Grilli R, Lomas J. Evaluating the message: the relationship between compliance rate and the subject of a practice guideline. *Med Care* 1994; 32: 202-213.
8. Grol R, Dalhuijsen J, Thomas S, in 't Veld C, Rutten G, Mookink H. Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. *BMJ* 1998; 317: 858-861.
9. Baker R. Is it time to review the idea of compliance with guidelines? *Br J Gen Pract* 2001; 51: 7.
10. Institute of Medicine. Field MJ, Lohr KN (eds). *Clinical practice guidelines: directions for a new agency*. Washington, DC: National Academy Press, 1990.
11. Cluzeau FA, Littlejohns P, Grimshaw JM, Feder G, Moran SE. Development and application of a generic methodology to assess the quality of clinical guidelines. *Int J Qual Health Care* 1999; 11: 21-28.
12. The AGREE Collaboration. *Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument*. www.agreecollaboration.org.
13. Spies TH, Mookink HGA. Toetsen aan standaarden. Het medisch handelen van huisartsen in de praktijk getoetst [eindrapportage]. Nijmegen: Centre for Quality of Care Research (WOK), 1999.
14. Grol R, Thomas S, Roberts R. Development and implementation of guidelines for family practice: lessons from The Netherlands. *J Fam Pract* 1995; 40: 435-439.
15. Grol R. National standard setting for quality of care in general practice: attitudes of general practitioners and response to a set of standards. *Br J Gen Pract* 1990; 40: 361-364.
16. Grol R, Zwaard A, Mookink 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: 135-140.
17. Grol R. Successes and failures in the implementation of evidence-based guidelines for clinical practice. *Med Care* 2001; 39(Suppl.): I146-I154.
18. Campbell S, Roland M, Wilkin D. Primary care groups. Improving the quality of care through clinical governance. *BMJ* 2001; 322: 1580-1582.
19. Baker R, Fraser RC. Development of review criteria: linking guidelines and assessment of quality. *BMJ* 1995; 311: 370-373.
20. Shekelle PG, Kravitz RL, Beart J, Marger M, Wang M, Lee M. Are non-specific practice guidelines potentially harmful? A randomized comparison of the effect of non-specific versus specific guidelines on physician decision making. *Health Serv Res* 2000; 34: 1429-1448.
21. Foy R, MacLennan G, Grimshaw J, Penney G, Campbell M, Grol R. Attributes of clinical recommendations that influence change in practice following audit and feedback. *J Clin Epidemiol* 2002; 55: 717-722.
22. Marshall DA, Simpson KN, Norton EC, Biddle AK, Youle M. Measuring the effect of clinical guidelines on patient outcomes. *Int J Technol Assess Health Care* 2000; 16: 1013-1023.

Chapter 7

The quality of Dutch clinical guidelines for general practice. Evaluation of 130 key recommendations from 28 guidelines

Jako S Burgers

Joost OM Zaat

Teun H Spies

Akke K van der Bij

Henk GA Mookink

Richard PTM Grol

Published as 'De kwaliteit van de NHG-standaarden. Beoordeling van 130 kernaankbevelingen uit 28 standaarden'. Huisarts en Wetenschap 2002; 45(7): 349-353.

Abstract

Introduction. For several years, the guidelines of the Dutch College of General Practitioners ('NHG Practice Guidelines') have provided a broadly-accepted scientific basis for general practice. However, the quality of the recommendations in the guidelines has never been studied systematically.

Methods. A panel of twelve general practitioners appraised 130 key recommendations selected from 28 Dutch guidelines for general practice (published in 1993-1997), using an instrument of ten items divided into three dimensions (scientific support, compatibility, and applicability).

Results. Only 44% of the recommendations were supported by a discussion of scientific evidence. Scientific evidence was provided for 67% of the therapeutic recommendations, compared with 35% of the recommendations for diagnosis and 29% of those for education. Most recommendations did not require new knowledge (75%), skills (86%), or organisational changes (88%). The applicability was still limited, for 43% of the recommendations required changes in routines and habits. Negative reactions in patients were expected in 39% of the recommendations.

Conclusions. The scientific support of the Dutch guidelines for general practice could be improved. The applicability of the recommendations is usually sufficient, but this could also mean that the recommendations reflect actual practice.

Introduction

Since 1989, the Dutch College of General Practitioners (NHG) has produced more than 70 clinical practice guidelines. Although the guidelines are broadly accepted, some authors have criticised the quality of the recommendations in the guidelines [1-4]. There have been several general discussions [5-9] but no systematic evaluation of the quality of the recommendations, in contrast to the numerous studies of their implementation and application [10].

A problem in assessing the quality of recommendations in guidelines is that a validated appraisal instrument is not available. The key question is: what are the criteria for good recommendations? The following basic criteria have often been considered [5,6,12-15]:

- The recommendations should be supported by scientific evidence.
- The recommendations should be compatible with existing norms and values in daily practice.
- The recommendations should be feasible and applicable in practice.

In this study we examined the extent to which recommendations in NHG practice guidelines meet these criteria. In addition, we were interested in potential differences among recommendations on diagnosis, education, treatment, monitoring, and referral to a specialist.

Methods

Development of appraisal criteria

Based on findings in the literature [12-15], we composed a list of criteria that operationalised the three dimensions listed above: (1) scientific support, (2) compatibility, and (3) feasibility.

Scientific support relates to the availability of empirical studies (*evidence*) and the discussion of the evidence linked to the recommendation. When evidence is lacking, a discussion of benefits, harms, and risks could provide a substitute. The completeness and precision of the scientific support was not assessed, because to do so requires specific clinical expert knowledge. Moreover, the NHG Practice Guidelines do not provide criteria for searching and selecting needed to assess the method of evidence collection [16].

Compatibility is concerned with the norms and values of the general practitioners as well as with the views and expectations of patients.

Feasibility is determined by the extent to which new knowledge, new skills, changes in the organisation, and changes in existing routines are required in applying the recommendations.

The criteria were formulated specifically and unambiguously, and overlapping between criteria was minimised. The final 'appraisal instrument' consisted of ten criteria (Table 1).

Table 1. Appraisal criteria of recommendations in clinical guidelines

<i>Scientific support</i>	
1	The recommendation is explicitly linked to the supporting evidence
2	The recommendation is primarily supported by scientific evidence
3	The recommendation is supported by a discussion of the benefits of applying the recommendation (health gain or quality of life)
4	The recommendation is supported by a discussion of the harms and risks of applying the recommendation
<i>Compatibility</i>	
5	The recommendation is compatible with existing norms and values in practice
6	The recommendation does not evoke negative reactions in patients
<i>Feasibility</i>	
7	Application of the recommendation requires no new knowledge
8	Application of the recommendation requires no new skills
9	Application of the recommendation requires no changes in the organisation
10	Application of the recommendation requires no changes in existing routines or habits

Selection of recommendations

In this study we used data of the research project 'Toetsen Aan Standaarden' (TAS project = evaluating NHG Practice Guidelines) [17]. This project was designed to determine the extent to which the key recommendations in the guidelines were followed in practice. A panel of guideline experts selected 324 concrete key recommendations from 29 NHG Practice Guidelines (published between 1993 and 1997). They aimed for an equal distribution of clinical topics, including most categories of the International Classification of Primary Care (ICPC) and including acute as well as chronic diseases. In our study, however, it was not feasible to assess all 324 key recommendations. Therefore we (JB and JZ) selected approximately one-third of the recommendations according to the following criteria:

- on average, four to five recommendations per guideline (with a maximum of eight)
- an equal distribution of recommendations on diagnosis, education, non-drug treatment, drug treatment, monitoring, and referral
- sufficient prevalence of the conditions of the recommendations deduced from the number of consultation registration forms collected in the TAS project.

Using these criteria, we selected 130 recommendations from 28 NHG Practice Guidelines (Table 2).

Table 2. Selected NHG Practice Guidelines (number of recommendations)

NHG-Standaarden voor de huisarts 1, edition 1993 (NHG Practice Guidelines for general practice)

- M01 Diabetes mellitus type 2 (6)
- M04 Ankle sprain (2)
- M09 Acute otitis media (4)
- M13 Peripheral arterial disease (5)
- M15 Acne vulgaris (2)
- M18 Otitis media with effusion (3)
- M19 Migraine (4)
- M20 Hypercholesterolaemia (5)
- M28 Vaginal blood loss (5)

NHG-Standaarden voor de huisarts 2, edition 1996

- M34 Acute diarrhoea (4)
- M37 Atopic eczema (5)
- M38 Vaginal discharge (5)
- M41 Rheumatoid arthritis (4)
- M43 Angina pectoris (8)
- M44 Depression (6)
- M45 TIA (5)
- M48 Allergic and hypersensitive rhinitis (5)
- M49 Tropical ear/tank ear (4)
- M51 Congestive heart failure (5)
- M54 Low back pain (5)
- M55 Herniated lumbar intervertebral disc (4)
- M57 The red eye (5)

Updated NHG guidelines

- M36 Gastritis and peptic ulcer disease, 1996 (5)
 - M17 Hypertension, 1997 (7)
 - M26 Asthma and COPD in adults – diagnosis, 1997 (5)
 - M27^b Asthma in adults – treatment, 1997 (5)
 - M27^a COPD in adults – treatment, 1997 (6)
 - M42 Benign prostatic hyperplasia, 1997 (5)
-

Appraisal of recommendations

We composed an external panel of twelve general practitioners to appraise the recommendations using the ten criteria. All panel members were experienced general practitioners and experts in general practice by reason of their academic affiliations. The recommendations were distributed among six groups. Each group consisted of recommendations covering the same clinical areas. Two panel members independently assessed one group of recommendations and determined whether they did or did not fulfil the criteria (dichotomous scale). The complete text of the NHG Practice Guidelines was enclosed to support the appraisal. We also provided additional information in a user guide adjacent to each item (Appendix D). The appraisal results were fed back to the pair of assessors, who were asked to discuss the contrasting results and to reach consensus in so far as possible. Before discussion, the proportion of agreement was 71% of all judgments (mean kappa 0.24; range 0-0.47). After discussion, the proportion of agreement was 99%.

Analysis

The dimension scores were determined by calculating the mean percentage of the individual criteria in a dimension. The scores were analysed in aggregate and according to consultation phase.

Results

General scores

Eighty (62%) of the 130 recommendations were explicitly supported by evidence, usually described in an explanatory note (Table 3). In 57 of these 80 recommendations, the evidence included scientific research, i.e., empirical studies such as clinical trials or observational studies. In more than half (56%) of all recommendations, scientific support was lacking. The benefits were discussed in 41% of the recommendations, while the harms and risks were only discussed in 14% of the recommendations.

The panel found 52% of the recommendations to be compatible with existing norms and values in practice as well as with the supposed expectations of the patient. Negative reactions in patients were assumed in 39% of the recommendations. The application of 48% of the recommendations required no new knowledge or skills, changes in organisation, or changes in routines (criteria

7-10 were all present). Feasibility was particularly limited by changes in routines, which were required in 43% of the recommendations.

Table 3. Appraisal of 130 recommendations from 28 NHG Practice Guidelines

Criterion	Criterion present (%)
<i>Scientific support (dimension score 40.7%)</i>	
1. explicit link	80 (62)
2. scientific evidence	57 (44)
3. discussion of benefits	54 (41)
4. discussion of harms and risks	18 (14)
<i>Compatibility (dimension score 72.3%)</i>	
5. compatible with norms and values	109 (84)
6. no negative reactions in patients	79 (61)
<i>Feasibility (dimension score 78.0%)</i>	
7. no new knowledge required	97 (75)
8. no new skills required	112 (86)
9. no changes in organisation required	114 (88)
10. no changes in routine required	74 (57)

Scores according to consultation phase

The relative distribution of the selected recommendations was 39% diagnosis, 32% therapy, 16% education, 9% monitoring, and 3% referral to a specialist. Scientific evidence was provided in support of 67% of the therapeutic recommendations, compared with 35% of the diagnostic recommendations and 29% of the recommendations on education and monitoring (Table 4). A discussion of the benefits was provided for 69% of the recommendations on non-drug treatment, while a discussion of harms and risks (e.g., side effects) was provided for 38% of the recommendations on drug treatment.

The compatibility of the recommendations on drug treatment was high (81%), according to the panel. In contrast, 69% of the recommendations on non-drug treatment could evoke negative reactions in patients and had a lower compatibility (59%). Feasibility was highest (87%) for the recommendations on drug treatment. The diagnostic recommendations required relatively more new knowledge (32%), while the recommendations on non-drug treatment and recommendations on monitoring and referral required relatively more changes in existing routines and habits (62% and 56%, respectively).

Table 4. Scores by consultation phase (%)

	Diagnosis (n=51)	Education (n=21)	Non-drug treatment (n=16)	Drug treatment (n=26)	Monitoring and referral criteria (n=16)
<i>Scientific support</i>	33.8	27.4	54.7	61.2	33.9
1. explicit link	63	48	75	82	50
2. scientific evidence	35	29	63	71	37
3. discussion of benefits	33	24	69	58	43
4. discussion of harms and risks	0	0	13	38	13
<i>Compatibility</i>	74.5	69.0	59.4	80.8	68.8
5. compatible with norms and values	90	81	87	77	75
6. no negative reactions in patients	59	57	31	85	63
<i>Feasibility</i>	77.1	75.0	68.3	87.5	78.1
7. no new knowledge required	68	81	69	81	87
8. no new skills required	90	80	80	96	94
9. no changes in organisation required	88	86	87	96	87
10. no changes in routine required	60	52	38	77	44

Discussion

The results of this study reveal that in general the key recommendations of the NHG Practice Guidelines are compatible with daily practice, being relatively feasible and applicable. However, the scientific support for the recommendations, especially on diagnosis, education, monitoring, and referral, could be improved. Although most recommendations referred to an explanatory note or literature citation, few were supported by empirical studies. This could be due to the lack of well-designed studies, in agreement with the study of Tasche et al. that identified on average 12.5 gaps in knowledge per guideline [18]. Further analysis of this study reveals that 42 research questions (resulting from the gaps in knowledge) concerned recommendations selected for our study. However, our panel judged that scientific evidence (criterion 2) was absent in only 14 of these 42 recommendations (33%). In contrast with the study of Tasche et al., our panel did not assess the quality of the evidence, which might explain the discrepancy. Moreover, the availability of scientific evidence does not exclude a gap in knowledge, in particular when the quality of the evidence is poor. Rigorous argument, including discussion of the benefits, harms, and risks of applying the recommendation, could often provide additional information, but this was done in

only a minority of the recommendations. Some recommendations, however, such as drinking more than usual in acute diarrhoea or not using contact lenses in case of conjunctivitis, are so self-evident that supporting evidence or arguments are less necessary.

The compatibility with existing norms and values in practice was high. However, the panel found that 40% of the recommendations could evoke negative reactions in patients. In particular, 69% of the recommendations on non-drug treatment were not patient friendly, possibly because of the high proportion of unpopular recommendations such as stopping smoking. However, we have not examined whether patients really dislike these recommendations. There is often a discrepancy between what practitioners suppose that patients think and what patients do think [19].

The high compatibility and feasibility of most of the recommendations can also be explained by the fact that they reflect actual practice and do not necessitate a change in behaviour. Nevertheless, we found that 40% of the recommendations required changes in routines. It should be noted that the agreement between the two appraisers before discussion was relatively low (60%) for this criterion. Routines and habits probably differ between individual practitioners.

Our findings should be taken with some reserve, for several reasons. We attempted to assess the quality of NHG Practice Guidelines using criteria derived from other instruments [12,14], but the validity of these criteria is unknown. Moreover, criteria testing is complicated because a gold standard is not available. This is a common problem in appraising guidelines [20,21].

Next, the assessment, particularly of compatibility and feasibility, is rather subjective. We anticipated this problem by asking the panel to base their judgment on the expected opinions of *general practitioners in general*, instead of on their own opinion. Nevertheless, we cannot exclude that the composition of the panel influenced the assessment.

Finally, we did not include recommendations from recent NHG Practice Guidelines. Half of the guidelines used in our study have now been updated, but since the scientific support of the recommendations may have been improved, we cannot simply extrapolate our findings to current NHG Practice Guidelines.

References

1. Dunk JHM. Pseudo-wetenschappelijk. (Pseudo-scientific) Huisarts Wet 1990; 33: 330-331
2. Te Dorsthorst JGJM. Standaarden. (Practice Guidelines) Huisarts Wet 1993; 36: 32.
3. Te Dorsthorst JGJM. De NHG-standaard Epicondylitis. Een voorbeeld van een onvoldoende wetenschappelijk gefundeerde richtlijn. (The NHG Practice Guideline Epicondylitis) Huisarts Wet 1997, 40: 671-673.
4. De Jongh TOH, Van Ruijven AGH. Waarom sluiten NHG-standaarden onvoldoende aan bij de diagnostiek van de huisarts? (Why are NHG Practice Guidelines insufficiently compatible with diagnose in general practice?) Huisarts Wet 2000; 43: 340-342.
5. Rutten GEHM. Criteria en waarborgen voor de kwaliteit van NHG-standaarden. (Criteria and guaranties for quality of NHG Practice Guidelines) Huisarts Wet 1990; 33: 338-339.
6. Meijman FJ. Publicatie van protocollen: ja, mits... (Publication of protocols: yes, providing that...) Huisarts Wet 1990; 33: 414.
7. Burgers JS. De wetenschappelijke onderbouwing van de NHG-standaarden. (The scientific base of the NHG Practice Guidelines) Huisarts Wet 1994; 37: 188-193.
8. Wiersma TJ, Meulenbergh F, Burgers JS. Helderheid en consistentie van standaarden: een kwestie van coherentie. (Clarity and consistency of guidelines: a matter of coherence) Huisarts Wet 1995, 38: 175-177.
9. Burgers JS. NHG-Standaarden en de invloed van evidence-based medicine. (NHG Practice Guidelines and the influence of evidence-based medicine) In: Geijer RMM, Burgers JS, Van der Laan JR, Wiersma TJ, Rosmalen CFH, Thomas S. NHG-Standaarden voor de huisarts, deel 1. (NHG Practice Guidelines for general practice, part 1) Maarssen: Elsevier/Bunge 1999: 11-23
10. Grol R, Wensing M, redactie. Implementatie. effectieve verandering in de patientenzorg. (Implementation: effective change in patient care). Maarssen: Elsevier gezondheidszorg, 2001
11. Graham ID, Calder LA, Hébert PC, Carter AO, Tetroe JM. A comparison of clinical practice guideline appraisal instruments. Int J Technol Assess Health Care 2000; 16: 1024-1038.
12. Institute of Medicine Field MJ, Lohr KN (eds). Guidelines for clinical practice. From development to use. Washington DC. National Academy Press, 1992.
13. Grol R, Dalhuijsen J, Thomas S, in 't Veld C, Rutten G, Mookink H. Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. BMJ 1998, 317: 858-861
14. The AGREE Collaboration. Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument www.agreecollaboration.org, 2001.
15. Burgers JS, Grol R, Zaat J. Richtlijnen en protocollen als hulpmiddel bij de verbetering van de zorg (Guidelines and protocols as tools for health care improvement) In: Grol R, Wensing M, redactie Implementatie. effectieve verandering in de patientenzorg (Implementation: effective change in patient care). Maarssen: Elsevier Gezondheidszorg, 2001.
16. Meijman F. Controleerbaarheid van informatie. meer dan een aantal voetnoten. (Check potential of data: more than a number of footnotes) Huisarts Wet 1994; 37: 173-175
17. Spies TH, Mookink HGA. Toetsen aan standaarden. Het medisch handelen van huisartsen in de praktijk getoetst: eindrapportage. (Evaluating practice guidelines. Audit of clinical performance of general practitioners in practice: final report) Nijmegen: Werkgroep Onderzoek Kwaliteit, Utrecht Nederlands Huisartsen Genootschap (NHG), 1999.
18. Tasche MJA, Oosterberg EH, Kolnaar BGM, Rosmalen CFH. Inventarisatie van lacunes in huisartsgeneeskundige kennis. Zeventig standaarden doorgelicht (Exploration of knowledge gaps in general practice. Seventy practice guidelines examined) Huisarts Wet 2001; 44: 91-94
19. Jung HP, Wensing M, Grol R. What makes a good general practitioner: do patients and doctors have different views? Br J Gen Pract 1997; 47: 805-809.
20. 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. 1900-1905.
21. Grilli R, Magrini N, Penna A, Mura G, Liberati A. Practice guidelines developed by specialty societies: the need for a critical appraisal. Lancet 2000; 355: 103-105

Chapter 8

Discussion: conclusions and recommendations

This thesis has described several research studies related to clinical guidelines. Both the process of development and the end product—the clinical guideline and the specific recommendations in the guideline—were examined. The main objective of the research was to formulate quality criteria for clinical guidelines and guideline development. This chapter summarises the major findings and provides an answer to the questions formulated in chapter 1. The main methodological issues and relations of the findings to other existing literature are discussed. Next, the implications of the results for guideline developers, for clinical practice, and for policy makers are examined. Finally, suggestions for future research and the framework in which such research can take place are presented.

Major findings and conclusions

Essential features of guideline programmes

We defined a clinical guideline programme as 'a structured and coordinated programme designed with the specific aim of producing several clinical practice guidelines'. Our hypothesis was that such a well-operated programme benefits the quality of guidelines [1]. This was confirmed in our study on characteristics of high-quality guidelines (chapter 4). Guidelines that were developed within a guideline programme scored higher in all quality domains than guidelines that were not developed within such a programme. The differences were, however, only statistically significant for the domains 'rigour of development' and 'clarity and presentation'. There also appeared to be important differences between different guideline programmes.

We used a systematic survey to examine the similarities and differences among eighteen guideline programmes (chapter 2). Principles of 'evidence-based medicine' dominated most programmes. In other words, rigorous guideline development starts with a systematic search of the literature for the best evidence [2,3]. The interpretation and application of research evidence in practice, however, is not easy and the formulation of many recommendations needs discussion and achievement of consensus among experts [4]. This was confirmed in our study: formal or informal consensus methods were used in almost all programmes.

Most guideline programmes aim at achieving consensus among all the relevant stakeholders concerned with the guidelines. The average guideline development group consists of 10 to 20 members from three to five disciplines. The programmes that target one specific professional group (for example, general

practitioners) have smaller groups with fewer disciplines. Patients are involved in the development process of ten programmes. The remaining programmes have plans to involve patients in the future.

There are important differences among the programmes in the area of implementation and evaluation of the guidelines. Some programmes assume that regional organisations are responsible for implementation, others consider it as part of the developmental process. The implementation strategies used also differ. In contrast, all programmes studied use the Internet for dissemination of their guidelines. The future plans of the guideline organisations are also similar. The main issues are improving the quality of the guidelines, including cost-effectiveness, focus on updating of the guidelines, and more international collaboration.

We conclude that there is a growing international consensus about the essential features of good guideline programmes. Hence, we may formulate some key features of good clinical guideline programmes, which are presented in Box 1.

Box 1. Key features of good clinical guideline programmes

People involved in guideline development

- credible organisation responsible for guideline development
- target users involved in guideline development ('ownership')
- balanced multidisciplinary guideline development group
- patient involvement at any stage of the development process

Methodology of guideline development

- systematic review of the literature, including existing high-quality guidelines
- combining evidence linkage and expert consensus in formulating recommendations
- external peer review
- formal update procedure
- use of quality criteria for guidelines and guideline development

Dissemination and implementation strategy

- production of different guideline formats, including patient versions, and tools for applications
- use of the Internet
- multiple implementation strategies

Criteria for good clinical guidelines

An important reason for starting our research project was the lack of internationally accepted quality criteria for clinical guidelines [5]. We participated in an international project with a group of researchers from thirteen countries—the AGREE Collaboration—resulting in the development and validation of the AGREE

Instrument (chapter 3). The AGREE instrument has been translated into various languages and is available on the Internet (www.agreecollaboration.org). The instrument, including the user guide and the Dutch translation, are presented in Appendix C. The items of the AGREE Instrument are presented in Box 2.

Box 2. Criteria for good clinical guidelines – the AGREE Instrument

Scope and purpose

- 1 The overall objective(s) of the guideline is (are) specifically described
- 2 The clinical question(s) covered by the guideline is (are) specifically described
- 3 The patients to whom the guideline is meant to apply are specifically described

Stakeholder involvement

- 4 The guideline development group includes individuals from all relevant professional groups
- 5 The patients' views and preferences have been sought
- 6 The target users of the guideline are clearly defined
- 7 The guideline has been piloted among target users

Rigour of development

- 8 Systematic methods were used to search for evidence
- 9 The criteria for selecting the evidence are clearly described
- 10 The methods used for formulating the recommendations are clearly described
- 11 The health benefits, side effects, and risks have been considered in formulating the recommendations
- 12 There is an explicit link between the recommendations and the supporting evidence
- 13 The guideline has been externally reviewed by experts prior to its publication
- 14 A procedure for updating the guideline is provided

Clarity and presentation

15. The recommendations are specific and unambiguous
- 16 The different options for management of the condition are clearly presented
- 17 The key recommendations are easily identifiable.
18. The guideline is supported with tools for application

Application

- 19 The potential organisational barriers in applying the recommendations have been discussed
20. The possible cost implications of applying the recommendations have been considered
- 21 The guideline presents key review criteria for monitoring and/or audit purposes

Editorial independence

- 22 The guideline is editorially independent from the funding body
23. Conflicts of interest of guideline development members have been recorded

In formulating the quality criteria for guidelines, we used existing appraisal instruments reflecting the principles of evidence-based medicine [6,7]. Criteria related to the applicability of guidelines were derived from other authors [8,9]. The final instrument consists of 23 items divided into six domains (scope and purpose, stakeholder involvement, rigour of development, clarity and presentation, applicability, editorial independence). A user guide adjacent to each item provides additional information. Each item is rated on a 4-point Likert scale. The scores should be analysed on the domain level using domain scores. The six domain scores are independent and should not be combined into a single quality score.

Draft versions of the instrument were tested in two validation rounds. More than 90% of the appraisers found the instrument useful for assessing guidelines. The median completion time per guideline was one hour and a half. The internal consistency of the quality domains measured by Cronbach's Alpha varied between 0.64 and 0.88 and the interrater reliability using intraclass correlations (ICC) on four appraisers varied between 0.57 and 0.91. We may conclude that the reliability is acceptable for most domains. However, if only one appraiser assesses the guideline, the reliability will be considerably lower (in our study, less than 0.50 for five of six domains). Therefore, at least two appraisers should assess each guideline to ensure sufficient reliability.

The instrument can be used by different groups for critical evaluation and comparison of existing clinical guidelines, for selection of good guidelines for local use in practice, and as a checklist in the development of guidelines. From further analysis of the validation data of the AGREE Instrument (chapter 4), it appears that the instrument is especially suited for comparison of guidelines developed by different organisations. Many items of the instrument are related to the guideline development method, which will not differ greatly for guidelines developed by the same organisation. On the domain 'rigour of development', the differences in scores between organisations were the greatest (81.3% variance), while the organisation had the least influence (28.1% variance) on the domain 'applicability'. This suggests that guideline organisations place more emphasis on the development process rather than the implementation of their guidelines.

Content analysis of clinical guidelines

Quality criteria of clinical guidelines focus mostly on the methodology of guideline development and how well this is reported. However, analysis of the clinical content is also required to judge guidelines and to explain similarities and

differences between guidelines. The study described in chapter 5 showed that the recommendations and underlying evidence in different guidelines on the management of type 2 diabetes mellitus were poorly correlated. Most of the guidelines referred to the most important trials in the field (i.e., DCCT trial and UKPDS studies), but in each guideline numerous studies were cited that were not discussed in other guidelines. Only 18% (185/1033) of all citations were shared. Important differences were also seen between different 'evidence-based' guidelines in this study (i.e., when recommendations were explicitly linked to the supporting evidence). Furthermore, there was a striking preference for citation of studies from the developers' own country.

In spite of the differences in evidence, it appeared that the concrete recommendations made in the diabetes guidelines were largely similar. Other factors, such as the influence of professional organisations (for instance, the American Diabetes Association) and international conferences, might have played an important role in the formulation of the recommendations. We conclude that expert opinions inevitably influence the process and outcome of guideline development. Formulation of recommendations solely based on—unequivocal—evidence is a utopian dream. This does not necessarily affect the quality of the guideline, provided that the methods used for formulating the recommendations, including the clinical considerations (i.e., health benefits, harms, and risks), are clearly described in the guideline.

Characteristics of effective clinical guidelines

Good guidelines are effective guidelines, i.e., guidelines that are actually used in clinical practice and have their influence on patient care. The study presented in chapter 6 suggested that recommendations with a high compliance are more compatible with the existing norms and values in practice, are more often supported with scientific evidence, require fewer new skills, and are less often part of a complex decision tree than recommendations with a low compliance. For diagnostic recommendations, the ease of applying them and the potential (negative) reactions of patients were also relevant, while this was not the case for therapeutic recommendations.

If recommendations in a guideline differ markedly from the existing practice, problems can arise in implementation [10]. If the potential facilitators and barriers of implementation are considered in formulating the recommendations, the guideline could discuss the measures that may facilitate the application of the

recommendations, for instance, which organisational changes are needed to use the guidelines appropriately [11,12]. Guideline users should be able to identify the key recommendations easily, for instance, by an attractive guideline design and by providing summaries, flowcharts, or computer aids [13]. Finally, the recommendations should be clearly and concretely formulated [8,14].

We summarise the characteristics of effective guidelines in Box 3. Five of these characteristics (1, 2, 3, 8, 11) are also covered by the AGREE Instrument (items 8, 9, 15, 17, 18, and 19).

Box 3. Characteristics of effective guidelines

- 1 the key recommendations in the guideline are easily identifiable
- 2 the key recommendations are based on scientific evidence
- 3 the key recommendations are concrete and specific
- 4 the key recommendations are easily to follow
- 5 the key recommendations are not part of a complex decision tree
- 6 the key recommendations are compatible with existing norms and values in practice
- 7 application of the key recommendations requires no new knowledge or skills
8. application of the key recommendations requires no changes in the organisation
9. application of the key recommendations requires no changes in existing routines and habits
- 10 application of the key recommendations do not provoke negative reactions in patients
- 11 the guideline is supported with tools for application

Quality of recommendations

The AGREE Instrument is an instrument for the evaluation of guidelines defined as 'documents that includes a set of statements (i.e., recommendations) about appropriate health care to support daily practice, based on evidence and its appraisal, aiming for explicitly formulating principles of good clinical care'. Evaluation of guidelines from one organisation that use the same methodology for all its guidelines is less informative, because of the limited discriminative value of most items of the AGREE Instrument. In such cases it will be more meaningful to focus on the individual recommendations in the different guidelines.

In the study described in chapter 7 we aimed to assess the key recommendations in guidelines developed by the Dutch College of General Practitioners (NHG Practice Guidelines). We developed a new instrument with 10 items (Box 4), including four items of the AGREE Instrument (reflected in item 1, 3, 4, 6, and 9) that could still discriminate between recommendations with high quality and recommendations with low quality (Appendix D). The items covered

three domains. The interrater reliability was rather low (mean kappa 0.24; range 0-0.47), but after discussion of the conflicting scores, agreement could be achieved in 99% of all judgements. The results of the assessment of the NHG Practice Guidelines suggested that the compatibility and feasibility of the recommendations were acceptable but that the scientific support could be improved.

Box 4. Quality criteria of recommendations in clinical guidelines

Scientific support

1. The recommendation is explicitly linked to the supporting evidence.
2. The recommendation is primarily supported with scientific evidence.
3. The recommendation is supported with a discussion of the benefits of applying the recommendation (health gain or quality of life).
4. The recommendation is supported with a discussion of the harms and risks of applying the recommendation.

Compatibility

5. The recommendation is compatible with existing norms and values in practice.
6. The recommendation does not evoke negative reactions in patients.

Feasibility

7. Application of the recommendation requires no new knowledge.
8. Application of the recommendation requires no new skills.
9. Application of the recommendation requires no changes in the organisation
10. Application of the recommendation requires no changes in existing routines or habits.

Methodological issues

We used a variety of research methods to examine the questions posed in this thesis. The methods were related to the analytic level of the study: (a) guideline programme, (b) clinical guideline, and (c) recommendation.

a. guideline programme survey

To examine guideline programmes, we conducted a systematic survey using a written questionnaire covering all relevant issues (Appendix B). We aimed to describe a broad range of programmes and hence we selected programmes from different countries and continents, 18 programmes in all. Initially, we considered semi-structured interviews as the research method, but this appeared to be infeasible due to the large number of programmes. The response to our survey questionnaire was 100%. We formulated additional specific questions when answers were unclear. Our summaries of the answers were checked and finally approved by the respondents. Draft tables of our study were presented at the

AGREE workshop in 2001. Half of the questionnaires were filled in by one key informant of the guideline programme. Most key informants were participants in the AGREE project and representatives of their own programmes. They might have had conflicts of interest (in particular, political interests). Therefore, we cannot exclude some socially desirable answers or the influence of political factors on the answers. On the other hand, we were not only interested in the 'real world' of guideline development but in the ideas and plans as well.

In principle, it is not possible to conclude from our cross-sectional study that there is a growing consensus on the essential features of guideline programmes. To identify changes or trends over time, prospective studies are needed. However, if we combine the findings of our study, information in the literature and from websites of guideline organisations, and the experiences in international conferences or workshops, our conclusion and our set of key features of guideline programmes can be upheld.

b. development and validation of appraisal instrument for clinical guidelines

Development and validation of an appraisal instrument for guidelines is similar to the development and validation of health measurement scales. Therefore, we used the basic concepts of the methodology described by Streiner & Norman [15]. Two validation rounds were needed to test different draft versions. In the first round difficulties arose with the appraisal of those aspects of the guideline for which no information was available in the guideline document (or technical background documents). This is also a common problem in assessing the quality of research studies such as randomised clinical trials [16]. As the quality of studies is necessarily reflected in the quality of reporting, we decided that when no information is available, the lowest score should be assigned.

Another problem in the use of the AGREE Instrument was that some items are more applicable to individual recommendations in the guideline (i.e., items 11, 12, 15, 16, 19, and 20) than to the guideline as a whole. For scoring these items, the recommendations must be easily identifiable, which was not always the case. Moreover, it is unclear which proportion of the recommendations should fulfil the criterion for a certain score. To address this problem, another substudy in the AGREE project examined the extent to which global assessment of a guideline using the AGREE Instrument corresponded with the assessment of the individual recommendations in this guideline [17]. From this study it appeared that the global assessment gave on average a 13% higher score than assessment of the

individual recommendations. The average correlation coefficient was 0.65. The appraisers indicated that the global assessment usually gave a too optimistic picture of the quality of the guideline. Consequently, for a valid evaluation of the quality of the guidelines with the AGREE Instrument, it may also be meaningful to examine the specific recommendations. Additional research work should be done to develop additional criteria.

For determining the characteristics of high-quality guidelines, we used the data of the first validation round of the AGREE Instrument. After this round, the AGREE Instrument was refined slightly by changing the formulation of some items and reordering the items and domains. It might be argued that our results would be different if we had used the final version of the AGREE Instrument.

c. analysis of recommendations in clinical guidelines

Quality assessment of clinical guidelines should also include clinical judgement [18]. Therefore, we also conducted a content analysis of different guidelines in the same clinical area to explore the similarities and differences between recommendations. We selected diabetes mellitus, for it is a common condition with evidence of variation in practice despite a substantial body of treatment trials and other studies [19,20]. To restrict the workload we limited our analysis to the outpatient care management of type 2 diabetes mellitus. Our results might have been different if we had selected another condition or clinical issue. The scope, aim, and format of the selected guidelines varied to a large extent. As a consequence, it was difficult to juxtapose the recommendations as well as the evidence, in particular of those guidelines that did not link the evidence directly to the recommendations. Therefore, all selected recommendations and evidence was cross-checked by another member of the study team. The overlap in citations was low. On the other hand, we did not take into account the quality and power of the studies. Almost all guidelines cited the most influential trials, which was not reflected in the weighted shared evidence score. However, it is not clear how to determine the weight of primary studies compared with reviews and other papers. Furthermore, some guidelines may cite contradictory evidence, some only the supporting evidence, and for widely accepted practice, evidence is often not cited. Thus, only a bibliometric analysis will not be sufficient to draw conclusions about the quality of evidence used in guidelines. Finally, it remains unclear how the clinical content of recommendations in guidelines is related to the methodology and process of guideline development. Recommendations do not emerge

spontaneously from the evidence and always require clinical judgement and consensus in the guideline development group [21]. Small group processes can influence the process of decision making and consensus development [22]. More qualitative studies are needed to examine this process.

The studies on the quality and effectiveness of recommendations used a sample of guidelines developed by the Dutch College of General Practitioners. A panel of twelve general practitioners (six pairs) assessed the recommendations using a list of 13 attributes, which were scored on a dichotomous scale. The reliability of the items was rather low, but after discussion the proportion of agreement was high (99%). We were aware that some discussion was needed to check the item interpretation of the other scoring partner. It is uncertain whether the results would be different if we had selected more than two appraisers per group of recommendations.

The characteristics of effective guidelines (i.e., recommendations) were formulated on the basis of differences between high compliance and low compliance recommendations. Our approach was similar to that of case-control studies, considering high compliance recommendations as cases and low compliance recommendations as controls. However, such studies only provide correlation figures, which cannot be used for identifying causality. Similarly, our study did not take into account the 'black box' between the introduction of guidelines and the actual behaviour of practitioners. Different guidelines may have received different attention in the media or in educational courses. Therefore, we do not know to what extent other factors may have influenced adherence to the recommendations. Prospective controlled trials are needed to minimise the bias of these factors.

Relations of findings to other literature

In recent years, concerns about the quality of clinical guidelines have been highlighted in several studies [7, 23-30]. In the absence of uniform standards for guideline quality [5], four of these studies used the Cluzeau instrument [7,28-30], and others developed their own criteria or adapted existing criteria (e.g., provisional assessment instrument of Field and Lohr, 1992 [6]). Only the Cluzeau and Shaneyfelt instruments were validated with external procedures [7,26]. Most existing criteria concern the methodology of development, for instance, the need for a systematic search for evidence and linkage of recommendations and evidence. The application and implementation of the guidelines, however, is only

modestly reflected in the existing criteria. The AGREE instrument covers these aspects properly, based on the assumption that good clinical guidelines should be also feasible and applicable in practice. Another strength of the instrument is that it was internationally developed and tested, involving many experts on guideline development in different countries. We may expect that the AGREE instrument will be the international standard for guideline quality for some time to come.

To ensure high quality, guidelines should be developed within a coordinated programme in which experience with developing guidelines has gradually been developed and sufficient support is available. This was our major finding from the further analysis of data from the validation of the AGREE Instrument. A recent French 'before-after' controlled study confirmed the positive impact of a newly established clinical practice guideline programme for cancer management on the quality of care [31]. Other studies showed lack of organisation and coordination in guideline development, nationally and internationally [32-34]. Our survey on guideline programmes is the first study that systematically examined the methods of guideline development, implementation, and evaluation in detail. We included several prominent programmes from different countries and different continents, in contrast to the Proguide survey and the AGREE survey conducted in 2000, which only included a number of European countries. Criteria for good guideline programmes were first formulated by Lohr [35], based on long experience with guideline development in the United States. Our study confirmed the relevance of most of these criteria.

It is almost self-evident that the methods used to develop guidelines influence the quality of the guidelines. A recent study of Cruse et al. confirmed that evidence-based guidelines had a higher quality than consensus-based guidelines [36]. However, it is unclear whether the guideline development methods can also explain differences in the clinical content of guidelines. Several guideline comparison studies showed substantial differences in recommendations in guidelines on the same clinical condition [37-46]. Some authors advocate more systematic approaches to guideline development, which might reduce variation between guidelines [38,43,45]. Others speculate on the role of cultural factors [38,40]. Our study on diabetes mellitus guidelines suggested that leading professional organisations, such as the American Diabetes Association, can influence the management of diabetes as reflected in guidelines worldwide. In contrast to previous studies, we also included a bibliometric analysis of the underlying evidence. This led to the surprising finding that recommendations can

be similar though the evidence differs. This was partly due to the preference of guideline developers for citing studies by authors in their own countries, which was also found by Grant et al. [47]. Furthermore, we found that little use was made of systematic reviews, which is consistent with the findings of Silagy et al. [48]. Finally, the potential influence of small group processes on the formulation of recommendations should not be underestimated [22].

High-quality guidelines are effective guidelines, i.e., guidelines that are used in practice. Therefore, it is interesting to study the characteristics of well-used guidelines. The study of Grol et al. was the first that related attributes of recommendations to compliance rates measured by audit [8]. Our study had the same design but was conducted on a larger scale (103 versus 48 recommendations). Moreover, we analysed the data differently by comparing recommendations having high compliance rates with those having low compliance rates. The findings of our study were largely consistent with the study of Grol et al. Both studies were limited by the cross-sectional design. In contrast, Foy et al. examined attributes of recommendations in a retrospective study using compliance rates before and after audit and feedback [49]. The results confirmed that recommendations compatible with existing norms and values, and not requiring changes to fixed routines, were associated with greater compliance. However, significant changes in compliance were only measured for recommendations seen as incompatible. In other words, the more compatible the recommendation the smaller the behavioural change. Qualitative studies and surveys concerning the reasons why physicians follow or do not follow guidelines can complement the knowledge about effective guidelines. Variables that affect the adoption of guidelines are related to characteristics of the health care professional, practice setting, legal or financial issues, regulation by accreditation or licensing bodies and, last but not least, patient factors [50]. Cabana et al. reviewed the literature on barriers to physician adherence to clinical guidelines and selected 76 articles, including 5 qualitative studies and 120 different surveys [11]. The barriers affected physician knowledge (lack of awareness, lack of familiarity), attitudes (e.g., lack of agreement with guidelines, lack of motivation), or behaviour (e.g., patient factors, organisational constraints). Based on this review they presented a framework for improving guideline adherence and for future research.

Implications and recommendations for guideline developers

Clinical guideline development is a real challenge. To develop good quality guidelines it is necessary to have sufficient budget and resources [3]. However, substantial savings can be obtained by cooperation between national and international guideline organisations. This can include exchange of existing guidelines and evidence reports, collaboration in literature searches for revision of guidelines, and joining together in commenting on draft guidelines [51]. Effective and efficient collaboration demands shared methodological principles. Our guideline programme study indicated that this is largely the case. The development of the AGREE Instrument, in which key figures of various guideline organisations were involved, also revealed an increased international consensus and willingness to work together. It must be kept in mind that each country has its own values that influence the content and presentation of guidelines. It is not so much a challenge to develop international guidelines as to reach agreement about the requirements for methodology and reporting of guidelines [52]. The advances in the area of guideline development are comparable to those in clinical research about five years ago. Concerns about the quality of reporting on randomised clinical trials have resulted in the CONSORT statement, which has recently been revised [53,16]. As a result, the reporting of RCTs in medical journals has improved substantially and is more uniform. In 1993, Hayward et al. initiated improvement in the reporting of guidelines [54], but this was not internationally adopted at that time.

In this thesis we have formulated quality criteria at different levels. A prerequisite for a clinical guideline to fulfil these criteria is to provide sufficient background information about the method of development. Therefore, in evaluation of guidelines the quality of reporting is as important as the clinical content of the recommendations. In the validation of the AGREE Instrument, in which 100 guidelines were assessed, as well as in our study of fifteen diabetes guidelines, it was apparent that the reporting in guidelines was extremely variable. Most guidelines limited themselves to clinical statements and gave little information about the methodology of their development. For a high-quality score, a good report is a *sine qua non*, and in order to improve the quality of guidelines, developers should take more account of the methodology of guideline development. The AGREE Instrument provides an excellent aid for that purpose. The Scottish Intercollegiate Guideline Network (SIGN) initiated this process on its website [55]. It provides a guide with examples derived from SIGN guidelines,

adjacent to each item of the AGREE Instrument, on how information can be made available. A possible drawback of a comprehensive report is that the document becomes too large and the recommendations become buried in an abundance of background information. However, if this information is organised in separate paragraphs, and the recommendations are presented in separate boxes, the readability of the guideline will not be affected and the recommendations can still be easily identified [56].

Uniform reporting gives a certain guarantee of quality. Moreover, it simplifies comparison of guidelines on the same clinical conditions. For the development (or revision) of clinical guidelines, the use of existing high-quality guidelines can save a lot of time and efforts [57,58]. The US National Guideline Clearinghouse is the first international database of clinical guidelines, with more than 1000 guidelines [59]. It only includes guidelines for which a systematic literature search and review of existing scientific evidence published in peer reviewed journals was performed during the development. In addition, the guideline should be English language and not be older than five years. The guidelines are presented using more than forty attributes. A 'guideline synthesis' is also offered regularly with a substantial analysis of the similarities and differences among guidelines on the same condition. Such documents can be very useful for guideline developers who have developed guidelines on the same subject or have them in the pipeline. For example, the literature review could be taken over, if it answers similar questions. Above all, it is inspiring to examine how other guideline development groups have collected and interpreted the evidence and how they have translated the evidence into recommendations. The NGC database is now dominated by American guidelines, but the proportion of non-American guidelines is gradually increasing. We expect that more guideline organisations throughout the world will aim for disseminating their guidelines via electronic databases.

After its dissemination the guideline must be implemented. The approaches and strategies depend on the objectives of the guidelines, the target users, and their setting. It is difficult to give general rules for this [60]. Nevertheless, a message of this thesis is that guideline developers should be aware that the content and format of a guideline can determine its use in practice. For actual improvement of health care, guideline developers should keep in mind the end user of the guideline working in practice. Hence, beyond the evidence, the applicability and feasibility should also play a role in the formulation of recommendations. If many changes are necessary in order to apply the recommendations, a separate

paragraph could be added with practical suggestions, for example, about the organisation of health care services. To facilitate the use of a guideline in practice, guideline developers should pay much attention to the format and presentation of the guideline, in particular by providing short summaries that can be easily used during the consultation [61,62]. Furthermore, tools for application should be developed, such as teaching materials, patient pamphlets, computer support, or indicators for monitoring the use of the guidelines.

A final concern is to keep guidelines up-to-date. Shekelle et al. presented a model for assessing the validity of guidelines based on a combination of multidisciplinary expert opinion and literature searches [63,64]. The use of recent systematic reviews can considerably limit the workload of literature searching [65, 48]. In principle, the update procedure should be performed every three years [66].

Implications for clinical practice

Many health care professionals feel they are flooded with clinical guidelines [67]. Whereas guidelines aim for reducing the information stream by presenting research evidence in a compact and readable document, there is a danger that an 'overdose' of guidelines will confuse the practitioners [68]. With the development of the AGREE Instrument, a serious effort was made to separate the wheat from the chaff, so that only high-quality guidelines could be selected and offered for application in practice. However, guideline selection may still need much deliberation, even when using the AGREE Instrument. Moreover, at least two appraisers are needed for a reliable assessment and the interpretation of the scores may need further instructions. Therefore, injudicious use of the AGREE Instrument by individual practitioners should be prevented.

For health care professionals clinical guidelines form one source of research evidence, but these are not the only source. Other sources, such as Clinical Evidence and the Cochrane Library, offer excellently summarised information about concrete, well-defined clinical topics [69,70]. Considering these sources, one may question whether practitioners still need guidelines. However, in contrast to evidence reports, guidelines also offer concrete recommendations when evidence is lacking or controversial, based on consensus between clinical experts and representatives of the relevant professional groups. As long as the arguments and decisions behind the recommendations are clearly described, guidelines will have a surplus value beyond evidence reports.

Why do physicians still not follow clinical guidelines? Reasons for non-compliance not only concern the quality or quantity of guidelines but also attitudinal and behavioural factors [11]. In particular, the unpredictability of daily practice, such as the varying views and preferences of patients, plays an important role in not adhering to guidelines [71,72]. The guideline assumes a standard patient in a standard setting, which is an artefact by definition. In some situations, diverging from guidelines is even desirable, considering the personal history, worries, and concerns of the individual patient [73]. Nevertheless, guidelines can help practitioners and patients in weighing the pros and cons in decision making, even when they ultimately do not follow them. If the underlying arguments can be made explicit, non-compliance with guidelines will have no legal consequences [74]. In contrast, rigid and uncritical adherence to guidelines without using clinical judgement might even be hazardous.

Implications for policy makers

Policy makers are interested in clinical guidelines because they offer the opportunity to influence and to evaluate health care processes. Clinical guidelines make the process of clinical decision making explicit and accessible to the outside world. In addition, the AGREE Instrument formulates explicit criteria for good guidelines. Policy makers can use this instrument to help them decide which guidelines could be recommended for use in practice. Then the instrument should be part of a formal assessment process. However, the AGREE scores should be interpreted cautiously. The clinical content of specific recommendations, applied in local settings, should help determine the selection of guidelines. Therefore, clinical experts and end users of the guideline should always be involved in this process.

Policy makers should also understand the need for national and international collaboration. The quality requirements for clinical guidelines increase and require more resources and funding. Collaboration allows tasks to be shared and savings to be made in the budget. Exchange of information about plans for guideline development or guideline revision should be encouraged. Internet can play an important role in this.

Future research and new developments

Evaluation of AGREE Instrument

The AGREE Instrument is being used in several European countries, Canada, the USA, Singapore, Japan, and Palestinian authority. The WHO has also adopted the instrument in order to evaluate its guidelines. It has been translated into Danish, Dutch, Finnish, French, German, Italian, and Spanish. Beyond Europe, the AGREE Instrument is used by academic institutions in the USA and in Canada to assess clinical guidelines. Evaluation of the use of the instrument is needed to examine the strengths and weaknesses of the instrument. Further refinement of the instrument and the user guide may be necessary if the interpretation of certain items proved to be ambiguous.

Clinical guideline comparison methodology

Now, many projects are in the pipeline—nationally and internationally—that aim to compare different guidelines on the same clinical condition using the AGREE Instrument. However, an assessment with the AGREE Instrument does not include a clinical judgement on the content of the recommendations and their supporting evidence. Qualitative and bibliographic methods are needed to study the clinical content of guidelines. Our comparative analysis of diabetes guidelines could be considered a pilot study on the methodology of guideline comparison. An assessment with the AGREE Instrument could be combined with the methods that we used. Furthermore, other features of guidelines, such as those used by the US National Guideline Clearinghouse, could be included in guideline comparison. The methods should be tested on other samples of guidelines covering various clinical conditions. If the methodology of guideline comparison is internationally shared, guideline reviews on a range of clinical topics could be developed. These reviews could then be used by guideline developers in different countries for ‘local’ adaptation.

Psychosocial and cultural factors in clinical guideline development

Guidelines on the same clinical condition developed by different organisations often contain different—sometimes even contradictory—recommendations. These differences may be due to different interpretations of evidence, small group processes, the influence of professional bodies, socio-economic factors, characteristics of health care systems and cultural factors such as differing

expectations of apparent risks and benefits. Qualitative studies are needed to understand the interaction of these factors with the process of guideline development.

Patient involvement

There is a growing acknowledgement of the need to involve patients in clinical guideline development. Patient representatives could participate in guideline development groups, focus group sessions could be organised to explore patient values and preferences, and literature describing patient experiences could be reviewed. However, it is unclear how bias should be minimised and which method, or combination of methods, is most successful to involve patients. Further research is needed on how guidelines can be tailored to patients' needs.

Effective use of clinical guidelines

Facilitators and barriers to the use of clinical guidelines need further study. However, the use of guidelines should not be equated with adherence to guidelines. Guidelines can also be used in decision making without following their recommendations. The ultimate goal of clinical guidelines is to help the practitioner and the patient in the process of health care, which should lead to better health outcomes and patient satisfaction. Audit studies should not only use compliance rates but also 'user rates' compared with the use of other sources of information. These user rates should reflect, for instance, to what extent the practitioner is aware of certain specific recommendations or whether the practitioner has used these in clinical decision making in concrete situations. We are aware that this approach will substantially change the focus of guideline implementation studies. Nevertheless, we think that it will contribute more to clinical practice and to seriously improving quality of care.

AGREE continued

Clinical guideline development, implementation, and evaluation is a challenging field of research that requires international collaboration. The AGREE Collaboration has been the first initiative toward collaboration in this field. In 2001 the first AGREE research project was finished. The second AGREE project, also financed by the European Union, started in September 2002 and includes further dissemination and implementation of the AGREE Instrument in 23 European

countries. This project will continue to 2004. The project has the following specific objectives:

1. To develop a standard training manual and evaluation questionnaire for disseminating and implementing the AGREE Instrument.
2. To disseminate and implement the AGREE Instrument across Europe to a targeted audience of user groups, national and local institutions, and policy makers.
3. To evaluate the impact of the dissemination in terms of acceptability, applicability, and usability of the AGREE instrument.
4. To provide a model for high-quality reporting of guideline development using electronic technologies, such as the Internet.
5. To establish an international network of excellence for the research and application of effective guideline settings in Europe.

The network will be the basis for information exchange and collaborative research. The information can concern existing clinical guidelines or guidelines under development, guidelines reviews, methodological information (for example, a guide for guideline developers), and tools for application and evaluation of guidelines. There are also plans for international cooperation between guideline organisations in disease-specific areas such as cancer. We may expect that guideline organisations throughout the world will join a future network. Although the concrete structure of the network is not clear at present, future collaboration on clinical guidelines is promising.

In conclusion, this thesis provides a framework for the evaluation of existing clinical guidelines and the development of new clinical guidelines. To ensure the highest quality, guidelines should meet rigorous criteria for quality. Obviously, it is difficult to meet all these quality criteria. However, guideline developers should not be discouraged in improving the quality of their guidelines and guideline users should not automatically disregard guidelines of sub-optimal quality. Quality improvement needs time and often moves on in small steps. Moreover, developing high-quality guidelines is not a goal in itself. We aim to improve the quality of care and, ultimately, to improve the quality of life and health of our patients. High-quality guidelines may contribute to this.

References

1. The AGREE Collaboration. Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument. www.agreecollaboration.org.
2. Grimshaw J, Eccles M, Russell I. Developing clinically valid practice guidelines. *J Eval Clin Pract* 1995; 1: 37-48.
3. Shekelle PG, Woolf SH, Eccles M, Grimshaw J. Developing guidelines. *BMJ* 1999; 318: 593-596.
4. Connis RT, Nickinovich DG, Caplan RA, Arens JF. The development of evidence-based clinical practice guidelines. Integrating medical science and practice. *Int J Technol Assess Health Care* 2000; 16: 1003-1012.
5. Graham ID, Calder LA, Hebert PC, Carter AO, Tetroe JM. A comparison of clinical practice guideline appraisal instruments. *Int J Technol Assess Health Care* 2000; 16: 1024-1038.
6. Institute of Medicine. Field MJ, Lohr KN (eds.). Guidelines for Clinical Practice: From Development to Use. Washington DC: National Academy Press, 1992.
7. Cluzeau FA, Littlejohns P, Grimshaw JM, Geder G, Moran SE. Development and application of a generic methodology to assess the quality of clinical guidelines. *Qual Health Care* 1999; 11: 21-28.
8. Grol R, Dalhuijsen J, Thomas S, in 't Veld C, Rutten G, Mookink H. Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. *BMJ* 1998; 317: 858-861.
9. Thomson R, Lavender M, Madhok R. How to ensure that guidelines are effective. *BMJ* 1995; 31: 237-242.
10. Jackson CL, De Jong IC. Clinicians and guidelines: leading a horse to water. *Med J Aust* 2001; 174: 267-268.
11. 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: 1458-1465.
12. Grol R. Improving the quality of medical care. Building bridges among professional pride, payer profit, and patient satisfaction. *JAMA* 2001; 286: 2578-2585.
13. Watkins C, Harvey I, Langley C, Gray S, Faulkner A. General practitioners' use of guidelines in the consultation and their attitudes to them. *Br J Gen Pract* 1999; 49: 11-15.
14. Shekelle PG, Kravitz RL, Beart J, Marger M, Wang M, Lee M. Are non-specific practice guidelines potentially harmful? A randomized comparison of the effect of non-specific versus specific guidelines on physician decision making. *Health Serv Res* 2000; 34: 1429-1448.
15. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. Second edition. Oxford: Oxford University Press, 1995.
16. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001; 357: 1191-1194.
17. Ten Have P, Klazinga N, Grimshaw J, Makela M, Varonen, Burnand B, et al. Appraising clinical guidelines: comparison of global assessment with assessment of individual recommendations [internal document].
18. Garfield FB. Garfield JM. Clinical judgement and clinical practice guidelines. *Int J Technol Assess Health Care* 2000; 16: 1050-1061.
19. Khuntia K, Goyder E, Baker R. Collation and comparison of multi-practice audit data: prevalence and treatment of known diabetes mellitus. *Br J Gen Pract* 1999; 49: 375-379.
20. Campbell SM, Hann M, Hacker J, Burns C, Oliver D, Thapar A, et al. Identifying predictors of high quality care in English general practice: observation study. *BMJ* 2001; 323: 1-6.
21. Feder G. Clinical guidelines for primary care: principles and problems. In: Miles A, Hampton J, Hurwitz B, eds. NICE, CHI and the NHS Reforms Enabling excellence or imposing control? London: Aesculapius Medical Press; 2000.
22. Pagliari C, Grimshaw J, Eccles M. The potential influence of small group processes on guideline development. *J Eval Clin Pract* 2001; 7: 165-173.
23. Ward JE, Grieco V. Why we need guidelines for guidelines: a study of the quality of clinical practice guidelines in Australia. *Med J Aust* 1996; 165: 574-576.
24. Varonen H, Mäkelä M. Practice guidelines in Finland: availability and quality. *Qual Health Care* 1997; 6: 75-79.
25. Helou A, Perleth M, Bitzer EM, Dörning H, Schwartz FW. Methodological quality of clinical practice guidelines in Germany: results of a systemic assessment of guidelines presented on the Internet. *Z Arztl Fortbild Qualitatssich* 1998; 92: 421-428.
26. 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: 1900-1905.
27. Grilli R, Magrini N, Penna A, Mura G, Liberati A. Practice guidelines developed by specialty societies: the need for a critical appraisal. *Lancet* 2000; 355: 103-105.
 28. Graham ID, Beardall S, Carter AO, Glennie J, Hebert PC, Tetroe JM, et al. What is the quality of drug therapy clinical practice guidelines in Canada? *CMAJ* 2001; 165: 157-163.
 29. Grimshaw GM, Khunti K, Baker R. Diagnosis of heart failure in primary care: an assessment of international guidelines. *Br J Gen Pract* 2001; 51: 384-386.
 30. Lacasse Y, Ferreira I, Brooks D, Newman T, Goldstein RS. Critical appraisal of clinical practice guidelines targeting chronic obstructive pulmonary disease. *Arch Intern Med* 2001; 161: 69-74.
 31. Ray-Coquard I, Philip T, de Laroche G, Froger X, Suchaud JP, Voloch A, et al. A controlled "before-after" study: impact of a clinical guidelines programme and regional cancer network organization on medical practice. *Br J Cancer* 2002; 86: 313-321.
 32. Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Potential benefits, limitations, and harms of clinical guidelines. *BMJ* 1999; 318: 527-530.
 33. The Appraisal of Guidelines, Research and Evaluation in Europe Collaborative Group (AGREE). Guideline development in Europe. An international comparison. *Int J Technol Assess Health Care* 2000; 16: 1039-1049.
 34. Engelbrecht R, Courté-Wienecke S, eds. A Survey on The Current State of Development, Dissemination and Implementation of Guidelines of Clinical Practice in European Countries. Neuherberg: GSF – National Research Center for Environment and Health, 1999. Available at: <http://www-mi.gsf.de/proguide/survey.htm>.
 35. Lohr KN. The quality of practice guidelines and the quality of health care. In: Selbmann HK. Guidelines in Health Care. Report of a WHO-Conference. Baden-Baden: Nomos Verlagsgesellschaft, 1998: 42-52.
 36. Cruse H, Winiarek M, Marshburn J, Clark O, Djulbegovic B. Quality and methods of developing guidelines. *BMC Health Services Research* 2002; 2: 1-7. Available from: <http://www.biomedcentral.com/1472-6963/2/1>
 37. Fahey TP, Peters TJ. What constitutes controlled hypertension? Patient based comparison of hypertension guidelines. *BMJ* 1996; 313: 93-96.
 38. Thomson R, McElroy H, Sudlow. Guidelines on anticoagulant treatment in atrial fibrillation in Great Britain. variation in content and implications for treatment. *BMJ* 1998; 316: 509-513.
 39. Psaty BM, Furberg CD. British guidelines on managing hypertension. Provide evidence, progress, and an occasional missed opportunity. *BMJ* 1999; 319: 589-590.
 40. Eisinger F, Geller G, Burke W, Holtzman NA. Cultural basis for differences between US and French clinical recommendations for women at increased risk of breast and ovarian cancer. *Lancet* 1999; 353: 919-920.
 41. Kroes RM, Numans ME, Jones RH, de Wit NJ, Verheij TJM. Gastro-oesophageal reflux disease in primary care. Comparison and evaluation of existing national guidelines and development of uniform European guidelines. *Eur J Gen Pract* 1999; 5: 88-97.
 42. Ärztliche Zentralstelle Qualitätssicherung. Leitlinien-Clearing-Bericht "Hypertonie". Kornwestheim: Brockhaus Commission Verlagsauslieferung, 2000.
 43. Beck C, Cody M, Souder E, Zhang M, Small GW. Dementia diagnostic guidelines: methodologies, results, and implementation costs. *J Am Geriatr Soc* 2000; 48: 1195-1203.
 44. Cornwall PL, Scott J. Which clinical practice guidelines for depression? An overview for busy practitioners. *Br J Gen Pract* 2000; 50: 908-911.
 45. Vogel N, Burnand B, Vial Y, Ruiz J, Paccaud F, Hohlfield P. Screening for gestational diabetes. variation in guidelines. *Eur J Obstet Gynecol Reprod Biol* 2000; 91: 29-36.
 46. Koes BW, van Tulder MW, Ostelo R, Kim Burton A, Waddell G. Clinical guidelines for the management of low back pain in primary care: an international comparison. *Spine* 2001; 26: 2504-2513.
 47. Grant J, Cottrell R, Cluzeau F, Fawcett G. Evaluating "payback" on biomedical research from papers cited in clinical guidelines: applied bibliometric study. *BMJ* 2000; 320: 1107-1111.
 48. Silagy CA, Stead LF, Lancaster T. Use of systematic reviews in clinical practice guidelines: case study of smoking cessation. *BMJ* 2001; 323: 833-836.
 49. Foy R, MacLennan G, Grimshaw J, Penney G, Campbell M, Grol R. Attributes of clinical recommendations that influence change in practice following audit and feedback. *J Clin Epidemiol* 2002; 55: 717-722.

50. Davis DA, Taylor-Vaisey A Translating guidelines into practice. A systematic review of theoretic concepts, practical experience and research evidence in the adoption of clinical practice guidelines. *CMAJ* 1997;157: 408-416.
51. Browman GP. Improving clinical practice guidelines for the 21st century. Attitudinal barriers and not technology are the main challenges. *Int J Technol Assess Health Care* 2000; 16: 959-968.
52. De Maeseneer J, Derese A European general practice guidelines: a step too far? *Eur J Gen Pract* 1999; 5: 86-104.
53. Moher D CONSORT: an evolving tool to help improve the quality of reports of randomized controlled trials Consolidated Standards of Reporting Trials *JAMA* 1998; 279: 1489-1491.
54. Hayward RSA, Wilson MC, Tunis SR, Bass EB, Rubin HR, et al. More informative abstract of article describing clinical practice guidelines. *Ann Intern Med* 1993; 118: 730-737.
55. Scottish Intercollegiate Guidelines Network. www.sign.ac.uk.
56. Scottish Intercollegiate Guidelines Network. SIGN 50: A guideline developers' handbook. SIGN publication No.50, 2001.
57. Baker R, Feder G. Clinical guidelines: where next? *Int J Qual Health Care* 1997; 9: 399-404.
58. Adams JL, Fitzmaurice DA, Heath CM, Loudon RF, Raiz A, Sterne A, et al. A novel method of guideline development for the diagnosis and management of mild to moderate hypertension. *Br J Gen Pract* 1999; 49: 175-179.
59. National Guideline Clearinghouse. www.guideline.gov/index.asp.
60. Grol R. Beliefs and evidence in changing clinical practice. *BMJ* 1997; 315: 418-421
61. Jackson R, Feder G. Guidelines for clinical guidelines; a simple, pragmatic strategy for guideline development. *BMJ* 1998; 317: 427-428
62. Hayward RSA, Guyatt GH, Moore KA, McKibbon KA, Carter AO. Canadian physicians' attitudes about and preferences regarding clinical practice guidelines. *CMAJ* 1997; 156: 1715-1723.
63. Shekelle P, Eccles MP, Grimshaw JM, Woolf SH When should clinical guidelines be updated? *BMJ* 2001; 323: 155-157
64. Browman GP. Development and aftercare of clinical guidelines; The balance between rigor and pragmatism. *JAMA* 2001; 286: 1509-1511.
65. Cook DJ, Greengold NL, Ellrodt AG, Weingarten SR. The relation between systematic reviews and practice guidelines. *Ann Intern Med* 1997; 127: 210-216
66. Shekelle PG, Ortiz E, Rhodes S, Morton SC, Eccles MP, Grimshaw JM, et al Validity of the agency for healthcare research and quality practice guidelines How quickly do guidelines become outdated? *JAMA* 2001; 286: 1461-1467
67. Hibble A, Kanka D, Pencheon D, Pooles F Guidelines in general practice the new Tower of Babel? *BMJ* 1998; 317: 862-863.
68. Grol R, Eccles M, Maisonneuve H, Woolf S. Developing clinical practice guidelines. the European experience. *Dis Manage Health Outcomes* 1998; 4: 255-266.
69. Godlee F, et al. (eds) *Clinical Evidence* 2nd ed London: BMJ, 1999
70. Cochrane Collaboration. The Cochrane Library [database on disk and CD-ROM]. Update Software, Oxford 2002. Updated quarterly.
71. Freeman AC, Sweeney K Why general practitioners do not implement evidence quality study. *BMJ* 2001; 323: 1-5
72. Schers H, Braspenning J, Drijver R, Wensing M, Grol R Low back pain in general practice: reported management and reasons for not adhering to the guidelines in the Netherlands. *Br J Gen Pract* 2000; 50: 640-645.
73. Sweeney KG, MacAuley D, Gray DP Personal significance: the third dimension. *Lancet* 1998; 351: 134-136.
74. Hurwitz B Legal and political considerations of clinical practice guidelines *BMJ* 1999; 318: 661-664.

Summary

The subject matter of this thesis is the quality of clinical practice guidelines. The goal of clinical guidelines is to improve the quality of health care and, ultimately, the quality of life. Good clinical guidelines should contribute to good patient care and thereby also to better health and quality of life. Generally-accepted quality criteria for clinical guidelines are, however, lacking. Therefore, in 1998 an international research project—the AGREE project (Appraisal Guidelines Research and Evaluation)—was started with the aim of formulating criteria for good clinical guidelines and guideline development. The project was funded by the European Union. The research aim of the AGREE project was the starting point of this thesis. Four out of six studies described in this thesis were carried out within the AGREE project. The remaining two studies concern Dutch research on the quality and effectiveness of guidelines for general practice.

Chapter 1 defines three analytic levels of clinical guideline research: (1) the level of the *guideline programme*, (2) the level of the *guideline*, and (3) the level of the specific *recommendations* as part of the guideline. These terms are defined as follows:

A *guideline programme* is a structured and coordinated programme designed with the specific aim of producing several clinical practice guidelines.

A *clinical guideline* is a document that includes a set of statements about appropriate health care to support daily practice, based on evidence and critical appraisal, aimed at the explicit statement of good medical practice.

A *recommendation* is a systematically developed statement to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances.

Further, '*quality of clinical guidelines*' is defined as 'the confidence that the potential biases of guideline development have been addressed adequately and that the recommendations are both internally and externally valid, and are feasible in practice'.

Developments in the field of clinical guidelines during the past twenty-five years are described. The first fifteen years were dominated by the development of consensus guidelines based on the guideline programme of the National Institutes of Health in the United States. In the 1980s many European countries adopted this programme to develop guidelines. In the 1990s the production of clinical guidelines increased exponentially, especially under the impetus of the evidence-based medicine movement. Both governmental agencies and numerous

professional organisations took the initiative in the systematic development of clinical guidelines. In recent years almost all western countries have become active in the field of clinical guideline development.

At present there is an overflow of guidelines and guideline activities—nationally and internationally—without clear coordination. Moreover, the quality of many guidelines is uncertain and hence there is an urgent need to achieve international consensus about the criteria for good clinical guidelines. Furthermore, guidelines on the same clinical condition sometimes contain contradictory recommendations, whereas the scientific evidence (such as the free access to Medline) is shared. Finally, the actual effect of guidelines on clinical practice is often limited.

These problems led to the following research questions:

1. What are the basic requirements for a guideline programme?
2. What are the criteria for good clinical guidelines?
3. How can differences between recommendations in clinical guidelines be explained?
4. What are the characteristics of effective clinical guidelines?
5. How can the quality of recommendations in clinical guidelines be assessed?

Chapter 2 presents the results of an international survey of the structure and working methods of clinical guideline programmes. Following a review of the literature, a questionnaire with 32 items covering relevant aspects of a guideline programme was developed (Appendix B). Eighteen prominent guideline programmes from thirteen countries were selected. The questionnaire was presented to key informants of the guideline organisations responsible for these programmes. The answers were tabulated and checked by the key informants. When answers were unclear, additional specific questions were asked. The response rate was 100%. The results suggest that clinical guideline development is generally guided by the principles of evidence-based medicine. All programmes use electronic databases to collect scientific evidence and most use systematic methods to analyse the evidence. Most programmes use both scientific evidence and consensus procedures in the formulation of the recommendations. External review of the guidelines prior to publication is used in almost all programmes. Both printed versions and the Internet are used to disseminate the guidelines. There are important differences among the programmes in implementation strategies. Some programmes leave the implementation of their guidelines to regional or local

organisations. Most programmes have a quality assurance system. However, about half of the programmes do not have a formal procedure for updating their guidelines. Their plans for the future reveal that guideline organisations aim at greater involvement of patients, increased attention to cost-effectiveness in guidelines, improving the update procedure, and more international collaboration. In conclusion, there is a growing consensus on the basic requirements of guideline programmes. Recent guideline programmes could benefit from the experiences and methods of long-standing guideline programmes. International collaboration between guideline organisations should be encouraged.

Chapter 3 describes the development and validation of the AGREE Instrument, the first internationally-tested instrument for assessing the quality of clinical guidelines. On the basis of relevant literature and existing appraisal instruments, a small working group of guideline experts generated a list of 34 items and a user guide describing the items. This version was circulated to the 30 partners of the AGREE project and to 15 international experts on guideline development. Based on their comments, the number of items was reduced. The result was the first draft instrument, comprising 24 items grouped into five domains (scope and purpose, stakeholder involvement, rigour of development, clarity and presentation, applicability). In the first validation round 100 guidelines from eleven countries were assessed by 194 appraisers. The results of this field test led to refinement of the instrument and user guide. After a second validation round using 33 guidelines and a new set of 70 appraisers, two items were removed and a sixth domain (editorial independence) and a new item were included. Thus the final instrument contains 23 items grouped into six domains with a 4-point Likert scale to score each item (Appendix C). The instrument was found useful for assessing guidelines by 95% of the appraisers. Its reliability was acceptable for most domains. Cronbach's Alpha ranged from 0.64 to 0.88 and the Intra Class Correlation (ICC) based on the means of four appraisers ranged from 0.57 to 0.91. A prerequisite for a valid appraisal is that sufficient background information is provided with the guideline, e.g., on the motivation and guideline development process. The domain scores were highly correlated with the overall assessment of the guidelines (0.67-0.88), which corroborated the criterion validity. However, the validity of the instrument needs further testing

Chapter 4 describes the results of further analyses of the validation data of the AGREE Instrument. We examined which characteristics of guidelines correlated with high scores on the domains of the AGREE Instrument. Six aspects of guideline development were considered: care level (primary/secondary care/both), scope (diagnosis/treatment/combination), type of guideline (new/update), year of publication, type of agency (governmental/professional/other), and whether the guideline was produced within a structured and coordinated program. Each variable was entered into a multilevel model in order to consider the clustering effect of the agency responsible for the guideline. Guidelines produced in a guideline programme had higher scores than those that were not, and guidelines produced by governmental agencies had higher scores than those produced by professional organisations. The influence of other characteristics on the quality scores was limited. The largest differences were found in the domain 'rigour of development', in which 81.3% of the variance was related to the guideline agency, while 18.7% was related to the guideline. Differences in the applicability of the guidelines could not be explained by the variables studied. The results of this study suggest that clinical guidelines should be developed within a well-operated programme.

Chapter 5 presents the results of a comparative analysis of 15 clinical guidelines on type 2 diabetes mellitus from 13 countries. Both the concrete recommendations and their supporting evidence were analysed with the aim of investigating the differences among guidelines and whether these could be explained by different use of scientific evidence. We selected diabetes mellitus as an example of a common condition with evidence of variation in practice. Bibliometric methods were used to measure the extent of overlap in citations used by different guidelines. The guidelines varied considerably in length (3-350 pages) and number of references (0-590). Nine out of 15 guidelines explicitly linked the recommendations and the evidence. The recommendations on the general management of patients with type 2 diabetes were largely similar. There were, however, some important differences in details. For example, the duration of a diet trial before drug treatment ranged from 2 to 9 months and target values for blood pressure control ranged from 130/80 to 160/90. Furthermore, there was remarkably little overlap in evidence cited by the guidelines. Only 18% (185/1033) of all citations used by the guidelines were shared with any other guideline and only ten citations appeared in six or more guidelines. Two percent of the citations

were systematic reviews or meta-analyses. Research originating from the same country as the guideline was significantly more likely to be cited. Two areas were selected for further analysis (use of metformin in obese patients and self-monitoring of blood glucose) to explore the similarities and differences between recommendations and their supporting evidence. Whereas the recommendations were largely concordant, the citations used differed substantially. If the primary studies in reviews were included, the overlap in citations was slightly greater. We conclude that recommendations do not automatically emerge from 'evidence' in literature. Various factors may influence the selection and interpretation of scientific evidence, such as the guideline development methods, professional and cultural values, and socio-economic factors.

Chapter 6 reports on a study that explored the characteristics of effective clinical guidelines. Performance data were derived from an audit study of 200 general practitioners in the Netherlands. Data were collected by using validated self-registration forms. The mean compliance with key recommendations selected from 29 guidelines developed by the Dutch College of General Practitioners ('NHG Practice Guidelines') was 71%. We compared the attributes of recommendations with high compliance rates (70-100%) with those of recommendations with low compliance rates (0-60%). A set of 12 attributes, including six potential facilitators and six potential barriers to guideline use, was formulated on the basis of literature and a pilot study. A panel of 12 general practitioners determined whether these attributes were present or not present in the selected recommendations. The correlation analysis indicated that high-compliance recommendations required fewer new skills than low-compliance recommendations (7% versus 22%) and were less often part of a complex decision tree (12% versus 25%). They were also more often supported by scientific evidence (47% versus 31%) and more compatible with existing norms and values in practice (87% versus 76%). The applicability of the recommendation and the potential negative reactions of patients seemed to be more important for diagnostic than for therapeutic recommendations. In conclusion, examination of facilitators and barriers to implementation can predict compliance with guidelines to some extent. Guideline developers should consider these in formulating recommendations.

Chapter 7 presents the results of an assessment of the quality of 130 recommendations from 28 guidelines developed by the Dutch College of General

Practitioners. The recommendations were assessed by a panel of 12 general practitioners using ten criteria formulated on the basis of literature (Appendix D). The criteria covered three dimensions: (1) scientific support, (2) compatibility with daily practice, and (3) feasibility. The panel found that less than half of the recommendations (44%) were supported by scientific evidence. The recommendations for therapy were more often supported by evidence (67%) than those for diagnosis (35%) or education (29%). The panel found that most recommendations (84%) were compatible with the existing norms and values in practice. Negative reactions in patients were, however, expected in 39% of the recommendations. The feasibility of the recommendations was usually sufficient, but was limited, for 43% of the recommendations required changes in routines and habits. We conclude that the quality of the scientific support of the recommendations in the Dutch guidelines for general practice can be improved. If evidence from well-designed studies is not available, a rigorous discussion of the benefits, harms, and risks of applying the recommendation may benefit the quality of the supporting evidence.

Chapter 8 provides the answers to the research questions formulated in chapter 1 and synthesises the conclusions drawn from the studies included in this thesis. The main methodological issues and relation of the findings to other published studies are also discussed.

The quality of clinical guidelines has been studied on three levels.

1. Based on the findings of the survey of clinical guideline programmes, we have formulated a number of *basic requirements for guideline programmes*. These concern the organisation and people involved in guideline development (such as target users of the guidelines and patients), the methodology of guideline development (e.g., systematic review of the literature, external peer review, formal update procedure), and the dissemination and implementation strategy (e.g., different guideline formats, including patient versions, use of the Internet, and combinations of different strategies).
2. The AGREE Instrument has been developed to appraise the quality of clinical guidelines. It contains 23 *criteria for good clinical guidelines* grouped into six domains: (1) scope and purpose, (2) stakeholder involvement, (3) rigour of development, (4) clarity and presentation, (5) applicability, and (6) editorial independence. A reliable assessment requires at least two appraisers. The instrument is especially suited for critical appraisal and comparison of clinical

guidelines developed by different organisations. It can also be used as a checklist in the development of new guidelines.

3. The *quality of recommendations* in clinical guidelines has been assessed using ten criteria, grouped into three dimensions (scientific support, compatibility, and feasibility). At least two appraisers are needed and scoring differences should be discussed. We have formulated characteristics of effective guidelines, which overlap to a large extent with the quality criteria for recommendations. This underlines the statement that good guidelines are also effective guidelines.

The content analysis of fifteen guidelines on diabetes mellitus revealed that guideline development is a complex process that is influenced by scientific as well as social factors. The selection and interpretation of scientific evidence is therefore not straightforward.

The results of our research including the AGREE Instrument, offer interesting opportunities to guideline developers, health care professionals, and policymakers. The implications for *guideline developers* are that the research has shown the importance of developing guidelines within a structured and coordinated programme. The quality of the guidelines can be improved by paying more attention to the reporting of the background and methods of guideline development, and to the format and presentation of the guideline. In this case the AGREE Instrument can be used as a practical tool for quality improvement.

The AGREE Instrument can also help *health care professionals* working in clinical practice to 'separate the wheat from the chaff' when they are confronted with too many clinical guidelines. One might be concerned that clinical guidelines are often not followed in practice. On the other hand, we suggest that guidelines do not need to be followed in all situations but can still be used in clinical decision making. Rigid and uncritical adherence to guidelines might even be more hazardous than deliberately diverging from guidelines.

Policymakers could use the AGREE Instrument in selecting clinical guidelines for use in practice. In such instances, the instrument should be part of a formal assessment procedure that includes clinical experts.

Finally, suggestions for international collaboration and future research are presented. Many countries are already using the AGREE Instrument. In a second AGREE project, also funded by the European Union, experience with the AGREE Instrument will be evaluated. Furthermore there are concrete proposals for establishing an international network, which could be the basis for information exchange and collaborative research in the field of clinical guidelines. Topics for

future research could be the methodology of guideline comparison, the role of psychosocial and cultural factors in clinical guideline development, the methods used to involve patients in guideline development, and the implementation of clinical guidelines.

In conclusion, this thesis provides a framework for the development and evaluation of clinical guidelines. Developing high-quality guidelines is, however, not a goal in itself. We should bear in mind that guidelines aim at improving the quality of care and, ultimately, improving the quality of life and health.

Samenvatting

Het onderwerp van dit proefschrift is de kwaliteit van klinische richtlijnen. Een abstract onderwerp. Het doel van richtlijnen—het verbeteren van de kwaliteit van de zorg en uiteindelijk de kwaliteit van leven—is daarentegen allerm minst abstract. Goede richtlijnen voor de klinische praktijk dienen bij te dragen tot een goede patiëntenzorg en daarmee ook tot een betere gezondheid en kwaliteit van leven. Algemeen geaccepteerde kwaliteitscriteria voor klinische richtlijnen ontbreken echter. Derhalve is in 1998 een internationaal onderzoeksproject—het AGREE project (Appraisal Guidelines Research and Evaluation)—gestart, gefinancierd door de Europese Unie, dat zich ten doel stelde criteria voor goede richtlijnen en richtlijnontwikkeling op te stellen. Dit was eveneens het uitgangspunt voor dit proefschrift. Vier van de zes studies zijn uitgevoerd binnen het AGREE project. De overige twee studies betreffen Nederlands onderzoek naar de kwaliteit en effectiviteit van richtlijnen voor de huisarts.

In **hoofdstuk 1** wordt onderscheid gemaakt tussen drie niveau's waarop richtlijnen kunnen worden bestudeerd: (1) het niveau van het *richtlijnprogramma*, (2) het niveau van de *richtlijn* en (3) het niveau van de concrete *aanbeveling*, als onderdeel van de richtlijn. De begrippen worden hierbij als volgt gedefinieerd:

Een *richtlijnprogramma* is 'een gestructureerd en gecoördineerd programma dat is opgezet met het specifieke doel om verschillende klinische richtlijnen te produceren'.

Een *klinische richtlijn* is 'een document met uitspraken over effectieve en efficiënte zorg ter ondersteuning van de dagelijkse praktijkvoering in de gezondheidszorg, berustend op de resultaten van wetenschappelijk onderzoek en een kritische beoordeling daarvan, gericht op het expliciteren van goed medisch handelen'.

Een *aanbeveling* is 'een systematisch ontwikkelde uitspraak ter ondersteuning van de zorgverlener en patiënt bij het nemen van beslissingen over effectieve en efficiënte zorg in specifieke klinische situaties'.

'*Kwaliteit van richtlijnen*' wordt vervolgens gedefinieerd als: 'het vertrouwen dat potentiële vertekening bij richtlijnontwikkeling adequaat is behandeld en dat de aanbevelingen zowel intern als extern valide zijn en haalbaar zijn in de praktijk'.

Vervolgens zijn een aantal ontwikkelingen op het gebied van richtlijnen uit de laatste vijftientig jaar beschreven. De eerste vijftien jaar zijn gedomineerd door de ontwikkeling van consensusrichtlijnen gebaseerd op het richtlijnprogramma van de National Institutes of Health uit de Verenigde Staten. Vele Europese landen hebben dit programma in de jaren tachtig overgenomen bij het maken van

richtlijnen. In de jaren negentig nam de productie van richtlijnen exponentieel toe, vooral door impulsen van de 'evidence-based medicine' beweging. Zowel overheidsinstellingen als talrijke beroepsorganisaties namen initiatieven tot het systematisch ontwikkelen van klinische richtlijnen. De laatste jaren zijn vrijwel alle westerse landen actief op het gebied van richtlijnontwikkeling.

Als we de balans opmaken stellen we vast dat er thans sprake is van een overvloed aan richtlijnen en richtlijnactiviteiten—nationaal en internationaal—zonder een duidelijke coördinatie. Ook zijn er twijfels over de kwaliteit van veel van deze richtlijnen en wordt het dringend noodzakelijk geacht internationale overeenstemming te krijgen over de criteria voor goede richtlijnen. Voorts bevatten richtlijnen over eenzelfde klinisch onderwerp nogal eens tegenstrijdige aanbevelingen, terwijl het wetenschappelijke bewijsmateriaal (onder andere door de vrije toegang tot Medline) wordt gedeeld. Tot slot is bekend dat het effect van richtlijnen in de praktijk in het algemeen beperkt is. Deze problemen hebben geleid tot de volgende onderzoeksvragen voor dit proefschrift:

1. Welke basiseisen kunnen gesteld worden aan een richtlijnprogramma?
2. Wat zijn de criteria voor goede klinische richtlijnen?
3. Hoe kunnen inhoudelijke verschillen tussen klinische richtlijnen worden verklaard?
4. Wat zijn de kenmerken van effectieve klinische richtlijnen?
5. Hoe kan de kwaliteit van aanbevelingen in klinische richtlijnen worden beoordeeld?

Hoofdstuk 2 beschrijft een onderzoek naar de structuur en methoden van richtlijnprogramma's. Hiertoe is op basis van literatuur een vragenlijst ontwikkeld met 32 items, die alle relevante aspecten van een richtlijnprogramma dekken (Appendix B). We selecteerden achttien uiteenlopende richtlijnprogramma's uit dertien landen, die in deze landen als toonaangevend bekend staan. De vragenlijst is voorgelegd aan sleutelfiguren van deze richtlijnorganisaties die verantwoordelijk zijn voor deze programma's. De antwoorden werden samengevat in tabellen en gecheckt door de sleutelfiguren. Bij onduidelijkheden werden een aantal specifieke aanvullende vragen gesteld. De respons was 100%. Uit de resultaten bleek dat de principes van 'evidence-based' geneeskunde in het algemeen de leidraad vormen bij de ontwikkeling van richtlijnen. Alle programma's gebruiken elektronische databestanden om wetenschappelijk bewijsmateriaal te verzamelen en de meeste gebruiken systematische methoden bij het analyseren

van dit materiaal. Bij het opstellen van de aanbevelingen wordt door de meeste programma's zowel gebruik gemaakt van wetenschappelijk bewijsmateriaal als van consensus-oordelen. Alle programma's laten hun richtlijnen voor publicatie door externe deskundigen beoordelen. Bij de verspreiding van de richtlijnen wordt zowel gebruik gemaakt van gedrukte publicaties als van elektronische versies en Internet. Wel zijn er forse verschillen in implementatiestrategieën. Sommige programma's laten de implementatie meer over aan regionale en lokale instanties. De meeste programma's hebben een kwaliteitsprocedure ingebouwd. Ongeveer de helft heeft echter geen formele procedure voor herziening van hun richtlijnen. De toekomstplannen wijzen op het streven naar het meer betrekken van patiënten in de richtlijnontwikkeling, meer aandacht voor de kosten-effectiviteit van richtlijnen, verbetering van de herzieningsprocedure en een toenemende behoefte aan internationale samenwerking. We concluderen dat er een toenemende consensus bestaat over de basiseisen van richtlijnprogramma's. Recente richtlijnprogramma's kunnen profiteren van de ervaringen en methoden van langer bestaande richtlijnprogramma's. Internationale samenwerking zou derhalve moeten worden aangemoedigd.

Hoofdstuk 3 beschrijft de ontwikkeling en validering van het AGREE Instrument, het eerste internationaal geteste instrument voor het beoordelen van de kwaliteit van richtlijnen. Op basis van de literatuur en bestaande beoordelingsinstrumenten werd door een werkgroep van richtlijndeskundigen een lijst van 34 items en een scoringshandleiding opgesteld. Na een commentaarronde waarin alle 30 deelnemers in het AGREE project en 15 internationale experts werden betrokken, werd de lijst gereduceerd tot 24 items, gerangschikt in vijf domeinen (onderwerp en doel, betrokkenheid van belanghebbenden, methodologie, helderheid en presentatie, toepassing). In de eerste valideringsronde werden 100 richtlijnen uit elf landen door 194 beoordelaars beoordeeld. Op grond van de resultaten werden het instrument en de handleiding aangepast. Na een tweede valideringsronde met 33 richtlijnen en 70 nieuwe beoordelaars vervielen twee items en werd er één item en één domein (onafhankelijkheid van de opstellers) toegevoegd. Zodoende bestaat het definitieve instrument uit 23 items ingedeeld onder zes domeinen (Appendix C). Bij het scoren van de items wordt gebruik gemaakt van een vierpunts Likert-schaal. Het instrument werd door 95% van de beoordelaars als nuttig ervaren bij het beoordelen van richtlijnen. De betrouwbaarheid was acceptabel voor de meeste domeinen. Cronbach's alpha varieerde van 0,64 tot

0,88 en de gemiddelde Intra Class Correlation (ICC), uitgaande van vier beoordelaars, varieerde van 0,57-0,91. Voorwaarde voor een betrouwbare beoordeling is dat er voldoende achtergrondinformatie in of bij de richtlijn wordt geleverd, bijvoorbeeld over de motivatie en methodologie. De domeinscores correleerden sterk met het algemene oordeel over de richtlijnen (0,67-0,88), hetgeen een aanwijzing is voor criterium validiteit. De validiteit dient nog verder getest te worden.

Hoofdstuk 4 beschrijft een onderzoek waarin we de resultaten van de valideringsstudie van het AGREE Instrument nader hebben geanalyseerd. We gingen na welke kenmerken van richtlijnen correleerden met hoge domeinscores met het AGREE instrument. Daarbij werden zes aspecten van richtlijnen onderzocht: 'echelon' (eerste, tweede lijn, beide), 'scope' (diagnostiek, behandeling, combinatie), 'type richtlijn' (nieuw, herziening), 'publicatiejaar', 'type richtlijnorganisatie' (overheidsinstelling, professionele organisatie, anders) en of de richtlijn al dan niet binnen een richtlijnprogramma was ontwikkeld. Bij de analyse werd gebruik gemaakt van een 'multilevel' model waarin het clustereffect van de richtlijnorganisatie werd verdisconteerd. Richtlijnen die binnen een richtlijnprogramma waren ontwikkeld hadden hogere scores dan richtlijnen die niet binnen een programma waren ontwikkeld. Ook hadden richtlijnen ontwikkeld door overheidsinstellingen hogere scores dan richtlijnen ontwikkeld door professionele organisaties. De invloed van andere kenmerken op de scores was beperkt. De grootste verschillen werden gevonden in het domein 'methodologie', waarbij 81,3% van de variantie samenhang met de richtlijnorganisatie (en 18,7% met de richtlijn). Verschillen in de toepasbaarheid van richtlijnen konden niet met de onderzochte variabelen worden verklaard. Op grond van de resultaten van deze studie komen we tot de aanbeveling dat richtlijnen bij voorkeur binnen een vastomlijnd programma zouden moeten worden ontwikkeld.

In **hoofdstuk 5** worden de resultaten gepresenteerd van een onderzoek waarin 15 richtlijnen uit 13 landen over diabetes mellitus type 2 zijn vergeleken en geanalyseerd. Zowel de concrete aanbevelingen als de achterliggende bewijsvoering zijn geanalyseerd met als doel na te gaan of verschillen tussen richtlijnen verklaard kunnen worden door een verschillend gebruik van wetenschappelijk bewijsmateriaal. We kozen voor diabetes mellitus, omdat dit een veelvoorkomende aandoening is waarbij ook bewijs bestaat voor praktijkvariatie.

We gebruikten bibliometrische methoden om de overlap in literatuurreferenties te meten. De onderzochte richtlijnen vertoonden grote verschillen in lengte (3-350 pagina's) en aantal referenties (0-590). Negen van de 15 richtlijnen hadden de referenties gekoppeld aan de aanbevelingen. De aanbevelingen over het beleid bij diabetes type 2 kwamen in grote lijnen overeen. Wel waren er enkele opvallende verschillen op een aantal deelgebieden. Zo varieerde de periode waarin een dieet wordt geprobeerd van twee tot negen maanden en varieerde het streefniveau van de bloeddruk van 130/80 tot 160/90. Daarnaast was er opvallend weinig overlap in de referenties. Slechts 18% (185/1033) van alle referenties die in de richtlijnen werden genoemd, werd gedeeld met een andere richtlijn en slechts tien studies werden geciteerd in zes of meer richtlijnen. Twee procent van alle referenties waren systematische reviews of meta-analyses. In het algemeen bestond er een opvallende voorkeur voor onderzoek uit het eigen land. Op twee deelgebieden (metforminegebruik bij obese patiënten en zelfcontrole van de bloedsuikers) is gedetailleerd gekeken naar de overeenkomsten en verschillen tussen de aanbevelingen en de onderzoeken die deze onderbouwen. Terwijl de inhoud van de aanbevelingen grotendeels overeenkwam, werden ook hier grote verschillen in referenties gevonden. Wel nam de overlap enigszins toe als we de studies die in reviews werden aangehaald meetelden. We concluderen dat er van een 'één-op-één-relatie' tussen aanbevelingen en 'evidence' beslist geen sprake is. Vermoedelijk spelen vele andere factoren een rol bij de keuze en interpretatie van het wetenschappelijk bewijsmateriaal, zoals de richtlijnontwikkelingsmethode, professionele en sociale invloeden, culturele en socio-economische factoren.

Hoofdstuk 6 beschrijft een onderzoek waarin we zijn nagegaan wat de kenmerken zijn van aanbevelingen uit richtlijnen die het beste worden opgevolgd in de praktijk. Hierbij is gebruik gemaakt van gegevens uit het Toetsen-aan-Standaarden-project waarbij de adherentie aan 29 Nederlandse richtlijnen voor de huisarts (NHG-standaarden) werd gemeten. Er deden 200 huisartsen mee aan deze studie. De gemiddelde adherentie was 71%, gemeten met behulp van zelfregistratielijsten. Wij vergeleken de kenmerken van aanbevelingen met een hoge adherentie (70-100%) met die van aanbevelingen met een lage adherentie (0-60%). Op basis van de literatuur en een pilotstudy stelden we een lijst samen van 12 kenmerken van aanbevelingen. De kenmerken bestonden uit zes potentieel bevorderende en zes potentieel belemmerende factoren bij de implementatie. Een panel van 12 huisartsen beoordeelde of de geselecteerde

aanbevelingen deze kenmerken al dan niet bezaten. Uit de correlatie-analyse bleek dat aanbevelingen met een hoge adherentie minder vaak nieuwe vaardigheden vereisten dan aanbevelingen met een lage adherentie (7% versus 22%) en minder vaak onderdeel waren van een complexe beslisboom (12% versus 25%). Daarnaast bleken ze vaker ondersteund te zijn met wetenschappelijk bewijsmateriaal (47% versus 31%) en meer aan te sluiten bij de bestaande normen en waarden in de praktijk (87% versus 76%). De toepasbaarheid van de aanbevelingen en de potentiële negatieve reacties bij patiënten bleken bij diagnostische aanbevelingen een grotere rol te spelen dan bij therapeutische aanbevelingen. Het nagaan van bevorderende en belemmerende factoren bij de implementatie van richtlijnen kan aldus de adherentie aan richtlijnen tot op zekere hoogte voorspellen. Richtlijnmakers zouden hier bij het opstellen van de aanbevelingen rekening mee kunnen houden.

In **hoofdstuk 7** wordt een onderzoek gepresenteerd waarbij de kwaliteit van 130 aanbevelingen uit 28 NHG-standaarden werd beoordeeld door een panel van 12 huisartsen met behulp van tien criteria (Appendix D). Deze criteria betreffen de wetenschappelijke onderbouwing, de compatibiliteit met de dagelijkse praktijk en de uitvoerbaarheid van de aanbevelingen en zijn op basis van de literatuur tot stand gekomen. Het panel vond dat een minderheid van de aanbevelingen (44%) werd onderbouwd met wetenschappelijk onderzoek. De therapeutische aanbevelingen waren vaker onderbouwd (67%) dan de aanbevelingen voor diagnostiek (35%) en voorlichting (29%). Het panel vond dat de meeste aanbevelingen (84%) aansloten bij de normen en waarden van de bestaande praktijk. Toch werden bij 39% van de aanbevelingen negatieve reacties van patiënten verwacht. De uitvoerbaarheid van de aanbevelingen was in het algemeen redelijk tot goed en werd vooral beperkt doordat er volgens het panel in 43% van de gevallen een verandering van bestaande routines en gewoontes vereist werd. We concluderen dat de kwaliteit van de wetenschappelijke onderbouwing van de aanbevelingen in NHG-Standaarden verbeterd kan worden. Bij gebrek aan wetenschappelijke onderzoek, kan een zorgvuldige bespreking van de voor- en nadelen van de interventie die wordt aanbevolen (of afgeraden) de kwaliteit van de onderbouwing ten goede komen.

In **hoofdstuk 8** worden de conclusies uit de verschillende studies in samenhang besproken en geven we antwoord op de eerder gestelde onderzoeksvragen.

Daarnaast worden de belangrijkste methodologische kwesties besproken alsmede de relatie van de bevindingen met in de literatuur beschreven onderzoeken.

Op drie niveau's hebben we de kwaliteit van richtlijnen bestudeerd.

1. Op grond van de bevindingen uit het onderzoek naar de inhoud van richtlijnprogramma's formuleren we een aantal *basiseisen voor een richtlijnprogramma*. Deze hebben betrekking op de organisatie en de mensen die bij de richtlijnontwikkeling betrokken zouden moeten zijn (zoals de beoogde gebruikers van de richtlijnen en patiënten), de methodologie van richtlijnontwikkeling (o.a. systematisch literatuuronderzoek, externe commentaarrronde, formele herzieningsprocedure) en de disseminatie en implementatiestrategie (o.a. verschillende versies van richtlijnen waaronder ook een versie voor de patient, gebruik van Internet en het combineren van meerdere strategieën).
2. Met het AGREE Instrument hebben we een beoordelingsinstrument ontwikkeld voor de beoordeling van de kwaliteit van klinische richtlijnen. Zij bevat 23 *criteria voor goede klinische richtlijnen* onderverdeeld in zes domeinen: (1) onderwerp en doel, (2) betrokkenheid van belanghebbenden, (3) methodologie, (4) helderheid en presentatie, (5) toepassing, en (6) onafhankelijkheid van de opstellers. Voor een betrouwbare beoordeling zijn ten minste twee beoordelaars vereist. Het instrument is vooral geschikt voor het kritisch beoordelen en vergelijken van de kwaliteit van bestaande richtlijnen ontwikkeld door verschillende richtlijnorganisaties maar kan ook gebruikt worden als checklist bij het ontwikkelen van nieuwe richtlijnen.
3. De *kwaliteit van aanbevelingen* in richtlijnen hebben we beoordeeld met tien criteria, verdeeld over drie dimensies (wetenschappelijke onderbouwing, compatibiliteit en uitvoerbaarheid). Ook hier geldt dat er ten minste twee beoordelaars nodig zijn, waarbij na een onafhankelijke beoordelingsronde bij voorkeur ook overleg plaatsvindt over de scoringsverschillen. Daarnaast hebben we kenmerken van effectieve aanbevelingen geformuleerd, die grotendeels overlappen met de kwaliteitscriteria van aanbevelingen. Dit onderstreept de stelling dat goede richtlijnen ook effectieve richtlijnen zijn.

De inhoudelijke analyse van vijftien richtlijnen over diabetes mellitus wijst uit dat richtlijnontwikkeling een complex gebeuren is waarin niet alleen wetenschaps-technische maar ook sociale factoren invloed hebben op de klinische inhoud van de richtlijn. De selectie en interpretatie van het wetenschappelijk bewijsmateriaal is allerminst eenduidig.

De resultaten van ons onderzoek, en met name het AGREE Instrument, bieden interessante mogelijkheden voor zowel richtlijnmakers en zorgverleners in de praktijk als voor beleidsmakers.

De consequentie van ons onderzoek voor *richtlijnmakers* is dat zij het belang aantoont van het ontwikkelen van richtlijnen binnen een vastomlijnd gecoördineerd programma. De kwaliteit van richtlijnen kan worden verbeterd door veel aandacht te besteden aan de verslaglegging van de achtergronden van de richtlijnontwikkeling, zoals de doelen en de methodologie, en aan de presentatie en de opmaak. Hierbij kan het AGREE Instrument een handig hulpmiddel zijn.

Ook voor *professionele zorgverleners* die werkzaam zijn in de praktijk kan het AGREE Instrument van pas komen. Dit instrument helpt in de berg van richtlijnen 'het kaf van het koren te scheiden'. Een punt van zorg zou kunnen zijn dat richtlijnen in de praktijk om diverse redenen vaak niet worden opgevolgd. Daarentegen wijzen we erop dat richtlijnen niet altijd hoeven te worden opgevolgd maar toch een handig hulpmiddel kunnen zijn bij het nemen van beslissingen in de praktijk. Een rigide toepassing van richtlijnen brengt wellicht meer gevaren met zich mee dan het gemotiveerd afwijken van richtlijnen.

Tot slot kunnen ook *beleidsmakers* het AGREE Instrument gebruiken bij het selecteren van richtlijnen voor gebruik in de praktijk. Er moet dan wel sprake zijn van een formele procedure waar ook klinische experts bij betrokken worden.

Dit proefschrift besluit met een aantal suggesties voor onderzoek en samenwerking in de toekomst. Talrijke landen gebruiken reeds het AGREE Instrument. In een tweede AGREE project, wederom gefinancierd door de Europese Unie, zullen de ervaringen met het AGREE Instrument worden geëvalueerd. Tevens zijn er concrete voorstellen voor het opzetten van een internationaal netwerk dat de basis zou kunnen vormen voor het uitwisselen van informatie en het verrichten van wetenschappelijk onderzoek op het gebied van richtlijnen. Onderwerpen voor toekomstig onderzoek zijn onder andere: het formuleren van een methodologie voor richtlijnvergelijking, de rol van psychosociale en culturele factoren bij richtlijnontwikkeling, de methoden waarop patiënten bij richtlijnontwikkeling worden betrokken en de implementatie van richtlijnen.

De conclusie is dat dit proefschrift een kader biedt voor de evaluatie van bestaande klinische richtlijnen en de ontwikkeling van nieuwe richtlijnen. Het nastreven van een hoge kwaliteit van richtlijnen is echter geen doel op zich. Het

gaat vooral om een verbetering van de kwaliteit van zorg en uiteindelijk om een verbetering van gezondheid en kwaliteit van leven.

Appendices

Appendix A. The AGREE Collaboration

The AGREE (Appraisal Guidelines Research and Evaluation) Collaboration consisted of a group of researchers from thirteen countries. From 1998 to 2001 they carried out a research project with the aim to provide a framework to create a coordinated international approach to the appraisal of clinical guidelines. The project was funded by the European Union BIOMED2 Programme. The table below presents the individuals and organisations that participated in the AGREE Collaboration.

Countries	Individuals	Organisations
<i>European countries</i>		
Denmark	Finn Kristensen, MD, PhD Pia Bruun Madsen Camilla Palmhøj-Nielsen	Danish Institute for Health Technology Assessment, Copenhagen
England	Françoise Cluzeau, MSc, PhD Gene Feder, MD, FRCGP Claire Hunt, MSc Peter Littlejohns, MBBS, MD	St George's Hospital Medical School, London Barts and The London Queen Mary's School of Medicine and Dentistry, University of London Institute of Psychiatry, London National Institute for Clinical Excellence (NICE), London
Finland	Marjukka Makela, MD, PhD, MSc	Finnish Office for Health Care Technology Assessment, Helsinki
France	Anne Bataillard, MD Béatrice Fervers, MD Isabelle Durand-Zaleski, PhD Pierre Durieux, MD	Fédération Nationale des Centres de Lutte Contre le Cancer Hôpital Henri Mondor, Cedex Hôpital Européen Georges Pompidou, Paris
Germany	Gunter Ollenschlager, MD, PhD	Agency for Quality in Medicine, Cologne
Italy	Roberto Grilli, MD	Agenzia Sanitaria Regionale, Bologna
Netherlands	Jako Burgers, MD Richard Grol, PhD Joost Zaat, MD, PhD Pieter ten Have, MD Kitty Rosenbrand, MD Niek Klazinga, MD, PhD	Centre for Quality of Care Research (WOK), University Medical Centre Nijmegen Dutch Institute for Healthcare Improvement CBO, Utrecht Academic Medical Centre, University of Amsterdam
Spain	José Asua, MD, PhD Rosa Rico-Iturriz, MD, MSc Albert Jovell, MD, PhD	Basque Office for Health Technology Assessment Fundacio Biblioteca Josep Laporte, Barcelona
Scotland	Juliet Miller, MA, MBA Safia Qureshi, PhD	Scottish Intercollegiate Guidelines Network (SIGN), Edinburgh
Switzerland	Bernard Burnand, MD, MPH John-Paul Vader, MD, MPH	Institut Universitaire de Médecine Sociale et Préventive, Lausanne

Countries	Individuals	Organisations
<i>Other countries</i>		
Canada	Melissa Brouwers, PhD	McMaster University and Cancer Care Ontario,
	Steven Hanna, PhD	Hamilton
	George Browman, MD	Hamilton Regional Cancer Centre
	Jeremy Grimshaw, MB, PhD	Ottawa Health Services Research Institute
New Zealand	Cindy Farquhar, MD, PhD	New Zealand Guidelines Group, Auckland
	Rod Jackson, PhD	Effective Practice Institute, University of Auckland
United States	Jean Slutsky	Agency for Healthcare Research and Quality, Rockville

Appendix B

Questionnaire for description of clinical guideline programmes

Basic characteristics

- 1 Name
- 2 Country
- 3 Website
- 4 Type of organisation
 - ☐ Academic Institution
 - ☐ Medical Specialty Society
 - ☐ Disease Specific Society
 - ☐ International Agency
 - ☐ Managed Care Organisation
 - ☐ Manufacturer
 - ☐ National Government Agency
 - ☐ Private Organisation
 - ☐ Professional Association
 - ☐ Regional/Local Government Agency
 - ☐ Other, please specify
- 5 Historical details
 - a Year of first guideline
 - b Reason for guideline development
- 6 Funding (more than one answer possible)
 - ☐ Own budget
 - ☐ Governmental support
 - ☐ Pharmaceutical sponsoring
 - ☐ Other, please specify
- 7 Estimated budget for guideline development (average budget in EUROS/US dollars per guideline)
 - ☐ 0 - 5,000
 - ☐ 5,000 - 10,000
 - ☐ 10,000 - 25,000
 - ☐ 25,000 - 50,000
 - ☐ 50,000 - 100,000
 - ☐ 100,000 - 200,000
 - ☐ > 200,000
- 8 Estimated budget for dissemination (average budget in EUROS/US dollars per guideline)
 - ☐ 0 - 5,000
 - ☐ 5,000 - 10,000
 - ☐ 10,000 - 25,000
 - ☐ 25,000 - 50,000
 - ☐ 50,000 - 100,000
 - ☐ 100,000 - 200,000
 - ☐ > 200,000

Purpose and topics

9. Objectives (more than one answer possible)
- ☐ Appropriate clinical care
 - ☐ Cost containment
 - ☐ Both
 - ☐ Other, please specify
10. Level of care (more than one answer possible)
- ☐ Public Health
 - ☐ Primary Care
 - ☐ Secondary Care
 - ☐ Tertiary Care
11. Target users (more than one answer possible)
- ☐ Physicians
 - ☐ Paramedical professions
 - ☐ Nurses
 - ☐ Patients
 - ☐ Health Care Organisations/Hospitals
 - ☐ Policymakers
12. Scope of guidelines (more than one answer possible)
- ☐ Screening
 - ☐ Prevention
 - ☐ Diagnosis
 - ☐ Treatment/management
13. Who selects topics?

People involved in guideline development group

14. Average number of members in a guideline development group
- ☐ 0 - 5
 - ☐ 5 - 10
 - ☐ 10 - 15
 - ☐ 15 - 20
 - ☐ > 20
15. Average number of disciplines in a guideline development group
- ☐ 0 - 3 disciplines
 - ☐ 3 - 5 disciplines
 - ☐ > 5 disciplines
16. Experts involved in guideline development (more than one answer possible)
- | | <i>always involved</i> | <i>only if necessary</i> |
|---|--------------------------|--------------------------|
| a. Informatics, library sciences | <input type="checkbox"/> | <input type="checkbox"/> |
| b. Clinical epidemiology | <input type="checkbox"/> | <input type="checkbox"/> |
| c. Statistics | <input type="checkbox"/> | <input type="checkbox"/> |
| d. Communication | <input type="checkbox"/> | <input type="checkbox"/> |
| e. Health economics | <input type="checkbox"/> | <input type="checkbox"/> |
| f. Social sciences (psychologist, sociologist, etc.), | <input type="checkbox"/> | <input type="checkbox"/> |
| g. Other, please specify | <input type="checkbox"/> | <input type="checkbox"/> |

17. Involvement of patients (more than one answer possible)
- ☐ Yes, by participation in development group
 - ☐ Yes, by surveys of patient views/preferences
 - ☐ Yes, by review by representatives for patient organisations
 - ☐ No
18. Who is responsible for editing the guideline? (more than one answer possible)
- ☐ All members of guideline development group
 - ☐ Chairman and/or secretary of the guideline development group
 - ☐ Standing editorial staff
 - ☐ Editorial committee that varies for different guidelines
 - ☐ Other, please specify ...

Methodology of guideline development

19. Is there methodological training for members of the guideline development group before starting with the guideline development?
- ☐ Yes, obligatory
 - ☐ Yes, optional
 - ☐ No
20. Method used to collect evidence (more than one answer possible)
- ☐ Hand searches of published literature (primary and/or secondary sources)
 - ☐ Searches of electronic databases
 - ☐ Searches of patient registry data
 - ☐ Searches of unpublished data
21. Methods used to analyse evidence (more than one answer possible)
- ☐ Decision analysis
 - ☐ Meta-analysis
 - ☐ Systematic review
 - ☐ Non-systematic review
 - ☐ Experience-based
22. Methods used to formulate recommendations (more than one answer possible)
- ☐ Subjective review
 - ☐ Informal expert consensus
 - ☐ Formal expert consensus (consensus conferences, nominal group technique or Delphi technique)
 - ☐ Evidence-linked (weighting according to a rating scheme)
23. Method of review (more than one answer possible)
- ☐ Clinical validation - pilot testing
 - ☐ Clinical validation - trial implementation period
 - ☐ Comparison with guidelines from other groups
 - ☐ External peer review
 - ☐ Internal peer review
24. Is there a process of guideline authorisation?
- ☐ Yes, formal authorisation by endorsement by professional organisation of the target users
 - ☐ Yes, authorisation otherwise, please specify
 - ☐ No

Products and deliveries

25. Total number of guidelines produced
- ☐ 0 - 10
 - ☐ 10 - 20
 - ☐ 20 - 30
 - ☐ 30 - 50
 - ☐ > 50
26. Average size of guideline
- ☐ 0 - 2 pages
 - ☐ 2 - 5 pages
 - ☐ 5 - 10 pages
 - ☐ 10 - 15 pages
 - ☐ 15 - 25 pages
 - ☐ 25 - 50 pages
 - ☐ > 50 pages
27. Different versions (more than one answer possible)
- ☐ Extensive version with notes/references
 - ☐ Short version
 - ☐ One or two pages summary
 - ☐ Patient version
28. Tools for application (more than one answer possible)
- ☐ No tools
 - ☐ Algorithms/flow charts
 - ☐ Balance sheets
 - ☐ Risk tables
 - ☐ Patient leaflets
29. Media used (more than one answer possible)
- ☐ Paper
 - ☐ CD-ROM
 - ☐ Internet

Implementation strategies

30. Health professional orientated interventions (more than one answer possible)
- ☐ Educational materials
 - ☐ Conferences
 - ☐ Local opinion leaders
 - ☐ Outreach visits
 - ☐ Patient mediated interventions
 - ☐ Audit and feedback
 - ☐ (Computer) reminder
31. Use of financial incentives
- ☐ Yes, please specify ..
 - ☐ No

32. Organisational interventions (more than one answer possible)
- ☐ Changes in settings/site of service delivery
 - ☐ Changes in physical structure, facilities and equipment
 - ☐ Changes in medical records systems
 - ☐ Changes in scope and nature of benefits and services
 - ☐ Presence and organisation of quality-monitoring mechanisms
 - ☐ Ownership, accreditation, and affiliation status
 - ☐ Staff organisation
 - ☐ Other, please specify.....

Evaluation and update procedure

33. Use of monitoring and documentation (i.e., systematic data collection)
- ☐ Yes
 - ☐ No
34. Is there any regular quality system for your guideline programme? (more than one answer possible)
- ☐ Yes, by developing and publishing criteria for good guideline development ('guidelines for guidelines')
 - ☐ Yes, by revising guidelines based on comments from the professional community
 - ☐ Yes, by appraising existing guidelines
 - ☐ Yes, we submit the guidelines to a guideline clearinghouse
 - ☐ Yes, otherwise, please specify ...
 - ☐ No
35. Procedure for updating guidelines (more than one answer possible)
- ☐ Updated on regular basis
 - ☐ Updated irregularly
 - ☐ Formal method, please specify
 - ☐ No formal method
 - ☐ Not updated
36. What are the plans for further development of your guideline programme for the near future?

Appendix C

1. AGREE Instrument - original English version

2. AGREE Instrument - translated Dutch version

The AGREE Collaboration

Writing group

Françoise Cluzeau^{1,2,3}, PhD, St George's Hospital Medical School, London, UK

Jako Burgers^{1,2,3}, MD, University Medical Centre St Radboud, Nijmegen, The Netherlands

Melissa Brouwers⁴, PhD, McMaster University and Cancer Care Ontario, Hamilton, Ontario, Canada

Richard Groi^{1,4}, PhD, University Medical Centre St Radboud, Nijmegen, The Netherlands

Marjukka Makela⁴, MD, PhD, MSc Finnish Office for Health Care Technology

Assessment, Finland

Peter Littlejohns^{1,4}, MBBS, MD, National Institute for Clinical Excellence, London, UK

Jeremy Grimshaw⁴, MB, PhD, University of Ottawa, Canada

Claire Hunt², MSc, Institute of Psychiatry, London, UK

¹ developed the first draft of the instrument and designed the field study

² analysed the data

³ drafted the paper presented in chapter three

⁴ helped write the final draft of the paper

Contributors

The following individuals provided input in the design and field testing of the AGREE Instrument

José Asua, MD, PhD, Basque Office for Health Technology Assessment, Spain

Anne Bataillard, MD, Federation Nationale des Centres de Lutte Contre le Cancer, Paris, France

George Browman, MD, Hamilton Regional Cancer Centre, Hamilton, Canada

Bernard Burnand, MD, MPH, Institut Universitaire de Médecine Sociale et Preventive, Lausanne, Switzerland

Pierre Dureux, MD, Hôpital Européen Georges Pompidou, Paris, France

Beatrice Fervers, MD, Federation Nationale des Centres de Lutte Contre le Cancer, Paris, France

Roberto Grilli, MD, Agenzia Sanitaria Regionale, Bologna, Italy

Steven Hanna, PhD, McMaster University, Hamilton, Ontario, Canada

Pieter ten Have, MD, CBO, Utrecht, The Netherlands

Albert Jovell, MD, PhD, Fundacio Biblioteca Josep Laporte, Barcelona, Spain

Niek Klazinga, MD, PhD, Academisch Medisch Centrum, University of Amsterdam, The Netherlands

Finn Kristensen, MD, PhD, Danish Institute for Health Technology Assessment, Copenhagen, Denmark

Pia Bruun Madsen, Danish Institute for Health Technology Assessment, DITHA, Copenhagen, Denmark

Juliet Miller, MA, MBA, Edinburgh, UK

Gunter Ollenschlager, MD, PhD, Agency for Quality in Medicine, Cologne, Germany

Safia Qureshi, PhD, Edinburgh, UK

Rosa Rico-Iturroz, MD, MSc, Basque Office for Health Technology Assessment, Spain

John-Paul Vader, MD, MPH, Institut Universitaire de Médecine Sociale et Preventive, Lausanne, Switzerland

Joost Zaat, MD PhD, University Medical Centre St Radboud, Nijmegen, The Netherlands

APPRAISAL OF GUIDELINES FOR RESEARCH & EVALUATION



AGREE

INSTRUMENT

The AGREE Collaboration

September 2001

INTRODUCTION

Purpose of the AGREE Instrument

The purpose of the Appraisal of Guidelines Research & Evaluation (AGREE) Instrument is to provide a framework for assessing the quality of clinical practice guidelines

Clinical practice guidelines are 'systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances'¹ Their purpose is 'to make explicit recommendations with a definite intent to influence what clinicians do'²

By quality of clinical practice guidelines we mean the confidence that the potential biases of guideline development have been addressed adequately and that the recommendations are both internally and externally valid, and are feasible for practice This process involves taking into account the benefits, harms and costs of the recommendations, as well as the practical issues attached to them Therefore the assessment includes judgements about the methods used for developing the guidelines, the content of the final recommendations, and the factors linked to their uptake

The AGREE Instrument assesses **both** the quality of the reporting, **and** the quality of some aspects of recommendations It provides an assessment of the predicted validity of a guideline, that is the likelihood that it will achieve its intended outcome It does not assess the impact of a guideline on patients' outcomes

Most of the criteria contained in the AGREE Instrument are based on theoretical assumptions rather than on empirical evidence They have been developed through discussions between researchers from several countries who have extensive experience and knowledge of clinical guidelines Thus the AGREE Instrument should be perceived as reflecting the *current* state of knowledge in the field

Which guidelines can be appraised with the AGREE Instrument?

The AGREE Instrument is designed to assess guidelines developed by local, regional, national or international groups or affiliated governmental organisations These include

- 1 New guidelines
- 2 Existing guidelines
- 3 Updates of existing guidelines

The AGREE Instrument is generic and can be applied to guidelines in any disease area including those for diagnosis, health promotion, treatment or interventions It is suitable for guidelines presented in paper or electronic format

¹ Lohr KN, Field MJ A provisional instrument for assessing clinical practice guidelines In Institute of Medicine Field MJ, Lohr KN (eds) *Guidelines for clinical practice From development to use* Washington D C National Academy Press, 1992

² Hayward RSA, Wilson MC, Tunis SR, Bass EB, Guyatt G, for the Evidence-Based Medicine Working Group Users' guides to the Medical Literature VIII How to Use Clinical Practice Guidelines A Are the Recommendations Valid? *JAMA* 1995, **274** 570-574

Who can use the AGREE Instrument?

The AGREE Instrument is intended to be used by the following groups:

- i) By *policy makers* to help them decide which guidelines could be recommended for use in practice. In such instances, the instrument should be part of a formal assessment process.
- ii) By *guideline developers* to follow a structured and rigorous development methodology and as a self-assessment tool to ensure that their guidelines are sound.
- iii) By *health care providers* who wish to undertake their own assessment before adopting the recommendations
- iv) By *educators* or *teachers* to help enhance critical appraisal skills amongst health professionals.

Key references

The following sources have been used for developing the AGREE Instrument criteria.

Lohr KN, Field MJ. A provisional instrument for assessing clinical practice guidelines. In: Institute of Medicine. Field MJ, Lohr KN (eds). Guidelines for clinical practice. From development to use. Washington D.C.: National Academy Press, 1992.

Cluzeau F, Littlejohns P, Grimshaw J, Feder G, Moran S. Development and application of a generic methodology to assess the quality of clinical guidelines. International Journal for Quality in Health Care 1999; 11: 21-28.

Grol R, Dalhuijsen J, Thomas S, in 't Veld C, Rutten G, Mookink H. Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. BMJ 1998; 317: 858-861.

Lohr KN. The quality of practice guidelines and the quality of health care. In: Guidelines in health care. Report of a WHO Conference. January 1997, Baden-Baden: Nomos Verlagsgesellschaft, 1998.

INSTRUCTIONS FOR USE

Please read the following instructions carefully before using the AGREE Instrument

1. Structure and content of the AGREE Instrument

AGREE consists of 23 key items organised in six domains. Each domain is intended to capture a separate dimension of guideline quality.

Scope and purpose (items 1-3) is concerned with the overall aim of the guideline, the specific clinical questions and the target patient population.

Stakeholder involvement (items 4-7) focuses on the extent to which the guideline represents the views of its intended users.

Rigour of development (items 8-14) relates to the process used to gather and synthesise the evidence, the methods to formulate the recommendations and to update them.

Clarity and presentation (items 15-18) deals with the language and format of the guideline.

Applicability (items 19-21) pertains to the likely organisational, behavioural and costs implications of applying the guideline.

Editorial independence (items 22-23) is concerned with the independence of the recommendations and acknowledgement of possible conflict of interest from the guideline development group.

2. Documentation

Appraisers should attempt to identify all information about the guideline development process prior to appraisal. This information may be contained in the same document as the recommendations or it may be summarised in a separate technical report, in published papers or in policy reports (e.g. guideline programmes). We recommend that you read the guideline and its accompanying documentation fully before you start the appraisal.

3. Number of appraisers

We recommend that each guideline is assessed by at least two appraisers and preferably four as this will increase the reliability of the assessment.

4. Response scale

Each item is rated on a 4-point scale ranging from 4 'Strongly Agree' to 1 'Strongly Disagree', with two mid points: 3 'Agree' and 2 'Disagree'. The scale measures the extent to which a criterion (item) has been fulfilled.

- If you are confident that the criterion has been fully met then you should answer '*Strongly Agree*'.
- If you are confident that the criterion has not been fulfilled at all or if there is no information available then you should answer '*Strongly Disagree*'.
- If you are unsure that a criterion has been fulfilled, for example because the information is unclear or because only some of the recommendations fulfil the criterion, then you should answer '*Agree*' or '*Disagree*', depending on the extent to which you think the issue has been addressed.

5. User Guide

We have provided additional information in the User Guide adjacent to each item. This information is intended to help you understand the issues and concepts addressed by the item. Please read this guidance carefully before giving your response.

6. Comments

There is a box for comments next to each item. You should use this box to explain the reasons for your responses. For example, you may 'Strongly Disagree' because the information is not available, the item is not applicable, or the methodology described in the information provided is unsatisfactory. Space for further comments is provided at the end of the instrument.

7. Calculating domain scores

Domain scores can be calculated by summing up all the scores of the individual items in a domain and by standardising the total as a percentage of the maximum possible score for that domain.

Example

If four appraisers give the following scores for Domain 1 (Scope and purpose)

	Item 1	Item 2	Item 3	Total
Appraiser 1	2	3	3	8
Appraiser 2	3	3	4	10
Appraiser 3	2	4	3	9
Appraiser 4	2	3	4	9
Total	9	13	14	36

Maximum possible score = 4 (strongly agree) x 3 (items) x 4 (appraisers) = 48

Minimum possible score = 1 (strongly disagree) x 3 (items) x 4 (appraisers) = 12

The standardised domain score will be

$$\frac{\text{obtained score} - \text{minimum possible score}}{\text{maximum possible score} - \text{minimum possible score}} =$$

$$\frac{36 - 12}{48 - 12} = \frac{24}{36} = 0.67 \times 100 = 67\%$$

Note:

The six domain scores are independent and should not be aggregated into a single quality score. Although the domain scores may be useful for comparing guidelines and will inform the decision as to whether or not to use or to recommend a guideline, it is not possible to set thresholds for the domain scores to mark a 'good' or 'bad' guideline.

8. Overall assessment

A section for overall assessment is included at the end of the instrument. This contains a series of options: 'Strongly recommend', 'Recommend (with provisos or alterations)', 'Would not recommend' and 'Unsure'. The overall assessment requires the appraiser to make a judgement as to the quality of the guideline, taking each of the appraisal criteria into account.

SCOPE AND PURPOSE

- 1. The overall objective(s) of the guideline is (are) specifically described.**

Strongly Agree

4	3	2	1
---	---	---	---

 Strongly Disagree

Comments

- 2. The clinical question(s) covered by the guideline is(are) specifically described.**

Strongly Agree

4	3	2	1
---	---	---	---

 Strongly Disagree

Comments

- 3. The patients to whom the guideline is meant to apply are specifically described.**

Strongly Agree

4	3	2	1
---	---	---	---

 Strongly Disagree

Comments

USER GUIDE

SCOPE AND PURPOSE

1. This deals with the potential health impact of a guideline on society and populations of patients. The overall objective(s) of the guideline should be described in detail and the expected health benefits from the guideline should be specific to the clinical problem. For example specific statements would be:
 - Preventing (long term) complications of patients with diabetes mellitus,
 - Lowering the risk of subsequent vascular events in patients with previous myocardial infarction,
 - Rational prescribing of antidepressants in a cost-effective way.

2. A detailed description of the clinical questions covered by the guideline should be provided, particularly for the key recommendations (see item 15). Following the examples provided in question 1:
 - How many times a year should the HbA1c be measured in patients with diabetes mellitus?
 - What should the daily aspirin dosage for patients with proven acute myocardial infarction be?
 - Are selective serotonin reuptake inhibitors (SSRIs) more cost-effective than tricyclic antidepressants (TCAs) in treatment of patients with depression?

3. There should be a clear description of the target population to be covered by a guideline. The age range, sex, clinical description, comorbidity may be provided. For example:
 - A guideline on the management of diabetes mellitus only includes patients with non-insulin dependent diabetes mellitus and excludes patients with cardiovascular comorbidity
 - A guideline on the management of depression only includes patients with major depression, according to the DSM-IV criteria, and excludes patients with psychotic symptoms and children
 - A guideline on screening of breast cancer only includes women, aged between 50 and 70 years, with no history of cancer and with no family history of breast cancer

STAKEHOLDER INVOLVEMENT

4. The guideline development group includes individuals from all the relevant professional groups.

Strongly Agree

4	3	2	1
---	---	---	---

 Strongly Disagree

Comments

5. The patients' views and preferences have been sought.

Strongly Agree

4	3	2	1
---	---	---	---

 Strongly Disagree

Comments

6. The target users of the guideline are clearly defined.

Strongly Agree

4	3	2	1
---	---	---	---

 Strongly Disagree

Comments

7. The guideline has been piloted among target users.

Strongly Agree

4	3	2	1
---	---	---	---

 Strongly Disagree

Comments

USER GUIDE

STAKEHOLDER INVOLVEMENT

4. This item refers to the professionals who were involved at some stage of the development process. This may include members of the steering group, the research team involved in selecting and reviewing / rating the evidence and individuals involved in formulating the final recommendations. This item excludes individuals who have externally reviewed the guideline (see Item 13). Information about the composition, discipline and relevant expertise of the guideline development group should be provided.
5. Information about patients' experiences and expectations of health care should inform the development of clinical guidelines. There are various methods for ensuring that patients' perspectives inform guideline development. For example, the development group could involve patients' representatives, information could be obtained from patient interviews, literature reviews of patients' experiences could be considered by the group. There should be evidence that this process has taken place.
6. The target users should be clearly defined in the guideline, so they can immediately determine if the guideline is relevant to them. For example, the target users for a guideline on low back pain may include general practitioners, neurologists, orthopaedic surgeons, rheumatologists and physiotherapists.
7. A guideline should have been pre-tested for further validation amongst its intended end users prior to publication. For example, a guideline may have been piloted in one or several primary care practices or hospitals. This process should be documented.

RIGOUR OF DEVELOPMENT**8. Systematic methods were used to search for evidence.**Strongly Agree

4	3	2	1
---	---	---	---

 Strongly Disagree*Comments***9. The criteria for selecting the evidence are clearly described.**Strongly Agree

4	3	2	1
---	---	---	---

 Strongly Disagree*Comments***10. The methods used for formulating the recommendations are clearly described.**Strongly Agree

4	3	2	1
---	---	---	---

 Strongly Disagree*Comments***11. The health benefits, side effects and risks have been considered in formulating the recommendations.**Strongly Agree

4	3	2	1
---	---	---	---

 Strongly Disagree*Comments*

USER GUIDE

RIGOUR OF DEVELOPMENT

8. Details of the strategy used to search for evidence should be provided including search terms used, sources consulted and dates of the literature covered. Sources may include electronic databases (e.g. MEDLINE, EMBASE, CINAHL), databases of systematic reviews (e.g. the Cochrane Library, DARE), handsearching journals, reviewing conference proceedings and other guidelines (e.g. the US National Guideline Clearinghouse, the German Guidelines Clearinghouse).

9. Criteria for including / excluding evidence identified by the search should be provided. These criteria should be explicitly described and reasons for including and excluding evidence should be clearly stated. For example, guideline authors may decide to only include evidence from randomised clinical trials and to exclude articles not written in English.

10. There should be a description of the methods used to formulate the recommendations and how final decisions were arrived at. Methods include for example, a voting system, formal consensus techniques (e.g. Delphi, Glaser techniques). Areas of disagreement and methods of resolving them should be specified.

11. The guideline should consider health benefits, side effects, and risks of the recommendations. For example, a guideline on the management of breast cancer may include a discussion on the overall effects on various final outcomes. These may include: survival, quality of life, adverse effects, and symptom management or a discussion comparing one treatment option to another. There should be evidence that these issues have been addressed.

RIGOUR OF DEVELOPMENT

12. There is an explicit link between the recommendations and the supporting evidence.

Strongly Agree

4	3	2	1
---	---	---	---

 Strongly Disagree

Comments

13. The guideline has been externally reviewed by experts prior to its publication.

Strongly Agree

4	3	2	1
---	---	---	---

 Strongly Disagree

Comments

14. A procedure for updating the guideline is provided.

Strongly Agree

4	3	2	1
---	---	---	---

 Strongly Disagree

Comments

USER GUIDE

RIGOUR OF DEVELOPMENT

12. There should be an explicit link between the recommendations and the evidence on which they are based. Each recommendation should be linked with a list of references on which it is based.

13. A guideline should be reviewed externally before it is published. Reviewers should not have been involved in the development group and should include some experts in the clinical area and some methodological experts. Patients' representatives may also be included. A description of the methodology used to conduct the external review should be presented, which may include a list of the reviewers and their affiliation.

14. Guidelines need to reflect current research. There should be a clear statement about the procedure for updating the guideline. For example, a timescale has been given, or a standing panel receives regularly updated literature searches and makes changes as required.

CLARITY AND PRESENTATION**15. The recommendations are specific and unambiguous.**Strongly Agree

4	3	2	1
---	---	---	---

 Strongly Disagree**Comments****16. The different options for management of the condition are clearly presented.**Strongly Agree

4	3	2	1
---	---	---	---

 Strongly Disagree**Comments****17. Key recommendations are easily identifiable**Strongly Agree

4	3	2	1
---	---	---	---

 Strongly Disagree**Comments****18. The guideline is supported with tools for application.**Strongly Agree

4	3	2	1
---	---	---	---

 Strongly Disagree**Comments**

USER GUIDE

CLARITY AND PRESENTATION

- 15.** A recommendation should provide a concrete and precise description of which management is appropriate in which situation and in what patient group, as permitted by the body of evidence
- An example of a specific recommendation is Antibiotics have to be prescribed in children of two years or older with acute otitis media if the complaints last longer than three days or if the complaints increase after the consultation despite adequate treatment with painkillers, in these cases amoxycillin should be given for 7 days (supplied with a dosage scheme)
 - An example of a vague recommendation is Antibiotics are indicated for cases with an abnormal or complicated course

However, evidence is not always clear cut and there may be uncertainty about the best management. In this case the uncertainty should be stated in the guideline

- 16.** A guideline should consider the different possible options for screening, prevention, diagnosis or treatment of the condition it covers. These possible options should be clearly presented in the guideline. For example, a recommendation on the management of depression may contain the following alternatives:

- a Treatment with TCA
- b Treatment with SSRI
- c Psychotherapy
- d Combination of pharmacological and psychological therapy

- 17.** Users should be able to find the most relevant recommendations easily. These recommendations answer the main clinical questions that have been covered by the guideline. They can be identified in different ways. For example, they can be summarised in a box, typed in bold, underlined or presented as flow charts or algorithms

- 18.** For a guideline to be effective it needs to be disseminated and implemented with additional materials. These may include for example, a summary document, or a quick reference guide, educational tools, patients' leaflets, computer support, and should be provided with the guideline

APPLICABILITY

19. The potential organisational barriers in applying the recommendations have been discussed.

Strongly Agree

4	3	2	1
---	---	---	---

 Strongly Disagree

Comments

20. The potential cost implications of applying the recommendations have been considered.

Strongly Agree

4	3	2	1
---	---	---	---

 Strongly Disagree

Comments

21. The guideline presents key review criteria for monitoring and/or audit purposes.

Strongly Agree

4	3	2	1
---	---	---	---

 Strongly Disagree

Comments

USER GUIDE

APPLICABILITY

- 19.** Applying the recommendations may require changes in the current organisation of care within a service or a clinic which may be a barrier to using them in daily practice. Organisational changes that may be needed in order to apply the recommendations should be discussed. For example:
- i. A guideline on stroke may recommend that care should be co-ordinated through stroke units and stroke services.
 - ii. A guideline on diabetes in primary care may require that patients are seen and followed up in diabetic clinics.
- 20.** The recommendations may require additional resources in order to be applied. For example, there may be a need for more specialised staff, new equipment, expensive drug treatment. These may have cost implications for health care budgets. There should be a discussion of the potential impact on resources in the guideline.
- 21.** Measuring the adherence to a guideline can enhance its use. This requires clearly defined review criteria that are derived from the key recommendations in the guideline. These should be presented. Examples of review criteria are:
- The HbA1c should be $< 8,0\%$
 - The level of diastolic blood pressure should be < 95 mmHg
 - If complaints of acute otitis media lasts longer than three days amoxycillin should be prescribed

22. The guideline is editorially Independent from the funding body.

Comments

Comments

FURTHER COMMENTS

USER GUIDE

EDITORIAL INDEPENDENCE

- 22.** Some guidelines are developed with external funding (e.g. Government funding, charity organisations, pharmaceutical companies). Support may be in the form of financial contribution for the whole development, or for parts of it, e.g. printing of the guidelines. There should be an explicit statement that the views or interests of the funding body have not influenced the final recommendations.

Please note: If it is stated that a guideline was developed without external funding, then you should answer 'Strongly Agree'.

- 23.** There are circumstances when members of the development group may have conflicts of interests. For example, this would apply to a member of the development group whose research on the topic covered by the guideline is also funded by a pharmaceutical company. There should be an explicit statement that all group members have declared whether they have any conflict of interest.

FURTHER COMMENTS

Would you recommend these guidelines for use in practice?

11

11

11

11

NOTES

Date	Time	Location	Weather	Wind	Temp	Humidity	Pressure	Visibility	Remarks

AGREE INSTRUMENT

VOOR BEOORDELING VAN RICHTLIJNEN

The AGREE Collaboration

September 2001

INLEIDING

Doel van het AGREE Instrument

Het doel van het AGREE Instrument is het bieden van een raamwerk om de kwaliteit van klinische richtlijnen te beoordelen

Klinische richtlijnen zijn 'systematisch ontwikkelde aanbevelingen om zorgverleners en patienten te helpen bij beslissingen over passende zorg in specifieke situaties'¹ Deze hebben als doel 'invloed uit te oefenen op het handelen van clinicus'²

Onder de kwaliteit van richtlijnen verstaan we het vertrouwen dat potentiële bronnen van vertekening bij het ontwikkelen van richtlijnen zo beperkt mogelijk zijn gebleven en dat de aanbevelingen zowel intern als extern valide zijn en haalbaar zijn in de praktijk Dit houdt ook in dat rekening is gehouden met de voordelen, nadelen en de kosten van de toepassing van de aanbevelingen, evenals met de praktische mogelijkheden en beperkingen die hiermee samenhangen De beoordeling van richtlijnen heeft betrekking op de methoden van richtlijnontwikkeling, de inhoud van de uiteindelijke aanbevelingen, maar ook op factoren die samenhangen met de acceptatie en invoering van de richtlijnen

Het AGREE Instrument beoordeelt **zowel** de kwaliteit van de verslaglegging **als** de kwaliteit van bepaalde aspecten van de aanbevelingen Het beoordeelt de kans dat een richtlijn zijn gewenste doel zal behalen, maar niet de daadwerkelijke impact op patientuitkomsten

De meeste criteria van het AGREE Instrument zijn meer gebaseerd op theoretische aannames dan op empirisch bewijsmateriaal Ze zijn vooral ontwikkeld op basis van discussies tussen onderzoekers afkomstig uit verscheidene landen met uitgebreide kennis en ervaring op het gebied van richtlijnen Daarom dient het AGREE Instrument te worden beschouwd als een afspiegeling van de huidige stand van kennis op dit gebied

Welke richtlijnen kunnen worden beoordeeld met het AGREE Instrument?

Het AGREE Instrument is gemaakt voor de beoordeling van richtlijnen ontwikkeld door lokale, regionale, nationale en internationale organisaties of groepen Met het instrument kunnen bestaande, nieuwe en herziene richtlijnen worden beoordeeld

Het AGREE Instrument is generiek van aard en kan worden toegepast op richtlijnen voor diagnostiek, voorlichting of behandeling van elk ziektebeeld Het is geschikt voor richtlijnen zowel in gedrukte als elektronische vorm

¹ Lohr KN, Field MJ A provisional instrument for assessing clinical practice guidelines In Institute of Medicine Field MJ, Lohr KN (eds) Guidelines for clinical practice From development to use Washington D C National Academy Press, 1992

² Hayward RSA, Wilson MC, Tunis SR, Bass EB, Guyatt G, for the Evidence-Based Medicine Working Group Users' guides to the Medical Literature VIII How to Use Clinical Practice Guidelines A Are the Recommendations Valid? JAMA 1995, 274 570-574

Wie kunnen het AGREE Instrument gebruiken?

Het AGREE Instrument is bedoeld voor de volgende groepen:

- i) *Beleidsmakers*, om hen te helpen te bepalen welke richtlijnen kunnen worden aanbevolen voor gebruik in de praktijk. In dergelijke gevallen dient het instrument deel uit te maken van een formele beoordelingsprocedure.
- ii) *Richtlijnmakers*, om een gestructureerde en zorgvuldige ontwikkelingsmethode te volgen. Hierdoor kunnen zij waarborgen dat hun richtlijnen van hoge kwaliteit zijn.
- iii) *Zorgverleners* die eerst de aanbevelingen zelf willen beoordelen alvorens ze over te nemen.
- iv) *Docenten*, ten behoeve van het onderwijs aan zorgverleners in het kritisch beoordelen van richtlijnen.

Literatuur

De volgende bronnen zijn gebruikt voor de ontwikkeling van het AGREE Instrument.

Lohr KN, Field MJ. A provisional instrument for assessing clinical practice guidelines. In: Institute of Medicine. Field MJ, Lohr KN (eds). Guidelines for clinical practice. From development to use. Washington D.C.: National Academy Press, 1992.

Cluzeau F, Littlejohns P, Grimshaw J, Feder G, Moran S. Development and application of a generic methodology to assess the quality of clinical guidelines. International Journal for Quality in Health Care 1999; 11: 21-28.

Grol R, Dalhuijsen J, Thomas S, in 't Veld C, Rutten G, Mokkink H. Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. BMJ 1998; 317: 858-861.

Lohr KN. The quality of practice guidelines and the quality of health care. In: Guidelines in health care. Report of a WHO Conference. January 1997, Baden-Baden: Nomos Verlagsgesellschaft, 1998.

INSTRUCTIES

Lees eerst zorgvuldig de volgende instructies voor gebruik van het AGREE Instrument

1. Opbouw en inhoud van het AGREE Instrument

Het instrument bestaat uit 23 items verdeeld over zes domeinen. Elk domein beslaat een aparte dimensie van kwaliteit van richtlijnen

Onderwerp en doel (items 1-3) betreft het doel van de richtlijn, de specifieke klinische vragen waarop de richtlijn een antwoord geeft en de patientenpopulatie waarop de richtlijn van toepassing is

Betrokkenheid van belanghebbenden (items 4-7) richt zich op de mate waarin de richtlijn de opvattingen van de beoogde gebruikers weerspiegelt

Methodologie (items 8-14) hangt samen met het proces waarin bewijsmateriaal is verzameld en samengesteld en met de gebruikte methoden om aanbevelingen op te stellen en te herzien

Helderheid en presentatie (items 15-18) gaat over het taalgebruik en de vorm van de richtlijn

Toepassing (items 19-21) houdt verband met de mogelijke organisatorische, gedragsmatige en financiële consequenties van het toepassen van de richtlijn

Onafhankelijkheid van de opstellers (items 22-23) betreft de onafhankelijkheid van de aanbevelingen en erkenning van mogelijke conflicterende belangen van leden van de werkgroep

2. Documentatie

Beoordelaars dienen vóór de beoordeling alle informatie over de totstandkoming van de richtlijnen proberen te achterhalen. Deze informatie kan in hetzelfde document staan als de aanbevelingen of samengevat zijn in een apart technisch rapport, gepubliceerde artikelen of beleidsrapporten (bijvoorbeeld richtlijnprogramma's). We raden u aan eerst de richtlijn en bijbehorende documenten en rapporten in zijn geheel te lezen voordat u aan de beoordeling begint.

3. Aantal beoordelaars

We raden aan elke richtlijn door ten minste twee - en bij voorkeur door vier - beoordelaars te laten beoordelen, omdat dit de betrouwbaarheid van de beoordeling vergroot.

4. Antwoordcategorieën

Elk item wordt gescoord op een vierpuntschaal die loopt van 4 ('Zeer Eens'), via 3 ('Eens') en 2 ('Oneens') naar 1 ('Zeer Oneens'). De schaal meet in hoeverre aan het criterium is voldaan.

- Als u er zeker van bent dat volledig aan het criterium is voldaan, antwoord dan 'Zeer Eens'
- Als u er zeker van bent dat helemaal niet aan het criterium is voldaan of als er geen informatie beschikbaar is, antwoord dan 'Zeer Oneens'
- Als u niet zeker bent of aan een criterium wordt voldaan, bijvoorbeeld omdat de informatie onduidelijk is of omdat alleen een deel van de aanbevelingen aan het criterium voldoet, antwoord dan 'Eens' of 'Oneens' afhankelijk van de mate waarin u denkt dat de kwestie is behandeld.

5. Handleiding

We hebben bij elk item aanvullende informatie in de handleiding vermeld. Deze informatie geeft uitleg over de gebruikte begrippen. Lees deze handleiding zorgvuldig alvorens een antwoord te geven.

6. Toelichting

Naast elk item is een kader voor commentaar beschikbaar. U dient dit kader te gebruiken om de reden van uw antwoord toe te lichten. Bijvoorbeeld: u antwoordt 'Zeer Oneens' omdat de informatie niet beschikbaar is, of omdat het item niet van toepassing is, of omdat de beschreven methodologie ontoereikend is. Aan het eind van het instrument is er ook ruimte voor nadere toelichting.

7. Berekening van domeinscores

Domeinscores kunnen worden berekend door alle scores van de individuele items in een domein op te tellen en het totaal te standaardiseren door het percentage te nemen van de maximaal mogelijke score voor dat domein.

Voorbeeld: als vier beoordelaars de volgende scores geven voor domein 1 (onderwerp en doel):

	Item 1	Item 2	Item 3	Totaal
Beoordelaar 1	2	3	3	8
Beoordelaar 2	3	3	4	10
Beoordelaar 3	2	4	3	9
Beoordelaar 4	2	3	4	9
Totaal	9	13	14	36

Maximaal mogelijke score = 4 (zeer eens) x 3 (items) x 4 (beoordelaars) = 48

Minimaal mogelijk score = 1 (zeer oneens) x 3 (items) x 4 (beoordelaars) = 12

De gestandaardiseerde domeinscore is dan:

$$\frac{\text{verkregen score} - \text{minimaal mogelijke score}}{\text{maximaal mogelijke score} - \text{minimaal mogelijke score}} =$$

$$\frac{36 - 12}{48 - 12} = \frac{24}{36} = 0.67 \times 100 = 67\%$$

NB: De zes domeinscores zijn onafhankelijk en dienen niet te worden opgeteld tot één kwaliteits-score. Hoewel de domeinscores nuttig kunnen zijn om richtlijnen te vergelijken en om te beslissen welke richtlijn al dan niet aan te bevelen, is het niet mogelijk om drempelwaarden vast te stellen die 'goede' of 'slechte' richtlijnen aanduiden.

8. Algemeen oordeel

Aan het eind van het instrument is een paragraaf voor een algemeen oordeel bijgevoegd. Deze bevat een reeks opties 'Sterk aan te bevelen', 'Aan te bevelen (onder voorwaarden of met veranderingen)', 'Niet aan te bevelen' en 'Onzeker'. Het algemene oordeel vereist dat de beoordelaar een oordeel geeft over de kwaliteit van de richtlijn waarbij elk beoordelingscriterium wordt meegenomen.

ONDERWERP EN DOEL

1. Het doel van de richtlijn is specifiek beschreven.

Zeer Eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

2. De klinische vraag/vragen die in de richtlijn aan de orde komt/komen, is/zijn specifiek beschreven.

Zeer Eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

3. De patiëntenpopulatie waarop de richtlijn van toepassing is, is specifiek beschreven.

Zeer Eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

HANDLEIDING

ONDERWERP EN DOEL

1. Dit betreft de mogelijke impact van een richtlijn op de samenleving en patientenpopulaties. Het doel van de richtlijn dient in detail te zijn beschreven. De te verwachten gezondheidswinst van de richtlijn dient specifiek te zijn voor het klinische probleem. Voorbeelden van specifieke formuleringen zijn:
 - Preventie van (lange termijn) complicaties van patienten met diabetes mellitus,
 - Verlagen van het risico van nieuwe vasculaire gebeurtenissen bij patienten met een doorgemaakt hartinfarct,
 - Rationeel en kosteneffectief voorschrijven van antidepressiva

2. De klinische vragen waarop de richtlijn een antwoord geeft, dienen gedetailleerd te zijn beschreven, vooral met betrekking tot de kernaanbevelingen (zie item 15). Uitgaande van de voorbeelden van vraag 1:
 - Hoe vaak per jaar moet het HbA1c bij patienten met diabetes mellitus worden gemeten?
 - Wat is de aanbevolen dagelijkse dosis aspirine voor patienten met een aangetoond myocardinfarct?
 - Zijn selectieve serotonine heropnameremmers (SSRI's) kosteneffectiever dan tricyclische antidepressiva (TCA's) bij de behandeling van patienten met een depressie?

3. Er dient een duidelijke beschrijving te zijn van de doelpopulatie waarop de richtlijn is gericht. Leeftijd, geslacht, klinisch beeld en co-morbiditeit kunnen vermeld zijn. Bijvoorbeeld:
 - Een richtlijn over de behandeling van diabetes mellitus gaat alleen over patienten met niet-insuline afhankelijke diabetes en sluit patienten met cardiovasculaire co-morbiditeit uit.
 - Een richtlijn over het beleid bij depressie is uitsluitend gericht op patienten met een ernstige depressie volgens de DSM-IV criteria, en sluit patienten met psychotische symptomen en kinderen uit.
 - Een richtlijn over de screening op borstkanker gaat alleen over vrouwen tussen de 50 en 70 jaar, zonder kanker in de voorgeschiedenis en zonder positieve familie-anamnese voor borstkanker.

BETROKKENHEID VAN BELANGHEBBENDEN

4. De leden van de werkgroep die de richtlijn heeft ontwikkeld komen uit alle relevante beroepsgroepen.

Zeer Eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

5. Het perspectief en de voorkeuren van patiënten zijn nagegaan.

Zeer Eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

6. De beoogde gebruikers van de richtlijn zijn duidelijk benoemd.

Zeer Eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

7. De richtlijn is getest onder de beoogde gebruikers.

Zeer Eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

HANDLEIDING

BETROKKENHEID VAN BELANGHEBBENDEN

4. Dit item verwijst naar de professionals die in een bepaald stadium bij de richtlijnontwikkeling betrokken waren. Dit kunnen de leden van de stuurgroep zijn, het onderzoeksteam dat betrokken was bij de selectie en beoordeling van het wetenschappelijke bewijsmateriaal, en degenen die de uiteindelijke aanbevelingen hebben geformuleerd. Het gaat hier niet om de externe personen die de conceptrichtlijn hebben beoordeeld (zie item 13). In de richtlijn dient informatie te staan over de samenstelling, discipline en relevante deskundigheid van de werkgroep.
5. Informatie over de ervaringen van patiënten en hun verwachtingen van de zorg dient bij de richtlijnmakers bekend te zijn. Er zijn diverse methoden waarop deze informatie vergaard kan worden. Bijvoorbeeld door vertegenwoordigers van patiënten in de werkgroep op te nemen, door interviews met patiënten, of literatuuronderzoek naar patiëntenervaringen. Het dient duidelijk te zijn dat dit proces heeft plaatsgevonden.
6. De beoogde gebruikers dienen duidelijk in de richtlijn te zijn benoemd, zodat zij onmiddellijk kunnen vaststellen of de richtlijn voor hen relevant is. De beoogde gebruikers van bijvoorbeeld een richtlijn over lage rugpijn kunnen huisartsen, neurologen, orthopedisch chirurgen, reumatologen en fysiotherapeuten zijn.
7. Vóór de publicatie dient een richtlijn voor verdere validering getest te zijn onder de beoogde gebruikers, bijvoorbeeld door een richtlijn uit te proberen in een of meerdere huisartspraktijken of ziekenhuizen. Dit proces dient gedocumenteerd te zijn.

METHODOLOGIE

8. Er zijn systematische methoden gebruikt voor het zoeken naar wetenschappelijk bewijsmateriaal.

Zeer Eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

9. De criteria voor het selecteren van het wetenschappelijk bewijsmateriaal zijn duidelijk beschreven.

Zeer eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

10. De gebruikte methoden om de aanbevelingen op te stellen, zijn duidelijk beschreven.

Zeer Eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

11. Gezondheidswinst, bijwerkingen en risico's zijn overwogen bij het opstellen van de aanbevelingen.

Zeer Eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

HANDLEIDING

METHODOLOGIE

8. De strategie waarmee de literatuur is verzameld dient in detail te zijn beschreven, inclusief zoektermen, geraadpleegde bronnen en de periode waarover artikelen werden verzameld. Mogelijke bronnen zijn elektronische databases (b.v. MEDLINE, EMBASE, CINAHL), databases van systematische reviews (bijv. Cochrane Library, DARE), handmatig geselecteerde tijdschriften, congresverslagen en andere richtlijnen (bijv. US National Guideline Clearinghouse, German Guidelines Clearinghouse).
9. Criteria voor het in- en uitsluiten van literatuur dienen te zijn vermeld. Deze criteria moeten expliciet zijn beschreven en de redenen voor in- en uitsluiting van literatuur moeten duidelijk zijn vermeld. De auteurs van richtlijnen kunnen bijvoorbeeld besluiten dat ze uitsluitend gerandomiseerde trials includeren en artikelen die in het Engels of Nederlands zijn geschreven.
10. De methoden die zijn gebruikt bij het opstellen van de aanbevelingen dienen te zijn beschreven evenals de wijze waarop men tot de uiteindelijke conclusies is gekomen. Voorbeelden van dergelijke methoden zijn een stemmingssysteem of formele consensustechnieken (bijv. Delphi, Glaser technieken). Punten waarover men van mening verschildte en hoe deze opgelost werden, dienen duidelijk te zijn omschreven.
11. De richtlijn dient de gezondheidswinst, bijwerkingen en risico's van de aanbevelingen te overwegen. Bijvoorbeeld, in een richtlijn over het beleid bij borstkanker kunnen de globale effecten op verschillende uitkomstmaten zijn beschreven. Deze kunnen zijn: de overleving, kwaliteit van leven, nadelige effecten, symptoombestrijding of een bespreking van verschillende behandelingsalternatieven. Het dient duidelijk te zijn dat deze punten zijn behandeld.

METHODOLOGIE

12. Er bestaat een expliciet verband tussen de aanbevelingen en het onderliggende wetenschappelijke bewijsmateriaal.

Zeer Eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

13. De richtlijn is voor publicatie door externe experts beoordeeld.

Zeer Eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

14. Een procedure voor herziening van de richtlijn is vermeld.

Zeer Eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

HANDLEIDING

METHODOLOGIE

12. Er dient een expliciet verband te bestaan tussen de aanbevelingen en het wetenschappelijke bewijs waarop zij zijn gebaseerd. Elke aanbeveling dient gekoppeld te zijn aan een referentielijst waarop zij is gebaseerd.

13. Een richtlijn dient extern te zijn beoordeeld voordat zij is gepubliceerd. Referenten dienen niet betrokken te zijn geweest bij de richtlijnwerkgroep en onder hen behoren zowel klinische experts op het gebied van de richtlijn als enkele methodologische experts aanwezig te zijn. Ook vertegenwoordigers van patiënten kunnen als referent optreden. Een beschrijving van de methodologie die bij de externe beoordeling is gebruikt dient aanwezig te zijn. Ook kan een lijst van referenten en de instellingen waaraan zij verbonden zijn, worden bijgevoegd.

14. Richtlijnen behoren de actuele stand van wetenschap weer te geven. In de richtlijn dient een duidelijke uitspraak te zijn gedaan over de procedure voor herziening van de richtlijn. Bijvoorbeeld, een geldigheidsduur is aangegeven of een vast panel ontvangt regelmatig bijgewerkte literatuursearces en brengt zo nodig wijzigingen aan.

HELDERHEID EN PRESENTATIE

15. De aanbevelingen zijn specifiek en ondubbelzinnig.

Zeer Eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

16. De verschillende beleidsopties zijn duidelijk vermeld.

Zeer Eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

17. De kernaanbevelingen zijn gemakkelijk te herkennen.

Zeer Eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

18. De toepassing van de richtlijn wordt ondersteund met hulpmiddelen.

Zeer Eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

HANDLEIDING

HELDERHEID EN PRESENTATIE

15. Een aanbeveling dient op grond van het beschikbare wetenschappelijke bewijsmateriaal een concrete en nauwkeurige beschrijving te geven over welk beleid geschikt is in bepaalde situaties bij een bepaalde patientengroep
- Een voorbeeld van een specifieke aanbeveling is 'Antibiotica dienen te worden voorgeschreven bij otitis media acuta bij kinderen van twee jaar of ouder indien de klachten langer duren dan drie dagen of indien de klachten toenemen na het consult ondanks adequate pijnstilling; in deze gevallen dient amoxicilline gedurende 7 dagen voorgeschreven te worden' (voorzien van een doseringsschema)
 - Een voorbeeld van een vage aanbeveling is 'Antibiotica zijn geïndiceerd bij een abnormaal of gecompliceerd verloop'
- Het wetenschappelijke bewijsmateriaal is echter niet altijd even duidelijk en er kan twijfel bestaan over het beste beleid. In dat geval dient deze twijfel in de richtlijn te zijn vermeld.
16. Een richtlijn dient de verschillende opties te overwegen voor screening, preventie, diagnostiek of behandeling van het betreffende klinische probleem. De keuzemogelijkheden dienen duidelijk in de richtlijn te zijn vermeld. Bijvoorbeeld, een aanbeveling voor het beleid bij depressie kan de volgende behandelingsalternatieven bevatten
- a Behandeling met TCA
 - b Behandeling met SSRI
 - c Psychotherapie
 - d Combinatie van farmacologische en psychologische therapie
17. Gebruikers van de richtlijn dienen in staat te zijn de meest relevante aanbevelingen gemakkelijk te vinden. Deze aanbevelingen geven antwoord op de belangrijkste klinische vragen die in de richtlijn aan de orde komen. Ze kunnen op verschillende manieren worden weergegeven, bijvoorbeeld samengevat in een kader, door vetdruk, door onderstreping of door ze te presenteren als stroomdiagrammen of algoritmen.
18. Voor een effectieve richtlijn zijn disseminatie- en implementatiematerialen nodig, bijvoorbeeld een samenvattingdocument of een 'quick reference guide', nascholingsmateriaal, patientenfolders of computerondersteuning. Deze middelen dienen bij de richtlijn geleverd te zijn.

TOEPASSING

19. De mogelijke organisatorische belemmeringen bij het toepassen van de aanbevelingen zijn besproken.

Zeer Eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

20. De mogelijke kostenimplicaties van het toepassen van de aanbevelingen zijn overwogen.

Zeer Eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

21. De richtlijn geeft de belangrijkste criteria om na te gaan en te toetsen of de richtlijn wordt gevolgd.

Zeer Eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

HANDLEIDING

TOEPASSING

19. Het toepassen van de aanbevelingen kan zodanige veranderingen vereisen in de huidige organisatie van de zorg binnen een instelling of praktijk dat deze een belemmering vormen om de aanbevelingen in de dagelijkse praktijk te gebruiken. Organisatorische veranderingen die nodig kunnen zijn om de aanbevelingen toe te passen dienen te zijn besproken. Bijvoorbeeld:

- Een richtlijn over beroerte kan adviseren dat de zorg moet worden gecoördineerd in stroke-units.
- Een richtlijn over diabeteszorg in de eerste lijn kan vereisen dat patiënten worden gezien en gecontroleerd in diabetespoliklinieken.

20. De toepassing van de aanbevelingen kan aanvullende middelen vereisen, bijvoorbeeld meer gespecialiseerd personeel, nieuwe apparatuur of behandeling met een duur geneesmiddel. Dit kan consequenties hebben voor het gezondheidszorgbudget. In de richtlijn dienen deze kostenimplicaties te zijn besproken.

21. Het meten van de naleving van de richtlijn kan haar gebruik bevorderen. Dit vereist helder gedefinieerde criteria die zijn afgeleid van de belangrijkste aanbevelingen uit de richtlijn. Deze criteria dienen te zijn vermeld. Voorbeelden van dergelijke criteria zijn:

- de HbA1c dient lager dan 8,0% te zijn
- de diastolische bloeddruk dient lager dan 95 mm Hg te zijn
- indien de klachten van een otitis media acuta langer duren dan drie dagen dient amoxicilline te worden voorgeschreven.

ONAFHANKELIJKHEID VAN DE OPSTELLERS

- 22. De richtlijn is niet beïnvloed door de opvattingen of belangen van de financierende instantie.**

Zeer Eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

- 23. Conflicterende belangen van leden van de werkgroep zijn vastgelegd.**

Zeer Eens

4	3	2	1
---	---	---	---

 Zeer Oneens

Toelichting

HANDLEIDING

ONAFHANKELIJKHEID VAN DE OPSTELLERS

- 22.** Sommige richtlijnen worden ontwikkeld met steun van externe financiering (bijv. van overheid, charitatieve instellingen, farmaceutische bedrijven). Deze steun kan een financiële bijdrage zijn voor de gehele richtlijnontwikkeling, of voor onderdelen ervan, bijv. het drukken van de richtlijnen. Er dient expliciet aangegeven te zijn dat de opvattingen of belangen van de financierende instantie de uiteindelijke aanbevelingen niet hebben beïnvloed.
Let op: indien is aangegeven dat een richtlijn zonder externe financiering is ontwikkeld, antwoord dan 'Zeer Eens'.

- 23.** Onder bepaalde omstandigheden kunnen leden van de richtlijnwerkgroep conflicterende belangen hebben, bijvoorbeeld als een werkgroep lid op het gebied van het onderwerp van de richtlijn onderzoek doet dat wordt gesponsord door een farmaceutisch bedrijf. Er dient expliciet te zijn vermeld dat alle werkgroep leden hebben verklaard of ze conflicterende belangen hebben.

NADERE TOELICHTING

ALGEMEEN OORDEEL

Zou u deze richtlijn aanbevelen voor gebruik in de praktijk?

Sterk aan te bevelen

Aan te bevelen (onder voorwaarden of met veranderingen)

Niet aan te bevelen

Onzeker

Appendix D

- 1. Handleiding voor de beoordeling van aanbevelingen in klinische richtlijnen (original Dutch version)**
- 2. User guide for appraisal of recommendations in clinical guidelines (translated English version)**

Handleiding voor de beoordeling van aanbevelingen in klinische richtlijnen

1. De onderbouwing van de aanbeveling wordt expliciet beschreven.

Het moet duidelijk en zichtbaar zijn waarop de aanbeveling gebaseerd is. De onderbouwing kan zowel in de hoofdtekst als in een bijbehorende noot worden beschreven of alleen uit een verwijzing naar een artikel bestaan. De onderbouwing hoeft niet per se een wetenschappelijk onderzoek te betreffen, maar kan ook een argumentatie of beslissing van de werkgroep zijn. De aanwezigheid van dit criterium wordt bepaald door het expliciet aanwezig zijn van een aan de aanbeveling gekoppelde onderbouwing.

2. De aanbeveling wordt hoofdzakelijk onderbouwd met wetenschappelijk onderzoek.

Met wetenschappelijk onderzoek worden empirische onderzoeken (veelal onderzoeken met patiënten) bedoeld, zoals experimentele trials, observationeel of case-control-onderzoek. Indien er geen onderzoek voorhanden is, wordt vaak alleen een argumentatie of consensusbesluit als onderbouwing gebruikt. De aanwezigheid van dit criterium wordt dus vooral bepaald door de beschikbaarheid van wetenschappelijk onderzoek, waar bij de onderbouwing vanuit wordt gegaan. Aanvullende argumentaties kunnen wel aanwezig zijn, maar vormen niet de kern van de onderbouwing.

3. De aanbeveling wordt ondersteund met een bespreking van de voordelen van het opvolgen van de aanbeveling (kans op toename van gezondheid of kwaliteit van leven).

De voordelen van het doen (of laten) van een interventie worden besproken. Bij een diagnostische test gaat het om de voorspellende waarde en de diagnostische winst van de test; bij therapeutische interventies kan het gaan om verkorting van de ziekteduur, vermindering van pijn, verhoging van kwaliteit van leven of vermindering van het risico op ziekten later in het leven.

4. De aanbeveling wordt ondersteund met een bespreking van de nadelen van het opvolgen van de aanbeveling (schadelijke effecten, bijwerkingen, risico's).

De nadelen van het doen (of laten) van een interventie worden besproken. Bij een diagnostische test kan het gaan om een belastende invasieve ingreep; bij een therapeutische interventie om bijwerkingen, schadelijke effecten of complicatierisico's.

5. De aanbeveling strookt met de normen en waarden van de bestaande klinische praktijk.

De aanbeveling sluit aan bij wat normaal en gangbaar wordt geacht in de huisartspraktijk. Het roept geen controverse of discussie op en sluit aan bij wat in het algemeen verwacht mag worden van de huisarts.

6 De aanbeveling roept geen negatieve reacties op bij patiënten

De aanbeveling kan negatieve reacties bij (een substantieel deel van de) patiënten oproepen als deze niet aansluit bij de verwachtingen of als het om belastende ingrepen gaat

7 De toepassing van de aanbeveling vereist geen nieuwe kennis

Bij dit criterium wordt uitgegaan van de kennis van de gemiddelde huisarts. Nieuwe kennis is bijvoorbeeld vereist bij betrekkelijk 'nieuwe ziektebeelden' zoals aids of RSI of bij de introductie van nieuwe diagnostische technieken of behandelingen

8 De toepassing van de aanbeveling vereist geen nieuwe vaardigheden

Bij dit criterium wordt uitgegaan van de vaardigheden zoals die in het basistakenpakket zijn omschreven. Scleroseren van varices, proctoscopie en het gebruik van een tympanometer vereisen bijvoorbeeld nieuwe vaardigheden

9 De toepassing van de aanbeveling vereist geen aanpassingen in de organisatie

De toepassing van de aanbeveling vraagt om aanpassingen in de organisatie als de logistiek van de praktijkvoering of de taken van de praktijkassistente drastisch veranderen. Voorbeelden zijn het opzetten van een driejaarlijkse screening op diabetes mellitus bij mensen ouder dan 60 jaar of een systematisch controlebeleid bij astma/COPD-patiënten

10 De toepassing van de aanbeveling vereist geen verandering van bestaande routines en gewoonten

Bij dit criterium wordt uitgegaan van de bestaande routines en gewoonten van de gemiddelde huisarts. Het aanbevelen van bijvoorbeeld de dipslide als eerste test bij urine-onderzoek of het ongevraagd aanbieden van onderzoek, zoals rectaal toucher bij oudere mannen of borstonderzoek bij vrouwen ten behoeve van vroege opsporing, doorbreekt de bestaande routines en gewoonten

User guide for appraisal of recommendations in clinical guidelines

1 The recommendation is explicitly linked to the supporting evidence

The evidence on which the recommendations are based should be made clear and visible. The supporting evidence may be described in the text or in an explanatory note, or it may only consist of a literature reference. The supporting evidence does not necessarily include scientific research and may also include arguments or decisions by the guideline development group. This criterion is present when there is an explicit link between the recommendation and the evidence.

2 The recommendation is primarily supported by scientific evidence

Scientific evidence is covered by empirical studies (often in patients), such as experimental trials, observational studies, or case-control studies. If no studies are available, arguments or group decisions are often used to support the recommendation. The presence of this criterion is particularly determined by the availability of scientific studies used to support the recommendation. Additional arguments may also be used, but these do not contribute to the main part of the supporting evidence.

3 The recommendation is supported by a discussion of the benefits of applying the recommendation (health gain or quality of life)

The benefits of the recommended intervention are discussed. For diagnostic tests the benefits can be expressed in measures such as the predictive value and diagnostic gain, for therapeutic interventions the benefits could include shortening the course of diseases, reduction of pain, improving quality of life, or prevention of risks.

4 The recommendation is supported by a discussion of the harms and risks of applying the recommendation

The harms and risks of the recommended intervention are discussed. Diagnostic interventions could be stressful and painful, while therapeutic interventions could induce serious side effects or complications.

5 The recommendation is compatible with existing norms and values in practice

The recommendation is compatible with usual general practice. It does not cause controversy and corresponds to general expectations of the practitioners.

6 The recommendation does not evoke negative reactions in patients

The recommendation might evoke negative reactions in a substantial proportion of the patients if it is not compatible with their expectations or if it includes stressful interventions.

7. *Application of the recommendation requires no new knowledge.*

The knowledge of the average general practitioner is assumed. New knowledge could be required in 'new clinical pictures', such as Aids or RSI, or with the introduction of new diagnostic or therapeutic techniques.

8. *Application of the recommendation requires no new skills.*

The skills described in the 'basistakenpakket' is assumed. For example, sclerosing of varices, proctoscopy, and the use of a tympanometer require new skills.

9. *Application of the recommendation requires no changes in the organisation.*

The application of the recommendation requires change in the organisation when it introduces changes in procedures or tasks of the personnel. For example, the introduction of triennial screening for diabetes mellitus in people above 60 years of age or regular monitoring of patients with asthma or COPD.

10. *Application of the recommendation requires no changes in existing routines or habits.*

The existing routines and habits of the average general practitioner are assumed. To recommend use of a dipslide for detection of bacteriuria as first line diagnostic test, or to offer rectal palpation in older men, or breast examination in women without complaints could disrupt existing routines and habits.

Dankwoord/Acknowledgement

Met de voltooiing van dit proefschrift is een einde gekomen aan een tumultueuze periode. Mijn dochters Femke en Lotte waren anderhalf jaar respectievelijk één maand oud toen Richard Grol mij vroeg om promotie-onderzoek te doen. Op dat moment wist ik ook dat er een druk jaar zat aan te komen wegens de nieuwbouw van een huis en een huisartspraktijk. Het promotie-onderwerp was echter zo'n schot in de roos dat ik deze kans niet wilde laten lopen. Na amper overleg met Karen, mijn dierbare eega, besloten we er samen voor te gaan.

Het geeft een enorme voldoening heelhuids uit deze periode tevoorschijn te zijn gekomen. Hoewel we menigmaal wensten het wat rustiger aan te doen, was het project zo inspirerend en fascinerend dat mijn motivatie geen moment ter discussie heeft gestaan. Ook kwam ik onderweg geen grote obstakels tegen en werd er elke week progressie geboekt, mede door de intensieve samenwerking en inspirerende contacten met collega's. Graag wil ik hierbij de gelegenheid nemen om een aantal personen expliciet te bedanken voor hun bijdrage aan de totstandkoming van dit proefschrift.

Richard Grol was de leidende kracht van mijn project. Het was een voorrecht om door zo'n kanjer in de internationale wetenschap begeleid te worden. Zijn optimisme, efficiëntie en werklust werkte zeer aanstekelijk. Ook zijn internationale optredens—en vooral de rust die hij daarbij uitstraalde—zijn een lichtend voorbeeld geweest. Richard, ik hoop dat we in de toekomst samen nog veel artikelen zullen schrijven en dat ik van je deskundige adviezen gebruik mag blijven maken.

Joost Zaat was van meet af aan als copromoter bij het project betrokken. Vooral in de eerste jaren hebben we intensieve contacten gehad. Joost was onnavolgbaar kritisch—soms zelfs nukkig—en wist feilloos de zwakke plekken in ons onderzoek te detecteren. Bovenal heb ik hem ervaren als een integere huisarts-maat die weet hoe het in de praktijk eraan toe gaat. Joost, je bent mijn sparring-partner; we zullen elkaar blijven uitdagen!

Kort nadat Niek Klazinga benoemd werd tot hoogleraar sociale geneeskunde aan de Universiteit van Amsterdam, heb ik hem gevraagd mede-promotor van mijn onderzoek te worden. Nooit zal ik zijn oprecht blijde en enigszins verraste reactie vergeten. Niek, je bent een wandelende bibliotheek op het gebied van richtlijnontwikkeling en nog veel meer. Je eruditie en warmte zijn van grote betekenis geweest voor mij.

As AGREE project leader, Françoise Cluzeau gave an excellent example of how an international project should be coordinated. She was involved in most of the studies included in this thesis. Her input to the development and validation of the AGREE Instrument was invaluable. I am also very grateful for her comments on drafts of several chapters of this thesis. It is a great prospect to keep working together with her in the next AGREE project.

Gene Feder was also involved in the AGREE project and was the supervisor of the diabetes guidelines study. It was a great pleasure to work with him, for we share the same philosophical and cultural interests. Our collaboration transformed into a warm friendship. Julia Bailey also participated in the diabetes guidelines study, going through much tough work. She was a master at keeping an eye on the details without losing control.

Marjukka Mäkelä was an active and inspiring member of the AGREE Collaboration. She was involved in the guideline programme study and her help in structuring the paper was very useful.

I am very grateful for the statistical support of Steve Hanna. He was able to explain complex analyses in simple, understandable terms.

Jeremy Grimshaw was one of the intellectual masters of the AGREE Collaboration. His warm personality and engagement gave me much inspiration. Robbie Foy did similar research on attributes of clinical guidelines and was supervised by Jeremy. He commented on drafts of the paper on characteristics of effective guidelines. I have enjoyed his great sense of humour, which was very helpful in the struggle for getting papers published.

I would like to thank all AGREE partners who contributed to the several papers included in this thesis. Like the weather was beautiful at the AGREE workshops, the atmosphere was great. The AGREE project was finished a year ago, but I am convinced that we will AGREE forever.

Teun Spies heeft het onderzoekspad voor mij geëffend. Dankbaar heb ik gebruik mogen maken van de data van het TAS-project waardoor de noeste arbeid van gegevensverzameling mij werd bespaard. De methodologische en statistische adviezen van 'good old' Henk Mookink waren van grote klasse. Akke van der Bij heeft in het derde jaar als onderzoeksassistente ervoor gezorgd dat het onderzoek in een stroomversnelling kwam. Ze bleek een allround-wetenschapper te zijn waar ik veel van heb geleerd.

Rob Dijkstra en Miranda Laurent waren mijn kamergenoten in Nijmegen. Hoewel ik de meeste tijd thuis heb gewerkt, heb ik een speciale band met hen opgebouwd. Rob heeft meegewerkt aan het beoordelen van de Nederlandse richtlijnen in de diabetes studie. Miranda heeft me opgevangen tijdens de eerste onwennige maanden op de WOK. Bovendien heeft zij belangrijk bijgedragen aan de perfecte organisatie van de AGREE workshop in Amsterdam in 2000.

In de periode voor het proefschrift hebben Siep Thomas en Frans Meulenberg als hoofd respectievelijk coördinator van de afdeling standaardenontwikkeling van het Nederlands Huisartsen Genootschap mij gecoacht. Zij hebben een sleutelrol vervuld in mijn wetenschappelijke en persoonlijke ontwikkeling. De collega-staffleden van 'ASO' hebben mij het plezier in het maken van richtlijnen verschaft.

Siep maakte ook deel uit van het huisartsenpanel dat de kenmerken van richtlijnen beoordeelde. Samen met Aria Biemond, Roeland Drijver, Rob Dijkstra, Bernard Frijling, Paul Janssen, Marijke Labots-Vogelesang, Vroon Pigmans, Ber Pleumeekers, Berend Terluin, Cees in 't Veld en Michel van Wijk ben ik hen veel dank verschuldigd voor hun inzet in hun kostbare vrije tijd.

Anita Oude Bos was als secretaresse gedurende het gehele project mijn steun en toeverlaat. Zij had aan weinig woorden genoeg en handelde zaken vlot en zelfstandig af. Jolanda van Haren heeft met eenzelfde voortvarendheid en met een scherp oog voor de details het manuscript en de druk van het proefschrift tot stand gebracht.

I would gratefully thank Bruce Belshaw for his excellent translations and editing, which were carried out in next to no time. Ook Bert te Velde, mijn zwager in Amerika, heeft een nuttig steentje bijgedragen aan het vertaalwerk.

De paranimfen, Ron van Doorn en Lex Goudswaard, weten net als ik welke overeenkomsten er bestaan tussen promoveren en hard fietsen. We gingen kop over kop en hielden de vaart erin. Ron bleek uiteindelijk de eerste promovendus maar fietst intussen wat minder hard. Lex is nog steeds kampioen fietsen maar moet daarentegen nog promoveren.

Douwe de Ruiter was als collega-huisarts een welkome bron van afleiding in de dagelijkse praktijk. Zijn vrouw Florien waakte over mijn gezondheid en wees mij geregeld op de wallen onder mijn ogen. Mieke Heikoop en Margriet Koot hebben mijn huisartspraktijk uitstekend draaiende weten te houden tijdens mijn afwezigheid. De praktijkassistenten, Carla Meijer-Rüger en Esther de Bruin, dank

ik voor het gedogen van een baas die tussen de spreekuren door steeds weer vluchtte naar zijn computer thuis. Bovendien hebben ze me zeer geholpen met het ordenen van de literatuur.

Mijn ouders en schoonouders dank ik voor hun zorg en liefde voor onze kinderen op de vaste woensdagen, ziektedagen en andere dagen. Jullie zijn van een onschatbare waarde geweest.

Karen, Femke en Lotte, jullie zijn mijn allergrootste steun geweest. Zonder jullie liefde en levensvreugde had ik niet kunnen volharden in deze proeve van bekwaamheid. Ik heb veel van jullie gevergd. Na deze zware investering is het nu tijd om samen heel veel te genieten.

Curriculum Vitae

Jako Burgers werd geboren op 3 september 1962 in Haarlem en groeide op in Aalsmeer. Hij behaalde in 1980 het VWO diploma aan het Christelijk Atheneum 'Adriaen Pauw' te Heemstede. Na een jaar gewijd te hebben aan de muziek (piano), startte hij in 1981 met de studie geneeskunde aan de Vrije Universiteit te Amsterdam (VU). In 1985 slaagde hij 'cum laude' voor het doctoraal examen en in 1988 studeerde hij af als basisarts. Van 1987 tot 1989 studeerde hij filosofie aan de Centrale Interfaculteit, eveneens aan de VU. Na zijn artsexamen werkte hij als assistent-docent bij de vakgroep filosofie en medische ethiek en als assistent-coördinator van het algemeen coschap, beide verbonden aan de VU.

Van 1989 tot 1991 volgde hij de huisartsopleiding aan de VU. Daarna heeft hij tot 1995 als waarnemend huisarts in Amsterdam en omgeving gewerkt. In deze periode volgde hij ook muziekcolleges bij het Randstedelijk Muziektheoretisch Onderwijs te Amsterdam, hetgeen in 1994 resulteerde in het behalen van het staatsexamen muziek.

In 1995 vestigde hij zich als huisarts in Gorinchem. Het betrof een gesteunde vrije vestiging die in drie jaar tijd uitgroeide tot een normpraktijk. Van 1992 tot 2002 heeft hij als huisarts-stafid gewerkt bij de Afdeling Standaarden Ontwikkeling van het Nederlands Huisartsen Genootschap te Utrecht. Hij heeft de redactie gevoerd over twaalf NHG-Standaarden en maakte deel uit van de redactie van het herziene Standaardenboek dat uitkwam in 1999.

In 1998 startte hij met de werkzaamheden voor dit proefschrift bij de Centre for Quality of Care Research (WOK) aan het Universitair Medisch Centrum St Radboud te Nijmegen. Zijn onderzoek maakte deel uit van het AGREE (Appraisal Guidelines Research and Evaluation) project, waarin onderzoekers uit dertien landen betrokken waren. Sinds 2002 is hij als vertegenwoordiger van de WOK betrokken bij het vervolg van dit project en bij het realiseren van een internationaal netwerk voor richtlijnorganisaties.

Hij is gehuwd met Karen te Velde en trotse vader van twee dochters, Femke (6 jaar) en Lotte (4 jaar).

Jako Burgers was born in Haarlem on 3 September 1962 and grew up in Aalsmeer. He received his secondary school education at the Christelijk Atheneum 'Adriaen Pauw' in Heemstede and graduated in 1980. After one year lessons in music (piano), he started to study medical science at the Vrije Universiteit in Amsterdam (VU). He graduated with honours in 1985 and took his doctor's degree in 1988. From 1987 to 1989 he studied philosophy at the Centrale Interfaculteit, also at the VU. After his doctor's degree he worked as assistant reader at the Department of Philosophy and Medical Ethics and as assistant coordinator of the general wards, both attached to the VU.

From 1989 to 1991 he did postgraduate training for general practitioner at the VU. Then he worked as acting general practitioner in Amsterdam and surroundings. In this period he also attended music theory classes in Amsterdam and took his State degree in 1994. In 1995 he settled as a general practitioner in Gorinchem. From 1992 to 2002 he worked as staff member at the Department of Guideline Development of the Dutch College of General Practitioners (NHG) in Utrecht. He was the editor of twelve NHG Practice Guidelines and was part of the editorial staff of the second edition of the NHG Practice Guidelines book that was published in 1999.

In 1998 he started with this PhD project at the Centre for Quality of Care Research (WOK) of the University Medical Centre in Nijmegen. The research was part of the AGREE (Appraisal Guidelines Research and Evaluation) project, in which researchers from thirteen countries were involved. Since 2002 he represents the WOK in the AGREE follow-up project and is involved in the establishment of an international network for guideline organisations.

He is married with Karen te Velde and proud of two daughters, Femke (6) and Lotte (4).

Stellingen

behorende bij het proefschrift

“Quality of clinical practice guidelines”

van Jako Burgers

-
1. Evidence-based guidelines cannot be developed without professional consensus (this thesis).
 2. The AGREE Instrument is the first internationally-validated instrument for the appraisal of clinical guidelines (this thesis).
 3. The guideline organisation is the most important predictor of guideline quality (this thesis).
 4. Similarities and differences between clinical guidelines can only be partly explained by the supporting evidence (this thesis).
 5. Good guidelines are not only evidence based but are also usable in practice (this thesis).
 6. Development of guidelines can be considerably shortened by making use of existing guidelines (this thesis)
 7. Guideline development is human work.
Thomas S. British Journal of General Practice, 1994.
 8. Good guidelines can only make you better.
 9. Internet raises the danger that patients are also going to develop ‘medical student’s disease’.
 10. The beauty of those born with a facial deformity will be found in the beauty of their children.
 11. The Pythagorean comma shakes belief in a harmonic world.
 12. The closer you come to the truth, the more opinions diverge.

