Methodological Quality and Validity of Content of Clinical Practice Guidelines in Laboratory Medicine

Ph.D. Thesis

Éva Nagy, MD

2010

Department of Laboratory Medicine,
Albert Szent-Györgyi Clinical Centre, Faculty of Medicine
University of Szeged, Hungary

Methodological Quality and Validity of Content of Clinical Practice Guidelines in Laboratory Medicine

Ph.D. Thesis

Éva Nagy, MD

Supervisor:
Prof. Andrea Rita Horváth
MD, PhD, FRCPath, EurClinChem

2010

LIST OF PUBLICATIONS

Publications related to Ph.D. thesis

Papers

I. Horvath AR, ***Nagy E***, Watine J. Quality of Guidelines for the Laboratory Management of Diabetes Mellitus
Scand J Clin Lab Invest 2005;65 (Suppl 240):41-50 (IF:1.235)

II. Watine J, Friedberg B, ***Nagy E***, Onody R, Oosterhuis W, Bunting P, Charet J-C, Horvath AR. Conflict between guideline methodological quality and recommendation validity: a potential problem for practitioners
Clin Chem 2006;1(52): 65-72 (IF:4.803)

III. ***Nagy E***, Watine J, Bunting PS, Onody R, Oosterhuis WP, Rogic D, Sandberg S, Boda K, and Horvath AR. Do Guidelines for the Diagnosis and Monitoring of Diabetes Mellitus Fulfill the Criteria of Evidence-Based Guideline Development?
Clin Chem 2008;54(11):1872-1882 (IF:4.803)

Cumulative impact factor (ISI JCR 2007): 10,841

Book chapter

I. Horvath AR, ***Nagy E***, Watine J. Critical appraisal of guidelines. In: Evidence-based Laboratory Medicine: From Principles to Practice. Price CP, Christenson RH (eds) AACC Press, Washington. 2nd edition, 2007. 295-319 pp.

Abstracts

I. ***Nagy É***, Ónody R, Nyári T, Horváth A. Bizonyítékokon alapul-e a diabetes mellitus diagnosztikája?
Klinikai és Kísérletes Laboratóriumi Medicina 2004, 31 (1): 26.

II. ***Nagy E***, Onody R, Horvath AR. Are diabetes mellitus guidelines evidence-based?
Clinica Chimica Acta 2005; 355, S340 (IF:1.633)

III. Horvath AR, ***Nagy E***, Watine J, Bunting P, Onody R, Oosterhuis W, Rogic D, Sandberg S, Boda K. Quality of Guidelines for Diabetes Mellitus: Can We Do Better? Clinical Chemistry and Laboratory Medicine 2007; 45, Special Supplement, ppS1-s473 (IF:1.725)

Lectures

I. ***Nagy É***, Ónody R, Nyári T, Horváth A. Bizonyítékokon alapul-e a diabetes mellitus diagnosztikája?
Magyar Laboratóriumi Diagnosztikai Társaság 52. kongresszusa 2004.09.04. Sopron,

II. Horvath A R, ***Nagy E***. Quality of Guidelines for the Diagnosis of Diabetes Mellitus
10th Bergmeyer Conference: Diabetes Mellitus and Cardiovascular Disease. 14-16 March 2005, Grainau Garmisch-Partenkirchen, Germany

III. Horvath AR, ***Nagy E.*** Quality of Guidelines in Laboratory Medicine

XIXth International Congress of Clinical Chemistry, 24-28 July 2005, Orlando, FL, USA

IV. Horvath AR, ***Nagy E***. Quality of Guidelines for the Laboratory Management of Diabetes Mellitus.
3rd G-I-N Conference, 5-7 December 2005, Lyon, France

V. ***Nagy É***, Ónody R, Horváth A Megbízható bizonyítékokon alapulnak-e a nemzetközi diabetes mellitus irányelvek?
Magyar Laboratóriumi Diagnosztikai Társaság 53. kongresszusa 2006.09.02. Szeged,

VI. ***Nagy E***, Watine J, Bunting PS, Onody R, Oosterhuis WP, Rogic D, Sandberg S, Boda K, and Horvath AR. A nemzetközi diabetes mellitus irányelvek módszertani minősége. Magyar Tudomány Ünnepe, 2007. 11. 13. Szeged

VII. Horvath AR, ***Nagy E***, Watine J, Bunting P, Onody R, Oosterhuis W, Rogic D, Sandberg S, Boda K. Quality of Guidelines for Diabetes Mellitus: Can We Do Better?
17th IFCC-FESCC European Congress of Clinical Chemistry and Laboratory Medicine3-7 June 2007, Amsterdam, The Netherlands

Publications relevant to the topic of Ph.D. thesis

Papers

I. Kulier R, Hadley J, Weinbrenner S, Meyerrose B, Decsi T, Horvath AR, ***Nagy E***, Emparanza JI, Coppus S FPJ, Arvanitis TN, Burls A, Cabello JB, Kaczor M, Zanrei G, Pierer K, Stawiarz K, Kunz R, Mol BWJ and Khan KS. Harmonising Evidence-based medicine teaching: a study of the outcomes of e-learning in five European countries.
BMC Medical Education 2008, 8:27 http://www.biomedcentral.com/1472-6920/8/27

II. Kulier R, Hadley J, Coppus S FPJ, Zamora J, Hadley J, Malick S, Das K, Weinbrenner S, Meyerrose B, Decsi T, Horvath AR, ***Nagy E***, Emparanza JI, Arvanitis TN, Burls A, Cabello JB, Kaczor M, Zanrei G, Pierer K, Stawiarz K, Kunz R, Mol WJ B and Khan SK. The effectiveness of a clinically integrated e-learning course in evidence-based medicine: A cluster randomised controlled trial.
BMC Medical Education 2009; 9:21 http://www.biomedcentral.com/1472-6920/9/21

III. Kunz R, ***Nagy E***, Coppus SFPJ, Emparanza JI, Hadley J, Kulier R, Weinbrenner S, Arvanitis TNRT, Burls A, Cabello JB, Decsi T, Horvath AR, Walzak J, P. Kaczor PM, Zanrei GE, Pierer K, Schaffler R, Suter K, Mol WJB, Khan SK. How far did we get? How far to go? A European survey on postgraduate courses in evidence-based medicine
Journal of Evaluation in Clinical Practice 2009; 15(6):1196-1204 (IF:1.843)

IV. Hadley J, Kulier R, Zamora J, Coppus SFPJ, Weinbrenner S, Meyerrose B, Decsi T, Horvath AR, ***Nagy E***, Emparanza JI, Arvantis TN, Burls A, Cabello JB, Kaczor M, Zanrei G, Pierer K, Kunz R, Wilkie V, Wall D, Mol BWJ, Khan KS. Effectiveness of an e-learning course in evidence-based medicine for foundation (internship) training.
J R Soc Med 2010; 103: 288-294 (IF:1.402)

V. Kis E, Szegesdi I, Dobos E, ***Nagy E***, Boda K, Kemény L, Horvath AR. Quality assessment of clinical practice guidelines for adaptation in burn injury.
Burns. 2010; 36 (5): 606-15. (IF:1.950)

VI. Kis E, Dobos E, ***Nagy E***, Kemeny L, Horvath AR. A bizonyítékokon alapuló orvoslás gyakorlata.
Bőrgyógyászati és Venerologiai Szemle. 2010; 86(4):103-107.

Book chapters

I. Watine J, Oosterhuis WP, ***Nagy E***, Bunting PS, Horvath AR. Formulating and using evidence-based guidelines. In: Evidence-based Laboratory Medicine: From Principles to Practice. Price CP, Christenson RH (eds) AACC Press, Washington. 2nd edition, 2007. 275-294 pp.

II. Dobos Éva, ***Nagy Éva***, Horváth Andrea 8. fejezet A szakmai irányelvek fejlesztése In: A klinikai hatékonyság fejlesztése az egészségügyben. Szerk.: Gődény Sándor Budapest 2007. 427-464. old.

III. ***Nagy Éva***, Horváth Andrea 3. fejezet Bizonyítékokon alapuló laboratóriumi medicina In: Gyakorlati Laboratóriumi Medicina Szerk.: Debreczeni Lóránd, Kovács L. Gábor 2008. 2. (átdolgozott) kiadás 127-138. old.

Abstracts

I. Horvath AR, ***Nagy E***. Evidence for monitoring hypothyroid patients on thyroxine replacement .
Clinica Chimica Acta 2005; 355, S25 (Lecture Abstract) (IF:1.633)

II. Seres E, ***Nagy E***, Miko TL, Watine J, Horvath AR. The role of evidence based laboratory medicine in meeting ISO 15189:2003 standards for accreditation. Clinica Chimica Acta 2005; 355, S412 (Poster Abstract) (IF:1.633)

Lectures

I. Horvath AR., ***Nagy E***. Guidelines in Central-Eastern European countries. AGREE/G-I-N International Meeting, Edinburgh, Scotland, November 14-15, 2003

II. Horvath AR, ***Nagy E***. Evidence Based Healthcare
Central European Perspective International Symposium of the German Agency for Quality in Medicine: 10 Jahre evidenzbasierte Gesundheitsversorgung (ÄZQ) 3 March 2005, Berlin

III. Watine J, Friedberg B, Charet J-C, ***Nagy E***, Onody R, Horvath AR. Conflicting Guideline Recommendations: a Practitioner's Dilemma.

XIX International Congress on Clinical Chemistry/2005 AACC Annual Meeting. 24-28 July 2005, Orlando, USA

IV. Horvath AR, ***Nagy E***. Evidence for monitoring hypothyroid patients on thyroxine replacement.
16th IFCC-FESCC European Congress of Clinical Biochemistry and Laboratory Medicine – EUROMEDLAB 8-12 May 2005, Glasgow

V. Horvath AR, ***Nagy E***, Onody R. What is the Evidence Base for Biochemistry Testing?
28th Annual Conference of the Association of Clinical Biochemists in Ireland. 14-15 October 2005, Limerick, Ireland

VI. ***Nagy E***, Watine J, Horvath AR. Are Guidelines in Laboratory Medicine Evidence Based? How to Assess Quality of Guidelines?
Evidence-based Laboratory Medicine Conference of Polish Laboratory Medicine College 26 – 27 June 2009, Kraków, Poland

CONTENTS

1. ABBREVIATIONS

AACE: American Association of Clinical Endocrinologists

ADA: American Diabetes Association

AAFP: American Academy of Family Physicians

ACP: American College of Physicians

ACCP: American College of Chest Physicians

AGREE Instrument: the Appraisal of Guidelines for Research and Evaluation for Europe Instrument

ALP: Alkaline Phosphatase

ANDEM: Agence Nationale pour le Développement de l'Évaluation Médicale

ASCO: American Society of Clinical Oncology

ATS-ERS: American Thoracic Society and European Respiratory Society

BTS-SCG: British Thoracic Society and Society of Cardiothoracic Surgeons of Great Britain and Ireland

CEA: Carcinoembryonic Antigen

CDA: Canadian Diabetes Association

CPG: clinical practice guideline

D: Domain of AGREE Instrument

DM: Diabetes Mellitus

DNA: Deoxyribonucleic Acid

DS: Domain Score

EBM: Evidence-Based Medicine

EGTM: European Group on Tumour Markers

FNCLCC: Fédération Nationale des Centres de Lutte Contre le Cancer

GGT: Gamma-Glutamyl Transferase

GPAC: Guidelines and Protocols Advisory Committee

I: Item of AGREE Instrument

ICC: Interclass Correlations

IDF: International Diabetes Federation

IFCC: International Federation of Clinical Chemistry and Laboratory Medicine

Kaiser P: Kaiser Permanente

LD: Lactate Dehydrogenase

NAC: North America Conference

NACB: National Academy of Clinical Biochemistry

NICE: National Institute for Clinical Excellence

NHMRC: National Health and Medical Research Council

NHS: the National Health Service of the United Kingdom

NSCLC: Non-Small Cell Lung Cancer

NSE: Neuron-Specific Enolase

NZG: New-Zealand Guidelines Group

PRODIGY: The NHS Clinical Knowledge Summaries

SE: Standard Error

SEMDSA: Society for Endocrinology, Metabolism and Diabetes of South Africa.

SGOT: Serum Glutamic-Oxaloacetic Transaminase

SIGN: Scottish Intercollegiate Guidelines Network

SOGC: Society of Obstetricians and Gynaecologists of Canada

SPLF: Société de Pneumologie de Langue Française

TPA: Tissue Polypeptide Antigen

USPSTF: U.S. Preventive Services Task Force

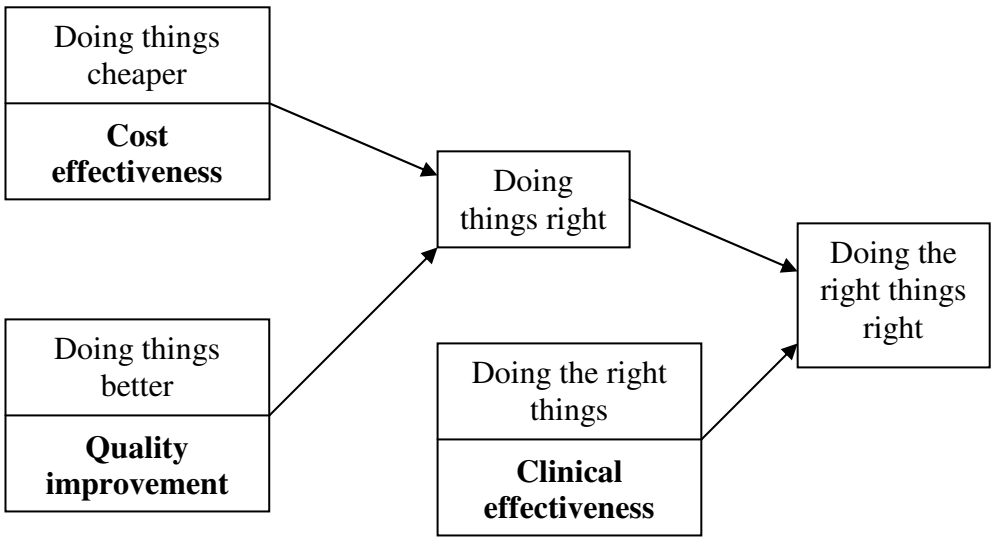
WHO: World Health Organisation

2. INTRODUCTION

The most commonly used definition of evidence-based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of the individual patients. The practice of evidence-based medicine (EBM) means integrating individual clinical expertise with the best available external clinical evidence from systematic research (1). Extending the application of the principles of evidence-based medicine to all professions associated with health care, including purchasing and management introduced a new term, the evidence-based health care.

The health care industry of the 21st century faces huge challenges worldwide, such as demographic explosion, increased incidence of chronic diseases, scientific and technological advancements, rapid development in information technology, as well as increasing expectations of society and growing personal responsibility for one's own health. Since resources are scarce costs and cost-effectiveness of health care have become top priority for society and governments all over the world (2). The science of evidence based medicine evolved in response to these challenges and expectations as shown in Figure 1 (3).

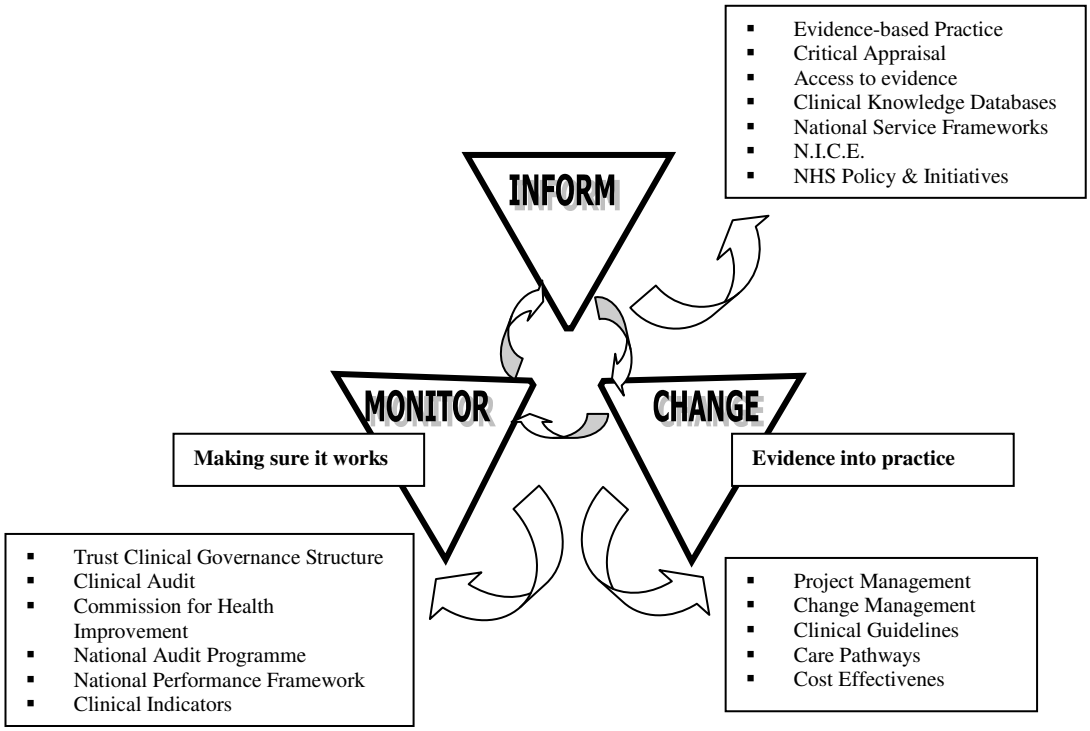
Figure 1 Evolution of evidence based health care (3)



In 1997 the Committee of Ministers of the Council of Europe recommended the evidence-based guidelines implementing the quality improving system (4). In 2001

they recommended a coherent and comprehensive national policy framework for the production, appraisal, updating and active dissemination of evidence-based clinical practice guidelines (CPGs). The main aim was to support and promote good clinical practice in the best interest of the patient and to improve the quality and effectiveness of health care (5). National Health Service of the United Kingdom (NHS), first in Europe, incorporated the evidence-based medicine methods into their strategy of clinical effectiveness (Figure 2) (6). They recognized that health policies and health care systems should be based on the best available evidence and clinical practice guidelines (CPGs) are useful tools for making more rational clinical decisions. The processes and outcomes of care should be regularly audited and should demonstrate that the delivery of care reflects adopted guidelines and protocols.

Figure 2 Clinical effectiveness strategy (6)



Clinical practice guidelines aim to for improve the quality of health care delivery and strengthen the position of the patient. According to the definition of the Institute of Medicine, clinical practice guidelines (CPGs) are systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances (7).

CPGs are developed to (7):

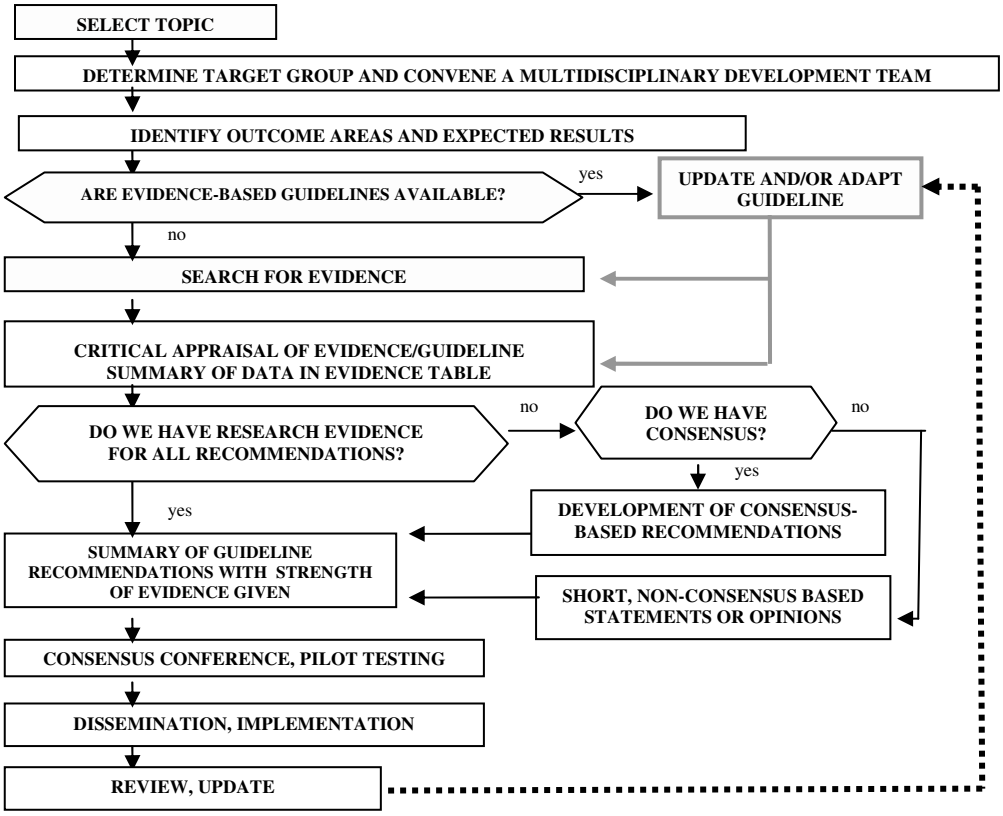
- disseminate best practice based on systematically appraised scientific evidence;
- decrease practice variation;
- improve the reliability of medical decisions by use of standardized criteria;
- improve quality of care and outcomes of patients;
- decrease harm to patients and the misfortune of professional misconduct and court cases;
- increase explicitness, transparency, patient information and autonomy of choice, thus facilitating ethical practice;
- facilitate training, education and continuous professional development;
- help target research to areas of uncertainty;
- inform policymakers, payers and managers;
- decrease costs and improve clinical and cost-effectiveness.

There is an exponentially rising interest toward CPGs in the medical literature and several organisations developed methodological manuals or so-called "Guidelines for guidelines". In 1997 an international initiative was set up by the European Community to compare the recommendations of guidelines covering the same topic area. Significant discrepancies were revealed in the process and reporting of the development of CPGs which led to the preparation of the AGREE (Appraisal of Guidelines for Research and Evaluation for Europe) Instrument (8) in 2000 (www.agreecollaboration.org). The AGREE Instrument is a standardized, generic and validated checklist (9-10) endorsed by WHO and the European Council for the critical appraisal of clinical practice guidelines. In 2001 the European Council initiated the harmonisation of health information and founded Guidelines International Network (11). A number of international organisations such as the WHO are also members but the numbers of organisational members and partners has grown to 86 representing 37 countries from all continents. Its aim is to improve the quality of health care by promoting systematic development of clinical practice guidelines and their application in clinical practice (http://www.g-i-n.net/).

We found only one relevant narrative review (12) in the literature, initiated by the Committee on Evidence-based Laboratory Medicine of the International

Federation of Clinical Chemistry and Laboratory Medicine (IFCC), which adapted methods of evidence-based guideline development to the field of laboratory medicine. This review provided an algorithm for the development process (Figure 3) and defined specific reporting standards related to the laboratory aspects of diagnostic recommendations (Table 1).

Figure 3 The process of evidence-based guideline development (12)



Due to these international efforts one would expect that the main steps of clinical practice guideline development are well harmonised and these methods would be implemented when diagnostic recommendations are developed in a multidisciplinary process and when high quality recommendations are adapted to national or local use.

Although methods for systematic reviewing of the literature and for the development of evidence-based recommendations, particularly in the field of therapeutics have been published and harmonised, the methodological quality of practice guidelines has been widely criticized (13-19). Most of these CPGs made therapeutic recommendations. Quality of CPGs in diagnostic fields and their impact in practice has been less well studied.

As effective treatment depends on effective diagnosis, diagnostic recommendations, based on poor quality evidence, may harm patients and contribute to inefficient health care delivery. Recommendations for the clinical use of tests, as diagnostic interventions, should therefore also fulfil the criteria of evidence-based guideline development. High methodological quality of CPGs requires that the potential biases of CPG development have been addressed adequately and that the recommendations, addressing both therapeutic and diagnostic interventions, are equally valid, and feasible in practice.

Table 1 Laboratory – related reporting standards by IFCC (12)

Preanalytical information • Prevalence of the conditions • When to request/not to request a test • Diagnostic algorithm • Patient preparation • Timing and frequency of testing • Sample type and handling of specimens • Biological variations
Postanalytical information • Medical decision limits • Laboratory-related outcomes • Diagnostic sensitivity, specificity • Predictive values • Interpretation of tests
Analytical information • Selections and validation of test methods • Detection limit, sensitivity, specificity • Imprecision, bias, quality goals • Standardization • Internal and external quality control • Interferences
Other information • Turnaround time • Where test is done (e.g. accreditation) • Quantifications and competence • Organizational and cost implications • Areas for further research

3. AIMS AND OBJECTIVES

Based on the known shortcomings of CPGs in therapeutic interventions, the key aims of my study were to investigate the methodological quality and content validity of CPGs primarily related to laboratory medicine. For objective assessments of the

quality of diagnostic CPGs we investigated the availability and applicability of generic guideline appraisal tools and methodologies to laboratory related recommendations. We investigated whether CPG development teams use appropriate and explicit methods for making diagnostic recommendations; and whether diagnostic CPGs meet basic reporting standards, or if not, what can be the reasons for such findings. We were also interested in differences between CPGs that are primarily diagnostic compared to those that are combined with therapeutic recommendations. We assessed whether there is a link between methodological rigour and quality and content validity of recommendation.

This thesis is a collection of work carried out over a period of just under 4 years.
For our aims we addressed the following key questions:

- Is there an easily applicable tool for the assessment of the quality diagnostic CPGs?
- What is the methodological quality of CPGs especially that of laboratory related recommendations?
- Are diagnostic and therapeutic CPGs different in their methodological quality?
- Is there any relationship between the characteristics and methodological quality of diagnostic CPGs?
- Do diagnostic CPGs meet basic reporting standards?
- Is there any correlation between methodological quality and validity of content of CPGs?

4. METHODS

4.1. Topic selection, search and selection strategy of clinical practice guidelines

For our investigations we have chosen two public health priority areas that have implications for laboratory medicine. One of them was the management of diabetes mellitus (DM). This global health care problem is facilitating many organisations and countries to improve and standardize the diagnosis and treatment of DM and to issue guidance on best practices. In this area there is a world-wide consensus and there are well-established and studied laboratory markers, such as

glucose, HbA1C, urinary albumin and protein for the diagnosis and monitoring of the condition. The other topic was related to oncology and focused on the management of non-small-cell-lung-cancer (NSCLC) patients. In this medical field there is more controversary about which laboratory test should be used for prognosing disease status. However existing systematic reviews (20-22) provided a good opportunity to compare the methodological quality and validity of content of recommendations on the management of NSCLC patients.

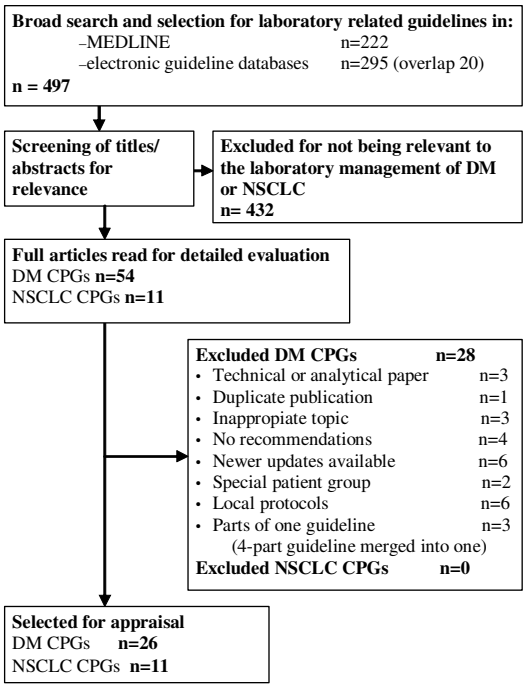
Systematic literature search was carried out to retrieve diagnostic CPGs. The aim of the search was to obtain a representative sample of CPGs, published in English between 1 January 1999 and 31 December 2007, that can be easily accessed and which are therefore likely to be read and used in many countries. We carried out manual and electronic searches for laboratory related guidelines in PubMed using broad search terms to capture as many relevant CPGs as possible (i.e. a high sensitivity search) strategy. In PubMed, one reviewer applied a broad search strategy, using the Clinical Queries filter "systematic[sb]", which is capable of retrieving systematic reviews and/or CPGs (23). This term was combined with the laboratory-specific MeSH terms of "Clinical Laboratory Techniques" [MeSH] AND systematic[sb] OR "Laboratory Techniques and Procedures" [MeSH] AND systematic[sb] (24). Another independent reviewer searched in electronic journals, using the keywords "guideline" AND "diabetes", and in dedicated GL databases and websites of professional organizations. The complete list of databases searched is shown in Table 2.

Then we restricted searching to CPGs for DM, published in English, and to CPGs for NSCLC published in English and French. Two independent reviewers applied the following inclusion criteria: the publication fulfilled the definition of CPGs (7) and dealt with the use of laboratory tests for the diagnosis or monitoring of DM and NSCLC; the CPGs were publicly available in a peer-reviewed journal; and/or in nationally- or internationally-endorsed CPG databases. If several updates of the CPG were available during the studied time period, only the latest version was selected (Figure 4).

Table 2 Databases used in searching for CPGs

Professional associations • The American Association of Clinical Endocrinologists (http://www.aace.com) • American College of Physicians (http://www.acponline.org) • Royal College of Physicians (http://www.rcplondon.ac.uk/index.asp) • American Academy of Family Physicians (http://www.aafp.org) • Canadian Diabetes Association (http://www.diabetes.ca) • American Diabetes Association (http://www.diabetes.org) • International Diabetes Federation (http://www.idf.org) • Alberta Medical Association (http://www.albertadoctors.org)
Electronic Guideline Databases • Clinical Practice Guidelines and Protocols in British Columbia (http://www.hlth.gov.bc.ca) • Scottish Intercollegiate Guidelines Network (http://www.sign.ac.uk) • Guidelines-International-Network (G.I.N) (http://www.g-i-n.net) • CDC Task Force on Community Preventive Services (http://www.thecommunityguide.org) • SCHARR database (http://www.shef.ac.uk/~scharr/ir/guidelin.html) • US National Guideline Clearing House (http://guideline.gov) • US Agency for Healthcare Research and Quality (http://www.ahrq.com) • The Canadian Task Force on Preventive Health Care (http://www.ctfphc.org) • German Agency for Quality in Medicine (http://www.aezq.de) • Guidelines Information Service (http://www.leitlinien.de) • New Zealand Guidelines Group (http://www.nzgg.org.nz) • Australian National Health and Medical Research Council http://www.health.gov.au/nhmrc/publications) • National Institute for Clinical Excellence (http://www.nice.org.uk) • Clinical Practice Guidelines (http://www.ogh.on.ca/library/cpg.htm)

Figure 4 Selection of CPGs



4.2. Evaluation of the methodological quality of clinical practice guidelines

4.2.1. Appraisal tool and its applicability to diagnostic guidelines (Paper I)

We chose the AGREE Instrument (8) a standardized, generic and validated checklist for the evaluation of the methodological quality of CPGs. We used this instrument with its accompanying Training Manual (8). The Instrument consists of

23 criteria (Items) (Table 3). These criteria are grouped into 6 domains (D) which assess scope and purpose, stakeholder involvement, rigor of development, clarity and presentation, applicability and editorial independence of CPGs. More than 2 reviewers need to assess the fulfilment of the AGREE criteria independently on a 4-point Likert scale. The assessors had formal training

in using the AGREE Instrument both at local and international workshop, and took part in

Table 3 The 23 items and 6 domains of AGREE Instrument (8)

Scope and purpose	
Item 1	The overall objective(s) of the guideline is (are) specifically described.
Item 2	The clinical question(s) covered by the guideline is (are) specifically described.
Item 3	The patients to whom the guideline is meant to apply are specifically described.
Stakeholder involvement	
Item 4	The guideline development group includes individuals from all relevant professional groups.
Item 5	The patients' views and preferences have been sought.
Item 6	The target users of the guideline are clearly defined.
Item 7	The guideline has been piloted among target users.
Rigour of development	
Item 8	Systematic methods were used to search for evidence.
Item 9	The criteria for selecting the evidence are clearly described.
Item 10	The methods used for formulating the recommendations are clearly defined.
Item 11	The health benefits, side effects and risks have been considered in formulating the recommendations.
Item 12	There is an explicit link between the recommendations and the supporting evidence.
Item 13	The guideline has been externally reviewed by experts prior to its publication.
Item 14	A procedure for updating the guideline is provided.
Clarity and presentation	
Item 15	The recommendations are specific and unambiguous.
Item 16	The different options for management of the condition are clearly presented.
Item 17	The key recommendations are easily identifiable.
Item 18	The guideline is supported with tools for application.
Application	
Item 19	The potential organisational barriers in applying the recommendations have been discussed.
Item 20	The possible cost implications of applying the recommendations have been considered.
Item 21	The guideline presents key review criteria for monitoring and/or audit purposes.
Editorial independence	
Item 22	The guideline is editorially independent from the funding body.
Item 23	Conflicts of interest of guideline development members have been recorded.

a pilot study to assess consistency of their ratings before the larger DM and NSCLC studies were carried out. Domain scores (DS) were calculated from the individual item scores of each assessors and expressed in percentages. Finally, all assessors had to select one out of 4 options (“strongly recommended”, “recommended with provisos or alterations”, “would not recommended”, “unsure”) judging the overall performance of the CPG and whether it can be recommended for use in practice.

In order to test the applicability of AGREE Instrument to diagnostic CPGs and to pilot test the use of this appraisal tool, we selected 4 most commonly cited and

used primarily diagnostic CPGs for DM. Each CPG was independently evaluated by seven assessors, and aggregated scores were calculated for each domain. We assessed the agreement between reviewers by statistical methods. Cronbach's alpha was calculated using standardized domain scores of assessors to measure internal consistency in each domain. Interclass correlations (ICC) were calculated to assess reliability within each domain.

We also tried to create a statistical null-based order for rating methodological quality of CPGs based on their final domain scores (DS). The four guidelines were compared using one-way ANOVA and ANOVA using repeated measurements. The level of significance was defined at $p<0.05$.

4.2.2. Method of appraisal of diabetes mellitus and non-small cell lung cancer guidelines using the AGREE Instrument

Each CPG was appraised by 4 trained assessors. We used the AGREE Instrument as described (8), but based on our results of pilot study we applied more stringent criteria for handling disagreements between assessors. We did not calculate aggregated domain scores automatically without comparing the individual item scores of assessors. If the difference in item scores between assessors was more than 2, the disagreements were resolved by discussion and consensus.

We appraised the methodological quality of NSCLC CPGs by the AGREE Instrument as described above except for changing the overall assessment terminology of AGREE ("strongly recommended", "recommended with provisos or alterations", "would not recommended", "unsure") to "very good", "good", "not so good", or "dubious" because we thought that this would lead to an easier understanding of the relation of methodological quality and content validity.

4.3. Statistical methods

4.3.1. Correlation between the characteristics and methodological quality of clinical practice guidelines

In order to assess any correlation we created subgrouped CPGs based on their source, scope, length, origin and whether they were supplemented with a guideline methods manual. We also investigated the quality of guidelines according to the date and type of publication. In the statistical analyses, the mean item and standardized domain scores of CPG subgroups were compared by the Kruskal -Wallis test. The

level of significance was set at $p \leq 0.01$ because of multiple comparisons. All analyses were performed using SPSS for Windows, version 13.

Furthermore, in the DM CPG study we investigated whether the CPG contained information that reflecting on evidence-based guideline methodology but was not covered in depth by the AGREE Instrument. This included questions whether the CPG contained 1) an evidence table, 2) a description of the grading system, 3) graded recommendations, 4) an expiry or review date? All reviewers checked the availability of these additional items and results were summarized by one independent assessor. We collated data in a table and used descriptive statistics (relative frequency).

4.3.2. Evaluation of differences between primarily diagnostic and combined clinical practice guidelines

In order to evaluate if methodological quality is dependent on the scope of CPGs, we assessed the 26 DM CPGs by scope. We also made an overview of published studies which used to AGREE Instrument irrespective of the scope of guidelines.

We created two subgroups of DM CPGs based on their scope (i.e. diagnostic if the CPG was primarily covering the laboratory diagnostic aspects of the condition, or combined if the CPG had both diagnostic and therapeutic recommendations). For investigating difference between “purely diagnostic” and “combined” CPGs in depth pair-wise comparisons were carried out using the Mann-Whitney U test with Bonferroni correction. The level of significance was set at $p \leq 0.01$ because of multiple comparisons. All analyses were performed using SPSS for Windows, version 13.

4.3.3. Evaluation of the fulfilment of basic diagnostic reporting standards

For investigating whether diagnostic CPGs meet additional reporting standards (25) that are not covered in depth by AGREE, we assessed the presence of the following items thought to be important for the correct use and interpretation of test results: 1) prevalence, 2) diagnostic accuracy of tests, 3) pre-analytical, and 4) analytical specifications. All reviewers checked the availability of these items and results were summarized by one independent assessor.

The frequency of reporting specific laboratory information in different guideline subgroups was compared with the Fisher's exact test. The level of

significance was set at $p \leq 0.01$ because of multiple comparisons. All analyses were performed using SPSS for Windows, version 13.

4.4. Systematic reviewing techniques to compare methodological quality in other medical fields

To compare the methodological quality of CPGs across all medical subspecialties, we systematically reviewed the literature that used the AGREE Instrument for such evaluation and compared our findings. We searched electronically in Medline in May 2007 with the following key word combinations: ((("Guideline "[Publication Type] OR "Guidelines as Topic"[Mesh] OR "Guideline Adherence"[Mesh] OR "Practice Guideline "[Publication Type]) AND quality) AND AGREE) without using any language limits. Data on the topics, origin, number and publication dates as well as the AGREE domain scores of each study and collected and presented in a summary table and a diagram.

4.5. Methods of the evaluation of relationship between methodological quality and validity of content of guidelines

For linking methodological quality to content validity two assessors extracted all laboratory-related recommendations from the 11 guidelines selected for review. Validity of recommendations was investigated regarding the use of tumor markers and other more global laboratory tests in NSCLC, based on a systematic review of the evidence (20), which was updated in 2002 and 2004 (21-22). Methodological quality was assessed as described before with slightly modified expression of scores (i.e. very good, good, not so good, or dubious).

5. RESULTS

5.1. Applicability of the AGREE Instrument to diagnostic guidelines (Paper I)

Our team (7 laboratory specialists) assessed four most well-known and widely used CPGs for the diagnosis and monitoring of DM. This exercise was carried out to test the applicability of the AGREE Instrument to diagnostic CPGs and as a training set for using this appraisal tool and investigated inter-rater variability (26-29). Item scores of the 7 assessors are shown in Table 4. The calculated domain scores and the overall assessments are presented in Table 5 with statistical data about the level of agreement of appraisers and their ratings of CPGs.

The agreement between assessors was acceptable based on statistical calculations (ICC: 0.68-0.91, Cronbach's alpha: 0.63-0.90), except for the 5th domain (ICC: 0.02, Cronbach's alpha: 0.23; for explanation see text below) (Table 5). We have noted some discrepancy between statistical judgements of agreement and the comparison of each item score of each appraiser. In some cases (e.g. for WHO in D1) in spite of the acceptable results of ICC and Cronbach's alpha there was a relevant difference in individual scores of appraisers in the same domain ranging from 11% to 78% and in the same item (from 1 to 4 scores) (Table 4). Such kind of disagreement was not acceptable in our study design therefore we decided not to use the calculation of ICC and Cronbach's alpha in subsequent analysis, but rather we reach consensus for each item where disagreement is grater than 2 scores/item.

When CPGs were rated according to domain scores with pair-wise comparisons by ANOVA (Table 5), the NICE guideline had a significantly higher quality rating ($p=0.003$) than the WHO guideline. A notable (but not significant) difference ($p=0.060$) was observed in the quality of the ADA guideline, which however was not significantly better ($p=0.097$) than the NACB recommendations. In spite of ANOVA calculating similar ranks for the NICE and NACB CPGs, by showing no significant differences between them, their overall assessments, based on AGREE Instrument, were very different (strongly recommend *vs*. recommend with alteration). Therefore the appraisers reached a consensus that we would hereafter use only the overall assessment method of the AGREE Instrument for characterizing the acceptance of the methodological quality of CPG, rather than the mentioned statistical methods.

Only the NICE 2001 (27) of the 4 widely-used DM CPGs was strongly recommended by assessors to use in practice, although its applicability scores were still low. The NACB 2002 (28), and the ADA 2003 (29) guidelines had some shortcomings, especially in the rigour of their development and stakeholder involvement domains therefore the appraisers recommended further improvements of these documents. Surprisingly the most well-known WHO 1999 (26) CPG had the most serious shortcomings in all appraised aspects, thus assessors did not recommend this CPG for use.

Assessors had judged that the AGREE Instrument is a useful tool and is applicable for the general assessment of methodological quality of CPGs in laboratory medicine as well.

Table 4 Item and domain scores of four Diabetes Mellitus CPGs of the pilot study.

		WHO 1999 (26)							NACB 2002 (27)							NICE 2001 (28)							ADA 2003 (29)						
		Assessors							Assessors							Assessors							Assessors						
	Item	A1	A2	A3	A4	A5	A6	A7	A1	A2	A3	A4	A5	A6	A7	A1	A2	A3	A4	A5	A6	A7	A1	A2	A3	A4	A5	A6	A7
Scope and purpose	1	3	3	4	1	1	1	2	3	4	4	2	2	2	3	4	4	4	4	3	4	4	3	4	4	3	3	4	3
	2	3	2	3	3	1	2	1	3	3	2	3	1	1	2	4	4	3	4	4	4	4	3	3	3	3	2	1	3
	3	3	3	3	2	2	3	1	3	4	3	3	1	4	1	4	3	4	4	4	4	4	3	3	3	4	3	2	3
Stakeholder involvements	4	3	2	1	2	1	1	1	3	4	3	1	3	3	3	3	3	4	4	4	4	4	1	3	1	2	1	1	1
	5	1	2	1	1	1	1	1	1	2	1	1	1	1	1	4	4	4	4	4	4	4	2	2	1	1	1	1	1
	6	2	2	1	1	1	1	1	3	3	4	3	1	2	1	4	4	4	4	4	4	4	4	3	4	4	3	2	3
	7	2	2	1	1	1	1	1	1	3	1	1	1	1	1	3	3	3	4	4	4	4	1	3	1	1	1	1	1
Rigour of developments	8	1	2	1	1	1	1	1	1	3	1	1	1	1	1	4	4	4	3	4	4	4	1	3	1	1	3	1	3
	9	1	2	1	1	1	1	1	1	3	1	3	1	3	2	2	4	3	3	4	4	4	1	3	1	3	1	1	1
	10	1	2	1	1	1	1	1	2	3	2	1	1	3	2	3	4	3	4	4	4	4	2	3	1	1	1	1	2
	11	1	3	2	1	1	1	1	1	3	2	1	3	4	3	2	4	4	4	3	4	3	2	3	3	2	3	4	4
	12	1	2	1	3	1	1	1	2	3	3	3	3	4	3	3	4	3	4	4	4	4	2	2	2	4	3	1	3
	13	2	2	1	1	1	1	1	2	2	3	4	4	3	4	3	3	2	4	4	4	4	1	2	1	1	1	1	1
	14	1	2	1	1	1	1	1	1	2	1	1	3	1	3	2	3	3	4	1	4	2	1	2	1	2	4	1	4
Clarity and presentation	15	3	2	3	3	3	4	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	4	4	4	4	4
	16	2	2	3	3	3	2	2	3	3	3	3	3	1	4	4	3	3	4	4	4	4	3	2	3	4	3	4	4
	17	2	2	3	1	2	3	1	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	4	4	4	4	3	4
	18	1	2	1	1	1	1	1	2	3	3	1	4	1	4	2	3	3	3	4	3	4	1	3	1	1	3	1	4
Applicability	19	2	3	1	1	1	1	1	2	3	1	3	1	1	1	2	4	1	1	1	1	1	1	3	2	1	1	4	1
	20	2	2	1	1	1	1	1	2	2	1	1	2	4	2	1	3	2	1	1	1	1	1	3	1	1	1	1	1
	21	1	2	1	3	1	1	1	1	3	1	4	1	1	2	1	3	4	4	4	4	3	1	2	1	4	1	2	2
Editorial Independence	22	1	3	2	2	3	2	3	1	3	2	1	1	4	2	1	4	3	2	4	4	4	1	3	1	1	1	1	1
	23	1	2	1	1	1	1	1	1	2	1	1	1	1	1	1	3	1	1	1	4	1	1	2	1	1	1	1	1

Table 5 Inter-rater agreement and rating quality of CPGs based on pair-wise comparison of domain scores in the DM pilot study

CPG and date of issue (ref)	Domain Score (%)						Overall assessment	Pair-ways comparison (P value)	Rank
	Scope and purpose	Stakeholder involvement	Rigor of development	Clarity and presentation	Applicability	Editorial independence			
WHO 1999 (26)	41	11	8	38	13	24	Not recommend	**0,03***	**3**
NICE 2001 (27)	95	94	82	90	36	52	Strongly recommend	**-**	**1**
NACB 2002 (28)	52	31	43	76	28	19	Recommend with alteration	**0,097**	**1**$^{+}$
ADA 2003 (29)	70	30	32	73	22	7	Recommend with alteration	**0,060****	**2**
Mean Domain Score (%)	70	39	45	63	28	35			
ICC	**0.78**	**0.91**	**0.86**	**0.76**	**0.02†**	**0.68**			
Cronbach's α	**0.78**	**0.88**	**0.90**	**0.77**	**0.23†**	**0.63**			

* significant (p≤0,05); ** notable difference (but not significant); †: not acceptable;
$^{+}$ difference is not statistically significant

5.2. Methodological quality of clinical practice guidelines

5.2.1. Diabetes mellitus (Paper III)

After the pilot study, we evaluated 26 CPGs which contained recommendations for the laboratory management of DM (30-55). The item and domain scores of individual CPGs and their overall assessments are shown in Table 6 and Table 7, respectively. Mean item scores and the proportion of CPGs scoring above 3 on the 4-point Likert scale are summarized and highlight the most common shortcomings of CPGs on DM in Table 6.

Based on the assessment of methodological quality, 22 CPGs were recommended by reviewers, of which only 11 were strongly recommended and the rest "with provisos and alterations". Four CPGs had 4 or 5 domains with scores <30%, thus reviewers did not recommend their use.

Overall, the best performing domains were D1 "Scope and purpose" (Table 7) with a high proportion of CPGs scoring above 3 for all items (Table 6). Although D4 "Clarity and presentation" scored highly (Table 7), item 18 (I18) within this domain performed poorly, as only 10 CPGs (38%) were supported with tools for application (Table 6).

In D3, which explored the rigour of development, 14 (54%) CPGs scored higher than 60% (Table 7). There are notable shortcomings in using systematic methods for searching the evidence and providing information on the literature retrieval and selection process (I8, I9); indicating the methods used for formulating recommendations (I10); and giving information on the peer reviewing (I13) and updating process (I14) (Table 6).

Domain 2, which explored stakeholder involvement, showed lower scores (Table 7). Nine CPGs (35%) scored higher than 60%. When investigating the scores of individual items in D2, only a small proportion of CPGs gave information about the composition and affiliations of the guideline development group (I4); provided some information on patient involvement in the development process (I5); defined their target users clearly (I6); and pilot tested the CPG by target users before publication (I7) (Table 6).

Low scores were achieved with in the "Applicability" (34%) and "Editorial independence" (39%) domains, in which each item performed very poorly (I19-I23).

We compared the minimum and maximum scores of each individual domain (Table 7) in order to evaluate heterogeneity in the variations of quality of these CPGs. The wide spread of these data in all domains (D1: 14-100%, D2: 6-88%, D3:6-92%, D4: 33-98%, D5:0-72%, D6: 0-100%) demonstrated unexpectedly large variation in CPG.

Table 6 Mean AGREE Item scores for 26 DM CPGs

CPG and date of issue (ref)	Items																						
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**
	Scope and purpose			**Stakeholder involvement**				**Rigour of development**							**Clarity and presentation**				**Applicability**			**Ed. indep.**	
AAFP 1999 (30)	3.75	3.75	3.50	2.50	2.50	2.50	1.25	4.00	3.75	2.50	3.50	3.50	3.25	1.00	2.75	2.50	1.75	1.25	1.75	1.75	1.00	2.25	1.00
SIGN 2001 (31)	2.75	1.75	3.50	3.75	4.00	2.50	2.75	3.25	2.75	3.25	2.50	3.50	3.75	3.50	4.00	3.50	3.75	1.25	1.25	1.25	1.25	3.25	3.00
NAC 2002 (32)	2.75	2.50	2.00	1.50	1.00	1.25	1.00	1.00	1.00	1.75	3.00	1.00	1.00	1.75	2.50	2.00	2.50	1.00	2.25	1.00	1.00	1.25	1.00
NACB 2002 (33)	3.25	2.50	2.00	2.25	1.00	2.50	1.00	1.25	1.25	1.25	2.50	3.00	3.00	1.25	3.25	3.25	4.00	1.50	1.50	1.50	1.00	2.00	1.00
NICE BG 2002 (34)	3.75	3.50	4.00	3.75	3.50	4.00	3.00	3.50	3.75	3.50	3.50	4.00	3.00	4.00	4.00	3.75	4.00	4.00	1.00	1.00	4.00	3.50	1.00
NICE L 2002 (35)	3.75	3.50	4.00	3.75	3.75	4.00	3.00	3.50	3.75	3.75	3.50	4.00	3.50	4.00	4.00	3.75	4.00	4.00	1.00	1.00	4.00	3.50	1.00
SEMDSA 2002 (36)	1.00	1.00	2.25	1.00	1.00	3.75	1.00	1.00	1.00	1.00	2.25	1.00	1.00	1.00	3.75	2.50	3.50	1.00	1.00	1.00	1.00	1.00	1.00
SOGC 2002 (37)	3.75	3.25	3.25	1.75	1.25	2.25	1.25	2.00	1.50	2.25	3.75	3.75	1.25	1.00	4.00	3.50	4.00	1.25	1.00	2.50	1.25	1.75	1.00
CDA 2003 (38)	3.75	3.50	3.50	3.00	1.50	2.50	1.00	2.00	3.75	3.25	3.25	3.75	2.50	1.00	3.50	3.25	4.00	4.00	2.25	2.00	1.00	3.50	1.00
NZG 2003 (39)	2.75	4.00	4.00	3.75	3.75	4.00	2.50	2.50	1.75	3.25	3.50	4.00	4.00	4.00	4.00	3.75	4.00	3.75	2.50	2.25	3.25	4.00	4.00
USPSTF T2 2003 (40)	4.00	3.75	4.00	2.50	1.25	1.75	1.00	4.00	3.75	2.75	4.00	4.00	3.50	1.25	3.75	3.75	4.00	3.25	1.50	3.75	1.25	3.75	3.50
USPSTF GDM 2003 (41)	3.75	3.75	4.00	2.50	1.25	1.75	1.25	4.00	4.00	2.25	4.00	3.75	3.00	1.50	4.00	3.75	4.00	2.00	1.50	3.75	1.50	3.50	3.50
WHO T2 2003 (42)	4.00	4.00	4.00	2.75	1.25	3.00	1.00	1.75	1.00	1.50	4.00	2.50	1.00	1.25	2.25	2.75	4.00	1.25	3.25	3.75	1.25	2.25	2.25
ACP 2004 (43)	3.75	4.00	4.00	1.50	1.00	3.75	1.00	3.50	3.00	1.25	3.75	3.50	2.50	3.00	3.75	3.50	4.00	2.25	1.25	1.00	1.25	3.75	2.75
Kaiser P 2004 (44)	1.50	1.50	3.75	1.25	1.25	3.75	1.00	1.25	1.25	1.50	1.25	1.00	1.00	1.00	3.50	3.50	3.75	1.00	1.00	1.00	1.00	1.00	1.00
NICE T1 2004 (45)	4.00	3.75	4.00	3.75	4.00	3.75	3.00	3.75	3.50	3.75	3.75	3.75	4.00	3.75	4.00	3.75	4.00	4.00	1.50	4.00	4.00	3.75	3.75
GPAC 2005 (46)	3.75	2.50	3.25	1.50	1.50	3.50	1.75	1.00	1.00	1.00	2.50	1.25	1.00	2.00	3.50	2.75	4.00	4.00	3.50	2.50	3.25	2.75	1.00
IDF T2 2005 (47)	2.75	2.50	3.00	2.50	2.50	1.25	3.25	2.25	2.25	2.75	2.50	2.75	4.00	2.00	3.50	3.75	4.00	2.25	3.50	2.50	1.00	3.75	4.00
NHMRC 2005 (48)	4.00	4.00	3.75	3.75	3.75	4.00	1.25	4.00	4.00	3.75	3.25	3.75	4.00	3.25	4.00	3.75	4.00	2.00	2.00	3.50	1.00	2.25	1.00
WHO DG 2006 (49)	3.75	3.50	2.75	1.25	1.00	2.50	1.00	1.50	1.00	1.25	2.75	3.50	1.50	1.00	3.75	3.00	4.00	1.50	1.75	2.25	1.25	1.50	1.75
AACE 2007 (50)	3.50	2.75	2.50	2.75	1.50	3.75	1.00	1.75	1.00	2.50	3.25	3.00	2.75	1.00	3.50	3.50	3.75	1.50	1.75	1.75	1.00	1.00	3.75
ADA 2007 (51)	3.00	2.50	3.00	2.00	1.50	3.25	1.00	1.00	1.00	1.75	2.50	4.00	2.25	2.75	3.75	4.00	4.00	3.25	3.75	1.75	1.00	1.00	1.00
IDF BG 2007 (52)	3.50	3.25	4.00	2.25	1.00	2.50	1.50	2.75	3.50	2.25	2.50	2.75	2.75	2.00	2.50	3.25	4.00	1.50	1.75	2.25	1.50	4.00	3.50
PRODIGY L 2007 (53)	3.75	4.00	4.00	3.00	3.00	3.50	3.00	2.75	2.75	2.00	3.50	3.25	2.25	4.00	3.75	3.50	3.50	4.00	1.25	2.75	4.00	2.25	1.00
PRODIGY R 2007 (54)	3.75	4.00	4.00	3.00	3.00	3.50	2.75	3.00	2.75	2.00	3.50	3.25	2.50	4.00	3.75	3.50	3.50	3.75	1.25	2.75	4.00	2.25	1.00
PRODIGY BG 2007 (55)	2.75	3.25	3.75	2.50	3.00	4.00	3.00	2.75	3.00	1.75	3.50	3.50	2.75	3.75	3.75	3.75	3.50	2.75	2.75	3.00	3.75	2.75	1.00
Mean Item Score	**3.34**	**3.16**	**3.45**	**2.53**	**2.12**	**3.03**	**1.75**	**2.50**	**2.42**	**2.30**	**3.14**	**3.12**	**2.58**	**2.31**	**3.57**	**3.37**	**3.75**	**2.43**	**1.88**	**2.21**	**1.95**	**2.60**	**1.95**
No with score ≥3	**19**	**17**	**21**	**9**	**9**	**15**	**6**	**10**	**11**	**7**	**17**	**19**	**11**	**10**	**22**	**21**	**24**	**10**	**4**	**6**	**8**	**11**	**8**
Rate with score ≥3 (%)	**73**	**65**	**81**	**35**	**35**	**57**	**23**	**38**	**42**	**27**	**65**	**73**	**42**	**38**	**85**	**81**	**93**	**38**	**15**	**23**	**31**	**42**	**31**

Table 7 Critical appraisal of diabetes mellitus guidelines by the AGREE Instrument

CPG and date of issue (ref)	Domain Score (%)						Overall assessment
	Scope and Purpose	Stakeholder involvement	Rigour of development	Clarity and Presentation	Applicability	Editorial indep.	
AAFP 1999 (30)	89	40	69	35	17	21	Recommend with alteration
SIGN 2001 (31)	56	75	74	71	8	71	Strongly recommend
NAC 2002 (32)	47	6	17	33	14	4	Would not recommend
NACB 2002 (33)	53	23	31	67	11	17	Recommend with alteration
NICE BG 2002 (34)	92	85	87	98	33	42	Strongly recommend
NICE L 2002 (35)	92	88	90	98	33	42	Strongly recommend
SEMDSA 2002 (36)	14	23	6	56	0	0	Would not recommend
SOGC 2002 (37)	81	21	40	73	19	13	Recommend with alteration
CDA 2003 (38)	86	33	60	90	25	42	Recommend with alteration
NZG 2003 (39)	86	83	76	96	56	100	Strongly recommend
USPSTF T2 2003 (40)	97	21	77	90	39	88	Strongly recommend
USPSTF GDM 2003 (41)	94	23	74	81	42	83	Strongly recommend
WHO T2 2003 (42)	100	33	29	52	58	42	Recommend with alteration
ACP 2004 (43)	97	27	64	79	6	75	Recommend with alteration
Kaiser P 2004 (44)	42	27	6	65	0	0	Would not recommend
NICE T1 2004 (45)	97	88	92	98	72	92	Strongly recommend
GPAC 2005 (46)	72	35	13	85	69	29	Recommend with alteration
IDF T2 2005 (47)	58	46	55	79	44	96	Recommend with alteration
NHMRC 2005 (48)	97	73	90	81	39	21	Strongly recommend
WHO DG 2006 (49)	78	15	26	69	25	21	Wouldn't recommend
AACE 2007 (50)	64	42	39	69	17	46	Recommend with alteration
ADA 2007 (51)	61	31	39	92	39	0	Recommend with alteration
IDF BG 2007 (52)	86	27	55	60	28	9	Recommend with alteration
PRODIGY L 2007 (53)	97	71	64	90	56	21	Strongly recommend
PRODIGY R 2007 (54)	97	69	67	88	56	21	Strongly recommend
PRODIGY BG 2007 (55)	75	71	67	81	72	29	Strongly recommend
Mean Domain Score (%)	**77**	**45**	**54**	**76**	**34**	**39**	
Range (%)	14-100	6-88	6-92	33-98	0-72	0-100	
No of CPGs with DS more than 60%	20	9	15	22	3	7	
Percentage of CPGs with DS more than 60%	77	35	58	85	11	27	

5.2.2. Non-small cell lung cancer (Paper II)

We appraised 11 selected CPGs (56-66) on NSCLC management by the AGREE Instrument. The mean domain score and overall assessment results are shown in Table 8.

Only 5 out of 11 CPGs (45%) were recommended for use by assessors and none achieved the best overall quality rating of„strongly recommend". Three CPGs (27%) were „wouldn't recommend" and for 3 CPGs the quality was difficult to assess („unsure") by appraisers.

Overall, the best performing domain was D1 “Scope and purpose” and D4 “Clarity and presentation” in which domains 8 and 5 of 11CPGs scored more than 60%, respectively.

There were notable shortcomings in other domains which explored the rigour of development, stakeholder involvement and editorial independence. The mean domain scores of these domains reached only about 30% because only 1 or 4 of 11 CPGs scored higher than 60%.

Domain 5, which explored the applicability of recommendations, showed the lowest scores. None of 11 CPGs scored higher than 60%.

In all domains, except Domain 5, there was a large of scores. The relatively narrow dispersion of results in this domain (0-33%) demonstrated that most CPGs had no tools, or additional information that would aid the application of recommendations in practice.

Table 8 Critical appraisal of NSCLC guidelines by the AGREE Instrument

CPG (ref)	Domain score (%)						Overall assessments
	Scope and purpose	Stakeholder involvement	Rigour of development	Clarity and presentation	Applicability	Editorial dependence	
ACCP (56)	61	46	60	46	6	75	Recommend with alterations
ANDEM (57)	89	25	10	71	0	25	Not recommend
ASCO (58)	94	50	71	67	17	75	Recommend with alterations
ATS-ERS (59)	44	4	5	29	0	8	Unsure
BTS-SCG (60)	100	33	60	79	6	83	Recommend with alterations
CIGNA (61)	67	13	12	54	11	8	Unsure
EGTM (62)	44	4	2	29	0	0	Unsure
FNCLCC (63)	94	54	57	79	17	33	Recommend with alterations
NACB (64)	50	17	29	54	11	25	Not recommend
SIGN (65)	89	75	76	75	33	25	Recommend with alterations
SPLF (66)	61	46	48	38	17	8	Not recommend
Mean Domain Score (%)	**72**	**33**	**39**	**56**	**11**	**33**	
Range (%)	44-100	4-75	2-76	29-79	0-33	0-83	
No of CPGs with DS more than 60%	8	1	4	5	0	3	
Percentage of CPGs with DS more than 60%	73	9	36	45	0	27	

5.3. Causes of poor methodological quality of diabetes mellitus guidelines

5.3.1. Correlation between the characteristics and methodological quality clinical practive guidelines (Paper III)

In order to explore the probable reasons for the observed methodological shortcomings and the large variations in quality of the 26 selected DM CPGs, we investigated the most important characteristics of CPGs (Table 9). Based on these features, we subgrouped CPGs and assessed the relationship of these characteristics to methodological quality.

Date of publication

Most CPGs were developed after 2002 and only 2 were developed between 1999 and 2001. The quality of DM CPGs was also investigated according to the date of publication in order to see whether any improvement can be observed over time. Only the highest scoring D1 and D4 showed some marginal development in quality over the time scale investigated (Table 7). CPGs seemed to have become more specific in stating their objectives (I1), in creating more focused clinical questions (I2), and the recommendations in CPGs have become more easily identifiable (I17) (Table 6). However, the poor performance in D6 showed further deterioration from 2005 onwards with failures to report editorial independence and conflict of interest in the majority of CPGs (Table 6).

Type of publication

We investigated how CPG developers defined the type of their publications and whether these reflected the methods used for their development. There was diversity in definitions: 19 publications were labelled as CPGs or recommendations, of which 7 stated that they were evidence-based, 4 were position statements or reports, and 3 guidance documents (Table 9). Amongst the 7 CPGs that claimed to be evidence-based 5 had evidence summaries and 6 graded their recommendations. Three CPGs that had evidence tables, however, did not define their publications as being evidence-based (40, 41 and 43). Over two thirds of CPGs (n=18, 69%) defined their grading system but only 16 (62%) graded their final recommendations (Table 9).

Table 9 Characteristics of DM CPGs

CPG (ref)	Date of issue	Source	Scope	Length (pages)	Guideline manual*	Origin	Type of publication as described by authors	Evidence table	Description of grading system	Graded recommendations	Review date (year)
AAFP (30)	1999	Both	Diagnostic	>100	no	USA	review of the evidence and recommendations	+	-	-	-
SIGN (31)	2001	Database	Combined	51-100	yes	UK	national clinical guidelines	-	+	+	3
NAC (32)	2002	Journal	Combined	1-10	no	North America	consensus report	-	-	-	1
NACB (33)	2002	Both	Diagnostic	11-50	no	USA	guidelines and recommendations	-	+	+	-
NICE BG (34)	2002	Database	Combined	>100	yes	UK	clinical guidelines and evidence review	+	+	+	4
NICE L (35)	2002	Database	Combined	>100	yes	UK	clinical guidelines and evidence review	+	+	+	4
SEMDSA (36)	2002	Database	Combined	1-10	no	South Africa	guideline	-	-	-	-
SOGC (37)	2002	Both	Diagnostic	1-10	no	Canada	clinical practice guidelines	-	+	+	-
CDA (38)	2003	Database	Combined	>100	no	Canada	clinical practice guidelines	-	+	+	-
NZG (39)	2003	Database	Combined	>100	yes	New Zealand	evidence-based best practice guidelines	-	+	+	3
USPSTF T2 (40)	2003	Database	Diagnostic	51-100	yes	USA	recommendation and rationale statement	+	+	+	-
USPSTF GDM (41)	2003	Database	Diagnostic	>100	yes	USA	recommendation and rationale statement	+	+	+	-
WHO T2 (42)	2003	Database	Diagnostic	51-100	yes	International	report	-	-	-	-
ACP (43)	2004	Both	Combined	1-10	yes	USA	clinical practice guidelines	+	-	-	5
Kaiser P (44)	2004	Database	Combined	1-10	no	USA, Canada	guidelines	-	-	-	-
NICE T1 (45)	2004	Database	Combined	>100	yes	UK	clinical guidelines and evidence review	+	+	+	4
GPAC (46)	2005	Database	Diagnostic	11-50	yes	Canada	guidelines and protocols	-	-	-	3
IDF T2 (47)	2005	Database	Combined	51-100	yes	International	global guideline	-	+	-	3-5
NHMRC (48)	2005	Database	Diagnostic	>100	yes	Australia	evidence based guidelines	+	+	+	3
WHO DG (49)	2006	Database	Diagnostic	51-100	yes	International	report	-	-	-	-
AACE (50)	2007	Both	Combined	>100	yes	USA	medical guidelines (evidence based)	-	+	+	-
ADA (51)	2007	Both	Combined	11-50	yes	USA	position statement	-	+	+	1**
IDF BG (52)	2007	Database	Diagnostic	0-50	yes	International	guideline	-	+	+	3
PRODIGY L (53)	2007	Database	Combined	51-100	yes	UK	guidance	-	+	+	Con.
PRODIGY R (54)	2007	Database	Combined	51-100	yes	UK	guidance	-	+	-	Con.
PRODIGY BG (55)	2007	Database	Combined	51-100	yes	UK	guidance	-	+	+	Con.
Percentage of CPGs fulfilling criteria								**31**	**69**	**62**	**58**

**: Guideline development manual or technical document was available before CPG publication.*
***: Information on updating is provided in a separate guideline development manual.*
Con.: Continuos

Procedure for updating guidelines

Item14 of the AGREE Instrument investigates whether CPG developers describe the procedures for updating recommendations, including the timescale, responsibilities and methods used. Fifteen CPGs (58%) gave a timescale or expiry date, of which one CPG provided this information in a separate CPG development manual of the issuing authority (Table 9). The most frequent review date was 3 and 4 years. Only 10 CPGs (38%) provided adequate information on the updating process (Table 9).

Subgroup analysis

CPGs were grouped according to source, scope, length, origin and availability of a guideline methods manual, to investigate whether there are statistically significant differences in CPG quality in these subsets. Results are shown in Table 10.

Sub-grouping by source

Grouping CPGs by source of publication revealed that one CPG was published in a peer-reviewed journal, 19 were available in electronic CPG databases and 6 in both sources. The CPG that was published exclusively in a peer-reviewed journal (32) was not recommended for use by the assessors. None of the 6 CPGs published both in peer-reviewed journals and CPG databases were strongly recommended. CPGs published in electronic guideline databases only, received a more favourable overall assessment. Notable difference, at a level of significance of $p \leq 0.05$, could be observed in the D5 Applicability domain only for the electronic CPGs (Table 10).

Sub-grouping by length

A clear relationship could be demonstrated between CPG length and methodological quality (Table-10). Most CPGs that were not recommended were shorter and all strongly recommended guidelines were longer than 50 pages. Significant differences between these subgroups could be found for most domains with higher quality of the longer CPGs. Moderate differences ($p \leq 0.05$) could be observed with the "Applicability" and "Clarity and presentation" domains. However, the best performing CPGs, scoring >50% in the "Applicability" domain (39, 40, 45, 46, 53-55) were generally longer than 50 pages and all were published in electronic databases (Table 7 and 9)

Sub-grouping by origin

Nine CPGs originated from the USA, 3 from Canada, 7 from the UK, one from Australia, New-Zealand and South Africa and 4 were international. The majority of the strongly recommended CPGs (7 out of 11) originated from the UK; the other four were from New Zealand, Australia and the USA (Table 10). Significant differences ($p \leq 0.01$) could be observed in fulfilling the criteria of the D2 "Stakeholder involvement" domain, with higher scores for the British CPGs. In the "Rigour of development", and "Clarity and presentation" domains the difference was moderate ($p \leq 0.05$) (Table 10).

Sub-grouping by the availability of guideline methods manual

Two thirds of CPGs had some accompanying manuals describing the methods of their development in some form. All strongly recommended CPGs had such a manual (Table 9 and 10). All mean domain scores were better in the subset where these manuals were available, and the differences were statistically highly significant ($p \leq 0.01$) in the D4, D5 and D6 domains. In D1, D2 and D3 domains the p values were somewhat greater than 0.01.

5.3.2. Methodological quality of primarily diagnostic and combined clinical practice guidelines (Paper III)

We investigated whether the scope of the guideline (i.e. diagnostic vs. combined diagnostic and therapeutic) correlates with methodological quality.

The rate of occurrence of strongly recommended CPGs was higher for the combined (50%), than for the diagnostic CPGs (30%), but the rate of CPGs not recommended was also higher in the combined group. The quality of purely diagnostic CPGs was not significantly different from that of combined CPGs (Table 10). The difference was moderate ($p \leq 0.05$) in the D2 domain only, with combined CPGs scoring higher (Table 10). Moderate differences were also found in four individual items (Table 11). Diagnostic CPGs defined their objectives better (I1) and considered the cost implications of the recommendations more frequently (I20), while combined CPGs defined their target users (I6) and their updating processes more precisely (I14) than diagnostic ones.

Table 10 Subgroup analysis

		Domains																Strongly recommend	Recommend with alteration	Wouldn't recommend
		Scope and purpose			Stakeholder involvement			Rigor of development			Clarity and presentation			Applicability			Editorial independence					
		DS (%)	SE	Range	DS (%)	SE	Range	DS (%)	SE	Range	DS (%)	SE	Range	DS (%)	SE	Range	DS (%)	SE	Range	No (%)	No (%)	No (%)
Source	Guideline database (n=19)	**80**	5.2	14-100	**52**	6.1	15-88	**58**	6.5	6-92	**80**	3.3	52-98	**40**	5.1	0-72	**45**	7.6	0-100	**11 (58)**	**5 (26)**	**3 (16)**
	Journal and GL database (n=7) [a]	**70**	7.1	47-97	**27**	4.6	6-42	**43**	6.8	17-69	**64**	8.3	33-92	**18**	3.9	6-39	**25**	10.0	0-75	**0 (0)**	**6 (86)**	**1 (14)**
	P=	***0.209***			***0.055***			***0.169***			***0.083***			***0.018*****			***0.152***					
Scope	Diagnostic (n=10)	**85**	4.5	53-100	**31**	5.2	15-73	**50**	8.2	13-90	**69**	5.3	35-90	**35**	5.8	11-69	**34**	8.9	9-88	**3 (30)**	**6 (60)**	**1 (10)**
	Combined (n=16)	**73**	6.2	14-97	**54**	6.8	6-88	**56**	6.9	6-92	**80**	4.5	33-98	**33**	6.1	0-72	**43**	8.8	0-100	**8 (50)**	**5 (31)**	**3 (19)**
	P=	***0.286***			***0.023*****			***0.660***			***0.097***			***0.776***			***0.551***					
Length	1-50 pages (n=9)	**61**	8.5	14-97	**24**	2.7	6-35	**30**	7.0	6-64	**68**	5.8	33-92	**21**	7.4	0-69	**16**	8.0	0-75	**0 (0)**	**6 (67)**	**3 (33)**
	>50 pages (n=17)	**86**	3.5	56-100	**56**	6.2	15-88	**67**	4.8	26-92	**80**	4.1	35-98	**41**	4.6	8-72	**52**	7.2	21-100	**11 (65)**	**5 (29)**	**1 (6)**
	P=	***0.009****			***0.003****			***0.001****			***0.051***			***0.018*****			***0.001****					
Origin	North America (n=12)	**74**	5.7	42-97	**27**	2.8	6-42	**44**	7.1	6-77	**72**	5.7	33-92	**25**	5.5	0-69	**35**	9.3	0-88	**2 (17)**	**8 (66)**	**2 (17)**
	British (n= 7)	**87**	5.9	56-97	**62**	3.2	69-88	**67**	4.5	64-92	**82**	3.8	71-98	**42**	8.9	0-72	**44**	10.1	0-92	**7 (100)**	**0 (0)**	**0 (0)**
	Other (n= 7)	**74**	11.3	14-100	**43**	9.8	15-83	**48**	11.2	6-90	**70**	5.9	52-96	**36**	7.6	0-58	**41**	15.4	0-100	**2 (28.5)**	**3 (43)**	**2 (28.5)**
	P=	***0.355***			***0.001****			***0.028*****			***0.037*****			***0.112***			***0.606***					
Manual	yes (n=19)	**84**	3.5	56-100	**53**	5.9	15-88	**62**	5.3	52-98	**82**	3.0	50-98	**42**	4.5	6-72	**49**	7.4	0-100	**11 (58)**	**7 (37)**	**1 (5)**
	no (n=7)	**59**	10.5	14-89	**25**	4.0	6-40	**33**	9.5	6-69	**60**	7.7	33-90	**12**	3.6	0-25	**14**	5.6	0-42	**0 (0)**	**4 (57)**	**3 (43)**
	P=	***0.013*****			***0.015*****			***0.022*****			***0.010****			***0.001****			***0.004****					

*[a] One guideline (32) was published in journal only; * p≤0.01; **p≤0.0*

Table 11 AGREE item scores in diagnostic and combined diabetes mellitus guidelines

	AGREE ITEMS	Diagnostic CPG's n=10		Combined CPG's n=16		P
		Mean Score	SE	Mean Score	SE	
Scope and purpose	1 The overall objective of the guideline is specifically described.	3.75	0.2	3.08	0.2	0.023**
	2 The clinical questions covered by the guideline are specifically described.	3.43	0.2	3.00	0.2	0.421
	3 The patients to whom the guideline is meant to apply are specifically described.	3.45	0.2	3.45	0.2	0.897
Stakeholder involvement	4 The guideline development team involves all relevant professional groups.	2.30	0.2	2.67	0.2	0.241
	5 The patients' views and preferences have been sought.	1.58	0.3	2.45	0.3	0.060
	6 The target users of the guideline are clearly defined.	2.63	0.2	3.28	0.2	0.047**
	7 The guideline has been piloted among target users.	1.23	0.2	2.08	0.2	0.220
Rigour of development	8 Systematic methods were used to search for evidence.	2.63	0.2	2.42	0.2	0.551
	9 The criteria for selecting the evidence are clearly described.	2.48	0.3	2.39	0.3	0.660
	10 The methods used for formulating the recommendations are clearly defined.	2.08	0.2	2.44	0.2	0.391
	11 The health benefits, side effects and risks have been considered.	3.28	0.2	3.06	0.2	0.391
	12 There is an explicit link between the recommendations and the supporting evidence.	3.18	0.3	3.08	0.3	0.938
	13 The guideline has been externally reviewed by experts prior to its publication.	2.43	0.2	2.67	0.2	0.856
	14 A procedure for updating the guideline is provided.	1.55	0.3	2.78	0.3	0.041**
Clarity and presentation	15 The recommendations are specific and unambiguous.	3.38	0.1	3.69	0.1	0.363
	16 The different options for management of the condition are clearly presented.	3.23	0.1	3.45	0.1	0.182
	17 The recommendations are easily identifiable.	3.78	0.1	3.73	0.1	0.150
	18 The guideline is supported with tools for application.	1.95	0.3	2.73	0.3	0.220
Applicability	19 The potential barriers in applying the recommendations have been discussed.	1.95	0.2	1.83	0.2	0.452
	20 The potential cost implications of applying the recommendations have been considered.	2.75	0.2	1.88	0.2	0.041**
	21 The guideline presents key review criteria for monitoring and/or audit purposes.	1.43	0.3	2.28	0.3	0.623
Editorial independence	22 The guideline is editorially independent from the funding body.	2.60	0.3	2.59	0.3	1.000
	23 Conflicts of interest of guideline development members have been recorded.	1.95	0.3	1.95	0.3	0.979

***$p \leq 0.05$*

5.3.3. Compliance of guidelines with basic diagnostic reporting standards (Paper III)

We investigated whether CPGs covered essential laboratory-specific information, such as prevalence/pre-test probability and diagnostic accuracy data or pre-analytical and analytical factors critical for the correct interpretation and application of laboratory results in clinical practice (Table 12). Only about 60 percent of the CPGs mentioned these issues in any detail.

Table 12 Qualitative analysis of reporting laboratory specific information in diabetes mellitus guidelines

		Percentage of guideli nes fulfilling criteria			
		Prevalence	**Diagnostic accuracy**	**Preanalytical information**	**Analytical information**
Source	Guideline database (n=19)	58	58	63	63
	Journal and database (n=7) [a]	57	57	57	43
	P=	*0.973*	*0.973*	*0.780*	*0.407*
Scope	Diagnostic (n=10)	70	80	70	50
	Combined (n=16)	50	44	56	63
	P=	*0.428*	*0.109*	*0.683*	*0.689*
Length	0-50 pages (n=9)	33	33	44	33
	>50 pages (n= 17)	71	71	71	71
	P=	*0.103*	*0.103*	*0.234*	*0.103*
Manual	Yes (n=19)	63	58	68	63
	No (n=7)	43	57	43	43
	P=	*0.407*	*0.973*	*0.369*	*0.407*
Origin	North America (n=12)	58	58	67	50
	British (n=7)	43	57	57	86
	Other (n=7)	71	57	57	43
	P=	*0.556*	*0.998*	*0.884*	*0.205*

[a] = *one guideline (32) published in journal only*

We looked for any correlation between reporting these laboratory specific aspects and the 5 predefined subgroup parameters (scope, source, length, manual, origin). Reporting these pieces of information was more frequent in diagnostic as compared to combined CPGs, but the difference was not statistically significant in the various CPG subgroups (Table 13).

5.4. Methodological quality of clinical practice guidelines in other medical fields (Book chapter I)

In order to investigate whether our findings regarding the quality of DM and NSCLC guidelines are generalizable, we systematically reviewed the relevant literature the assessment of the methodological quality of CPGs in other medical fields. We found 21 studies up till 2007which, have investigated the quality of CPGs using the AGREE

Instrument in various clinical fields. These 21 studies included our 3 published studies, too. The collected data and calculated mean domain scores of each study are shown in Table 14

Table 13 Diabetes mellitus guidelines reporting laboratory-specific criteria

CPG and date of issue (ref)	Prevalence / Pre-test probability	Diagnostic accuracy	Preanalytical information	Analytical information
AAFP 1999 (30)	-	+	-	-
SIGN 2001 (31)	+	+	-	+
NAC 2002 (32)	-	-	-	-
NACB 2002 (33)	+	+	+	+
NICE BG 2002 (34)	-	+	-	+
NICE L 2002 (35)	-	-	-	-
SEMDSA 2002 (36)	-	-	-	-
SOGC 2002 (37)	+	+	+	+
CDA 2003 (38)	+	+	+	+
NZG 2003 (39)	+	-	+	+
USPSTF T2 2003 (40)	+	+	+	+
USPSTF GDM 2003 (41)	+	+	+	+
WHO T2 2003 (42)	+	+	-	-
ACP 2004 (43)	-	-	-	-
Kaiser P 2004 (44)	-	-	-	-
NICE T1 2004 (45)	-	+	+	+
GPAC 2005 (46)	-	-	+	-
IDF T2 2005 (47)	+	+	+	+
NHMRC 2005 (48)	+	+	+	+
WHO DG 2006 (49)	+	+	+	-
AACE 2007 (50)	+	-	+	-
ADA 2007 (51)	+	+	+	+
IDF BG 2007 (52)	-	-	-	-
PRODIGY L 2007 (53)	-	-	+	+
PRODIGY R 2007 (54)	+	+	+	+
PRODIGY BG 2007 (55)	+	-	+	+
Percentage of CPGs fulfilling criteria	**58**	**58**	**62**	**58**

The altogether 712 CPGs appraised in the 21 reviewed studies had recommendations in diverse medical fields. None assessed the diagnostic or therapeutic nature of CPGs, but majority of these CPGs were predominantly therapeutic. The relatively high number of CPGs and the wide spectra of origin enabled us to compare our finding to independent assessments of the overall methodological quality of international CPGs and to draw some generalizable conductions.

The maximum and minimum values of mean domain scores from the 21 studies are presented in a diagram in conjunction with our own results for the DM and NSCLC CPGs (Figure 5). These data show that in most guidelines the scope and purpose of recommendations (D1) are clearly defined and guidance is given in a clear format (D4). There are serious shortcomings, however, in the multidisciplinary nature of guideline teams, including patient involvement (D2) and in the rigour (or reporting) of an evidence-based CPG methodology (D3). Guidelines often fail the criteria of editorial independence,

Table 14 Studies investigating the methodological quality of CPGs by the AGREE Instrument (Book chapter I)

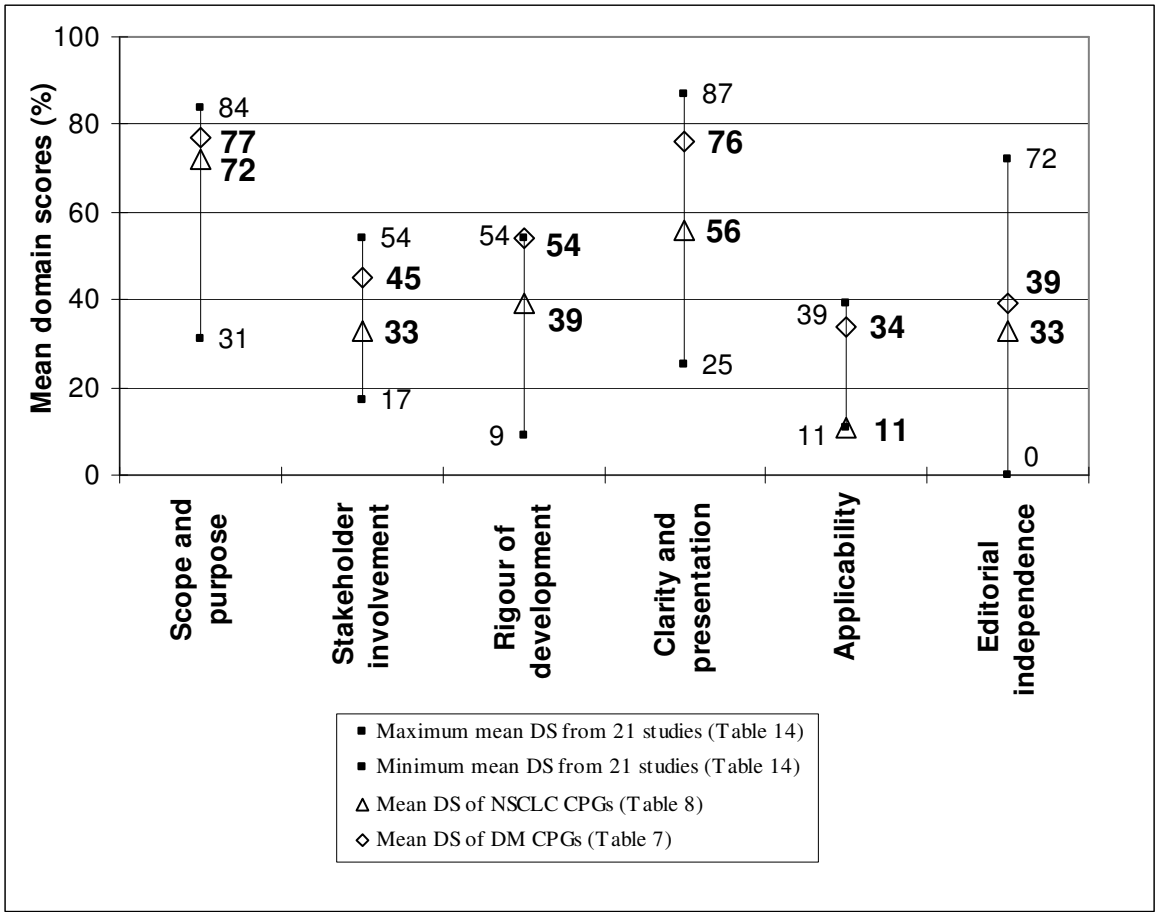
Study (ref)	Topics covered by the CPGs	Publication date of CPGs	Origin of CPGs	No. of CPG	Mean Domain Scores (%)						Comments
					D1	D2	D3	D4	D5	D6	
AGREE 2003 (9)	mixed	1992-1999	international*	33	69	36	41	66	37	30	*:CPG origin : 10 from European countries and 1 from Canada
Burgers *et al.* 2003 (67)	mixed	1992-1999	international*	86	66	34	37	57	31	48	*: CPG origin :10 European countries and Canada (62 deifferent agencies and organisations)
Harpole *et al*. 2003 (17)	lung cancer	1989-2001	international	51	72	35	52	57	20	24	
Brosseau *et al.* 2004 (68)	musculosceletal physiotherapy	1998-2002	French or English	9	64	54	49	60	29	24	
Burgers *et al.* 2004 (18)	non-oncolgy	1992-1999	international	68	65	30	29	52	30	41	
van Tulder *et al*. 2004 (19)	acute low back pain in primary care	1987-2001	international*	17	79	50	52	76	28	28	*CPG origin:: 4 USA, 3 Canada, 1 UK, 1 Israel, 2 Netherlands, 1 Germany, 1 Sweden, 1 New Zealand, 1 Finland, 1 Switzerland, 1 Denmark,
Burgers *et al.* 2004 (18) **Fervers *et al.* 2005 (69)**	oncology	1992-1999	international*	32	63	34	42	57	26	47	*CPG origin: 13 countries
Boluyt *et al*. 2005 (70)	pediatrics	1990-2005	mostly North American *	17	84	42	54	78	19	40	*CPG origin: 13 US A, 3 Canada, 1 Scotland
Horvath *et al*. 2005 (PaperI)	diagnosis of DM	1990-2003	international*	4	64	41	41	69	25	25	*CPG origin: 2 USA, 1 UK, 1 WHO
Lindberg *et al*. 2005 (71)	Swedish CPG on diabetes mellitus	not stated	local*	1	78 (73)	30 (67)	14 (37)	61 (72)	31 (60)	72 (60)	* Östergötland county; percentages in parentheses represent evaluation by lay persons
Lindberg *et al*. 2005 (71)	Swedish CPG on asthma /allergy	not stated	local*	1	50 (67)	25 (25)	9 (13)	71 (56)	26 (30)	8 (0)	* Östergötland county; percentages in parentheses represent evaluation by lay persons
Navarro Puerto *et al*. 2005 (72)	Spanish CPGs	1999-2002	national	61	31	18	18	25	13	38	Domain scores are an approximation of the true mean, reconstructed from the available original data.
Presztoczki *et al*. 2005 (73)	Hungarian CPGs on management of DM	1993-2004	national	9	78	17	12	54	39	0	Thesis, unpublished
Stiegler *et al.* 2005 (74)	psychiatric treatment	1998-2003	international*	61	33	31	48	71	23	20	*CPGs origin: 14 European countries
Arnau *et al.* 2006 (75)	diagnosis and treatment of low back pain	1994-2002	international	17*	63	38	32	53	21	22	*11 guidelines common with the study by van Tulder *et al.* (19) 2004
Cates *et al.* 2006 (76)	occupational health	2004	USA	1	80	46	27	87	31	29	
Ministry of Health, Hungary 2006 (77)	Hungarian care pathway protocols	2005-2006	national	180	72	30	28	74	37	8	Internal evaluation for the Ministry of Health; unpublished
Vervey *et al.* 2006 (78)	suicide attempts	1995-2005*	local in the Netherlands	27	43	22	12	65	15	**	*starting date is not precisely defined; **not used in this study for being considered irrelevant
Watine *et al.* 2006 (Paper III)	laboratory tests in lung cancer	1997-2003	international*	11	72	33	39	56	11	33	*CPG origin: 5 USA, 3 France, 2 UK, 1 EU
Nagy *et al.* 2007 (Lecture VII)	diagnosis and monitoring of DM	1999-2005	international*	26	74	41	50	70	27	35	*CPG origin: 13 USA, 3 Canada 6 UK, 1 Australia, 1 New Zealand, 1 South Africa, 1 WHO (unpublished)
TOTAL/MEAN				**712**	**65**	**34**	**34**	**63**	**26**	**30**	

D1: Scope and purpose; D2: Stakeholder involvement; D3: Rigour of development; D4: Clarity and presentation; D5: Application, D6: Editorial independence.

i.e. reporting on funding and conflicts of interest (D6), and most recommendations lack external validity, *i.e.* applicability in practice (D5).

The heterogeneity is very large (Figure 5) between these 21 studies because individual domain scores span a wide range for each domain (D1: 31- 81%, D2: 17-54%, D3:9-54%, D4: 25-87%, D5:11-39%, D6: 0-72%). Our results for CPGs related to laboratory medicine were similar to there international findings (Figure 5). Despite the heterogeneity of the published data, the major shortcomings were very similar in each study and domain. The only notable difference in our finding was that the NSCLC CPGs reached lower scores in each domain, than the DM CPGs.

Figure 5 Fulfilment of AGREE criteria of CPGs based on mean domain scores of 21 studies in different medical fields.



5.5. Correlation between guideline methodological quality and validity of content (Paper II)

To assess the relationship between quality and content of CPGs the methodological quality of 11 NSCLC CPGs we collected the recommendations about the use of laboratory tests in these CPGs (Table 15) and information from existing systematic reviews (20-22) on the laboratory parameters to be measured during the pre-treatment evaluation of NSCLC patients (Table 16).

Table 15 Recommendations of eleven CPGs for use of laboratory tests in the pretreatment management of NSCLC patients.

CPG (ref)	Recommended	Unclear recommendation	Not recommended
ACCP (56)	Hematocrit, ALP, calcium, electrolytes, glucose, GGT, SGOT	Other routine laboratory tests	None
ANDEM (57)	Leucocyte count, albumin, SR, calcium, ALP, LD	None	Tumor markers
ASCO (58)	Hemoglobin, leucocyte counts, LD, ALP, calcium	Other routine chemistries, liver function tests	LASA, CA 19-9, DNA index, DNA flow cytometric proliferation analysis, p53 tumor supressor gene, as oncogene
ATS-ERS (59)	Blood counts, electrolytes, albumin, calcium, ALP, transaminases, bilirubin, creatinine	None	Tumor markers
BTS-SCG (60)	Albumin, creatinine, glucose	None	None
CIGNA (61) *	None	None	CEA, NSE, cyfra 21-1
EGTM (62) *	cyfra 21-1, CEA**	CA 125, TPA	None
FNCLCC (63)	Hemoglobin, leucocyte counts with differential, LD, albumin, calcium	None	Tumor markers
NACB (64) *	None	cyfra 21-1, CEA, NSE	None
SIGN (65)	ALP, calcium	Other biochemistry and hematology tets, liver function tets	None
SPLF (66) *	None	cyfra 21-1	CEA

**: CPGs intended for tumor markers only*
***: Only in cases of adenocarcinoma or large cell carcinoma*

Recent systematic reviews provide no evidence that measurement of tumor markers in routine practice would improve NSCLC patients' outcomes (20-22). Only 4 CPGs (57, ANDEM), (59, ATS-ERS), (61, CIGNA), (63, FNCLCC), which did not recommend the use of tumor markers, were scored for validity of content as "good". Most CPGs which either recommended clearly (62, EGTM) or gave unclear recommendations the use of tumor markers (64, NACB) (66, SPLF) or did not mention tumor markers at all (56, ACCP) (58, ASCO) (60, BTS-SCG) (65, SIGN) were scored as "not so good" for validity of content.

Table 16 Laboratory variables that should be measured for the pre-treatment evaluation of NSCLC patients based on previously published systematic reviews

Purpose of test	Variables to be measured
Evaluation of toxicity (or tolerance) to treatments	**In all patients:** hemoglobin, leucocyte counts with differential, platelets, electrolytes, glucose, creatinine, transaminases, bilirubin, albumin
Pretreatment prognostic evaluation	**In all patients:** hemoglobin (if radiation therapy), leucocyte counts with differential, lactate dehydrogenase, albumin, calcium **In patients participating in therapeutic trials:** hemoglobin, leukocyte counts with differential, lactate dehydrogenase, albumin, calcium, NSE

Regarding the other laboratory tests, only 5 CPGs (56, ACCP) (57, ANDEM) (58, ASCO) (59, ATS-ERS) (63, FNCLCC) recommended clearly most of the laboratory tests which were found to be useful by previously published systematic reviews. These CPGs were scored "good" for validity of content about laboratory tests. Six CPGs out of 11 either did not mention any other laboratory tests (65, CIGNS) (62, EGTM) (64, NACB) (66, SPLF) or did not recommend enough laboratory tests (60, BTS-SCG) or did not recommend these clearly (65, SIGN).

Results of our comparison of guideline quality *versus* content for each guideline are shown in Table 17. For ease of interpretation we expressed CPG quality as "very good", "good", "not so good" or "dubious" instead of the original terminology of the AGREE Instrument (see Method Section).

Table 17 Correlation between methodological quality and validity of content of NSCLC CPGs

CPGs (ref)	Methodological quality	Validity of content of recommendation	
		Tumor markers	Other laboratory tests
ACCP (56)	Good	Not so good	Good
ANDEM (57)*	Not so good	Good	Good
ASCO (58)	Good	Not so good	Good
ATS-ERS (59)*	Dubious	Good	Good
BTS-SCG (60)*	Good	Not so good	Not so good
CIGNA (61)*	Dubious	Good	
EGTM (62)*	Dubious	Not so good	
FNCLCC (63)	Good	Good	Good
NACB (64)*	Not so good	Not so good	
SIGN (65)	Good	Not so good	Not so good
SPLF (66)*	Not so good	Not so good	

**: diagnostic CPG only*

The methodological quality of those CPGs which were good for validity of content, i.e. did not recommend tumor markers in routine practice, is very heterogeneous (2 “dubious”, 1 “not so good”, 1 “good”). Similar results were found for other laboratory tests: 5 CPGs of valid content reached a wide range of scores for quality (1 “dubious”, 1 “not so good”, 3 “good”). Only in one CPG out of 11 (63, FNCLCC) was there correlation between methodological quality and content validity for the use of laboratory tests, especially of tumor markers.

We did not find any relationship between the quality and validity of content and scope of CPGs (containing only diagnostic or therapeutic recommendations, as well). Our results did not confirm any relationship between the date of publications and the scores of quality or the validity of content.

6. DISCUSSION

6.1. AGREE Instrument as a critical appraisal tool for diagnostic clinical practice guidelines

Findings of our pilot study confirmed that the AGREE Instrument is an easy-to-learn and easy-to-use critical appraisal tool for assessing methodological quality of CPGs and is applicable to laboratory related CPGs as well. Our observations highlighted the need for at least 2 but preferably 3 or 4 independent reviewers. Assessor should reach consensus to avoid bias due to subjectivity of judgement in the rating of performance in each domain and for overall acceptance of guidelines for use in practice.

6.2. Methodological quality of clinical practice guidelines

Irrespective of the topic of CPGs we found large variation in the way diagnostic recommendations in CPGs are developed and how methodological quality is incorporated in the development process. We found severe shortcomings of methodological quality in CPGs for the management of DM and NSCLC. Most of the guideline groups did not use systematic and rigorous development processes and did not involve target users and patients in formulating recommendations. There were serious weaknesses in applicability and editorial independence of recommendations.

A notable number of diagnostical CPGs (15% in DM and 27% in NSCLC), including some of the most accepted CPGs worldwide (e.g. WHO in DM and NACB in NSCLC), were not recommended for use in practice by the AGREE evaluation as they failed to meet

basic quality criteria. These findings raise concern about both the internal and the external validity of international recommendations even when they are issued by highly reputed authorities. The heterogeneity in quality highlights the need for critical evaluation of every document before recommendations are used in clinical practice.

Our evaluation revealed that CPGs developed by prestigious authorities in many other disciplines suffer from the same methodological weaknesses as diagnostic recommendations in the field of DM and NSCLC (Figure 5). Our literature review (Book chapter I) revealed relatively high number of CPGs (n=712) critically appraised by the AGREE Instrument. Large proportion of these CPGs predominantly had therapeutic recommendations. Therefore we can conclude that the quality of laboratory related CPGs did not differ from therapeutic CPGs. Since some studies did not report the scores of individual CPGs but only quoted mean scores of their evaluations, we could only compare the mean domain scores of published studies. Despite these limitations our findings depict a similar picture across studies in many medical fields independently of date of publication or origin of recommendations. Shortcomings in methodological quality are mostly due to lack of rigour or inappropriate reporting of the CPG development process, and lack of applicability and declaration of editorial independence. Our results are the first in the field of diagnostic because no studies published so far (13-19) investigated the quality of diagnostic recommendations.

6.3. Causes of poor methodological quality of diagnostic guidelines for diabetes mellitus

Some studies investigated the probable reasons of methodological shortcomings (13-19, 67-78) but not one has evaluated these reasons in diagnostic CPGs and not one studied the reporting of laboratory related information in guidelines yet. This question was addressed by subgroup analyses of our study on DM CPGs. Our findings demonstrated that longer and electronically published CPGs and the availability of CPG development manuals yielded higher methodological scores in most AGREE domains (Table 10). One simple explanation is the lack of space available for detailed and accurate reporting of CPG methodology in journals (79). Paradoxically, lengthy CPGs are thought to be less practical for daily use (79), so one may argue that the length of CPGs adversely affects implementation. In our case, CPGs that achieved high scores for “Applicability” were indeed longer documents, but they also covered additional information on organization, cost implications and monitoring of the use of recommendations in practice. All these tools

help CPG implementation and thus, at least in principle, we cannot confirm that lengthy CPGs are not applicable in practice. The Conference on Guideline Standardization defined a standard for CPG reporting in order to promote quality and facilitate implementation (25). Such CPG reporting standards (25) have not yet been adopted by most journals, and peer-reviewers also rarely use the AGREE or other criteria for systematic assessment of recommendations prior to publication (19, 69, 80). These shortcomings highlight the need to use CPG reporting standards and clear publication and peer reviewing policies for CPGs by major medical journals.

In our study the quality of purely diagnostic CPGs was not significantly different from that of combined diagnostic and therapeutic CPGs (Table 11). Our additional evaluation has shown that nearly half of all diagnostic CPGs do not report pre-analytical, analytical and diagnostic accuracy data (Table 12), which may lead to inappropriate requesting and interpretation of tests in clinical practice (81). Fulfilling these criteria would be desirable in any CPGs that provide laboratory testing-related recommendations, since it is expected that practice guidelines are developed in a multidisciplinary process (82). Unfortunately this could not be confirmed by our study as only 41% of the criteria were fulfilled in D2 which explored the involvement of all relevant stakeholders in the CPG development process.

All CPGs that achieved higher scores in the comparison by origin were from agencies that had detailed CPG manuals which provided a clear description and standards for the development process (Table 9 and 10). The availability of a CPG manual, however, does not always guarantee that CPG teams follow those processes consistently, and it has been shown that it is often not clear how decisions are made by the CPG team when arriving at final recommendations (27). The substantial heterogeneity, both in how the type of publication is defined and the adherence to this definition in the final presentation of the CPG, suggests that there is likely to be a disparity between the methodology CPG developers described and what is actually followed in practice (Table 9). We found, for example, several CPGs that described a grading system but did not grade their final recommendations. The lack of evidence tables in CPGs that claim to be evidence-based may also point to potential deviations from the processes set in CPG manuals (Table 9). Therefore it is advisable that diagnostic CPG development teams adhere to pre-set methodology and document the procedures followed explicitly and report those transparently.

Our study evaluated different publications that were defined in various ways by their

authors. Such heterogeneity of definitions (such as guideline, guidance, protocols, position statement, recommendation and rationale statement, consensus report, *etc.*) may highlight different approaches in formulating recommendations for practice. We also found several CPGs that, while having proof of using evidence-based methods, failed to define their publication as such (43, 44, and 46). This suggests that the definitions used in the international guideline community may be confusing for both guideline developers and users, and that simplification and standardization of terminology is needed.

Even though guideline development methods have gradually improved and were published by several organisations, we could not demonstrate major improvements in CPG quality for most domains (Table 7) and in the "Editorial independence" domain even deterioration in scores was observed over time. We further evaluated the improvement of quality of CPGs over time in some cases where the authorities issued several CPGs (e.g. NICE, WHO, IDF) within the time scale investigated. The NICE CPG in 2004 was of higher quality than the NICE 2002 version due to improvements in the scores for the "Applicability" and "Editorial independence" domains. Recognising there methodological problems, many international organizations are now moving towards international standardization of guideline methods (83, 84-86). Surprisingly, the international WHO and IDF CPGs in 2006 and 2007 had lower scores in most domains than the 2003 and 2005 versions despite the fact that both agencies released guideline development manuals in 2003 (http://whqlibdoc.who.int/hq/2003/EIP_GPE_EQC_2003_1.pdf, http://www.idf.org). Therefore we assume that the lower AGREE scores are due to the lack of reporting some methodological details rather than the lack of following the methodology described in the manuals. Explicit reporting of methodology and adherence to that methodology is particularly important for influential agencies (e.g. ADA and WHO) whose recommendations are universally followed or adopted and adapted worldwide.

There are several limitations in our study. By evaluating English publications only, our results may suffer from language bias. However, several publications, including our own review of the topic (Table 14), confirm no significant differences in the quality of English *versus* non-English publications of guidelines or trials (87-88). Since most national DM CPGs are based on or strongly influenced by international recommendations primarily published in English, we believe our results are likely to be generalizable.

The AGREE Instrument or other CPG appraisal tools can neither investigate the accuracy of the content of recommendations nor their impact on patient outcomes (89-90). Another shortcoming of all critical appraisal tools is that they do not differentiate between

whether the publication fails certain criteria due to lack of reporting or to poor methodology and design. Therefore, our results should not be interpreted as criticisms of the truth of scientific statements or the validity of recommendations made in a given publication about DM. However, the demonstrated shortcomings in reporting and/or the methodology applied by different CPG developers could lead to distrust in and/or misuse of recommendations (91). With such shortcomings, the energy put into developing scientifically accurate but otherwise poorly presented CPGs could end up being wasted, whereas, inaccurate but otherwise nicely presented CPGs might be promoted and used widely. This is why we advise that CPGs be critically evaluated for methodology before recommendations are used in clinical practice (92).

6.4. Correlation between guideline methodological quality and validity of content

A number of studies confirm the assumption that CPGs of poor methodological quality potentially transmit biased opinions that may cause unnecessary burden to patients and costs to society (93-96). Others, however, demonstrated that despite the high inconsistencies in formulating recommendations and the great variation in the supporting evidence cited, the agreement in the content of recommendations was remarkable (97, 98). Our results have shown that guidelines with poor methodological quality are not necessarily invalid in their content and *vice versa*; high quality CPGs do not necessarily provide the best recommendations.

The discrepancy between methodological quality and clinical validity of recommendations could be explained by the authors using different pieces of evidence or differing judgements to base their statements on. The reasons for this could be manifold: (a) non-systematic searching for the evidence, (b) ignoring findings that confirm the beliefs and assumptions or the experience and practice of the guidelines development group, (c) other competing interests as priorities, or (d) considered judgements taking into account other influencing factors such as costs, organizational barriers, patients' preferences, ethics, and safety. It has to be acknowledged that the evidence is only one element in formulating recommendations (18). Guideline developers may down- or upgrade the strength of evidence in final recommendations if other reasons (e.g. social, economical, organizational, societal, ethical, patient perspectives, safety or legal) strongly justify it (99-101). However, considered judgement and grading should be a well-documented and transparent process so that users of CPGs understand the rationale and

reasoning behind final recommendations and why and to what extent guideline teams decided to direct from research findings.

The other reason of this discrepancy might be that the quality of a guideline depends not only on the rigour of its development but also on the quality of the evidence base underlying the recommendations. A number of studies confirmed this assumption demonstrating that poor of high quality evidence was used for CPGs in different medical fields (102) and especially in oncology (103, 104), such as in CPGs for lung cancer (105).

Our study has a limitation because it focused on a small part of an oncology topic. Therefore our data cannot be generalized to other medical topics. The common failure of the currently available guideline appraisal instruments, such as the AGREE Instrument used in our study too, is that they are unsuitable for assessing either the accuracy of the content of recommendations, or the impact on patient outcomes (106). Nevertheless the discrepancies found in our study between quality and content highlight the need for critical appraisal of not only the methodology but also the content of recommendation before their use in practice.

Conflicting recommendations on the use of laboratory tests are likely to lead to a waste of laboratory resources and might even cause harm to patients (107). Effective treatment depends on the effective use of diagnostic tests, and if diagnostic recommendations are not evidence based, it is reasonable to assume that therapeutic interventions will sometimes be initiated and monitored inappropriately.

6.5. Recommendations for the future

Our findings identified several shortcomings in current guideline development initiatives as well as gap in our current knowledge and understanding of guideline processes. These findings call for further research and improvements that could provide tools both for guideline developers and uses in their endeavour of delivery improved care and outcomes for their patients.

In conclusion:

- Our studies highlight the need for clear guideline development and reporting standards for making diagnostic recommendations in CPGs.
- Guideline development must be a multidisciplinary process and the laboratory profession must take a nurse proactive role in joining clinical guideline teams in order to influence diagnostic recommendations.

- Guideline terminology and development methods from systematic literature retrieval to a uniform grading system should be standardized, and training in evidence-based CPG methods should be provided to all engaged in formulating recommendations.
- Until more advanced assessment tools become available that can evaluate the quality and reliability of the content of CPGs, we recommend that all CPGs are critically reviewed before clinical application.

7. SUMMARY

In our studies we could demonstrate that:

- There is large variation in the way diagnostic recommendations in guidelines for clinical practice are developed and how methodological quality is incorporated in the development process.
- The methodological shortcomings of DM and NSCLS CPGs are very similar to those in other medical fields.
- There are serious shortcomings in involving all relevant stakeholders in the guideline development process, in the rigour of development, applicability and editorial independence and these raise concern about both the internal and the external validity of recommendations.
- The quality of purely diagnostic CPGs was not significantly different from that of combined CPGs for DM.
- Subgroup analyses of our DM study demonstrated that longer and electronically published CPGs and the availability of CPG development manuals yielded better overall methodological quality with higher scores in most AGREE domains.
- Nearly half of all DM CPGs do not report pre-analytical, analytical and diagnostic accuracy data, which may lead to inappropriate reporting and interpretation of tests in clinical practice.
- Diagnostic recommendations about tumor markers are conflicting in CPGs for the managements of NSCLS patients.
- We did not find any straight forward relationship between methodological quality and validity of content of NSCLS CPGs.
- Our findings highlight the need for critical evaluation of both the methodology and content of any CPG before recommendations are put in clinical practice.

In conclusions, we make the following recommendations for the future:

- There is a need for systematically developed, explicit recommendations based on evidence-based guideline development and reporting standards in laboratory medicine.
- To overcome the methodological shortcomings of current guidelines standardized methods for making evidence-based guideline recommendations need to be disseminated more effectively in laboratory medicine.
- Evidence should always be assessed in close collaboration between clinicians and specialists in laboratory medicine. Evidence should be only one element in formulating recommendations. Interpretation of the evidence and its translation to practical recommendations should be documented explicitly and transparently and must be free from any form of vested interest or bias
- There is a need for simplification and standardization of CPGs terminology.
- A unified system for grading diagnostic recommendations might help to improve the validity of resulting recommendations.
- Further studies are needed to explore in depth the relationship between the scientific validity and the methodological quality of diagnostic recommendations.
- All CPGs should be critically evaluated for methodology and content before recommendations are used in clinical practice.

8. ACKNOWLEDGEMENTS

I'm deeply grateful to my supervisor, **Prof. Dr. Andrea Rita Horváth**, for providing taking part in IFCC C-EBLM and other international, national team work; and providing working facilities in the Department of Laboratory Medicine (Clinical Chemistry).

I would like to express my special thanks to my co-authors **Dr. Jean-Christophe Charet, Dr. Bruno Friedberg, Dr. Rita Ónody, Dr. Wytze Oosterhuis, Dr. Dunja Rogic, Dr. Swere Sandberg** for pleasant cooperation and my warmest thaks to **Dr. Joseph Watine** and **Dr. Peter S Bunting** for their professional help.

I am especially grateful to **Dr. Krisztina Boda** and **Dr. Tibor Nyári** for their professional help in statistical analyses.

I wish to thank **Dr. Erika Kis** for fruitful discussions and cooperation in other EBM topic.

Finally, I would like to express my special thanks to **my parents** who supported my studies and my work in and all of my family whose love helped me to overcome the recent stressful and complicated period of my life, and thus enabled me to accomplish my doctoral thesis.

9. REFERENCES

1. Sackett DL, Rosenberg WC, Gray JAM, Haynes RB, Richardson WS. Evidence-based medicine: What it is and what it isn't. BMJ 1996; 312:71-2

2. Irvine D, Irvine S The practice of quality. Radliffe Medical Press, Albington, Oxon, 1996

3. JA Muir Gray eds. Evidence-based Healthcare. Churchill Livingstone 2nd edition, 2001; 124 p

4. Recommendation No R(97) 17 of the Council of Europe Committee of Ministers to members states on the development and implementation of quality improvement systems (QIS) in health care. Committee of Ministers on 30 September 1997, at the 602nd meeting of the Ministers' Deputies

5. Council of Europe (2001) Recommendation of the Committee of Ministers to Member States on developing a methodology for drawing up guidelines on best medical practices. Rec. No. R (01) 13. www.coe.int.

6. Kerry Lyon, Clinical effectiveness Department, NHS, 2003

7. Field MJ, Lohr KN, eds. Guidelines for clinical practice. From development to use. Washington, DC: National Academy Press, 1992; 426 pp

8. The AGREE Collaboration. Appraisal of Guidelines for Research and Evaluation (AGREE) Instrument. http://www.agreecollaboration.org (Accessed December 2007)

9. The AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical guidelines: the AGREE project. Qual Saf Health Care 2003;12:18-23.

10. MacDermid JC, Brooks D, Solway S, Switzer-McIntyre S, Brosseau L, Graham ID. Reliability and validity of the AGREE instrument used by physical therapists in assessment of clinical practice guidelines. BMC Health Services Research 2005;5:18. (http://www.biomedcentral.com/1472-6963/5/18) (Accessed December 2007)

11. Ollenschläger G, Marshall C, Quershi S, Rosenbrand K, Burgers J, Mäkelä M, Slutsky J and for the Board of Trustees 2002, Guidelines International Network Improving the quality of health care: using international collaboration to inform guideline programmes by founding the Guiddelines International Network (G-I-N) Qual Saf Health Care 2004;13:455-460

12. Oosterhuis WP, Bruns DE, Watine J, Sandberg S, Horvath AR. Evidence-based guidelines in laboratory medicine: principles and methods. Clin Chem 2004; 50:806-18.

13. Graham ID, Beardall S, Carter AO, Glennie J, Hebert PC, Tetroe JM, McAlister FA, Visentin S, Anderson GM. What is the quality of drug therapy clinical practice guidelines in Canada? CMAJ 2001; 165(2):157-63. http://www.cmaj.ca/cgi/reprint/165/2/157.pdf

14. McAlister FA, Campbell NR, Zarnke K, Levine M, Graham ID. The management of hypertension in Canada: a review of current guidelines, their shortcomings and implications for the future. CMAJ. 2001; 164(4):517-22.

15. Cranney A, Waldegger L, Graham I, Man-Son-Hing M, Byszewski A, Ooi D. Systematic assessment of the quality of osteoporosis guidelines. BMC Musculoskelet Disord. 2002; 3:20.
http://www.biomedcentral.com/content/pdf/1471-2474-3-20.pdf

16. McCormack J, Perry Jr T, Rangno R, van Breemen C, Wright JM, Baltzan M, Graham ID, Beardal S. Assessing the quality of clinical practice guidelines. CMAJ 2002; 166(2):168-9.
17. Harpole LH, Kelley MJ, Schreiber G, Toloza EM, Kolimaga J, McCrory DC. Assessment of the scope and quality of clinical practice guidelines in lung cancer. Chest 2003; 123(1 Suppl):7S-20S.
18. Burgers JS, Fervers B, Haugh M, Brouwers M, Browman G, Philip T, Cluzeau FA. International assessment of the quality of clinical practice guidelines in oncology using the Appraisal of Guidelines and Research and Evaluation Instrument. J Clin Oncol. 2004; 22(10):2000-7.
19. van Tulder MW, Tuut M, Pennick V, Bombardier C, Assendelft WJ. Quality of primary care guidelines for acute low back pain. Spine. 2004; 29(17):E357-62.
20. Watine J. Prognostic evaluation of primary non-small cell lung carcinoma patients using biological fluid variables. A systematic review. Scand J Clin Lab Invest 2000;60:259–74.
21. Watine J. Are laboratory investigations recommended in current medical practice guidelines supported by available evidence? Clin Chem Lab Med 2002;40:252–5.
22. Watine J, Friedberg B, Charet JC. Laboratory variables and stratification of lung cancer patients: recommendations for the future (Therapeutic Trials and Clinical Practice Guidelines). In: Carafaro RL, ed. Focus on lung cancer research. Hauppauge, NY: Nova Science Publishers, 2004:173–203.
23. Hunt DL, McKibbon KA. Locating and appraising systematic reviews. Ann Intern Med 1997;126:532-8.
24. Horvath AR. Pewsner D. Systematic reviews in laboratory medicine: principles, processes and practical considerations. Clin Chim Acta 2004;342:23-39.
25. Shiffman RN, Shekelle P, Overhage JM, Slutsky J, Grimshaw J, Deshpande AM. Standardized Reporting of Clinical Practice Guidelines: A Proposal from the Conference on Guideline Standardization. Ann Intern Med 2003;139:493-8.
26. WHO/NCD/NCS/99.2/ Definition, diagnosis and classification of diabetes mellitus and its complication. Report of a WHO Consultation. Part 1: Diagnosis and Classification of Diabetes Mellitus. WHO. Department of Noncommunicable Disease Surveillance. Geneva, Switzerland, 1999. pp 65.
27. McIntosh A, Hutchinson A, Home PD, Brown F, Bruce A, Damerell A, Davis R, Field R, Frost G, Marshall S, Roddick J, Tesfaye S, Withers H, Suckling R, Smith S, Griffin S, Kaltenthaler E, Peters J, Feder G. Clinical guidelines and evidence review for Type 2 diabetes: management of blood glucose. Sheffield: ScHARR, University of Sheffield. 2001.
28. Sacks DB, Bruns DE, Goldstein DE, MacLaren NK, McDonald JM, Parrott M Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus Clin Chem 2002; 48:436-72.
29. American Diabetes Care Standards of medical care for patients with diabetes mellitus. Diab Care 2003; 26(Suppl.1):S33-50.
30. Woolf SH, Davidson MB, Greenfield S, Bell HS, Ganiats TG, Hagen MD et al. American Academy of Family Physicians and American Diabetes Association. The benefits and risk of controlling blood glucose levels in patients with type 2 diabetes mellitus. A review of the evidence and recommendations. AAFP Policy Action April 1999. http://www.aafp.org/online/etc/medialib/aafp_org/documents/clinical/clin_recs/diabetespolicy.Par.0001.File.tmp/clinicalrecs_diabetespolicy05.pdf (Accessed December 2007)

31. Scottish Intercollegiate Guidelines Network. Management of diabetes CPG55. 2001. http://www.sign.ac.uk/guidelines/fulltext/55/index.html (Accessed December 2007)
32. Reece EA, Homko C, Miodovnik M, Langer O. A consensus report of the Diabetes in Pregnancy Study Group of North America Conference, Little Rock, Arkansas, May 2002. J Matern Fetal Neonatal Med 2002;12:362-4.
33. Sacks DB, Bruns DE, Goldstein DE, Maclaren NK, McDonald JM, Parrott M. Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. Clin Chem 2002;48:436-72.
34. McIntosh A, Hutchinson A, Home PD, Brown F, Bruce A, Damerell A, Davis R et al. (2001) Clinical guidelines and evidence review for Type 2 diabetes: management of blood glucose. Sheffield: ScHARR, University of Sheffield. http://www.nice.org.uk/cat.asp?c=36733 (Accessed December 2007)
35. McIntosh A, Hutchinson A, Feder G, Durrington P, Elkeles R, Hitman GA et al. (2002) Clinical guidelines and evidence review for Type 2 diabetes: Lipids Management. Sheffield: ScHARR, University of Sheffield. http://www.nice.org.uk/cat.asp?c=38551 (Accessed December 2007)
36. Society for Endocrinology, Metabolism and Diabetes of South Africa. Revised SEMDSA Guidelines for diagnosis and management of type 2 diabetes mellitus for primary health care in 2002. http://www.semdsa.org.za/guidelines.htm (Accessed December 2007)
37. Berger H, Crane J, Farine D, Armson A, De La Ronde S, Keenan-Lindsay L et al. Screening for gestational diabetes mellitus. J Obstet Gynaecol Can 2002;24:894-912.
38. Canadian Diabetes Association Clinical Practice Guidelines Expert Committee. Canadian Diabetes Association 2003 Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada. http://www.diabetes.ca /cpg2003/downloads/cpgcomplete.pdf (Accessed December 2007)
39. New-Zealand Guidelines Group. Management of Type 2 Diabetes. http://www.nzgg.org.nz/guidelines/dsp_guideline_popup.cfm?guidelineCatID=32 &guidelineID=36 (Accessed December 2007)
40. U.S. Preventive Services Task Force. Screening for Type 2 Diabetes Mellitus in Adults: Recommendations and Rationale. February 2003. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/clinic/ uspstf/uspsdiab.htm (Accessed December 2007)
41. U.S. Preventive Services Task Force. Screening for gestational diabetes mellitus: recommendations and rationale. http://www.ahrq.gov/clinic/uspstf/uspsgdm.htm (Accessed December 2007)
42. World Health Organisation: Screening for type 2 diabetes. Report of a World Health Organization and International Diabetes Federation meeting. Geneva 2003 http://whqlibdoc.who.int/hq/2003/WHO_NMH_MNC_03.1.pdf (Accessed December 2007)
43. Snow V, Aronson MD, Hornbake ER, Mottur-Pilson C, Weiss KB. Clinical Efficacy Assessment Subcommittee of the American College of Physicians. Lipid control in the management of type 2 diabetes mellitus: a clinical practice guideline from the American College of Physicians. Ann Intern Med 2004;140:644-9.
44. Kaiser Permanente, Care Management Institute's Diabetes Guidelines workgroup. Guidelines for the Management of Adult Diabetes in Primary Care http://members.kaiserpermanente.org/kpweb/pdf/feature/247clinicalpracguide/C

MI_AdultDiabetesGuideline_public_web_033104.pdf (Accessed December 2005)
45. National Collaborating Centre for Women's and Children's Health (NICE guidance). Type 1 diabetes: diagnosis and management of type 1 diabetes in children and young people. http://www.nice.org.uk/page.aspx?o=213575 (Accessed December 2007)
46. Guidelines and Protocols Advisory Committee (British Columbia Ministry of Health). Diabetes Care, revised 2005. http://www.health.gov.bc.ca/gpac/guideline_diabetes.html (Accessed December 2007)
47. IDF Clinical Guidelines Task Force. Global guideline for Type 2 diabetes. Brussels: International Diabetes Federation, 2005. http://www.idf.org/home/index.cfm?node=1457 (Accessed December 2007)
48. National Health and Medical Research Council. National evidence based guidelines for the management of type 2 diabetes mellitus. http://www.health.gov.au/nhmrc/publications/pdf/cp86.pdf (Accessed November 2007)
49. World Health Organisation: Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia. Report of a WHO/IDF Consultation. Geneva, 2006. http://www.who.int/diabetes/publications/Definition%20and%20diagnosis%20of%20diabetes_new.pdf (Accessed December 2007)
50. American Association of Clinical Endocrinologists. American Association of Clinical Endocrinologists medical guidelines for clinical practice for the management of diabetes mellitus. Endocr Pract 2007;13:Suppl1:1-68.
51. American Diabetes Association. Standards of medical care in diabetes. Diabetes Care 2007;30:S4-41
52. IDF Clinical Guidelines Task Force. Guideline for management of postmeal glucose Brussels: International Diabetes Federation, 2007. http://www.idf.org/home/index.cfm?node=1620 (Accessed December 2007)
53. Sowerby Centre for Health Informatics at Newcastle. PRODIGY Guidance: Diabetes type 2 – lipid management. http://www.prodigy.nhs.uk/guidance.asp?gt= Diabetes%20-%20lipid%20management (Accessed December 2007)
54. Sowerby Centre for Health Informatics at Newcastle: PRODIGY Guidance: Diabetes type 2 – renal disease http://www.prodigy.nhs.uk/ guidance.asp?gt= Diabetes%20-%20renal%20disease (Accessed December 2007)
55. Sowerby Centre for Health Informatics at Newcastle: PRODIGY Guidance: Blood glucose. http://www.prodigy.nhs.uk/guidance.asp?gt=Diabetes%20-%20renal%20disease (Accessed December 2007)
56. American College of Chest Physicians. Lung cancer guidelines. Chest 2003;123(1 Suppl):1S–156S.
57. Agence Nationale pour le Développement de l'Évaluation Médicale. Conférence de Consensus. Bilan d'extension pré-thérapeutique du cancer bronchique non á petites cellules. http://www.uvp5.univ-paris5.fr/UV_MED/AC/Impression.asp?Nsuj_98
58. American Society of Clinical Oncology. Clinical practice guidelines for the treatment of unresectable non-small-cell lung cancer. Adopted on May 16, 1997 by the American Society of Clinical Oncology. J Clin Oncol 1997;15:2996–3018.
59. American Thoracic Society and European Respiratory Society. Pre-treatment evaluation of non-small-cell lung cancer. Am J Respir Crit Care Med 1997;156:320–32.
60. British Thoracic Society and Society of Cardiothoracic Surgeons of Great Britain and Ireland. BTS guidelines: guidelines on the selection of patients with lung

cancer for surgery. Thorax 2001;56:89–108. http://thorax.bmjjournals.com/cgi/content/full/56/2/89 (accessed November 8, 2005).

61. CIGNA HealthCare Medicare Administration. Tumor markers for diagnosis and management of cancer. Coverage position number 0172. http://www.cigna.com/health/provider/medical/procedural/coverage_positions/medical/mm_0172_coveragepositioncriteria_tumor_markers_for_diagnosis_mgmt_cancer.pdf (Accessed November 2005).

62. European Group on Tumour Markers. Tumour markers in lung cancer: EGTM recommendations. Anticancer Res 1999;19:2817–9.

63. Fédération Nationale des Centres de Lutte Contre le Cancer. Standards, Options et Recommandations 2000 pour la prise en charge des patients atteints d'un cancer bronchopulmonaire non ápetites cellules. http://www.fnclcc.fr/ (Accessed November 2005).

64. National Academy of Clinical Biochemistry. Practice guidelines and recommendations for use of tumor markers in the clinic. Laboratory medicine practice guidelines. http://www.nacb.org/LMPG/Monograph_TumorMarkers.pdf (Accessed November 2005).

65. Scottish Intercollegiate Guidelines Network. Management of lung cancer, pilot edition, February 1998. http://www.sign.ac.uk (this pilot edition is available from the corresponding author upon request).

66. Société de Pneumologie de Langue Française. Recommandations concernant l'utilisation des marqueurs tumoraux sériques dans la prise en charge des cancers bronchiques primitifs. Rev Mal Resp 1997;14(Suppl 3):5–39.

67. Burgers JS, Cluzeau FA, Hanna SE, Hunt C, Grol R. Characteristics of high-quality guidelines: evaluation of 86 clinical guidelines developed in ten European countries and Canada. Int J Technol Assess Health Care. 2003;19:148-57.

68. Brosseau L, Graham ID, MacLeay L, Cleaver S, Dumont AM, Gravel M, et al. What is the quality of clinical practice guidelines accessible on the World Wide Web for the treatment of musculoskeletal conditions in physiotherapy? Physiother Theory Pract 2004;20:91-105

69. Fervers B, Burgers JS, Haugh MC, Brouwers M, Browman G, Cluzeau F, Philip T. Predictors of high quality clinical practice guidelines: examples in oncology. Int J Qual Health Care 2005;17:123-32.

70. Boluyt N, Lincke CR, Offringa M. Quality of evidence-based pediatric guidelines. Pediatrics 2005;115:1378-91.

71. Lindberg M, Schöld A_K, Östgren CJ, Tilling B, Tunér K. AGREE-instrumentet praktisk metod för utvärding av kliniska riktlinjer. Kvalitetsgranskning av riktlinjer för astma/allergi och diabetes I Östergötland. (The AGREE-instrument, a practical method for evaluation of clinical guidelines. Quality assessment of guidelines for asthma/allergy and diabetes in Ostergotland) Läkartidningen 2005;41:2942-5.

72. Navarro Puerto MA, Romero FR, DomínguezAR, Ibarlucea IG, Gago TH, Ortiz del Río CA, et al. Las guías que nos guían son fiables? Evaluación de las giías de práctica clínica espaenolas. (Are the guidelines guiding us? Evaluation of the Spanish clinical practice guidelines) Rev Clin Esp 2005;205:533-40.

73. Prestoczki B. Bizonyítékokon alapuló irányelvfejlesztés: hazai szakmai ajánlások módszertanának értékelése (Doktori tézis) (Evidence-based guideline development: evaluation of the methodology of Hungarian guidelines recommendations (Thesis) Szeged, Hungary: University of Szeged, Medical Faculty, 2005.

74. Stiegler M, Rummel C, Wahlbeck K, Kissling W, Leucht S. European psychiatric treatment guidelines: ist he glass half full or half empty? Eur Psych 2005;20:554-8.
75. Arnau JM, Vallano A, Lopez A, Pellisé F, Delgado MJ, Prat N. A critical review of guidelines for low back pain treatment. Eur Spine J 2006;15:53-53.
76. Cates JR, Young DN, Bowerman DS, Porter RC. An independent AGREE evaluation of the occupational medicine practice guidelines. Spine J 2006;6:72-7.
77. Ministry of Health and TUDOR network, Hungary. Evaluation of Hungarian care pathway protocols. 2007
78. Vervey B, van Waarde JA, van Rooij IALM, Gerritsen G, Zitman FG. Availability, content and quality of local guidelines for the assessment of suicide attempters in university and general hospitals in the Netherlands. Gen Hosp Psych 2006;28:336-42.
79. Deeks JJ. Word limits best explain failings of industry supported meta-analyses [Letter]. BMJ 2006;333:1021.
80. Miller J, Petrie J. Development of practice guidelines [Commentary]. Lancet 2000;355:82-3.
81. Skeie S, Nordin G, Oosterhuis WP, Araczki A, Horvath AR, Perich C et al. Post-analytical external quality assurance of blood glucose and HbA1c: an international survey. Clin Chem 2005;51:1145–53.
82. Grimshaw GM, Khunti K, Baker R. Diagnosis of heart failure in primary care: an assessment of international guidelines. Br J Gen Pract 2001;51:384-6.
83. Van der Wees PJV, Hendriks EJM, Custers JWH, Burgers JS, Dekker J, de Bie RA. Comparison of international guideline programs to evaluate and update the Dutch program for clinical guideline development in physical therapy. BMC Health Services Research 2007;7:191 http://www.biomedcentral.com/1472-6963/7/191 (Accessed December 2007)
84. Raine R, Sanderson C, Black N. Developing clinical guidelines: a challenge to current methods. BMJ 2005;331:631-33.
85. Oxman AD, Fretheim A, Schünemann HJ, SURE. Improving the use of research evidence in guideline development: Introduction. Health Res Policy Syst 2006;4:12.
86. Schünemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: I. Guideline for guidelines. Health Res Policy Syst 2006;4:13.
87. Moher D, Fortin P, Jadad AR, Juni P, Klassen T, Le Lorier J et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. Lancet 1996;347:363-6.
88. Burgers JS, Grol R, Klazinga NS, Makela M, Zaat J. Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs. Int J Qual Health Care 2003;15:31-45.
89. Vlayen J, Aertgeerts B, Hannes K, Sermeus W, Ramaekers Dirk. A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. Int J Qual Health Care 2005;17:235-42.
90. Burgers JS. Guidelines quality and guidelines content: are they related [Editorial]. Clin Chem 2006;52:3-4.
91. Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Potential benefits, limitations, and harms of clinical guidelines. BMJ 1999;318:527-30.
92. Qaseem A, Vijan S, Snow V, Cross T, Weiss KB, Owens DK, for the Clinical Efficacy Assessment Subcommittee of the American College of Physicians.

Glycaemic control and type 2 diabetes mellitus: The optimal Hemoglobin A1c targets. A guidance statement from the American College of Physicians. Ann Intern Med 2007;147:417-22.

93. Vogel N, Burnand B, Vial Y, Ruiz J, Paccaud F, Hohlfeld P. Screening for gestational diabetes: variation in guidelines. Eur J Obs Gynecol Reprod Biol 2000;91:29-36.

94. Cruse H, Winiarek M, Marshburn J, Clark O, Djulbegovic B. Quality and methods of developing practice guidelines. BMC Health Services Research 2002;2:1 www.biomedcentral.com/1472-6963/2/1(Accessed February 2007)

95. Savoie I, Kazanjian A, Bassett K. Do clinical practice guidelines reflect research evidence? J Health Serv Res Policy 2000;5:76-82

96. 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;316:509-13.

97. FionaCampbell, Heather O Dickinson, JuliaVF Cook, Fiona R Beyer, Martin Eccles and James M Mason. Methods underpinning national clinical guidelines for hypertension: describing the evidence shortfall BMC Health Services Research 2006;6:47

98. Burgers JS, Bailey JV, Klazinga NS, Vander Bij AK, Grol R, Feder G. Inside guidelines. Comparative analysis of recommendations and evidence in diabetes guidelines from 13 countries. Diabetes Care 2002;25:1933–9

99. GRADE WorkingGroup Grading quality of evidence and strength of recommendations BMJ 2004;328;1490

100. K Verkerk, H van Veenendaal, JL. Severens, E J M Hendriks and J S Burgers. Considered judgement in evidence-based guideline development. International Journal for Quality in Health Care 2006;18:365–369.

101. Gordon Guyatt, David Gutterman, Michael H. Baumann, Doreen Addrizzo-Harris, Elaine M. Hylek, Barbara Phillips, Gary Raskob, Sandra Zelman Lewis and Holger Schünemann Grading Strength of Recommendations and Quality of Evidence in Clinical Guidelines: Report From an American College of Chest Physicians Task Force Chest2006;129;174-181

102. Eeva Ketola, Minna Kaila, Mari Honkanen Qual Saf Health Care 2007;16:308-312.

103. Susan Mallett, Jonathan J Deeks, Steve Halligan, Sally Hopewell, Victoria Cornelius, Douglas G Altman Systematic reviews of diagnostic tests incancer: review of methods and reporting doi:10.1136/bmj.38895.467130.55 2006;333;413-; originally published online 18 Jul 2006; BMJ

104. F.Vigna-Taglianti, P Vineis, A Liberati, F Faggiano Quality of systematic reviews used in guidelines for oncology practice Annals of Oncology 2006;17:691-701

105. Linda H. Harpole, Michael J. Kelley, Gilbert Schreiber, Eric M. Toloza, Jane Kolimaga, Douglas C. McCrory Assessment of the Scope and Quality of Clinical Practice Guidelines in Lung Cancer MHS Chest 2003; 123:7S–20S

106. Vlayen J, Aertgeerts B, Hannes K, Sermeus W, Ramaekers Dirk. A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. Int J Qual Health Care 2005;17:235–42.

107. Shekelle PG, Kravitz RL, Beart J, Marger M, Wang M, Lee M. Are nonspecific practice guidelines potentially harmful? A randomized comparison of the effect of nonspecific versus specific guidelines on physician decision making. Health Serv Res 2000;34:1429–48

10. ANNEX Copies of papers and book chapter related to Ph.D. thesis