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

Summary of Ph.D. thesis

Éva Nagy, MD

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

2010

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 (EBM) 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 means integrating individual clinical expertise with the best available external clinical evidence from systematic research. In 1997 the Committee of Ministers of the Council of Europe recommended the evidence-based guidelines implementing the quality improving system. 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. There is an exponentially rising interest toward CPGs in the medical literature and several organisations developed methodological manuals or so-called "Guidelines for guidelines". We found only one relevant narrative review 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 and defined specific reporting standards related to the laboratory aspects of diagnostic recommendations. 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. Most of these CPGs made therapeutic recommendations. Quality of CPGs in diagnostic fields and their impact in practice has been less well studied.

3. AIMS AND OBJECTIVES

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?

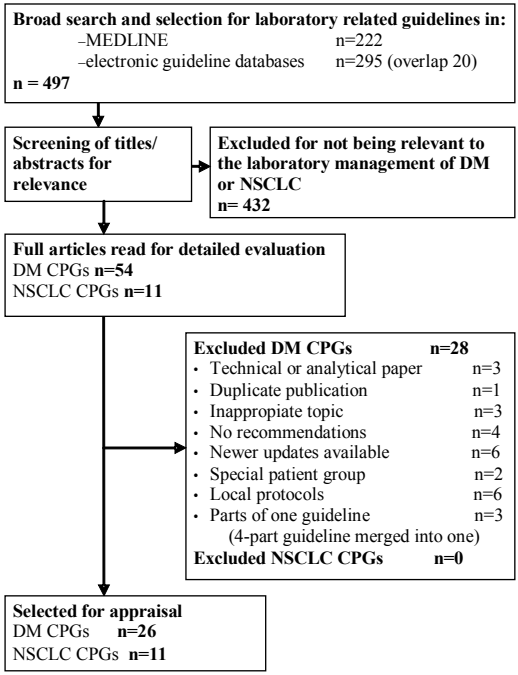
4. METHODS

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). The other topic was related to oncology and focused on the management of

non-small-cell-lung-cancer (NSCLC) patients. Systematic literature search was carried out in PubMed, in dedicated CPG databases and websites of professional to retrieve diagnostic CPGs published in English between 1 January 1999 and 31 December 2007. Two independent reviewers selected 26 DM and 11 NSCLC CPGs (Figure1).

Figure 1 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 a standardized, generic and validated checklist for the evaluation of the methodological quality of CPGs. 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 we assessed the agreement between reviewers by statistical methods (Cronbach's alpha, Interclass correlations (ICC)). 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 by the AGREE Instrument as described in its manual. We assessed fulfilment of 23 criteria (I) grouped into 6 domains on a 4-point Likert scale and calculated Domain Scores (DS) in percentages, than judged overall performance of CPGs with one of 4 options. We appraised the methodological

quality of NSCLC CPGs in this way, except for we changed the overall assessment terminology of AGREE ("strongly recommend", "recommend with provisos or alterations", "would not recommend", "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

We created subgrouped DM 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 (I) and standardized domain scores (DS) 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 1) an evidence table, 2) a description of the grading system, 3) graded recommendations, 4) an expiry or review date? 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

We created two subgroups of DM CPGs based on their scope 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

We assessed the presence of 1) prevalence, 2) diagnostic accuracy of tests, 3) pre-analytical, and 4) analytical specifications. The frequency of this 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

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

Two assessors extracted all laboratory-related recommendations from the 11 NSCLC guidelines selected for review. Validity of recommendations was investigated based on a published systematic review regarding the use of tumor markers and other more global laboratory tests in NSCLC. Methodological quality was assessed by AGREE Instrument with slightly modified expression of overall assessments.

5. RESULTS

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

The agreement between assessors was acceptable based on statistical calculations (Table 1). We have noted some discrepancy between statistical judgements of agreement and the comparison of each item score of each appraiser 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. 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. 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. 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. Surprisingly the well-known CPGs had some serious shortcomings in all appraised aspects, thus reflected need for appraising of methodological quality of all CPGs before their use.

Table 1 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	41	11	8	38	13	24	Not recommend	**0,03***	**3**
NICE 2001	95	94	82	90	36	52	Strongly recommend	-	**1**
NACB 2002	52	31	43	76	28	19	Recommend with alteration	**0,097**	**1**$^{+}$
ADA 2003	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 ($p \leq 0,05$); ** 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)

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" (Table 2). Overall, the best performing domains were D1 "Scope and purpose". Although D4 "Clarity and presentation" scored highly only 10 CPGs (38%) were supported

Table 2 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	

with tools for application. In D3, which explored the rigour of development, there are notable shortcomings in using systematic methods for searching the evidence and providing information on the literature retrieval and selection process; indicating the methods used for formulating recommendations; and giving information on the peer reviewing and updating process. Domain 2, which explored stakeholder involvement, showed lower scores. Low scores were achieved with in the "Applicability" (34%) and "Editorial independence" (39%) domains, in which each item performed very poorly. The wide spread of the minimum and maximum scores of each individual domain these data in all domains demonstrated unexpectedly large variation in CPG.

Table 3 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	

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

Only 5 out of 11 CPGs were recommended for use by assessors and none achieved the best overall quality rating of„strongly recommend" (Table 3). Three CPGs 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". There were notable shortcomings in other domains which explored the rigour of development, stakeholder involvement and editorial independence. Domain 5, which explored the applicability of recommendations, showed the lowest scores. In all domains, except Domain 5, there was a large of scores.

5.3. Causes of poor methodological quality of diabetes mellitus guidelines

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

Date of publication

Most CPGs were developed after 2002 and only 2 were developed between 1999 and 2001. Only the highest scoring D1 and D4 showed some marginal development in quality over the time scale investigated (Table 2). 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.

Type of publication

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 4). Amongst the 7 CPGs that claimed to be evidence-based 5 had evidence summaries and 6 graded their

Table 4 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.: Continuous*

Table 5 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			
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 was published in journal only; * p≤0.01; **p≤0.05*

recommendations. Three CPGs that had evidence tables, however, did not define their publications as being evidence-based. Over two thirds of CPGs defined their grading system but only 16 graded their final recommendations.

Procedure for updating guidelines

Fifteen CPGs (58%) gave a timescale or expiry date (Table 4). The most frequent review date was 3 and 4 years. Only 10 CPGs (38%) provided adequate information on the updating process.

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 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 5).

Sub-grouping by length

A clear relationship could be demonstrated between CPG length and methodological quality (Table 5). 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. However, the best performing CPGs, scoring >50% in the

"Applicability" domain were generally longer than 50 pages and all were published in electronic databases (Table 2 and 4)

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 originated from the UK. 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.

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. All mean domain scores were better in the subset where these manuals were available.

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

The rate of occurrence of strongly recommended CPGs and the not recommended CPGs was also higher for the combined, than for the diagnostic CPGs (Table 5). The quality of purely diagnostic CPGs was not significantly different from that of combined CPGs based on their domain scores (Table 6).

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

Only about 60 percent of the CPGs mentioned essential laboratory-specific information (Table 7) in any detail. 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 8).

Table 6 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≤0.05*

Table 7 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**

Table 8 Qualitative analysis of reporting laboratory specific information in diabetes mellitus guidelines

		Percentage of guidelines 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 published in journal only*

Table 9 Studies investigating the methodological quality of CPGs by the AGREE Instrument

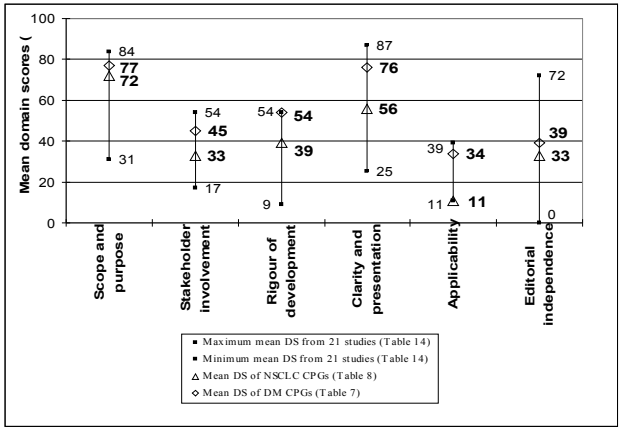
Study (ref)	Topics covered by the CPGs	Publication date of CPGs	Origin of CPGs	No. of CPG	Mean Domain Scores (%)						Comments
					D1	D2	D3	D4	D5	D6	
AGREE 2003 (9)	mixed	1992-1999	international*	33	69	36	41	66	37	30	*:CPG origin : 10 from European countries and 1 from Canada
Burgers *et al.* 2003 (67)	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.

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

We found 21 studies up till 2007 which, have investigated the quality of CPGs using the AGREE Instrument in various clinical fields (Table 9). The altogether 712 CPGs had recommendations in diverse medical fields. Majority of these CPGs were predominantly therapeutic. 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 2). The heterogeneity is very large between these 21 studies. Our results for CPGs related to laboratory medicine were similar to there international findings. 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 2 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)

We collected the recommendations about the use of laboratory tests in these NSCLC CPGs (Table 10) and information from existing systematic reviews on the laboratory parameters to be measured during the pre-treatment evaluation of NSCLC patients (Table 11). Recent systematic reviews provide no evidence that measurement of tumor markers in routine practice would improve NSCLC patients' outcomes. Only 4 CPGs, which did not recommend the use of tumor markers, were scored for validity of content as "good".

Regarding the other laboratory tests, only 5 CPGs 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. Results of our comparison of guideline quality *versus* content for each guideline are shown in Table 12. 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.

Table 10 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*

Table 11 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

Table 12 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*

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. Our literature review 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 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 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. One simple explanation is the lack of space available for detailed and accurate reporting of CPG methodology in journals. Paradoxically, lengthy CPGs are thought to be less practical for daily use, 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. Such CPG reporting standards 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.

In our study the quality of purely diagnostic CPGs was not significantly different from that of combined diagnostic and therapeutic CPGs. Our additional evaluation has shown that nearly half of all diagnostic CPGs do not report pre-analytical, analytical and diagnostic accuracy data, which may lead to inappropriate requesting and interpretation of tests in clinical practice. 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. 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. 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. 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. 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.

Such heterogeneity of definitions of guidelines 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. 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 and in the "Editorial independence" domain even deterioration in scores was observed over time.

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, confirm no significant differences in the quality of English *versus* non-English publications of guidelines or trials. 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. 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.

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. 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. 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. 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. 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 and especially in oncology, such as in CPGs for lung cancer.

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 AGREE Instrument or other CPG appraisal tools can neither investigate the accuracy of the content of recommendations nor their impact on patient outcomes. 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. 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.

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 n 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 thanks 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.

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 ed. 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, US

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 (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 (IF:1.633)

Lectures

I. Horvath AR., ***Nagy E***. Guidelines in Central-Eastern European countries. AGREE/G-I-N International Meeting, Edinburgh, Scotland, Nov14-15, 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 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 2627 June 2009, Kraków, Poland