Applicability of the AGREE II Instrument in Evaluating the Development Process and Quality of Current National Academy of Clinical Biochemistry Guidelines

Andrew C. Don-Wauchope,[1,2,3*] John L. Sievenpiper,[1] Stephen A. Hill,[1,2] and Alfonso Iorio[3,4]

BACKGROUND: Laboratory medicine practice guidelines (LMPGs) are an important part of clinical laboratory medicine. The Appraisal of Guidelines for Research and Evaluation II (AGREE II) instrument has been developed to evaluate the process of practice-guideline development and the quality of reporting. We assessed the applicability of AGREE II in assessing the National Academy of Clinical Biochemistry (NACB) LMPGs.

METHODS: The NACB website was searched for all available LMPGs up to December 2011. Two independent appraisers used the AGREE II instrument to assess each LMPG identified by the search. Quality was assessed across 6 domains (scope and purpose, stakeholder involvement, rigor of development, clarity of presentation, applicability, and editorial independence), comprising a total of 23 items and 2 overall assessments, each scored on a 7-point scale (1, strongly disagree, to 7, strongly agree). All scores were expressed as AGREE II calculated percentages (100% indicates that all items scored 7 by all appraisers).

RESULTS: Eleven LMPGs were identified. All of the LMPGs provided some information seen as applicable to clinical practice by the appraisers. Only 5 of the LMPGs had overall scores ≥50%, with a median score of 42% (range: 8%–92%). Individual domain scores varied considerably from 0% to 100%. One guideline achieved a very high score on the instrument.

CONCLUSIONS: The AGREE II instrument is applicable and useful to evaluate LMPGs. All domains were evaluated as being useful to assess LMPGs, some were addressed well (e.g., clarity of presentation), whereas others could be improved (e.g., applicability).

© 2012 American Association for Clinical Chemistry

The National Academy of Clinical Biochemistry (NACB)[5] has produced several laboratory medicine practice guidelines (LMPGs) intended to inform clinical laboratory practice. These guidelines are the most relevant source of information regarding evidence-based practice in laboratory medicine for clinical chemistry. There are other guidelines produced at local, regional, and national levels that are specific for laboratory testing or that include aspects of laboratory testing.

Standardized protocols, methodology, and ways of reporting evidence for guidelines have been developed (1–3). Following a standardized methodology should result in higher quality clinical practice guidelines (CPGs). The standardization of guideline methodology is intended to make it easier for all healthcare providers and the public to appraise, agree with, and implement the guideline (4, 5). A good example of guideline standardization is provided by the National Guideline Clearing House, which produces standardized summaries of guidelines, reformatting their content and verifying critical information to further support implementation and use of the CPG (6). Standardizing the production of guidelines at the source by emphasizing good methodology is the ultimate goal (7). Initially, it is important to use appropriate tools to ensure a reliable and reproducible selection and appraisal of the evidence used for CPGs during the systematic review step of the guideline development process (2, 4, 5). Subsequent steps will ensure the transparent and accountable translation of that evidence into the guideline recommendations (8, 9).

Because most guidelines are issued in the fields of treatment and prevention, the applicability of the general guideline methodology to LMPG must be considered, because several specific aspects may be included in laboratory-related guidelines that might not be considered in guidelines for producing CPGs or in evalua-

---

[1] Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, Canada; [2] Hamilton Regional Laboratory Medicine Program, Hamilton, Ontario, Canada; [3] Department of Medicine, McMaster University, Hamilton, Ontario, Canada; [4] Departments of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada.
* Address correspondence to this author at: Juravinski Hospital and Cancer Centre, Rm. A0–120. Fax 905-381-7066; e-mail donwauc@mcmaster.ca.

[5] Nonstandard abbreviations: NACB, National Academy of Clinical Biochemistry; LMPG, laboratory medicine practice guideline; CPG, clinical practice guideline; AGREE, Appraisal of Guidelines for Research and Evaluation.

tion instruments designed for CPGs *(10)*. For example CPGs for diabetes mellitus have been reviewed on 2 occasions, and in both occurrences the methodology used and the reporting of laboratory testing was found to be variable and not standardized *(11, 12)*. This finding raised the question of the validity of the clinical laboratory testing component of these CPGs *(12)*. In addition, the suitability of all-purpose guideline methods for producing LMPGs may be questioned.

The Appraisal of Guidelines for Research and Evaluation (AGREE) collaboration has been formed with the purpose of improving the quality and effectiveness of CPGs *(13)*. The AGREE collaboration has developed an instrument for use in evaluating guidelines, the AGREE II instrument *(8)*. This instrument has been used to evaluate CPGs covering a range of clinical topics around the world. The AGREE *(9)* or AGREE II instrument also has been used to evaluate guidelines that relate to aspects of laboratory medicine *(11, 12, 14–17)*. In addition to the reviews of diabetes CPGs *(11, 12)*, Watine et al. published an example of evaluating CPGs with laboratory information using the AGREE instrument *(17)*. However, no publication evaluating a series of LMPGs has been produced to confirm the applicability of AGREE II to LMPGs. The current study was designed to demonstrate the applicability of the AGREE II instrument to LMPGs by using it to evaluate a series of LMPGs produced by the NACB. This work has enabled us to report the quality of NACB LMPGs and make recommendations for improving future LMPGs.

Methods

GUIDELINES SELECTION
The NACB website was searched for all LMPGs published up to December 2011. All LMPGs identified as current and available for downloading as full files were eligible for inclusion *(18)*. Other published versions were not used. When available, associated supplemental files were also collected from the NACB website.

GUIDELINES APPRAISAL
The AGREE II instrument *(8)* was used to evaluate each of the identified LMPGs produced by the NACB *(18)*. The AGREE II is organized into 6 quality domains (scope and purpose, stakeholder involvement, rigor of development, clarity of presentation, applicability, and editorial independence) comprising a total of 23 items *(9)*. Each item targets 1 key aspect of the practice guideline quality. The instrument also includes 2 overall rating items requiring the appraiser to make overall judgments of the practice guideline *(9)*. Each of the items and the 2 overall rating items were assessed for their applicability to the LMPGs and then

rated on a 7-point scale (1, strongly disagree to 7, strongly agree). A score of 1 was given when relevant information was very poorly reported or not provided. Scores from 2 to 6 were assigned when the reporting did not meet the full criteria or considerations for an item, with scores increasing as more criteria and considerations were met. A score of 7 was given when the quality of reporting was exceptional and all criteria and considerations were met in full for an item. Four appraisers (A.C. Don-Wauchope, J.L. Sievenpiper S.A. Hill, A. Iorio) were trained to use the AGREE II instrument with the on-line training instruments provided on the AGREE trust website *(13)*. Two appraisers independently reviewed each LMPG using an online or printed version of the AGREE II instrument. The eligible LMPGs were randomly allocated to appraisers with a random number generator. The independent appraisals were discussed by the team at a face-to-face meeting of all of the appraisers, and if the individual scores varied by 4 or more points on the 7-point scale the individual scores were reconsidered after the documentation was looked at. There was no assumption that the 2 scores should agree, nor was there any further adjustment of the revised scores *(9)*.

DATA ANALYSES
Data are expressed as calculated percentage scores. The AGREE II instrument *(8)* calculates domain and overall rating scores by summing up all of the individual item scores for each appraiser (obtained scores) minus the minimum possible score [minimum possible score per item (1, strongly disagree) $\times$ n items $\times$ n appraisers]. The total is expressed as a percentage of the maximum possible score [maximum possible score per item (7, strongly agree) $\times$ n items $\times$ n appraisers] minus the minimum possible score, according to the formula:

$$\frac{(\text{Obtained score} - \text{Minimum possible score})}{(\text{Maximum possible score} - \text{Minimum possible score})}.$$

Agreement on individual items among the various appraisers was determined by weighted (quadratic) κ analysis of the final scores (Analyze-It, version 2.22, Analyze-It Software). We performed a sensitivity analysis by using the Mann–Whitney test to compare guidelines published before or after 2009 (publication of AGREE II).

Results

GUIDELINES SELECTION
Eleven LMPGs were identified *(19–29)* as current on December 5, 2011. For 4 *(23, 25–27)* of 11 LMPGs, reports included consideration of the methodological
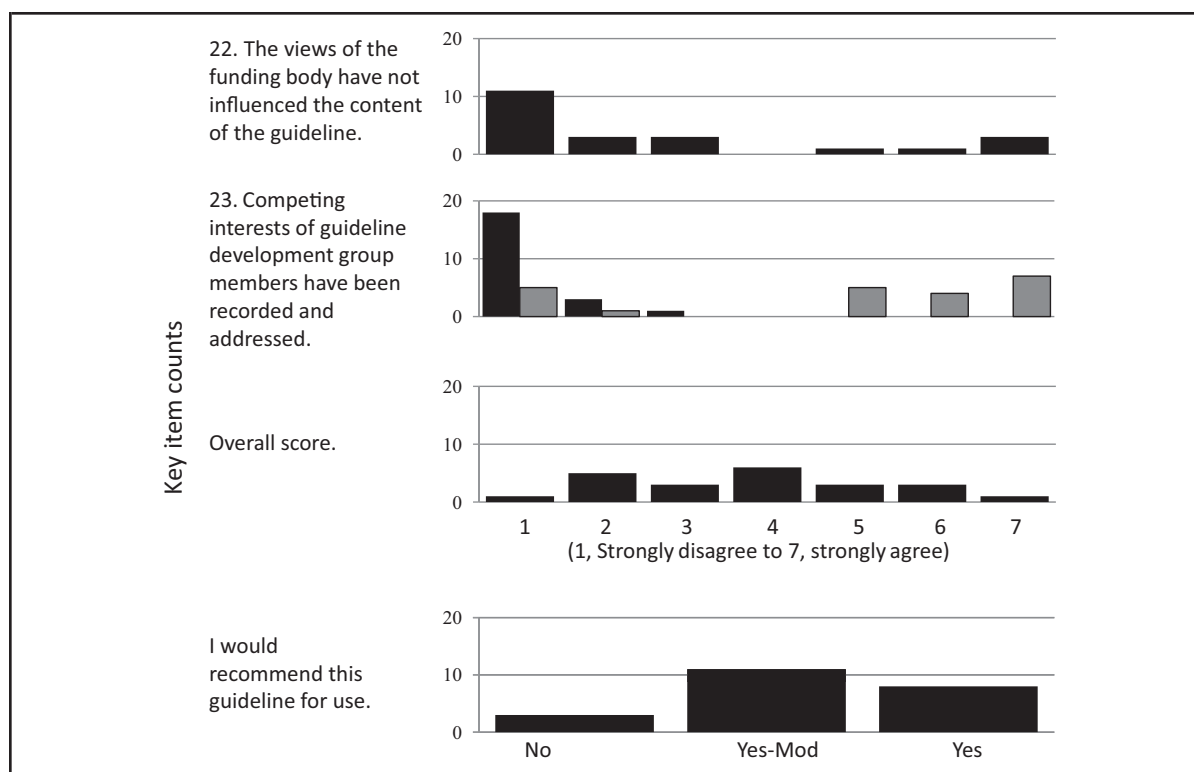
Fig. 1. Counts of individual scale (1–7) scores for Domain 6 (editorial independence) and the overall guideline assessment, showing key items 22 and 23 and the final 2 questions.

Item 23 was scored twice, once from the full guideline (black bar) and a second time with the supplemental file (grey bar). This indicator demonstrates that this item had been considered but was not well reported for the LMPGs. The *y* axis is set to 20 individual counts. Each item was scored twice for eleven LMPGs (maximum count of 22). The final question was scored on No, yes with modifications (Yes-Mod), and Yes.

quality criteria of the AGREE/AGREE II instrument in the development of their guideline. Eight *(19– 21, 23, 25–28)* of the LMPGs had an additional supplemental file that contained disclosures for conflicts of interest of the LPMG team. In our opinion these supplemental files were not easy to locate and were often not linked or referenced in the main version. The scoring of the item on editorial independence was done initially without, and then after considering, the supplemental files (Fig. 1). Table 1 has an additional row for editorial independence showing the supplemental scores.

GUIDELINES APPRAISAL

The AGREE II instrument was applied successfully to all 11 of the LMPGs. Item 16 (the different options for management of the condition or health issue are clearly presented) was excluded in the Tumor Marker Quality Requirements LMPG *(26)*. It was felt by both appraisers and agreed with by the team that this item was not applicable in the context of this guideline. All other items were deemed applicable to all LMPGs.

The overall scores for each guideline and the domain scores are presented in Table 2, and the individual appraiser scores are shown in Table 1. The quality of the guidelines, as scored by use of the AGREE II instrument, was generally poor. Only 5 *(20–23, 25)* of the 11 guidelines had overall scores ≥50%, with a median score of 42% (range: 8%–92%). Domains, which contributed to lower overall scores, were "Domain 3. Rigor and development" [28% (18%–85%)] (Fig. 2), and "Domain 5. Applicability" [25% (6%–67%)]. On the other hand, "Domain 1. Scope and purpose" [64% (31%–92%)] "Domain 2. Stakeholder involvement" [50% (11%–100%)], and "Domain 4. Clarity of presentation" [71% (19%–94%)] contributed to higher overall scores.

The Diabetes LMPG *(23)* scored the highest, reflecting the improvement in the way in which the guideline addressed the domains in the AGREE II instrument. It was also 1 of 2 guidelines to get "Yes" from both appraisers to the final item ("I would recommend this guideline for use") of the overall evaluation (Table 1

Table 1. Data used for the AGREE II calculations of domain score for the Assessment of the NACB LMPGs.[a]

	Guideline									
	Guidelines and Recommendations for Laboratory Analysis in the Diagnosis and Management of Diabetes Mellitus		Use of Tumor Markers in Liver, Bladder, Cervical, and Gastric Cancers		Laboratory Analysis and Application of Pharmacogenetics to Clinical Practice		Use of Tumor Markers in Testicular, Prostate, Colorectal, Breast, and Ovarian Cancers		Emerging Biomarkers for Primary Prevention of Cardiovascular Disease and Stroke	
Reviewer[b]	A.I.	S.A.H.	S.A.H.	J.L.S.	A.C.D.	J.L.S.	A.C.D.	A.I.	S.A.H.	A.C.D.
Domain 1. Scope and purpose	**21**	**17**	**11**	**10**	**15**	**14**	**13**	**21**	**19**	**17**
1. The overall objective(s) of the guideline is (are) specifically described.	7	6	6	5	6	5	4	7	5	6
2. The health question(s) covered by the guideline is (are) specifically described.	7	4	3	2	7	7	4	7	7	5
3. The population (e.g., patients, public) to whom the guideline is meant to apply is specifically described.	7	7	2	3	2	2	5	7	7	6
Domain 2. Stakeholder involvement	**21**	**21**	**10**	**8**	**12**	**10**	**14**	**16**	**11**	**16**
4. The guideline development group includes individuals from all the relevant professional groups.	7	7	2	2	4	2	6	6	3	5
5. The views and preferences of the target population (patients, public, etc.) have been sought.	7	7	2	1	1	1	2	3	5	6
6. The target users of the guideline are clearly defined.	7	7	6	5	7	7	6	7	3	5
Domain 3. Rigor of development	**52**	**46**	**24**	**19**	**26**	**17**	**20**	**21**	**36**	**32**
7. Systematic methods were used to search for evidence.	5	7	2	2	1	1	2	2	3	3
8. The criteria for selecting the evidence are clearly described.	7	4	2	2	1	1	1	2	4	2
9. The strengths and limitations of the body of evidence are clearly described.	7	6	4	3	4	2	3	4	5	3
10. The methods for formulating the recommendations are clearly described.	7	6	3	2	4	3	5	1	6	6
11. The health benefits, side effects, and risks have been considered in formulating the recommendations.	5	4	4	3	4	2	2	1	3	5
12. There is an explicit link between the recommendations and the supporting evidence.	7	7	5	3	6	4	4	7	6	6
13. The guideline has been externally reviewed by experts prior to its publication.	7	7	2	2	5	3	1	1	6	5
14. A procedure for updating the guideline is provided.	7	5	2	2	1	1	2	3	3	2
Domain 4. Clarity of presentation	**20**	**20**	**14**	**13**	**17**	**12**	**17**	**18**	**14**	**15**
15. The recommendations are specific and unambiguous.	7	7	5	4	6	4	5	7	6	5
16. The different options for management of the condition or health issue are clearly presented.	7	7	4	5	5	4	6	5	3	5

Table 1. Data used for the AGREE II calculations of domain score for the Assessment of the NACB LMPGs.[a] *(Continued from page 1429)*

Guideline											
Follow-up Testing for Metabolic Diseases Identified by Expanded Newborn Screening Using Tandem Mass Spectrometry		Use of Tumor Markers in Clinical Practice: Quality Requirements		Biomarkers of Acute Coronary Syndromes and Heart Failure		Evidence-Based Practice For Point-of-Care Testing		Maternal-Fetal Risk Assessment and Reference Values in Pregnancy		Recommendations for the Use of Laboratory Tests to Support Poisoned Patients Who Present to the Emergency Department	
S.A.H.	J.L.S.	S.A.H.	A.C.D.	A.C.D.	J.L.S.	A.I.	J.L.S.	A.C.D.	A.I.	A.I.	J.L.S.
11	**6**	**13**	**11**	**16**	**15**	**21**	**18**	**13**	**11**	**12**	**12**
4	2	6	4	6	5	7	6	3	4	5	5
4	2	2	4	5	4	7	6	4	1	5	3
3	2	5	3	5	6	7	6	6	6	2	4
6	**4**	**12**	**15**	**15**	**11**	**14**	**10**	**10**	**8**	**11**	**10**
2	2	3	6	6	5	5	4	7	6	6	3
2	1	3	3	2	1	4	2	1	1	1	1
2	1	6	6	7	5	5	4	2	1	4	6
20	**15**	**19**	**28**	**29**	**26**	**44**	**34**	**18**	**17**	**14**	**19**
2	1	1	2	2	3	7	5	1	1	1	1
2	1	1	1	1	2	7	6	1	1	1	1
2	2	1	4	3	3	7	4	3	1	1	1
2	2	2	5	7	5	7	6	1	1	2	2
5	2	1	3	5	4	5	4	4	4	3	6
3	5	2	5	4	6	7	5	3	5	3	5
2	1	6	6	5	2	2	3	4	2	2	2
2	1	5	2	2	1	2	1	1	2	1	1
11	**9**	**4**	**9**	**17**	**15**	**17**	**17**	**16**	**18**	**14**	**16**
5	3	2	5	5	5	5	6	5	6	5	6
1	2	1	1	6	6	6	6	5	6	3	6

Table 1. Data used for the AGREE II calculations of domain score for the Assessment of the NACB LMPGs.[a] *(Continued from page 1430)*

	Guideline									
	Guidelines and Recommendations for Laboratory Analysis in the Diagnosis and Management of Diabetes Mellitus		Use of Tumor Markers in Liver, Bladder, Cervical, and Gastric Cancers		Laboratory Analysis and Application of Pharmacogenetics to Clinical Practice		Use of Tumor Markers in Testicular, Prostate, Colorectal, Breast, and Ovarian Cancers		Emerging Biomarkers for Primary Prevention of Cardiovascular Disease and Stroke	
Reviewer[b]	A.I.	S.A.H.	S.A.H.	J.L.S.	A.C.D.	J.L.S.	A.C.D.	A.I.	S.A.H.	A.C.D.
17. Key recommendations are easily identifiable.	6	6	5	4	6	4	6	6	5	5
Domain 5. Applicability	**20**	**20**	**11**	**6**	**12**	**13**	**11**	**11**	**11**	**10**
18. The guideline describes facilitators and barriers to its application.	7	4	3	1	1	1	4	1	5	3
19. The guideline provides advice and/or tools on how the recommendations can be put into practice.	5	6	2	2	5	6	4	3	4	5
20. The potential resource implications of applying the recommendations have been considered.	1	3	3	1	5	2	1	4	1	1
21. The guideline presents monitoring and/or auditing criteria.	7	7	3	2	1	4	2	3	1	1
Domain 6. Editorial independence	**8**	**8**	**6**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
22. The views of the funding body have not influenced the content of the guideline.	7	7	3	1	1	1	1	1	1	1
23. Competing interests of guideline development group members have been recorded and addressed.	1	1	3	1	1	1	1	1	1	1
S23. Competing interests of guideline development group members have been recorded and addressed.[c]	S7	S7	S7	S7	S6	S5	S7	S5	S6	S5
Overall Score										
Rate the overall quality of this guideline	7	6	4	3	4	2	4	6	5	5
I would recommend this guideline for use[d]	Y	Y	M	M	Y	M	Y	M	Y	Y

[a] Each key item scored from 1–7. Domain score is the total of the key item scores for the domain.
[b] A.C.D., Andrew C. Don-Wauchope; J.L.S., John L. Sievenpiper; S.A.H., Stephen A. Hill; A.I., Alfonso Iorio.
[c] S before a number reflects the assessment of the supplemental file as described in the text.
[d] Y, yes; N, no; M, yes, with modifications.

and Fig. 1). The overall score for the Emergency Toxicology LMPG *(29)* was very low but it was still felt that it had some potential for use in a clinical context, with modifications by 1 of the 2 appraisers. A number of the other LMPGs *(24, 26)* received a "No" and "Yes, with modifications" on the final item of the overall evaluation. However, no LMPG was found not to be useful for clinical laboratory practice by both assessors. All had aspects that contributed to good laboratory practice and were considered to have a role in clinical practice.

There were 40 of 253 scores (16%) that differed by more than 4 on the 7-point scale. These required consideration for revision. Two domains [stakeholder involvement (27%) and editorial independence (22%)] had a higher percentage of differing scores (range (11%–27%). Most often the appraiser had not identified an item in the LMPG or had misinterpreted the information presented. After rescoring these 40 items, weighted (quadratic) κ demonstrated agreement of >0.5 for 9 of 11 LMPGs (Table 2).

Table 1. Data used for the AGREE II calculations of domain score for the Assessment of the NACB LMPGs.[a]
(Continued from page 1431)

Guideline											
Follow-up Testing for Metabolic Diseases Identified by Expanded Newborn Screening Using Tandem Mass Spectrometry		Use of Tumor Markers in Clinical Practice: Quality Requirements		Biomarkers of Acute Coronary Syndromes and Heart Failure		Evidence-Based Practice For Point-of-Care Testing		Maternal-Fetal Risk Assessment and Reference Values in Pregnancy		Recommendations for the Use of Laboratory Tests to Support Poisoned Patients Who Present to the Emergency Department	
S.A.H.	J.L.S.	S.A.H.	A.C.D.	A.C.D.	J.L.S.	A.I.	J.L.S.	A.C.D.	A.I.	A.I.	J.L.S.
5	4	1	3	6	4	6	5	6	6	6	4
7	6	10	12	12	8	4	7	5	10	7	6
2	3	2	4	5	3	1	2	2	4	1	3
2	1	4	6	5	2	1	2	1	2	2	1
2	1	2	1	1	2	1	2	1	3	2	1
1	1	2	1	1	1	1	1	1	1	2	1
3	2	3	2	8	7	3	3	3	6	5	4
1	1	1	1	7	6	2	2	2	5	3	3
2	1	2	1	1	1	1	1	1	1	2	1
S7	S7	S6	S5	S5	S6	n/a	n/a	n/a	n/a	n/a	n/a
3	2	2	3	4	4	6	5	2	4	1	2
M	M	N	M	Y	M	Y	M	N	M	N	M

SENSITIVITY ANALYSIS

When we compared domain scores for guidelines issued before or after 2009 using the Mann–Whitney test no significant changes were noted (Table 3). However, the direction of change for many of the domains was toward improvement, suggesting that a more standardized guideline development methodology was being followed. For example, the median applicability score improved from 13% to 29% ($P = 0.07$).

Discussion

INTERPRETATION OF DOMAIN SCORES

Interpreting the domain scores is subjective, but the percentage score can allow the user of the LMPG to gauge the overall utility of the LMPG.

Domain 1 (items 1–3) covers the overall aim of the document. The best example of a well-described aim was found in the Point-of-Care Testing LMPG (22) (92%), whereas the Expanded Newborn Screening LMPG (19) received the lowest score (31%). The item that was least clear in this domain was the population to whom the guideline is meant to apply (Table 1). The Diabetes guideline (23) describes this item very well (maximum score) whereas the Pharmacogenetics guideline (28) does not (minimum score). Domain 1 is important for LMPGs because it defines the area in which the LMPG should be found applicable in both laboratory and clinical practice. Without clear definition of this domain, the potential users of the guideline will find it difficult to assess its applicability to their situation.

Table 2. Assessment of the NACB LMPGs with the AGREE II Instrument.

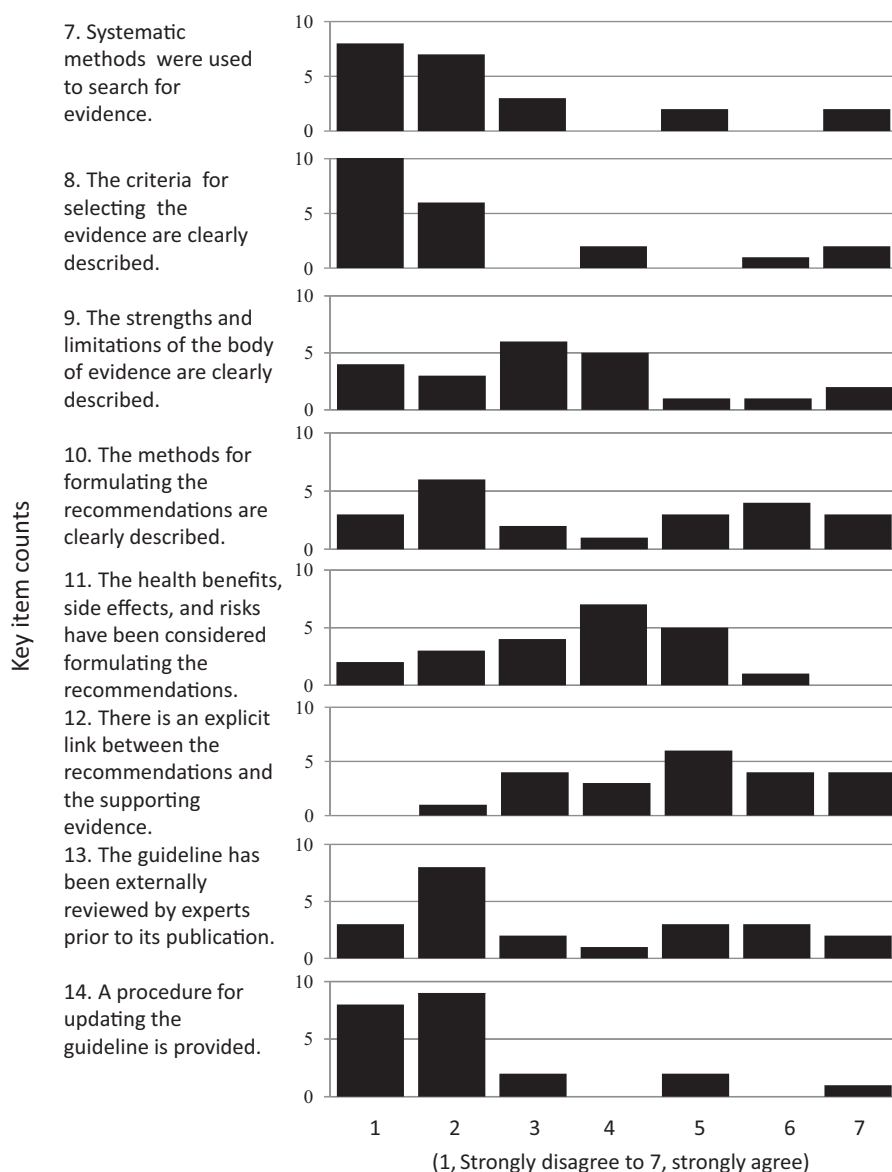| Guideline	Year of publication	Domain 1. Scope and purpose	Domain 2. Stakeholder involvement	Domain 3. Rigor of development	Domain 4. Clarity of presentation	Domain 5. Applicability	Domain 6. Editorial independence	Overall score	Weighted κ
Guidelines and Recommendations for Laboratory Analysis in the Diagnosis and Management of Diabetes Mellitus.	2011	89%	100%	85%	94%	67%	100%	92%	0.45
Use of Tumor Markers in Liver, Bladder, Cervical, and Gastric Cancers	2010	42%	33%	28%	58%	19%	58%	42%	0.76
Laboratory Analysis and Application of Pharmacogenetics to Clinical Practice	2010	64%	44%	28%	64%	35%	38%	33%	0.75
Use Of Tumor Markers In Testicular, Prostate, Colorectal, Breast, And Ovarian Cancers	2009	78%	67%	26%	81%	29%	42%	67%	0.60
Emerging Biomarkers for Primary Prevention of Cardiovascular Disease and Stroke	2009	83%	58%	54%	71%	27%	38%	67%	0.72
Follow-up Testing for Metabolic Diseases Identified by Expanded Newborn Screening Using Tandem Mass Spectrometry	2009	31%	11%	20%	39%	10%	50%	25%	0.50
Use of Tumor Markers in Clinical Practice: Quality Requirements	2009	50%	58%	32%	19%	29%	38%	25%	0.47
Biomarkers of Acute Coronary Syndromes and Heart Failure	2007	69%	56%	41%	72%	25%	83%	50%	0.70
Evidence-Based Practice for Point-of-Care Testing	2007	92%	50%	65%	78%	6%	8%	75%	0.84
Maternal-Fetal Risk Assessment and Reference Values in Pregnancy	2006	50%	33%	20%	78%	15%	21%	33%	0.75
Recommendations for the Use of Laboratory Tests to Support Poisoned Patients Who Present to the Emergency Department	2005	50%	42%	18%	67%	10%	21%	8%	0.61

7. Systematic methods were used to search for evidence.

8. The criteria for selecting the evidence are clearly described.

9. The strengths and limitations of the body of evidence are clearly described.

10. The methods for formulating the recommendations are clearly described.

11. The health benefits, side effects, and risks have been considered formulating the recommendations.

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

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

14. A procedure for updating the guideline is provided.

Key item counts

(1, Strongly disagree to 7, strongly agree)

Fig. 2. Counts of individual scale (1–7) scores for Domain 3 (Rigor of Development) showing key items 7–14.
The *y* axis is set to 10 individual counts (max 22 counts). Each item was scored twice for the 11 LMPGs.

Domain 2 (items 4–6) focuses on the inclusion of appropriate stakeholders in the development of the guideline to ensure that the guideline represents the view of the intended users of the guideline. With the exception of the diabetes guideline this section was not that well addressed in the LMPGs. Domain 2 had a higher percentage (27%) of individual scores that required reconsideration, suggesting that the appraisers found it difficult to locate or interpret the information presented. The recording of the affiliation and roles of each of the committee members and the views and

preferences of the public were poorly recorded in many of the LMPGs (Table 1). Domain 2 is important for LMPGs because the reader needs to be made aware that appropriate experts were involved in producing the document and that appropriate public representation and expertise were sought in the development of LMPGs. The list should include an affiliation, area of expertise, and what entity or organization the individual is representing on the guideline committee.

Domain 3 (items 7–14) assesses the methodology used to develop the guideline. Three items were partic-

Table 3. Comparison of LMPGs reported prior to 2009 against those reported after 2009.

	Pre-2009 (n = 4)		2009 onwards (n = 7)		Mann–Whitney test
	Mean (SD)	Median (IQR)[a]	Mean (SD)	Median (IQR)	P
Domain 1. Scope and purpose	65% (20%)	60% (32%)	63% (20%)	64% (39%)	0.79
Domain 2. Stakeholder involvement	45% (10%)	46% (16%)	53% (28%)	58% (30%)	0.53
Domain 3. Rigor of development	36% (22%)	30% (36%)	39% (23%)	28% (24%)	0.79
Domain 4. Clarity of presentation	74% (5%)	75% (9%)	61% (25%)	64% (34%)	0.41
Domain 5. Applicability	14% (8%)	13% (13%)	31% (18%)	29% (14%)	0.07
Domain 6. Editorial independence	33% (34%)	21% (19%)	52% (22%)	42% (16%)	0.16
Overall score	42% (28%)	42% (46%)	50% (25%)	42% (40%)	0.93

[a] IQR, interquartile range.

ularly poorly addressed in most of the LMPGs (Fig. 2). These were: (*a*) the use of systematic methods to find evidence; (*b*) descriptions of the methods used to evaluate and select the evidence; and (*c*) a description of the explicit plan to update the guideline. Domain 3 is crucial for the reader to assess the validity of the LMPG by enabling judgment of the underlying quality of evidence.

Domain 4 (items 15–17) covers the clarity of the presentation and structure of the guideline. The importance of domain 4 in any guideline is straightforward. Most of the guidelines were well written and clearly presented (Table 1). Item 16, describing the different options for management of the condition or health issue, was the item for which appraisers struggled to reconcile the explicit criteria for the AGREE II instrument with the laboratory-specific quality requirements guideline (*26*). This difficulty might be partly related to the specific field of clinical chemistry, but some attention to the offering and discussion of alternatives, such as screening, diagnostic, and prognostic test use, would certainly add value to the guidelines. However, item 16 is mostly based on the depth and breadth of the question for the LMPG. If the question is very laboratory specific it would probably be more difficult to meet the criteria for item 16. As reported in the results, this item was deemed not applicable in 1 guideline, because the question being addressed was focused on the area of quality of testing and not the utility of the laboratory test in clinical practice (*26*). However, if an LMPG addresses a clinical utility question, then explicitly stating the question should result in the ability to adequately address this domain.

Domain 5 (items 18–21) reviews the way in which the guideline describes the applicability of the recommendations to clinicians and clinical laboratory practice, including the identification of barriers and facilitators to implementation. This domain scored quite low for many of the LMPGs. Item 20 ("The potential resource implications of applying the recommendations have been considered.") was very poorly addressed in all the LMPGs (Table 1). Item 21 ("The guideline presents monitoring and/or auditing criteria.") was poorly addressed with the exception of the Diabetes LMPG. Domain 5 is an important section to consider in the development of LMPGs and we would encourage developers of LMPGs to address this domain in future guidelines.

Domain 6 (items 22 and 23) is concerned with the editorial independence and risk of bias of the document. Items 22 and 23 that deal with funding and conflicts of interest were initially scored low in most of the LMPGs (Fig. 1 and Table 1). This finding reflects the fact that online supplemental files reporting conflict of interest were not included or linked in the main report, and it was our opinion that this did not allow an easy assessment for the average user of their editorial independence. In addition the layout of the website has changed from December 2011, making the supplemental files a little easier to locate in May 2012. Table 1 demonstrates that scores for item 23 improved for 8 of 11 LMPGs when the supplemental files were considered. Reporting the conflicts of interest is important for readers to understand the relationship between the funding agency and the authors of the guideline as well as to determine independence. Explicit listing of potential conflicts of interest is a requirement for most publications and presentations. Production of guidelines can potentially influence a wide range of practice and it would be desirable to have any potential conflicts reported explicitly in the guideline. It would be preferable if the committee was selected from individuals who did not have substantive conflicts of interest. Representation from industry is important but should not influence guideline development.

The final evaluation of the guideline summarizes the overall opinion of the appraiser for the quality of the guideline and if the guideline is recommended for clinical use. This result represents the overall opinion of the appraiser (Fig. 1). It is important to note that the AGREE II instrument does not assess content validity of the LMPG (7, 17), let alone the quality of the evidence. What is important is that the methods by which the evidence was sought and weighted are clearly and reproducibly reported in the LMPG. The AGREE II instrument is designed to improve the reporting of LMPG methodology and to encourage clarity of presentation and transparent reporting of conflicts of interest. These improvements enable the reader to judge the validity of the CPG. A high score on the AGREE II instrument indicates that the information meets these criteria. A low score indicates that the information is not well presented or the methodology is not well reported. Of course a high-quality method does not eliminate the risk of relying on poor evidence, but a poor method does not allow the reader to appraise the value of evidence at all. We judged the quality of the process, including the quality of evidence selection and appraisal, and the quality and completeness of the process leading from evidence to recommendation. Direct assessment of the quality of evidence content is beyond the AGREE II instrument and was not assessed in this investigation.

UTILITY OF THE AGREE II INSTRUMENT

We found that the AGREE II instrument was useful in appraising 11 NACB LMPGs. The AGREE II domains and items were all applicable and appropriate in the evaluation of most LMPGs that consider the clinical application of laboratory tests. The AGREE II instrument was possibly less useful for the appraisal of LMPGs that considered purely technical, analytical, organizational, or quality aspects of clinical laboratory testing. It would be recommended that all groups writing LMPGs consider these criteria as they plan and prepare the LMPG. These findings concur with the previous use of the AGREE instrument in a laboratory medicine context (17).

Most of the LMPGs had overall scores <50%, which suggests that they were of lower quality. Although some of the domains were well covered and had higher mean scores, others were not. Table 3 demonstrates that the median overall score did not change, although many domains had changed. The most recent LMPG (Diabetes) (23), which reported the use of the AGREE II instrument in the development of the guidelines, scored the highest, reflecting the improvement in the way in which the guideline addressed the domains in the AGREE II instrument. One can argue that this guideline scored better as the obvious result of the instrument having been used in the planning phase of the guideline production, which is of course true. On the other hand, this assessment is not, in our opinion at least, biased, because the AGREE II is the only instrument available on a wide scale to improve the quality of CPGs. That we found an LMPG that successfully implements the AGREE II recommendations is actually a robust empirical demonstration that AGREE II can be used in the field of laboratory medicine.

RECOMMENDATIONS

For future LMPGs the following items should be considered more carefully, because these were items that could have been addressed but that scored poorly in this evaluation.

1. A specific description of the population to which the guideline (recommendation) applies is critical to the interpretation of the LMPG.
2. It is important for LMPGs to specifically describe who has produced the guideline and who has been consulted. Both expert and public opinion must be actively sought and reported. Detailed lists of funding sources and potential conflicts of the individuals and organizations involved in the LMPG should be made available as an embedded or linked appendix to the LMPG.
3. It is essential that every effort is made to ensure that all NACB LMPGs present their methodology in an explicit manner, with transparency and assessable external validity so that they become more accessible to both clinicians and clinical laboratories. Domain 3 requires the most improvement for future LMPGs. It would be recommended that clear and reproducible strategies for obtaining and appraising evidence are used and described.
4. A strategy for updating the LMPG must be clearly described in the document.
5. Writers of the LMPG should address factors that implicate resource use in relation to the recommendations of the guideline as described in item 20. This will facilitate translation of guidelines to local practice. In addition, the LMPG should suggest ways of monitoring the recommendations and the implementation of the recommendations.
6. Setting more focused questions for LMPGs may result in a document that is easier to digest and put into practice, but this approach might limit appraisal of the role and the value of a specific diagnostic procedure in the full scope of healthcare.

Conclusion

The NACB has selected topics relevant and important to laboratory medicine, and people working in clinical laboratories or clinical practice have considered these

LMPGs a source of good information. The AGREE and AGREE II instruments have been extensively adopted in other clinical fields to evaluate guidelines. These evaluations can be made available on the on-line database of the AGREE trust to people who wish to adapt a given guideline to their clinical practice *(13)*. Our findings show that the AGREE II instrument is useful for the purposes of evaluating LMPGs, as demonstrated by the variability of the scores we obtained and by 4 guidelines that were reported to have been formulated considering the AGREE instrument in the development process *(23, 25–27)*. This evaluation of the NACB LMPGs has also identified areas for improvement in most aspects of LMPG development and reporting, which could be taken into account in the NACB guideline process.

References

1. SIGN 50 a guideline developer's handbook. First revision. Edinburgh: Scottish Intercollegiate Guidelines Network; 2011.
2. Brozek JL, Akl EA, Jaeschke R, Lang DM, Bossuyt P, Glasziou P, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines: Part 2 of 3. The GRADE approach to grading quality of evidence about diagnostic tests and strategies. Allergy 2009;64:1109–16.
3. Davis D, Goldman J, Palda VA. Handbook on clinical practice guidelines. Ottawa: Canadian Medical Association; 2007.
4. Methods Guide for Medical Test Review. Rockville (MD): Agency for Healthcare Research and Quality (US); 2010. http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=558. (Accessed May 2012).
5. Higgins JPT, Green S, eds. Cochrane handbook for systematic reviews of interventions, version 5.1.0. The Cochrane Collaboration; 2011. http://www.cochrane-handbook.org/. (Accessed July 2012).
6. National Guideline Clearing House. Rockville (MD): Agency for Healthcare Research and Quality (US). http://www.guideline.gov. (Accessed May 2012).
7. Horvath AR, Kis E, Dobos E. Guidelines for the use of biomarkers: principles, processes and practical considerations. Scand J Clin Lab Invest Suppl 2010;70(Suppl 242):109–16.
8. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. CMAJ 2010;182:E839–42.
9. AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. Qual Saf Health Care 2003;12:18–23.
10. 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.
11. Horvath A, Nagy E, Watine J. Quality of guidelines for the laboratory management of diabetes mellitus. Scand J Clin Lab Invest Suppl 2005;65(Suppl 240):41–50.
12. Nagy E, Watine J, Bunting PS, Onody R, Oosterhuis WP, Rogic D, et al. Do Guidelines for the Diagnosis and Monitoring of Diabetes Mellitus Fulfill the Criteria of Evidence-Based Guideline Development? Clinical Chemistry 2008;54:1872–82.
13. AGREE Trust. http://www.agreetrust.org/ (Accessed February 2012).
14. Brozek J, Jankowski M, Placzkiewicz-Jankowska E, Jaeschke R. International Diabetes Federation document concerning postmeal glycemic control: assessment of quality of clinical practice guidelines using AGREE instrument. Pol Arch Med Wewn 2009;119:18–24.
15. De Weggheleire A, Bortolotti V, Zolfo M, Crowley S, Colebunders R, Riedner G, et al. Challenges in developing national HIV guidelines: experience from the eastern Mediterranean. Bull World Health Organ. 2011;89:442–50.
16. van Diermen DE, Aartman IHA, Baart JA, Hoogstraten J, van der Waal I. Dental management of patients using antithrombotic drugs: critical appraisal of existing guidelines. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2009;107:616–24.
17. Watine J, Friedberg B, Nagy E, Onody R, Oosterhuis W, Bunting P, et al. Conflict between guideline methodologic quality and recommendation validity: a potential problem for practitioners. Clin Chem 2006;52:65–72.
18. National Academy of Clinical Biochemistry laboratory medicine practice guidelines. http://www.aacc.org/members/nacb/pages/default.aspx# (Accessed May 2012).
19. Bennett MJ, ed. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines: Follow-up testing for metabolic diseases identified by expanded newborn screening using tandem mass spectrometry. Washington (DC); AACC Press; 2009.
20. Christenson RH, ed. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines: biomarkers of acute coronary syndromes and heart failure. Washington (DC): AACC Press; 2007.
21. Myers GL, ed. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines: emerging biomarkers For primary prevention of cardiovascular disease and stroke. Washington (DC): AACC Press; 2009.
22. Nichols JH, ed. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines: evidence-based practice for point-of-care testing. Washington (DC): AACC Press; 2006.
23. Sacks DB, ed. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines: guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. Washington (DC): AACC Press; 2010.
24. Sherwin JE, Lockitch G, Rosenthal P, Ashwood ER, Geaghan S, Magee LA, et al. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines: maternal-fetal risk assessment and reference values in pregnancy. Washington (DC): AACC Press; 2006.
25. Sturgeon CM, Diamandis E, eds. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines: use of tumor markers in testicular, prostate, colorectal, breast, and ovarian cancers. Washington (DC): AACC Press; 2009.
26. Sturgeon CM, Diamandis E, eds. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines: use of tumor markers in clinical practice: quality requirements. Washington (DC): AACC Press; 2009.
27. Sturgeon CM, Diamandis E, eds. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines: use of tumor markers in liver, bladder, cervical, and gastric cancers. Washington (DC): AACC Press; 2010.
28. Valdes R, Payne D, Linder MW, eds. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines: laboratory analysis and application of pharmacogenetics to clinical practice. Washington (DC): AACC Press; 2010.
29. Wu AHB, Broussard LA, Hoffman RS, Kwong TC, McKay C, Moyer TP, et al. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines: recommendations for the use of laboratory tests to support the impaired and overdosed patients from the emergency department. Clin Chem 2003;49:357–79.