

Conflict between Guideline Methodologic Quality and Recommendation Validity: A Potential Problem for Practitioners

JOSEPH WATINE,^{1*†} BRUNO FRIEDBERG,² EVA NAGY,³ RITA ONODY,³ WYTZE OOSTERHUIS,^{4‡}
PETER S. BUNTING,^{5†} JEAN-CHRISTOPHE CHARET,⁶ and ANDREA RITA HORVATH^{3§}

Background: It is not clear if good methodologic quality in current practice guidelines necessarily leads to more valid recommendations, i.e., those that are supported with consistent research evidence or, when evidence is conflicting or lacking, with sufficient consensus among the guideline development team. To help clarify this issue, we assessed whether there is a link between methodologic quality and recommendation validity in practice guidelines for the use of laboratory tests in the management of patients with non-small cell lung cancer (NSCLC).

Methods: We conducted a systematic review of data on laboratory tests in NSCLC published in English or in French within the last 10 years and retrieved 11 practice guidelines for the use of these tests. The guidelines

were critically appraised and scored for methodologic quality and recommendation validity based on the Appraisal of Guidelines Research and Evaluation (AGREE) criteria and on the systematic review.

Results: Overall, these 11 guidelines had considerable shortcomings in methodologic quality and, to a lesser extent, in recommendation validity. Practice guidelines with the best methodologic quality were not necessarily the most valid in their recommendations, and conversely.

Conclusions: Poor methodologic quality and lack of recommendation validity in laboratory medicine call for methodologic standards of guideline development and for international collaboration of guideline development agencies. We advise readers of guidelines to critically evaluate the methods used as well as the content of the recommendations before adopting them for use in practice.

© 2006 American Association for Clinical Chemistry

¹Laboratoire de Biologie Polyvalente, Centre Hospitalier Général de Rodez, Rodez, France.

²Laboratoire de Biologie Polyvalente, Centre Hospitalier Général de Wissembourg, Wissembourg, France.

³Department of Clinical Chemistry, University of Szeged, Szeged, Hungary.

⁴Department of Clinical Chemistry, Atrium Medical Centre, Heerlen, The Netherlands.

⁵Department of Pathology and Laboratory Medicine, The Ottawa Hospital, Ottawa, Ontario, Canada.

⁶Service de Pneumologie, Centre Hospitalier Général de Rodez, Rodez, France.

* Address correspondence to this author at: Laboratoire de Biologie Polyvalente, Hôpital Général, F-12027 Rodez Cédex 9, France. Fax 33-5-6575-1973; e-mail j.watine@ch-rodez.fr-watine61@hotmail.com.

† Member of, ‡ consultant to, and § Chair of the Committee on Evidence-Based Laboratory Medicine (C-EBLM) of the Education and Management Division (EMD) of the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC-LM). <http://www.ifcc.org/divisions/emd/c-eblm/aboutus.asp#2>.

Received July 1, 2005; accepted October 25, 2005.

Previously published online at DOI: 10.1373/clinchem.2005.056952

The methodologic quality of current practice guidelines must be improved (1–20). For many reasons, the methodologic quality of diagnostic guidelines is poorer than that of therapeutic guidelines, particularly in the field of laboratory medicine (10, 21, 22). It is not clear, however, whether these shortcomings in the methodology of guideline development necessarily lead to invalid recommendations, i.e., those that are not supported with consistent research evidence or sufficient consensus among the guideline development team when evidence is conflicting or lacking. We assessed to what extent methodologic quality is linked to recommendation validity in practice guidelines for the use of laboratory tests in the management of patients with non-small cell lung cancer

(NSCLC),⁷ specifically laboratory tests measuring quantities in biological specimens, thus excluding tissue (or anatomic) pathology tests.

Surgery performed at the early stages of NSCLC (I, II, or IIIA to a lesser extent) offers patients a reasonable chance of long-term survival, but this option is available to only a small minority of patients. In more advanced NSCLC (IIIB or higher), chemotherapy alone and chemotherapy with radiotherapy are options, but these therapies mostly aim to prolong patient survival, and the overwhelming majority of patients relapse. In many treatment facilities, no standard therapeutic schemes exist; therefore, controlled trials that include as many patients as possible are needed to assess the potential contribution of new drugs and new therapeutic schemes (23–26). Demonstrating the superiority of a given protocol over another is difficult to accomplish if the prognostic features of different patient subgroups cannot be compared. Consequently, independent prognostic factors must be identified before valid therapeutic trials can be designed, conducted, and interpreted (27). The medical and scientific communities have developed methods for conducting and reporting such therapeutic trials (28).

Materials and Methods

SEARCH FOR AND SELECTION OF GUIDELINES

The general principles of our manual and electronic search strategies have been described previously (21). All practice guidelines published in English or French within the last 10 years that provided advice for the use of laboratory tests in the management of NSCLC patients were selected for appraisal (29–39).

RECOMMENDATIONS IN PRACTICE GUIDELINES

Two of us (J.W. and B.F.) extracted all laboratory-related recommendations from the 11 guidelines selected for review (Table 1). In Table 1, the term “unclear recommendation” indicates either that the clinical decisions to be made based on the results of the recommended laboratory investigations were not precisely specified or that the names of the recommended laboratory tests themselves were not specified (e.g., the general term “biochemistry tests” was used).

Four organizations, the CIGNA HealthCare Medicare Administration (CIGNA), the European Group on Tumor

Markers (EGTM), the National Academy of Clinical Biochemistry (NACB), and the Société de Pneumologie de Langue Française (SPLF), focused on tumor markers only.

SPLF classifies 2 tumor markers in NSCLC, carcinoembryonic antigen (CEA) and cyfra 21-1, according to their “levels of scientific evidence”. SPLF considers the CEA level of scientific evidence as “not sufficient” for prognosis, staging, or surveillance, whereas the cyfra 21-1 level of evidence is considered sufficient for prognosis but not for staging or surveillance.

EGTM recommends the measurement of cyfra 21-1 before therapy and during posttherapy follow-up in NSCLC patients, and of CEA in cases of adenocarcinoma or large cell carcinoma. EGTM also stresses the independent prognostic value of cyfra 21-1, CEA, and CA 125 and of cyfra 21-1, CEA, and tissue-polypeptide antigen for monitoring therapy efficacy in NSCLC.

The NACB guidelines are very similar to those of EGTM except that they do not as clearly recommend that cyfra 21-1 or CEA be measured in NSCLC patients. CIGNA does not recommend the measurement of CEA, neuron-specific enolase (NSE), or cyfra 21-1.

The 7 other organizations, the American College of Chest Physicians (ACCP), Agence Nationale pour le Développement de l’Evaluation Médicale (ANDEM), American Society of Clinical Oncology (ASCO), American Thoracic Society, and European Respiratory Society (ATS-ERS), British Thoracic Society and Society of Cardiothoracic Surgeons of Great Britain and Ireland (BTS-SCG), Fédération Nationale des Centres de Lutte Contre le Cancer (FNCLCC), and the Scottish Intercollegiate Guidelines Network (SIGN) recommend the measurement of several different laboratory variables for the pretreatment evaluation of NSCLC patients (Table 1). ATS-ERS stresses the pretreatment prognostic significance of serum albumin and, to a lesser extent, that of serum calcium, particularly in case of advanced disease, whereas SIGN stresses the pretreatment prognostic significance of calcium, alkaline phosphatase (ALP), and liver function tests, and ASCO stresses the importance of lactate dehydrogenase (LD), hemoglobin, and leukocyte counts. ANDEM, ATS-ERS, and FNCLCC do not recommend the routine measurement of any laboratory variables other than those mentioned in Table 1, including serum tumor markers, in NSCLC patients.

SYSTEMATIC REVIEW OF THE EVIDENCE

We previously carried out a systematic review of the evidence (40, 41), which we updated last year (42). Recommendations in the 11 guidelines about the use of laboratory tests in NSCLC (Table 1) were compared with the findings of our systematic reviews (40–42).

In the management of NSCLC patients, laboratory tests can be useful in relation either to the disease itself or to the therapies administered. In relation to the therapies administered, if “routine chemistries and hematological tests” were to be taken into account, as summarized in

⁷ Nonstandard abbreviations: NSCLC, non-small cell lung cancer; EGTM, European Group on Tumor Markers; NACB, National Academy of Clinical Biochemistry; SPLF, Société de Pneumologie de Langue Française; CEA, carcinoembryonic antigen; NSE, neuron-specific enolase; ACCP, American College of Chest Physicians; ANDEM, Agence Nationale pour le Développement de l’Evaluation 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; FNCLCC, Fédération Nationale des Centres de Lutte Contre le Cancer; SIGN, the Scottish Intercollegiate Guidelines Network; ALP, alkaline phosphatase; LD, lactate dehydrogenase; and AGREE, Appraisal of Guidelines Research and Evaluation.

Table 1. Recommendations made in 11 clinical practice guidelines providing advice for the use of laboratory variables in the pretreatment management of NSCLC patients (29–39).

Guidelines	Recommended	Unclear recommendation	Not recommended
ACCP	Hematocrit, ALP, calcium, electrolytes, glucose, GGT, ^a SGOT	Other routine laboratory tests	None
ANDEM	Leukocyte count, albumin, SR, calcium, ALP, LD	None	Tumor markers
ASCO	Hemoglobin, leukocyte counts, LD, ALP, calcium	Other routine chemistries; liver function tests	LASA, CA 19-9, DNA index, DNA flow cytometric proliferation analysis, p53 tumor suppressor gene, ras oncogene
ATS-ERS	Blood counts, electrolytes, albumin, calcium, ALP, transaminases, bilirubin, creatinine	None	Tumor markers
BTS-SCG	Albumin, creatinine, glucose	None	None
CIGNA ^b	None	None	CEA, NSE, cyfra 21-1
EGTM ^b	cyfra 21-1, CEA ^c	CA 125, TPA	None
FNCLCC	Hemoglobin, leukocyte counts with differential, LD, albumin, calcium	None	Tumor markers
NACB ^b	None	cyfra 21-1, CEA, NSE	None
SIGN	ALP, calcium	Other biochemistry and hematology tests; liver function tests	None
SPLF ^b	None	cyfra 21-1	CEA

^a GGT, γ -glutamyl transferase; SGOT, glutamic-oxaloacetic transpeptidase; SR, sedimentation rate; LASA, lipid-associated sialic acid; TPA, tissue-polypeptide antigen.

^b Guidelines intended for tumor markers only.

^c Only in cases of adenocarcinoma or large cell carcinoma.

Table 1, virtually all 7 practice guidelines dealing with nontumor markers (ACCP, ANDEM, ASCO, ATS-ERS, BTS-SCG, FNCLCC, and SIGN) would probably agree that to evaluate toxicity or tolerance to the therapies administered to NSCLC patients, it may be necessary to measure hemoglobin, leukocyte counts, platelets, electrolytes, glucose, creatinine, transaminases, bilirubin, and albumin. On the basis of such a consensus, we have therefore considered that it is valid to recommend the measurements of these variables in NSCLC patients, particularly in patients suffering from advanced stages of disease.

In relation to the disease itself, almost all authors agree that laboratory tests (excluding pathology tests) currently have no clinical utility for NSCLC screening or diagnosis. In addition, 2 prognostic covariables are universally used in NSCLC patients: disease stage and performance status (23–25). Among other prognostic covariables that can be used for the stratification of NSCLC patients in trials, some authors use patient age and sex. Our systematic reviews of the evidence (40–42) indicated that the pretreatment prognostic values of blood hemoglobin, leukocyte counts with differential, serum LD, albumin, calcium, and, to a lesser extent, NSE are very likely to be independent of the aforementioned other covariables. There thus is sufficient evidence for recommending the measurement of at least all of these laboratory variables in all NSCLC patients participating in therapeutic trials (42). We also consider it valid, based on the evidence, to recommend the measurement of blood hemoglobin in patients treated

with radiotherapy (either inside or outside therapeutic trials), because the outcomes of patients with low blood hemoglobin are very likely to improve if they receive erythropoietin before radiation therapy (42, 43).

These guidelines also suggest that laboratory tests might be useful for the staging (pretreatment prognostic evaluation) and surveillance (posttreatment prognostic evaluation) of NSCLC. Seven guidelines (ACCP, ANDEM, ASCO, ATS-ERS, BTS-SCG, FNCLCC, and SIGN) thus recommend the use of diverse biochemistry and/or hematology tests for the staging of NSCLC. Abnormal result(s) in a patient with otherwise resectable (stage IIIA, or lower) NSCLC might indicate the presence of metastases and unresectable disease; therefore, all putative metastatic sites must be carefully investigated. All of the guidelines do not agree, however, on which laboratory tests are necessary under these circumstances. Some recommend the use of calcium and/or albumin, whereas others recommend ALP and/or LD (Table 1). Other guidelines (e.g., those of EGTM, NACB, or SPLF to a lesser extent) extend this recommendation to cyfra 21-1 (44). According to our own systematic review, however, there is more evidence to support the use of routine biochemical and hematologic tests (i.e., leukocyte counts with differential, LD, albumin, and calcium), or even to a lesser extent NSE, than cyfra 21-1, thus confirming that some new laboratory tests (e.g., cyfra 21-1 or other tumor markers for the management of NSCLC) may be introduced into routine practice before they are demonstrated to have greater validity than older, less expensive tests

(45,46). We therefore consider there to be sufficient evidence for recommending the measurement of leukocyte counts with differential, serum LD, albumin, and calcium, but that it is not valid to recommend the measurement of tumor markers in NSCLC patients, except perhaps for NSE in patients in chemotherapy trials.

In summary, available evidence suggests that the laboratory tests indicated in Table 2 should be performed for the pretreatment evaluation of NSCLC patients. Some laboratory tests recommended by one or several experts in the 11 practice guidelines are not mentioned in Table 2, e.g., ALP for the staging of NSCLC, because according to the systematic review of the evidence, it is quite clear that the prognostic value of ALP is inferior to, and is not independent of, that of LD, as summarized in Table 3.

APPRAISAL OF GUIDELINES

Scores for methodologic quality were assigned to each of the 11 guidelines (29–39), based on their critical appraisal using the Appraisal of Guidelines Research and Evaluation (AGREE) Instrument. The AGREE Instrument comprises 23 criteria, arranged in 6 domains (as shown in Table 4), covering the key elements of the guideline development process (47). Among the published appraisal checklists for practice guidelines, we chose the AGREE Instrument because it has shown the greatest potential as a tool for assessing recommendations for clinical pathways (48). This instrument has been endorsed by the WHO and the European Commission (21). In accordance with the AGREE recommendations, we assigned to each guideline 1 of 4 possible overall final scores: “very good” (the equivalent of “strongly recommend” of the AGREE Instrument) if the guideline rated high on the majority of items and most domain scores were >60%, indicating that the guideline had a high

Table 2. Laboratory variables that should be measured for the pretreatment evaluation of NSCLC patients, taking into account both the systematic review of the evidence and the consensual opinions of the development teams of current practice guidelines (29–39).

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

^a Recommendation validated by most experts and not contradicted by systematic review of the evidence.

^b Recommendation based on systematic review of the evidence.

Table 3. Pretreatment prognostic value of LD vs that of ALP in NSCLC patients.

	Results of multivariate analysis ^a		
	Not significant	Unspecified significance ^b	Significant
LD ^c	16 (2450)	1 (189)	9 (4900)
ALP	18 (7252)	0	1 (207) ^d

^a Number of studies (number of patients included in the studies).

^b Significant in multivariate analysis that did not take into account performance status or disease stage.

^c The prognostic value of LD seems to be more obvious in unresected patients than in resected patients.

^d In this study, LD was not included in the multivariate analysis, whereas all studies that compared the prognostic value of LD with that of ALP concluded that LD performed better than ALP [for explanations and more details, see Refs. (40–42)].

overall quality and could be considered for use in practice without alterations; “good” (the equivalent of “recommend with provisos or alterations” of the AGREE Instrument) if the guideline rated high or low on a similar number of items and most domain scores were 30%–60%, indicating that the guideline had a moderate overall quality; “not so good” (the equivalent of “would not recommend” of the AGREE Instrument) if the guideline rated low on the majority of items and most domain scores were <30%, indicating that the guideline had a low overall quality and serious shortcomings and thus should not be recommended for use in practice; and finally, “dubious” (the equivalent of “unsure” of the AGREE Instrument) if the guideline did not give sufficient information to enable us to assess its quality. We have chosen to use the scores “very good”, “good”, “not so good”, or “dubious”, rather than the original terminology of the AGREE Instrument, as indicated above, because we thought that this would lead to an easier understanding of our review.

Scores for validity of recommendations were also assigned to each guideline, based on a systematic review of the evidence (40), which has been updated twice (41, 42), also taking into account the consensual opinions of the guideline development teams, as summarized above in the section *Systematic Review of the Evidence*. The scale of ratings that we used was the same as for methodologic quality, consisting of 4 possible scores (very good, good, not so good, or dubious), as explained in more detail in the *Results* section. Two scores for validity of recommendations were assigned to each guideline: one for recommendations regarding tumor markers and one for recommendations regarding other laboratory tests.

HOW DISAGREEMENTS WERE RESOLVED

During the whole process of the study described in the 4 sections above (*Search for and Selection of Guidelines, Recommendations in Practice Guidelines, Systematic Review of the Evidence, and Appraisal of Guidelines*), disagreements between the 2 assessors (J.W. and B.F.) were resolved by

Table 4. AGREE scores for methodologic quality of NSCLC guidelines.^a

Guidelines	Domain scores, %					
	Scope and purpose	Stakeholder involvement	Rigor of development	Clarity and presentation	Applicability	Editorial independence
ACCP	61	46	60	46	6	75
ANDEM	89	25	10	71	0	25
ASCO	94	50	71	67	17	75
ATS-ERS	44	4	5	29	0	8
BTS-SCG	100	33	60	79	6	83
CIGNA ^b	67	13	12	54	11	8
EGTM ^b	44	4	2	29	0	0
FNCLCC	94	54	57	79	17	33
NACB ^b	50	17	29	54	11	25
SIGN	89	75	76	75	33	25
SPLF ^b	61	46	48	38	17	8

^a Maximum score for best quality = 100%.
^b Guidelines intended for tumor markers only.

consensus, and if necessary, a third person (J.C.C.) was available as a referee (this was never necessary). For methodologic quality, the consensual scores thus obtained were validated by an independent set of assessors (E.N. and R.O.), and when necessary a third person (A.R.H.) was used as a referee (this was necessary only once).

Results

SCORES FOR METHODOLOGIC QUALITY

Each guideline was scored for each of the 6 domains of the AGREE Instrument, as shown in Table 4. As can be seen in Table 4, some guidelines are better in some aspects and others in other aspects of guideline development method-

ology. The overall final scores obtained for each guideline are indicated in Table 5.

SCORES FOR VALIDITY OF RECOMMENDATIONS

The only guideline that clearly recommended the use of tumor markers (EGTM) was scored as not so good because there is no evidence that measurement of tumor markers in routine practice would improve NSCLC patient outcomes. The EGTM guideline was not attributed the worst possible score (dubious) because, as already stressed, tumor markers might be useful in therapeutic trials. The 4 guidelines that clearly did not recommend the use of tumor markers in routine practice (ANDEM, ATS-ERS, CIGNA, and FNCLCC) were scored as good. These 4 guidelines were not attributed the best possible score (very good) because they did not allude to the needs of patients in therapeutic trials. Guidelines that gave unclear recommendations regarding the use of tumor markers (NACB and SPLF) or that did not mention tumor markers at all (ACCP, ASCO, BTS-SCG, and SIGN) were also scored as not so good (Table 5).

Regarding the other laboratory tests, the 5 guidelines (ACCP, ANDEM, ASCO, FNCLCC, and ATS-ERS) in which only a few laboratory tests were missing among those recommended (compared with the reference list of tests in Table 2) were scored as good. SIGN and BTS-SCG guidelines were scored as not so good because their lists of recommended laboratory tests were clearly less evidence based than those in the 5 other guidelines (as can be seen in Table 1).

Discussion and Conclusions

It is universally acknowledged that the methodologic quality of practice guidelines must improve (1–22). For example, in a study that was probably one of the largest surveys ever done on the subject, Grilli et al. (7) assessed the quality of 431 practice guidelines produced by specialty societies and published in English-language

Table 5. Scores for methodologic quality and for validity of content of recommendations (scales adapted from the AGREE Instrument).^a

Guidelines	Methodologic quality	Validity of content of recommendations	
		Tumor markers	Other laboratory tests
ACCP ^b	Good	Not so good	Good
ANDEM ^{c,d}	Not so good	Good	Good
ASCO ^c	Good	Not so good	Good
ATS-ERS ^{c,d}	Dubious	Good	Good
BTS-SCG ^d	Good	Not so good	Not so good
CIGNA ^{b,d}	Dubious	Good	
EGTM ^{c,d}	Dubious	Not so good	
FNCLCC	Good	Good	Good
NACB ^d	Not so good	Not so good	
SIGN ^c	Good	Not so good	Not so good
SPLF ^{c,d}	Not so good	Not so good	

^a The best possible score (i.e., very good, the equivalent of strongly recommend of the AGREE instrument) was never attributed.

^b Guidelines published less than 3 years ago.

^c Guidelines published more than 5 years ago.

^d Diagnostic guidelines (the 4 other guidelines provide both diagnostic and therapeutic recommendations).

peer-reviewed journals in 1988–1998. Their assessment included 3 criteria: whether the report provided information on the types of professionals and stakeholders involved in the development process; the strategy to identify primary evidence; and an explicit grading of recommendations according to the quality of supporting evidence (7). These 3 criteria were not met at all in 67%, 88%, and 82% of these 431 guidelines, respectively, and all 3 criteria for quality were met in only 22 guidelines (5%) (7). It is not clear whether this situation will improve in the near future (49).

When we used the AGREE Instrument to assess the methodologic quality of 11 practice guidelines providing advice for the use of laboratory tests for the management of NSCLC, many fell short of basic quality criteria, a result that confirms the aforementioned observations (1–22).

Regarding our judgment that some of the recommendations made in the 11 guidelines are not entirely valid, it could be argued that our judgment might be incorrect because even evidence-based guidelines can contain different recommendations. Scientific evidence is only one of many factors that may influence the translation of research findings to the context of use. The process of considered judgment is essential in guideline development and often requires extensive discussions and consensus among experts (50). Availability of services, resources, and cost-effectiveness are important considerations (22) but do not apply to our study (Table 1). Recommendations about tumor markers are conflicting, and availability of services, resources, and cost-effectiveness do not support the use of any test that lacks evidence of clinical utility. In addition, some recommendations made in these 11 guidelines are unclear. Regarding these unclear recommendations, the 3 guidelines (ACCP, ASCO, and SIGN) in which the measurement of “other (routine) laboratory tests” was recommended (quite a vague recommendation indeed) offered both therapeutic and diagnostic recommendations; it is therefore possible that guideline development teams followed the right guideline development methods regarding therapeutic recommendations but did less well when formulating diagnostic ones. The AGREE Instrument, as a generic appraisal toolbox, did not allow us to investigate this possibility in more depth, and it was not possible for us to conclude whether the methodologic quality or validity of recommendations was better or worse in the 4 guidelines that offer both therapeutic and diagnostic recommendations than in the 7 other, purely diagnostic, guidelines, although the 3 guidelines that obtained the lowest possible scores regarding methodologic quality (dubious) were purely diagnostic guidelines (Table 5).

In summary, the clinical validity of the recommendations regarding laboratory tests made in some of the 11 guidelines can be questioned (at least in those scoring not so good in Table 5). Some authors have shown that guidelines of poor methodologic quality are more likely to provide invalid diagnostic recommendations than guide-

lines of high methodologic quality (10,18), but other authors have shown that guidelines of poor methodologic quality can provide diagnostic recommendations as valid as guidelines with high methodologic quality (2). In our study, practice guidelines with the best methodologic quality were not necessarily those that were the most valid in the content of their recommendations, and conversely. For example, ATS-ERS and CIGNA guidelines were valid in their recommendations, whereas their methodologic quality was poor, and SIGN and BTS-SCG guidelines were not valid in their recommendations, whereas their methodologic quality was good (Table 5). Because the AGREE Instrument does not involve the use of global scores to assess the methodologic quality of guidelines, we also looked at the individual scores obtained not only in each of the 6 domains (Table 4) but also in each of the 23 questions (data not shown), and again we were not able to establish any correlation between validity of content and methodologic quality in any of the 6 domains or in any of the 23 questions.

This result is worrisome because the busy practitioner confronted with conflicting guideline recommendations has no easy means to help in deciding which guideline should be trusted. 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 (51). 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. Fortunately, the shortcomings in methodologic quality seemed to be more frequent than those of the content validity (see Table 5). The discrepancy between methodologic guideline quality and clinical validity of recommendations is perhaps less obvious in therapeutic recommendations, in which the quality of evidence from randomized trials is higher, than in diagnostic recommendations, in which the level of evidence is generally much poorer. Another possibility is that in other areas of medicine, more valid laboratory recommendations in practice guidelines are available than in NSCLC. Whatever the true situation is, the results of our study call for the critical appraisal of guidelines providing both diagnostic and therapeutic recommendations in various medical areas. Such work is in progress within our team in the field of diabetes mellitus (21). On the basis of our studies and reports from the literature, however, we strongly advise colleagues to do similar studies in other areas of medicine before guideline recommendations are used in local practice.

Because FNCLCC guidelines obtained the best scores in all items used for comparison (Table 5), one could argue that the French authors of the present review were biased toward guidelines in their own language. Taking into account the fact that we had the opportunity of expert discussions with some of the authors of the FNCLCC guidelines before their guidelines were published [these

discussions have partly been published (52)], we rather believe that the authors of the FNCLCC guidelines are much more likely than the authors of the other guidelines [except perhaps for the authors of NACB guidelines who quoted us (37)] to have read and (partly) taken into account the results of systematic reviews available on this topic (40–42, 52).

According to Shekelle et al. (53), the point at which no more than 90% of the guidelines published by the US Agency for Healthcare Research and Quality are still valid is 3.6 years (95% confidence interval, 2.6–4.6 years). To assess this hypothesis, we checked whether there was better correlation between methodologic quality and validity of recommendations in the most recently published guidelines. We failed to find any difference (Table 5), which suggests that the hypothesis of Shekelle et al. (53) may be valid for US Agency for Healthcare Research and Quality guidelines and similar sorts of guidelines, particularly in those areas in which the evidence base of recommendations develops faster than in the field we investigated.

In conclusion, to overcome the methodologic shortcomings of current practice guidelines and to improve the validity of resulting recommendations, standardized methods for making evidence-based guideline recommendations in laboratory medicine must be disseminated. In particular, we need a unified system for grading diagnostic recommendations [such a work is in progress within the GRADE group (54)], as well as common standards for guideline reporting [such a work also seems to be in progress (55)], together with appropriate tools for guideline implementation. Finally, we need to educate our profession about the principles of evidence-based laboratory medicine and guideline development methods (22). We advise that guidelines be critically evaluated for methodology and content before recommendations are used in clinical practice.

This work was presented at the Edutrak session entitled “The Role of Evidence-Based Laboratory Medicine in Guidelines”, developed with AACC’s EBM Group (R. Christenson, Chairman) and IFCC C-EBLM (A.R. Horvath, Chairwoman), at the XIXth International Congress on Clinical Chemistry (July 25–27, 2005, Orlando, FL).

References

1. Boluyt N, Lincke CR, Offringa M. Quality of evidence-based pediatric guidelines. *Pediatrics* 2005;115:1378–91.
2. Burgers JS, Bailey JV, Klazinga NS, Van der Bij AK, Grol R, Feder G. Inside guidelines. Comparative analysis of recommendations and evidence in diabetes guidelines from 13 countries. *Diabetes Care* 2002;25:1933–9.
3. Burgers JS, Fervers B, Haugh M, Brouwers M, Browman G, Philip T, et al. International assessment of the quality of clinical practice guidelines in oncology using the Appraisal of Guidelines and Research and Evaluation Instrument. *J Clin Oncol* 2004;22:2000–7.
4. Cates JR, Young DN, Guerriero DJ, Jahn WT, Armine JP, Korbett AB, et al. An independent assessment of chiropractic practice guidelines. *J Manipulative Physiol Ther* 2003;26:282–6.
5. Cranney A, Waldegger L, Graham I, Man-Son-Hing M, Byszewski A, Ooi D. Systematic assessment of the quality of osteoporosis guidelines. *BMC Musculoskelet Disord* 2002;3:20. (available free of charge at <http://www.biomedcentral.com/content/pdf/1471-2474-3-20.pdf>).
6. Graham ID, Beardall S, Carter AO, Glennie J, Hebert PC, Tetroe JM, et al. What is the quality of drug therapy clinical practice guidelines in Canada? *CMAJ* 2001;165:157–63. (available free of charge at <http://www.cmaj.ca/cgi/reprint/165/2/157.pdf>).
7. Grilli R, Magrini N, Penna A, Mura G, Liberati A. Practice guidelines developed by specialty societies: the need for a critical appraisal. *Lancet* 2000;355:103–6.
8. Grimshaw GM, Khunti K, Baker R. Diagnosis of heart failure in primary care: an assessment of international guidelines. *Br J Gen Pract* 2001;51:384–6.
9. Harpole LH, Kelley MJ, Schreiber G, Toloza EM, Kolimaga J, McCrory DC. Assessment of the scope and quality of clinical practice guidelines in lung cancer. *Chest* 2003;123(1 Suppl):7S–20S.
10. Irani J, Brown CT, van der Meulen J, Emberton M. A review of guidelines on benign prostatic hyperplasia and lower urinary tract symptoms: are all guidelines the same? *BJU Int* 2003;92:937–42.
11. Littlejohns P, Cluzeau F, Bale R, Grimshaw J, Feder G, Moran S. The quantity and quality of clinical practice guidelines for the management of depression in primary care in the UK. *Br J Gen Pract* 1999;49:205–10.
12. McAlister FA, Campbell NR, Zarnke K, Levine M, Graham ID. The management of hypertension in Canada: a review of current guidelines, their shortcomings and implications for the future. *CMAJ* 2001;164:517–22.
13. McCormack J, Perry T Jr, Rangno R, van Breemen C, Wright JM, Baltzan M, et al. Assessing the quality of clinical practice guidelines. *CMAJ* 2002;166:168–9.
14. Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA* 1999;281:1900–5.
15. Sudlow M, Thomson R. Clinical guidelines: quantity without quality. *Qual Health Care* 1997;6:60–1.
16. van Tulder MW, Tuut M, Pennick V, Bombardier C, Assendelft WJ. Quality of primary care guidelines for acute low back pain. *Spine* 2004;29:E357–62.
17. Varonen H, Makela M. Practice guidelines in Finland: availability and quality. *Qual Health Care* 1997;6:75–9.
18. Vogel N, Burnand B, Vial Y, Ruiz J, Paccaud F, Hohlfeld P. Screening for gestational diabetes: variation in guidelines. *Eur J Obstet Gynecol Reprod Biol* 2000;91:29–36.
19. Ward JE, Grieco V. Why we need guidelines for guidelines: a study of the quality of clinical practice guidelines in Australia. *Med J Aust* 1996;165:574–6.
20. Wegman A, van der Windt D, van Tulder M, Stalman W, de Vries T. Nonsteroidal antiinflammatory drugs or acetaminophen for osteoarthritis of the hip or knee? A systematic review of evidence and guidelines. *J Rheumatol* 2004;31:344–54.
21. Horvath AR, Nagy E, Watine J. Quality of guidelines for the laboratory management of diabetes mellitus. *Scand J Clin Lab Invest Suppl* 2005;240:41–50.
22. 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.
23. Bruderman I. Bronchogenic carcinoma. In: Baum GL, Wolinsky E, eds. *Textbook of pulmonary diseases*, 5th ed. New York: Little, Brown and Company, 1994:1345–92.
 24. Ginsberg RJ. Continuing controversies in staging NSCLC: an analysis of the revised 1997 staging system. *Oncology* 1998;12(Suppl 2):51–4.
 25. Ihde DC, Pass HI, Glatstein EJ. Small cell lung cancer. In: De Vita VT, Hellman S, Rosenberg SA, eds. *Cancer. Principles and practice of oncology*, 4th ed. Philadelphia: JB Lippincott, 1993:723–58.
 26. Lebeau B. Analyse bibliographique quantitative de la recherche clinique française en thérapeutique des cancers broncho-pulmonaires. *Rev Mal Respir* 1998;15:185–9.
 27. Fielding LP, Fenoglio-Preiser CM, Freedman LS. The future of prognostic factors in outcome prediction for patients with cancer. *Cancer* 1992;70:2367–77.
 28. Moher D, Schulz KF, Altman DG. CONSORT GROUP (Consolidated Standards of Reporting Trials). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med* 2001;134:657–62.
 29. American College of Chest Physicians. Lung cancer guidelines. *Chest* 2003;123(1 Suppl):1S–156S.
 30. Agence Nationale pour le Développement de l'Évaluation Médicale. Conférence de Consensus. Bilan d'extension pré-thérapeutique du cancer bronchique non à petites cellules. http://www.uvp5.univ-paris5.fr/UV_MED/AC/Impression.asp?Nsubj=98.
 31. American Society of Clinical Oncology. Clinical practice guidelines for the treatment of unresectable non-small-cell lung cancer. Adopted on May 16, 1997 by the American Society of Clinical Oncology. *J Clin Oncol* 1997;15:2996–3018.
 32. American Thoracic Society and European Respiratory Society. Pre-treatment evaluation of non-small-cell lung cancer. *Am J Respir Crit Care Med* 1997;156:320–32.
 33. British Thoracic Society and Society of Cardiothoracic Surgeons of Great Britain and Ireland. BTS guidelines: guidelines on the selection of patients with lung cancer for surgery. *Thorax* 2001;56:89–108. <http://thorax.bmjournals.com/cgi/content/full/56/2/89> (accessed November 8, 2005).
 34. CIGNA HealthCare Medicare Administration. Tumor markers for diagnosis and management of cancer. Coverage position number 0172. http://www.cigna.com/health/provider/medical/procedural/coverage_positions/medical/mm_0172_coveragepositioncriteria_tumor_markers_for_diagnosis_mgmt_cancer.pdf (accessed November 8, 2005).
 35. European Group on Tumour Markers. Tumour markers in lung cancer: EGTM recommendations. *Anticancer Res* 1999;19:2817–9.
 36. Fédération Nationale des Centres de Lutte Contre le Cancer. Standards, Options et Recommandations 2000 pour la prise en charge des patients atteints d'un cancer bronchopulmonaire non à petites cellules. <http://www.fnclcc.fr/> (accessed November 8, 2005).
 37. National Academy of Clinical Biochemistry. Practice guidelines and recommendations for use of tumor markers in the clinic. Laboratory medicine practice guidelines. http://www.nacb.org/LMPG/Monograph_TumorMarkers.pdf (accessed November 8, 2005).
 38. Scottish Intercollegiate Guidelines Network. Management of lung cancer, pilot edition, February 1998. <http://www.sign.ac.uk/> (this pilot edition is available from the corresponding author upon request).
 39. Société de Pneumologie de Langue Française. Recommandations concernant l'utilisation des marqueurs tumoraux sériques dans la prise en charge des cancers bronchiques primitifs. *Rev Mal Respir* 1997;14(Suppl 3):5–39.
 40. Watine J. Prognostic evaluation of primary non-small cell lung carcinoma patients using biological fluid variables. A systematic review. *Scand J Clin Lab Invest* 2000;60:259–74.
 41. Watine J. Are laboratory investigations recommended in current medical practice guidelines supported by available evidence? *Clin Chem Lab Med* 2002;40:252–5.
 42. Watine J, Friedberg B, Charet JC. Laboratory variables and stratification of lung cancer patients: recommendations for the future (Therapeutic Trials and Clinical Practice Guidelines). In: Carafaro RL, ed. *Focus on lung cancer research*. Hauppauge, NY: Nova Science Publishers, 2004:173–203.
 43. Watine J, Bouarioua N. Anemia as an independent prognostic factor for survival in patients with cancer. *Cancer* 2002;94:2793–6; author reply 2796–7.
 44. Pujol JL, Molinier O, Ebert W, Daures JP, Barlesi F, Buccheri G, et al. CYFRA 21-1 is a prognostic determinant in non-small-cell lung cancer: results of a meta-analysis in 2063 patients. *Br J Cancer* 2004;90:2097–105.
 45. Sandberg S, Oosterhuis W, Freedman D, Kawai T. Systematic reviewing in laboratory medicine. Position paper from the IFCC committee on systematic reviewing in laboratory medicine. *J Int Fed Clin Chem* 1997;9:154–5.
 46. Durand-Zalesky I, Rymer JC, Roudot-Thoraval F, Revuz J, Rosa J. Reducing unnecessary laboratory use with new test request form: example of tumour markers. *Lancet* 1993;342:150–3.
 47. Appraisal of Guidelines for Research and Evaluation 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. (the AGREE Instrument is available at www.agreecollaboration.com).
 48. Vluyen J, Aertgeerts B, Hannes K, Sermeus W, Ramaekers D. A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. *Int J Qual Health Care* 2005;17:235–42.
 49. Hasenfeld R, Shekelle PG. Is the methodological quality of guidelines declining in the US? Comparison of the quality of US Agency for Health Care Policy and Research (AHCPR) guidelines with those published subsequently. *Qual Saf Health Care* 2003;12:428–34.
 50. Burgers JS, van Everdingen JJ. Beyond the evidence in clinical guidelines. *Lancet* 2004;364:392–3.
 51. Shekelle PG, Kravitz RL, Beart J, Marger M, Wang M, Lee M. Are nonspecific practice guidelines potentially harmful? A randomized comparison of the effect of nonspecific versus specific guidelines on physician decision making. *Health Serv Res* 2000;34:1429–48.
 52. Watine J, Charet JC. Faut-il suivre les sociétés savantes de pneumologie dans leurs recommandations concernant l'utilisation des marqueurs tumoraux sériques dans l'évaluation pronostique des cancers bronchiques primitifs non à petites cellules? *Rev Mal Respir* 1999;16:139–49.
 53. Shekelle PG, Ortiz E, Rhodes S, Morton SC, Eccles MP, Grimshaw JM, et al. Validity of the Agency for Healthcare Research and Quality clinical practice guidelines: how quickly do guidelines become outdated? *JAMA* 2001;286:1461–7.
 54. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490.
 55. Shiffman RN, Shekelle P, Overhage JM, Slutsky J, Grimshaw J, Deshpande AM. Standardized reporting of clinical practice guidelines: a proposal from the Conference on Guideline Standardization. *Ann Intern Med* 2003;139:493–8.