Critical Evaluation of Oncology Clinical Practice Guidelines

Bradley N. Reames, Robert W. Krell, Sarah N. Ponto, and Sandra L. Wong

See accompanying editorial on page 2530

All authors: University of Michigan, Ann Arbor, MI.

Published online ahead of print at www.jco.org on June 10, 2013.

Presented as a poster discussion at the 48th Annual Meeting of the American Society of Clinical Oncology, Chicago, IL, June 1-5, 2012.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: Sandra L. Wong, MD, MS, Division of Surgical Oncology, Department of Surgery, 1500 E. Medical Center Dr, 3310 Cancer Center, Ann Arbor, MI 48109-5932; e-mail: wongs@umich.edu

© 2013 by American Society of Clinical Oncology

0732-183X/13/3120w-2563w/\$20.00

DOI: 10.1200/JCO.2012.46.8371

ABSTRACT

Purpose

Significant concerns exist regarding the content and reliability of oncology clinical practice guidelines (CPGs). The Institute of Medicine (IOM) report "Clinical Practice Guidelines We Can Trust" established standards for developing trustworthy CPGs. By using these standards as a benchmark, we sought to evaluate recent oncology guidelines.

Methods

CPGs and consensus statements addressing the screening, evaluation, or management of the four leading causes of cancer-related mortality in the United States (lung, breast, prostate, and colorectal cancers) published between January 2005 and December 2010 were identified. A standardized scoring system based on the eight IOM standards was used to critically evaluate the methodology, content, and disclosure policies of CPGs. All CPGs were given two scores; points were awarded for eight standards and 20 subcriteria.

Results

No CPG fully met all the IOM standards. The average overall scores were 2.75 of 8 possible standards and 8.24 of 20 possible subcriteria. Less than half the CPGs were based on a systematic review. Only half the CPG panels addressed conflicts of interest. Most did not comply with standards for inclusion of patient and public involvement in the development or review process, nor did they specify their process for updating. CPGs were most consistent with IOM standards for transparency, articulation of recommendations, and use of external review.

Conclusion

The vast majority of oncology CPGs fail to meet the IOM standards for trustworthy guidelines. On the basis of these results, there is still much to be done to make guidelines as methodologically sound and evidence-based as possible.

J Clin Oncol 31:2563-2568. © 2013 by American Society of Clinical Oncology

INTRODUCTION

Clinical practice guidelines (CPGs) and consensus statements are widely used to inform decisions about evaluation and treatment. Health care providers rely on these documents to implement evidence-based practice. To date, the Agency for Healthcare Research and Quality National Guideline Clearinghouse contains more than 2,500 CPGs, and the Guidelines International Network contains more than 6,500 CPGs.^{1,2} Studies have shown that within the field of oncology, CPGs can have a measureable positive effect on clinical practice and outcomes,³ and the demand for more guidelines on a broader range of topics continues to increase.⁴

Accordingly, the scrutiny of the methodology used to create CPGs and their content has increased. Previous evaluations of CPG development processes have concluded that many CPGs lack quality by failing to meet widely accepted standards, which

may undermine their clinical utility.^{5,6} To address this, several organizations have proposed standards for guideline development,⁷⁻¹¹ and in 2011, the Institute of Medicine (IOM) published a set of standards for developing rigorous, trustworthy CPGs.¹²

Given a growing emphasis on evidence-based medicine together with increasing measurement of adherence to guidelines, the importance of ensuring high-quality CPG production is clear. In light of the IOM standards for developing trustworthy guidelines, we sought to evaluate recent oncology guidelines to determine their clinical trustworthiness and to inform directions for future guideline development.

METHODS

A systematic MEDLINE literature search (Table 1) was performed to identify oncology CPGs published between

Table 1. Systematic Search Strategy for Identification of Clinical Practice Guidelines and Consensus Statements

Query	Search Terms
1.	Recommendation(s) OR consensus, in title
2.	Society(s) OR college(s) OR association(s), in title
3.	Query 1 AND 2
4.	Practice guideline OR consensus development conference (MeSH or publication type)
5.	Query 3 OR 4
6.	Limit 5 to English language, humans, and year = 2005-2010
7.	Limit 6 to breast neoplasms, colorectal neoplasms, thoracic neoplasms, or prostatic neoplasms

Abbreviation: MeSH: Medical Subject Heading.

January 2005 and December 2010 for the leading causes of cancer-related mortality: non-small-cell lung cancer (NSCLC), prostate cancer, and colorectal cancer for men; NSCLC, breast cancer, and colorectal cancer for women.¹³ The initial search strategy used to identify CPGs has been described previously.⁶ For breast, colorectal, and prostate cancers, only adenocarcinomas were considered, and other histologic subtypes were excluded. To be included, a CPG had to contain information and recommendations pertaining to the screening, diagnosis, treatment, or follow-up of one of the four selected cancer types. CPGs limited to primary prevention, those not written in English, and duplicates were excluded. If a development group produced updated guidelines addressing the same subject during the study period, only the most recent guideline was included. Selected guidelines were independently reviewed.

We scored each guideline by using the eight standards outlined by the IOM's "Clinical Practice Guidelines We Can Trust."¹² All CPGs were given two scores by each reviewer; points were awarded for each of a possible eight standards and for each of 20 subcriteria within the standards (Table 2). Fulfillment of a particular criterion was judged on a dichotomous basis. A CPG was considered to have met a particular standard only if it fulfilled all of a given standard's subcriteria. References made to online resources, documentation, or methods set forth by the guideline development groups (GDGs) were used to determine fulfillment of subcriteria if there was explicit mention of additional information or documentation within the guideline. Additional online information was used to assess fulfillment as long as a relatively straightforward search for the material (what a reasonable clinician could elucidate with investigation on the GDG's Web site) could locate the information.

Credit was awarded if the reviewer believed a reasonable attempt at fulfillment was made. For example, for standard 5 (and subcriteria 5.1), "Established evidence foundations for and rating strength of recommendations," a point was awarded if the guideline provided a grade or score for level of evidence and assessed the strength of the recommendation, but a point was not awarded if such statements were not made. For standard 2 (with four subcriteria), "Management of conflict of interest (COI)" disclosure statements were taken at face value. If the CPG stated "No conflicts of interest to report," then the standard was met in full, including points for each of the four subcriteria. If there were COIs reported, then subcriteria 2.2 and 2.3 were fulfilled only if the CPG contained COI reporting statements and addressed divestment, respectively.

The scores from each independent reviewer were tabulated, and summary statistics were generated. The level of agreement of the four independent reviews was determined by calculating a weighted kappa to account for the ordinality of the scoring system.¹⁴ At the end of the review process, questions about scores were adjudicated by all reviewers. Compliance with a given subcriterion by a CPG was defined as fulfillment of that subcriterion as determined by three of four independent reviewers.

RESULTS

A systematic MEDLINE search identified 593 potentially eligible CPGs. Applying our defined inclusion criteria resulted in the exclu-

Table 2. IOM Standards and Subcriteria

Eight IOM Standards and 20 Subcriteria

1. Establishing transparency
 - 1.1 Funding and development should be explicitly stated
2. COIs
 - 2.1 COIs should be declared prior to GDG formation
 - 2.2 All COIs should be reported and discussed
 - 2.3 GDG members should divest COIs
 - 2.4 Members who have COIs should be < 50% of the GDG
3. GDG composition
 - 3.1 GDG should be multidisciplinary and balanced
 - 3.2 Patients and public should be represented on GDG
 - 3.3 These representatives should be trained
4. Systematic review
 - 4.1 Systematic reviews should be used
 - 4.2 GDG and systematic review team (if used) should communicate
5. Evidence foundations for and rating strength of evidence
 - 5.1 Strength of recommendations and grading of evidence should be explicitly stated
6. Articulation of recommendations
 - 6.1 Articulate recommendations in standard form
 - 6.2 Strong recommendations should be worded as such
7. External review
 - 7.1 External review should include full spectrum of stakeholders
 - 7.2 Authorship of external review is confidential
 - 7.3 GDG should consider all external review comments
 - 7.4 Final draft of CPGs should be available for public comment
8. Updating
 - 8.1 Proposed date of future CPG review should be documented
 - 8.2 Literature pertaining to CPG should be monitored regularly
 - 8.3 CPG should be updated if new literature suggests modification

NOTE. Data adapted.¹²

Abbreviations: COI, conflict of interest; CPG, clinical practice guideline (and consensus statement); GDG, guideline development group; IOM, Institute of Medicine.

sion of 424 documents, resulting in 169 CPGs for review (complete bibliography shown in the Data Supplement). Reasons for exclusion included subject matter other than the four selected cancer types or a focus on something other than cancer diagnosis, treatment, or follow-up (287 [67.7%]); determination that the document was not a guideline (103 [24.3%]); an out-of-date CPG (27 [6.4%]); or duplicate publication (seven [1.7%]). Baseline characteristics of included studies are as shown in Table 3.

The majority of included CPGs were actually designated as "practice guidelines," but 47 (27.8%) of the publications were labeled as consensus statements. There was an increase in CPG publication over time, with the majority (60.4%) of CPGs published after 2007. There was an even distribution of CPGs from US groups (75 [44.4%]) and international groups (94 [55.6%]) over the study period. Evaluation and/or treatment of breast adenocarcinoma accounted for 59 (34.9%), colorectal for 41 (24.3%), NSCLC for 37 (21.9%), and prostate adenocarcinoma for 32 (18.9%) of CPGs. Nine professional societies or CPG development groups produced at least four CPGs during the study period, accounting for 70 (41.4%) of all documents. Forty-nine (29.0%) of the CPGs addressed multiple aspects of care, and the remaining 71% addressed a single aspect of cancer treatment.

Results of aggregate CPG scoring are shown in Table 4. Scores from independent reviewers demonstrate minimal inter-rater variability. Paired weighted kappa results ranged between 0.86 and 0.94

Table 3. Baseline Characteristics of Included CPGs

Characteristic	No.	%
Total	169	100
Type of publication		
Clinical practice guideline	122	72.2
Consensus statement	47	27.8
Time of publication		
2005-2007	67	39.6
2008-2010	102	60.4
Cancer type		
Breast	59	34.9
Colorectal	41	24.3
Lung	37	21.9
Prostate	32	18.9
Location of publication		
United States	75	44.4
International	94	55.6
Production by CPG development group		
Four or more CPGs produced during study period	70	41.4
Less than four CPGs produced during study period	99	58.6
Scope of CPG		
Multi-aspect guideline	49	29.0
Single-aspect guideline	120	71.0
Risk	2	1.7*
Screening	22	18.3*
Diagnosis	21	17.5*
Staging	5	4.2*
Treatment	66	55.0*
Follow-up	4	3.3*

Abbreviation: CPG, clinical practice guideline (and consensus statement).

*Percentage of 120 single-aspect CPGs.

across the two scoring systems. The light kappa statistic, an average of the pairwise weighted kappas, was 0.87 for the rating of eight standards and 0.89 for 20 subcriteria of the standards. No single CPG fully met all the IOM standards, and the overall average scores $\pm$ standard deviation were 2.75 ± 1.72 of eight possible major criteria and 8.24 ± 4.45 of 20 possible subcriteria. CPGs published before 2008 had slightly higher scores than those published after 2008.

Subgroup analyses demonstrated that guidelines focused on lung cancer had higher average scores compared with other cancer types (mean, 11.4 points for 20 criteria met *v* 7.4 points for breast cancer, 7.5 for colorectal cancer, and 7.0 for prostate cancer). CPG compliance with each standard and subcriteria was reviewed (Fig 1). CPGs from US groups scored higher than those from international groups ($9.9 \nu 7.0$ of 20; Fig 2A). CPGs for lung cancer had better overall scores than CPGs for other cancers; this difference was most pronounced for standards 1, 2, and 3 (Fig 2B), which related to transparency of the development process, management of conflicts of interest, and composition of the GDGs. Examination of CPG development based on a group's productivity revealed a substantial difference in CPG scores. CPGs produced by groups that published at least four guidelines during the study period had higher overall scores compared with those produced by groups that published fewer CPGs ($11.0 \nu 6.3$ of 20 points; Fig 2C). If consideration is given only to formal practice guidelines, they performed better than the subgroup of consensus statements, and the differences are particularly pronounced for standards

Table 4. Scoring Summary of Included CPGs

CPG Characteristic	Score $\pm$ SD of Eight Standards	Score $\pm$ SD of 20 Subcriteria
Total	2.75 ± 1.72	8.24 ± 4.45
Type of publication		
Clinical practice guideline	3.10 ± 1.70	9.10 ± 4.50
Consensus statement	1.84 ± 1.39	5.99 ± 3.46
Year of publication		
2005-2007	3.01 ± 2.1	8.78 ± 5.5
2008-2010	2.58 ± 1.4	7.88 ± 3.6
Cancer type		
Breast	2.46 ± 1.49	7.44 ± 3.95
Colorectal	2.57 ± 1.56	7.51 ± 4.03
Lung	3.80 ± 2.03	11.36 ± 4.72
Prostate	2.41 ± 1.52	7.03 ± 3.90
Location of publication		
International	2.28 ± 1.53	6.95 ± 3.83
United States	3.40 ± 1.77	9.89 ± 4.66
Production of CPG development group		
Four or more CPGs	3.71 ± 1.58	10.96 ± 4.19
Less than four CPGs	2.11 ± 1.51	6.31 ± 3.54

Abbreviations: CPG, clinical practice guideline (and consensus statement); SD, standard deviation.

3, 5, and 6, which relate to the multidisciplinary composition of the GDGs, strength of evidence, and articulation of recommendations (Fig 2D).

There was generally poor compliance with IOM standards for all of the CPGs evaluated, and only four of the 20 subcriteria (2.1, 3.1, 6.1, and 6.2) were met by more than half of all CPGs (Fig 1). As previously mentioned above, groups producing higher numbers of CPGs during the study period had higher overall scores than those producing fewer than four guidelines.

DISCUSSION

CPGs are used frequently in the care of the oncology patient and are thought to have substantial effects on practice patterns and patient outcomes.³ As the body of evidence for cancer treatment expands and there is more emphasis on algorithms to reduce unwarranted variations in care, patient and provider demand for reliable guidelines will continue to increase.⁴ The recent IOM publication of standards for CPG development may be a major step toward establishing "trustworthy" guidelines. However, our results find that none of the currently existing CPGs that address the screening, evaluation, or management of the four leading causes of cancer-related mortality in the United States fulfilled all of the IOM standards or subcriteria.

The IOM standards emphasize numerous widely accepted essentials of guideline development: transparency in funding, processes, and conflicts of interest to fully inform potential readers; multidisciplinary guideline panel composition to ensure a comprehensive examination of the subject; systematic reviews of the evidence; precise articulation of the recommendations made; and regular updating to ensure relevancy. Our analysis used these standards as a benchmark and revealed that more than half the current CPGs evaluated fall short in most of these areas (Fig 1), noting that no attempt was made to

Fig 1. Compliance with Institute of Medicine (IOM) subcriteria for all included clinical practice guidelines (CPGs) and consensus statements.

interpret the quality of the IOM standards. Particularly poor performance was noted in patient representative inclusion (standard 3), formalized systematic review of the literature (standard 4), peer and public review (standard 7), and regular updates of guidelines (standard 8).

Interestingly, a recent study by Kung et al¹⁵ also evaluated the performance of 130 randomly selected CPGs with regard to the IOM standards. Although their methodology was slightly different because of a modified scoring system that excluded seven “vague and subjective” IOM standards (eg, use of external review and regular

Fig 2. Clinical practice guideline (CPG) and consensus statement compliance with Institute of Medicine (IOM) standards, stratified by (A) location of publication, (B) type of cancer, (C) CPG and consensus statement development group productivity, and (D) type of document. US, United States.

monitoring of the literature by GDGs), their overall findings were similar to ours. Both evaluations of CPGs found that there was poor adherence to IOM standards, particularly with regard to COI management. The CPGs in their study had an overall median score of 44.4%, which is similar to the aggregate scores in this study, even with our scoring system, which included all IOM standards. Although we focused on an exhaustive review of all CPGs for only four common cancers, it is not surprising that guidelines across all diseases would fare similarly in terms of the guideline development process.

Although the aggregate results of this evaluation demonstrate lack of fulfillment of the IOM standards, there was considerable variation among individual CPGs. Several CPG development groups warrant mention as exemplars, given their relatively high adherence to IOM standards. For example, the American College of Chest Physicians (ACCP) produced 16 guidelines that met inclusion criteria within the defined 6-year period. All 16 guidelines scored in the top 20% of guidelines evaluated. This largely explains why guidelines addressing NSCLC performed better than those of other groups. Similar exemplars include the six guidelines produced by the American Society of Clinical Oncology (ASCO), which also scored in the top 20%. Eight guidelines published by the National Comprehensive Cancer Network (NCCN) also scored reasonably well. Notably, even among these exemplars, the ACCP guidelines inadequately discussed updating measures, the ASCO and NCCN guidelines failed to include the public in the external review process, and the NCCN guidelines formulated their recommendations on the basis of expert consensus instead of using systematic reviews.

This study has several limitations. One limitation is our heavy reliance on accurate representation of the actual CPG development process and content within the CPG document. It is possible that CPGs within this study fulfilled more standards and subcriteria than they were given credit for, although reasonable efforts were made by reviewers to find additional pertinent information via Web sites and external searches. A second limitation is the use of a dichotomous scoring of standard and subcriteria fulfillment. No points were given for partial fulfillment of stated standards, which could result in a relative dampening of overall scores toward more conservative (lower) overall values. However, a far more common event was that CPG processes were ambiguous, but acknowledged, and credit was awarded in these situations. Giving some CPG processes the benefit of the doubt actually resulted in higher overall scores than a more stringent evaluation might have allowed. Despite this, overall performance was poor. The unavoidable component of subjectivity present in the scoring process could also be considered a minor limitation. However, inter-rater reliability was high, confirming the ability of reviewers to effectively and consistently grade CPGs.

Although the methodology previously mentioned yields an evaluation of the quality of current oncology guidelines based on IOM standards, it begs the question of how important strict adherence to the IOM standards is for clinicians seeking trustworthy guidelines. Future studies will be able to evaluate the impact of the IOM publication on CPG development going forward, although studies have shown that there has been little improvement in adherence to guidelines over time.^{5,15,16} Certain requirements of the IOM, such as divestment for management of COIs, involvement of laypeople in the development panel, and requirement of public comment during the external review process, were missed by almost all CPGs in this study. It may be impractical for GDGs that address highly specialized treatment modalities (target volumes for postresection radiation fields, for example) to include laypeople on their development and review committees. These requirements add substantially to the administrative and financial burden of guideline development without direct evidence to support that meeting the IOM standards would lead to improved guidelines and improved clinical care. Consideration should be given to development of high-quality guidelines using more pragmatic standards or to a more thorough evaluation of the standards in order to determine which ones are most likely to produce the desired results when strictly adhered to.

In conclusion, on the basis of recently published IOM standards, there are substantial deficiencies in the current body of oncology CPGs for the four leading causes of cancer-related mortality in the United States. Although the vast majority of recently published oncology CPGs fail to meet IOM standards for trustworthy guidelines, many quality guidelines do exist and could serve as models for future CPG development. On the basis of these results, there is still much to be done to make current and future guidelines as methodologically sound and evidence-based as possible.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The author(s) indicated no potential conflicts of interest.

AUTHOR CONTRIBUTIONS

Conception and design: Sandra L. Wong

Administrative support: Sandra L. Wong

Collection and assembly of data: All authors

Data analysis and interpretation: All authors

Manuscript writing: All authors

Final approval of manuscript: All authors

REFERENCES

1. Agency for Healthcare Research and Quality: National Guideline Clearinghouse Guideline Index. <http://guideline.gov/browse/index.aspx?alpha=A>
2. Guidelines International Network: Guidelines International Network homepage. <http://www.g-i-n.net/library>
3. Winn RJ: Oncology practice guidelines: Do they work? *J Natl Compr Canc Netw* 2:276-282, 2004

4. Somerfield MR, Einhaus K, Hagerty KL, et al: American Society of Clinical Oncology clinical practice guidelines: Opportunities and challenges. *J Clin Oncol* 26:4022-4026, 2008
5. Shaneyfelt TM, Mayo-Smith MF, Rothwangl J: Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA* 281:1900-1905, 1999
6. Grilli R, Magrini N, Penna A, et al: Practice guidelines developed by specialty societies: The need for a critical appraisal. *Lancet* 355:103-106, 2000

7. Lohr KN, Field MJ: A provisional instrument for assessing clinical practice guidelines (Appendix B), in *Guidelines for Clinical Practice: From Development to Use*. Washington, DC, National Academies Press, 1992
8. Hayward RS, Wilson MC, Tunis SR, et al: More informative abstracts of articles describing clinical practice guidelines. *Ann Intern Med* 118:731-737, 1993
9. Liddle J, Williamson M, Irwig L: Method for Evaluating Research Guideline Evidence. New South Wales Health Department, Sydney, Australia, 1996

10. Cluzeau FA, Littlejohns P, Grimshaw JM, et al: Development and application of a generic methodology to assess the quality of clinical guidelines. *Int J Qual Health Care* 11:21-28, 1999

11. Fervers B, Burgers JS, Haugh MC, et al: Predictors of high quality clinical practice guidelines: Examples in oncology. *Int J Qual Health Care* 17: 123-132, 2005

12. Institute of Medicine: Clinical Practice Guidelines We Can Trust. Washington, DC, National Academies Press, 2011, pp 1-300

13. Siegel R, Naishadham D, Jemal A: Cancer statistics, 2012. *CA Cancer J Clin* 62:10-29, 2012

14. Banerjee M, Capozzoli M, McSweeney L, et al: Beyond kappa: A review of interrater agreement measures. *Can J Stat* 27:3-23, 1999

15. Kung J, Miller RR, Mackowiak PA: Failure of clinical practice guidelines to meet Institute of Medicine standards: Two more decades of little, if any, progress. *Arch Intern Med* 172:1628-1633, 2012

16. Shaneyfelt T: In guidelines we cannot trust: Comment on "Failure of Clinical Practice Guidelines to Meet Institute of Medicine Standards". *Arch Intern Med* 172:1633-1634, 2012

ASCO's Quality Care Symposium

Each year, ASCO, in conjunction with our cosponsors, organizes a wide array of high-quality meetings, providing educational and scientific programs to advance your understanding of cancer. Join us for one or more of ASCO's meetings to interact with oncology experts, network with colleagues, and earn CME credit.

ASCO's Quality Care Symposium (November 1-2, 2013, San Diego, CA), brings together top leaders in the field to address critical questions, methods, and strategies in the expanding field of health services research and oncology outcomes.

To learn more, visit quality.asco.org.

American Society of Clinical Oncology