

Evaluation of the NCCN guidelines using the RIGHT Statement and AGREE-II instrument: a cross-sectional review

Cole Wayant,¹ Craig Cooper,² D'Arcy Turner,³ Matt Vassar²

10.1136/bmjebm-2018-111153

¹Biomedical Sciences, Oklahoma State University Center for Health Sciences, Tulsa, Oklahoma, USA

²Psychiatry and Behavioral Sciences, Oklahoma State University Center for Health Sciences, Tulsa, OK, USA

³Medical Education, Kansas City University of Medicine and Biosciences, Kansas City, Kansas, USA

Correspondence to: **Cole Wayant**, Oklahoma State University Center for Health Sciences, Tulsa, OK 74107, USA; cole.wayant@okstate.edu



© Author(s) (or their employer(s)) 2019. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Wayant C, Cooper C, Turner D'A, et al. *BMJ Evidence-Based Medicine* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bmjebm-2018-111153

Abstract

Introduction Robust, clearly reported clinical practice guidelines (CPGs) are essential for evidence-based clinical practice. The Reporting Items for practice Guidelines in HealthCare (RIGHT) Statement and Appraisal of Guidelines for Research and Evaluation (AGREE) II instrument were published to improve the methodological and reporting quality in healthcare CPGs.

Methods We applied the RIGHT Statement checklist and AGREE-II instrument to 48 National Comprehensive Cancer Network (NCCN) guidelines. Our primary objective was to assess the adherence to RIGHT and AGREE-II items. Since neither RIGHT nor AGREE-II can judge the clinical usefulness of a guideline, our study is designed to only focus on the methodological and reporting quality of each guideline.

Results The NCCN guidelines demonstrated notable strengths and weaknesses. For example, RIGHT Statement items 19 (conflicts of interest), 7b (description of subgroups) and 13a (clear, precise recommendations) were fully reported in all guidelines. However, the guidelines inconsistently incorporated patient values and preferences and cost. Regarding the AGREE-II instrument, the NCCN guidelines scored highly on the domains 4 (clear, precise recommendations) and 6 (handling of conflicts of interest), but lowest on domain 2 (inclusion of all relevant stakeholders).

Conclusions In this investigation, we found that NCCN CPGs demonstrate key strengths and weaknesses with respect to the reporting of key items essential to CPGs. We recommend the continued use of NCCN guidelines and improvements to weaknesses in reporting and methods. Doing so serves to improve the evidence delivered to healthcare providers, thus potentially improving patient care.

Introduction

Robust, clearly reported clinical practice guidelines (CPGs) are essential for evidence-based clinical practice. The Institute of Medicine recognises CPGs as necessary reference material for physicians seeking to optimise patient care.¹ CPGs are capable of increasing the quality of patient care and improving patient outcomes,² but the adoption of low-quality guidelines may result in widespread

use of ineffective treatments, inefficient practices and harm to patients.^{3,4} Even though they are an essential resource, CPGs have historically exhibited low-quality reporting.⁵ The ramifications of low reporting quality in CPGs are broad, but most pressing is the lack of a distinction between poor methods and poorly reported methods. In practice, the two may be indistinguishable. For example, if CPG developers perform a narrow, inadequate search of the literature, their subsequent recommendations may not be reproducible or trustworthy. Similarly, if the CPG developers do not report their search strategy, the question remains as to whether the recommendations are trustworthy. The quality of CPG reporting is as important as its methodological quality.

In oncology, new drug approvals may result in rapid changes to patient care. Articulating the available evidence, its strength and its limitations to physicians is vital. The National Comprehensive Cancer Network (NCCN)—arguably the premier guideline organisation in the USA⁶—has a policy to update their CPGs ‘at least annually’.⁷ This policy of annual updates highlights the urgent need for clear reporting of current and future CPGs.

Two popular instruments exist for assessing the quality of CPGs in healthcare: The Reporting Items for practice Guidelines in HealthCare (RIGHT) Statement⁸ and the Appraisal of Guidelines for Research and Evaluation (AGREE) II instrument.⁹ The AGREE-II instrument includes items related to the methodological (eg, quality of search strategy, inclusion of stakeholder preferences) and reporting quality of CPGs, whereas the RIGHT Statement focuses solely on reporting quality (eg, providing a summary of recommendations, disclosure of funding source). Neither was created as a handbook for developing guidelines. According to the RIGHT Statement authors, the RIGHT Statement is not designed to assess the inherent quality of a guideline.⁸ Rather, the RIGHT Statement is designed to complement tools that are designed to assess the inherent quality of a guideline, such as the AGREE-II instrument.

Given the comprehensiveness and importance of the NCCN CPGs to oncology practice,⁶ the aim of this investigation is to highlight the strengths and weaknesses in the reporting of NCCN guidelines. By doing so, we aim to improve the delivery of oncology evidence to oncologists and improve patient care. In this study, we applied the RIGHT

Statement and AGREE-II instrument to 49 NCCN guidelines for the treatment of cancer by cancer site.

Methods

A version of this manuscript is available as a preprint via bioRxiv.¹⁰ Since NCCN guidelines update frequently throughout a calendar year, we downloaded the Portable Document Format of all 49 NCCN treatment guidelines on 21 March 2018 from the NCCN website under the heading 'NCCN Guidelines for the Treatment of Cancer by Site'. To be included in this study, a guideline must have a written Discussion section, which is equivalent to the guideline narrative. Prior to data extraction, CW, CC and DT reviewed the RIGHT statement and AGREE-II instrument manuals to become familiar with the checklist items.^{8,9} We met and devised a Google Form for both tools. CW, CC and DT extracted data for all items from each tool independently while masked to each other's decisions. Since the NCCN does not detail their full methods in each CPG and provides a full explanation of many aspects of their methods on their website (www.nccn.org), we extracted data from the CPG and website policy documents. Any discrepancies in data extraction were resolved via consensus discussion. After extraction and validation of all Google Form responses, we exported these responses to a Google Sheet. We used this Google Sheet to calculate summary statistics. We correlated the RIGHT and AGREE-II scores using Stata V.15.1 and the commands *pwcorr*, for a Pearson's *r*, and *graph twoway scatter* for a two-way scatter plot. Raw AGREE-II scores were used, rather than scaled scores, with a maximum value of 161 (23 items, 7-point Likert scale) indicating a judgement of perfect methodological quality across all domains for a CPG.

The design of the RIGHT Statement parallels other statements and reporting guidelines, such as Consolidated Standards of Reporting Trials for clinical trials or Preferred Reporting Items for Systematic Reviews and Meta-Analyses for systematic reviews, and consists of a 35-item checklist and an Explanation and Elaboration document.⁸ For each of the items, we assigned a numeric score of 1 (full adherence), 0.5 (partial adherence) or 0 (no adherence). An example of partial adherence may be if a guideline provides a partial explanation of cancer epidemiology, explaining only the prevalence and incidence of the disease. Full explanation includes a description of prevalence/incidence, morbidity, mortality and burden (including financial). We present summary data using the described scoring convention for each of the 35 items. Rather than dichotomising the data in an attempt to separate CPGs into high, medium or low reporting quality groups, we present data as continuous and out of the maximum possible score of 35. This decision was made because there is no guidance for what constitutes high-, medium- or low-quality reporting quality in CPGs.

The AGREE-II instrument is organised differently, and consists of 23 items divided into six domains, with each item scored on a 1 (strongly disagree) to 7 (strongly agree) Likert-type scale. In accordance with the AGREE-II manual,⁹ we calculated a scaled domain score for each domain for each CPG. The scaled domain score is calculated as follows:

$$\frac{(\text{Obtained Score} - \text{Minimum Score})}{(\text{Maximum Score} - \text{Minimum Score})}$$

The scaled domain score can be converted to an average rating (1 to 7 scale) by multiplying the scaled domain score by 7. The obtained score is calculated for each domain and is the sum of all rater scores in that domain. The minimum score is calculated by

multiplying the minimum item score (1, strongly disagree), the number of raters (3, in this study) and the number of items in the domain. The maximum score is calculated similarly, but substitutes the maximum item score (7, strongly agree) for the minimum item score. Lastly, we made a consensus judgement about whether the CPG should be used in practice or not based on the six-scaled domain scores for each CPG. We based our judgement of each NCCN CPG off the AGREE-II manual, which suggests answering whether a CPG should be used with 'yes', 'yes with modifications' or 'no'. We rendered our judgements by looking at the full scope of domain scores, rather than using dichotomous decision rules. The rationale for this decision was that each domain has been shown to independently associated with CPG quality.¹¹ Our primary objective was to assess CPG scores on the RIGHT Statement and AGREE-II instrument. Since all NCCN guidelines were published after the RIGHT Statement and AGREE-II instrument were published, they are all eligible for analysis. As neither the RIGHT Statement nor AGREE-II instrument can judge the clinical usefulness of a guideline, our study is designed to only focus on the methodological and reporting quality of each guideline.

Results

We identified 49 NCCN CPGs for the treatment of cancer by site. The uveal melanoma CPG was excluded because the Discussion section (the narrative section of NCCN guidelines) was under development and not written. All of our data, including data for each individual item on the RIGHT Statement and AGREE-II instrument, are publicly available via the Open Science Framework.¹²

RIGHT Statement

The NCCN guidelines were largely homogeneous, and many key methodological items were reported clearly in policy documents on the NCCN website. Table 1 shows each NCCN guideline and its adherence to all RIGHT Statement items. Notable strengths of the NCCN CPGs were the reporting of conflicts of interest for all authors (items 19a and 19b), complete description of pertinent subgroups (item 7b) and the clarity of CPG recommendations (item 13a). Notable deficiencies were the description of stakeholder involvement (eg, patient views and preferences) (item 14a), the cost and resource implications of therapies (item 14b), which outcomes were prioritised when formulating recommendations (item 10b) and the approach to assess the certainty of the quality of evidence (item 12).

Agree-II

Table 2 shows the scaled domain scores for each NCCN CPG. Using the AGREE-II instrument, we were able to assess CPG scores in six domains, each essential to a methodologically robust CPG. No guideline scored extremely low for any domain. The fourth domain (clarity of presentation) and sixth domain (editorial independence) scored the highest, overall. The clarity of presentation domain asks whether the recommendations are specific and unambiguous, if alternative treatment options were mentioned, and if the key recommendations are easily identifiable. The sixth domain asks questions about the influence of the funding source on CPG development and whether conflicts of interest were disclosed. The lowest, individual domain score was 36.1% in the applicability domain for the acute lymphoblastic leukaemia CPG. This score indicates that average score (1 to 7 scale) for this domain was approximately 2.5. With respect to overall domain scores across all guidelines, the stakeholder involvement domain scored the lowest with an average score of 48.6% (eg, 3.4 out of 7).

Table 1 Adherence to RIGHT Statement items (n=35) by each included NCCN guideline (n=48)

Guideline	RIGHT Statement domain									
	Version	Date of publication	Basic information (n=6)	Background (n=8)	Evidence (n=5)	Recommendations (n=7)	Review and quality assurance (n=2)	Funding and interests (n=4)	Total score (%) (n=35)	
Acute lymphoblastic leukaemia	1.2018	3/12/2018	4.0	6.5	1.5	3.5	1.0	4.0	21 (60.0)	
Acute myeloid leukaemia	1.2018	2/7/2018	4.0	5.0	1.5	3.5	1.0	4.0	19.5 (55.7)	
Amyloidosis	1.2018	11/3/2017	5.0	5.0	1.5	3.5	1.0	4.0	20.5 (58.6)	
Anal	1.2018	2/6/2018	5.0	5.0	1.5	3.5	1.0	4.0	20.5 (58.6)	
B cell	2.2018	2/26/2018	4.0	6.0	1.5	3.5	1.0	4.0	20.5 (58.6)	
Basal cell	1.2018	9/18/2017	4.0	5.0	1.5	3.5	1.0	4.0	19.5 (55.7)	
Bladder	3.2018	3/14/2018	4.0	5.0	1.5	3.5	1.0	4.0	19.5 (55.7)	
Bone	1.2018	8/29/2017	4.0	5.0	1.5	3.5	1.0	4.0	19.5 (55.7)	
Breast	4.2017	2/7/2018	4.5	5.0	1.5	3.5	1.0	4.0	20 (57.1)	
Cervical	1.2018	3/20/2018	4.0	4.5	1.5	3.5	1.0	4.0	19 (54.3)	
Chronic lymphocytic leukaemia	1.2018	10/25/2017	4.0	5.0	1.5	3.5	1.0	4.0	19.5 (55.7)	
Chronic myeloid leukaemia	4.2018	3/14/2018	4.0	5.0	1.5	3.5	1.0	4.0	19.5 (55.7)	
Central nervous system	4.2018	1/24/2018	5.0	5.0	1.5	3.5	1.0	4.0	20.5 (58.6)	
Colon	2.2018	3/14/2018	5.0	5.0	1.5	3.5	1.0	4.0	20.5 (58.6)	
Cutaneous B cell	1.2018	9/18/2017	4.0	3.5	1.5	3.5	1.0	4.0	18 (51.4)	
Dermato-protuberans	1.2018	3/16/2018	4.0	5.0	1.5	3.5	1.0	4.0	19.5 (55.7)	
Esophageal	1.2018	3/16/2018	4.0	5.0	1.5	3.5	1.0	4.0	19.5 (55.7)	
Gastric	2.2018	9/26/2017	5.0	5.0	1.5	3.5	1.0	4.0	20.5 (58.6)	
Hairy cell	1.2018	2/15/2018	5.0	5.0	1.5	3.5	1.0	4.0	20.5 (58.6)	
Head/neck	1.2018	2/14/2018	4.0	4.0	1.5	3.5	1.0	4.0	18.5 (52.9)	
Hepatobiliary	1.2018	12/20/2017	4.0	5.0	1.5	3.5	1.0	4.0	19.5 (55.7)	
Hodgkin	3.2018	2/6/2018	4.0	5.5	1.5	3.5	1.0	4.0	20 (57.1)	
Kaposi	2.2018	2/26/2018	4.0	5.0	1.5	3.5	1.0	4.0	19.5 (55.7)	
Kidney	2.2018	1/19/2018	4.0	5.0	1.5	3.5	1.0	4.0	19.5 (55.7)	
Myelodysplastic syndrome	1.2018	9/18/2017	4.5	5.0	1.5	3.5	1.0	4.0	20 (57.1)	
Melanoma	4.2018	2/12/2018	5.0	5.0	1.5	3.5	1.0	4.0	20.5 (58.6)	
Merkel	2.2018	2/15/2018	4.5	5.0	1.5	3.5	1.0	4.0	20 (57.1)	
Mesothelioma	2.2018	9/7/2017	4.0	5.0	1.5	3.5	1.0	4.0	19.5 (55.7)	
Myeloproliferative Neoplasms	3.2017	6/13/2017	4.0	5.0	1.5	3.5	1.0	4.0	19.5 (55.7)	
Myeloma	3.2018	2/12/2018	4.0	5.0	1.5	3.5	1.0	4.0	19.5 (55.7%)	
Neuroendocrine	1.2018	9/9/2017	4.0	5.0	1.5	3.5	1.0	4.0	19.5 (55.7)	
Non-small-cell lung	1.2018	10/10/2017	4.0	5.0	1.5	3.5	1.0	4.0	19.5 (55.7)	
Occult primary	3.2017	9/11/2017	4.0	5.5	1.5	3.5	1.0	4.0	20 (57.1)	
Ovarian	1.2018	1/8/2018	4.0	5.0	1.5	3.5	1.0	4.0	19.5 (55.7)	
Pancreatic	2.2018	8/10/2018	4.0	5.0	1.5	3.5	1.0	4.0	19.5 (55.7)	
Penile	2.2018	3/8/2018	5.0	5.0	1.5	3.5	1.0	4.0	20.5 (58.6)	

Continued

Table 1 Continued

Guideline	RIGHT Statement domain										Total score (%) (n=35)
	Version	Date of publication	Basic information (n=6)	Background (n=8)	Evidence (n=5)	Recommendations (n=7)	Review and quality assurance (n=2)	Funding and interests (n=4)			
Prostate	1.2018	3/14/2018	4.0	6.0	1.5	3.5	1.0	4.0			20.5 (58.6)
Rectal	2.2018	1/17/2018	4.0	5.0	1.5	3.5	1.0	4.0			19.5 (55.7)
Sarcoma	1.2018	10/31/2017	4.0	5.0	1.5	3.5	1.0	4.0			19.5 (55.7)
Squamous cell	2.2018	10/5/2017	4.5	5.0	1.5	3.5	1.0	4.0			20 (57.1)
Small-cell lung	1.2018	9/6/2017	4.0	4.0	1.5	3.5	1.0	4.0			18.5 (52.9)
T cell	3.2018	2/22/2018	5.0	6.0	1.5	3.5	1.0	4.0			21.5 (61.4)
Testicular	2.2018	2/16/2018	4.0	5.0	1.5	3.5	1.0	4.0			19.5 (55.7)
Thymus	2.2018	2/16/2018	5.0	5.0	1.5	3.5	1.0	4.0			20.5 (58.6)
Thyroid	2.2017	5/17/2017	4.0	5.0	1.5	3.5	1.0	4.0			19.5 (55.7)
Uterine	1.2018	10/13/2017	4.0	5.0	1.5	3.5	1.0	4.0			19.5 (55.7)
Vulvar	1.2018	10/27/2017	5.0	5.0	1.5	3.5	1.0	4.0			20.5 (58.6)
Waldenström	1.2018	3/7/2018	4.0	5.0	1.5	3.5	1.0	4.0			19.5 (55.7)

NCCN, National Comprehensive Cancer Network; RIGHT, Reporting Items for practice Guidelines in HealthCare.

The stakeholder involvement domain asks questions related to the description of guideline development members, the incorporation of target population views and preferences and the identification of target users of the guidelines.

Correlation of RIGHT and AGREE-II scores

There was a low correlation between RIGHT and AGREE-II scores ($r=-0.25$) (figure 1). The negative correlation is likely driven by the four guidelines that adhered to only 19/35 (54.2%) of RIGHT items while maintaining relatively high AGREE-II scores. Overall, most data clustered between RIGHT scores of 19.5–20.5 and AGREE-II scores of approximately 105–115. Visual inspection of our data shows that many CPGs had identical RIGHT scores, with slight variations in their AGREE-II scores.

Certain outliers are visible in the scatter plot, which have been labelled with the CPG name. Notable outliers are the guidelines for Merkel Cell Carcinoma and Primary Cutaneous B-Cell Lymphoma. The Merkel Cell Carcinoma guideline scored lowest on AGREE-II, but average on RIGHT. This guideline was judged to score relatively low on three methodological domains: stakeholder involvement, rigour of development and applicability. None of these items had direct overlap with RIGHT Statement items, so the Merkel Cell Carcinoma guideline was still capable of achieving an average score in terms of reporting quality. On the other hand, the Primary Cutaneous B-Cell Lymphoma guideline scored lowest on the RIGHT Statement, but above average on AGREE-II. In absolute terms, the Primary Cutaneous B-Cell Lymphoma guideline only scored 2 items lower than most other guidelines.

Discussion

In this investigation, we found that NCCN CPGs demonstrate key strengths and weaknesses with respect to the reporting of key items essential to CPGs. For example, the NCCN CPGs require conflicts of interest disclosure, clearly describe all pertinent subgroups and delineate key recommendations. On the other hand, the NCCN CPGs did not consistently describe how patient values and preferences were incorporated into recommendations, the financial burden of the recommendations or describe the approach used to assess the certainty of the evidence underpinning the recommendations. The NCCN guidelines were incredibly uniform in how they are reported and conducted, which resulted in similar (or identical, in the case of the RIGHT Statement) scores for most CPGs. This uniformity is reflected in the scatter plot. Across all NCCN guidelines, certain items, such as providing a summary of recommendations, were always reported. On the other hand, some items, such as describing the approach to assessing the certainty of the evidence, were never reported. The slight variation in AGREE-II scores for identical RIGHT scores is a product of 1–7 Likert scale format, which allows more variation in judgements than the RIGHT Statement scoring system of full, partial or no adherence. In light of the uniformity of our data, our findings should be interpreted to mean that there are significant shortcomings in the reporting and development of NCCN guidelines, but all of these shortcomings could be addressed at once by updating the central NCCN policies and procedures.

Nonetheless, compared with other CPGs scored with the AGREE-II instrument, those published by the NCCN appear to have as good or stronger methodological quality.^{11 13–15} A recent evaluation in *JAMA Internal Medicine* of CPGs for the pharmacological management of non-communicable diseases in primary care found that three CPG characteristics are associated with high-quality CPGs: greater than 20 authors, development at a government institution and reported funding.¹⁶ The NCCN is a

Table 2 AGREE-II scores for all domains and overall judgement for guideline use

Guideline	Scope and purpose	Stakeholder involvement	Rigour of development	Clarity of presentation	Applicability	Editorial independence	Overall assessment
	Scaled per cent	Scaled per cent	Scaled per cent	Scaled per cent	Scaled per cent	Scaled per cent	
Acute lymphoblastic leukaemia	77.8	42.6	61.8	85.2	36.1	94.4	Recommend with improvement
Acute myeloid leukaemia	79.6	38.9	59.0	87.0	40.3	94.4	Recommend with improvement
Amyloidosis	42.6	50.0	57.6	74.1	48.6	94.4	Recommend with improvement
Anal	72.2	37.0	54.9	81.5	40.3	94.4	Recommend with improvement
B cell	74.1	50.0	56.9	81.5	52.8	94.4	Recommend with improvement
Basal cell	77.8	53.7	61.8	85.2	62.5	94.4	Recommend with improvement
Bladder	74.1	42.6	57.6	90.7	51.4	94.4	Recommend with improvement
Bone	77.8	51.9	67.4	85.2	59.7	94.4	Recommend with improvement
Breast	79.6	50.0	70.8	79.6	62.5	94.4	Recommend with improvement
Cervical	68.5	51.9	63.9	87.0	63.9	94.4	Recommend with improvement
Chronic lymphocytic leukaemia	83.3	51.9	57.6	83.3	65.3	94.4	Recommend with improvement
Chronic myeloid Leukaemia	70.4	40.7	65.3	87.0	45.8	94.4	Recommend with improvement
Central nervous system	77.8	53.7	58.3	81.5	66.7	94.4	Recommend with improvement
Colon	83.3	42.6	61.1	70.4	51.4	94.4	Recommend with improvement
Cutaneous B cell	63.0	50.0	43.8	77.8	61.1	94.4	Recommend with improvement
Dermato-protuberans	77.8	50.0	65.3	88.9	63.9	94.4	Recommend with improvement
Esophageal	81.5	50.0	66.0	87.0	63.9	94.4	Recommend with improvement
Gastric	66.7	40.7	61.1	83.3	50.0	94.4	Recommend with improvement
Hairy cell	70.4	42.6	63.9	85.2	50.0	94.4	Recommend with improvement
Head/neck	74.1	51.9	65.3	81.5	65.3	94.4	Recommend with improvement
Hepatobiliary	77.8	48.1	64.6	83.3	59.7	94.4	Recommend with improvement
Hodgkin	70.4	42.6	68.1	87.0	52.8	94.4	Recommend with improvement
Kaposi	70.4	48.1	63.9	83.3	54.2	94.4	Recommend with improvement
Kidney	74.1	53.7	61.8	85.2	65.3	94.4	Recommend with improvement
Myelodysplastic syndrome	63.0	40.7	43.8	68.5	40.3	94.4	Recommend with improvement
Melanoma	72.2	51.9	62.5	81.5	61.1	94.4	Recommend with improvement
Merkel	68.5	48.1	65.3	85.2	61.1	94.4	Recommend with improvement
Mesothelioma	81.5	40.7	62.5	87.0	51.4	94.4	Recommend with improvement
Myeloproliferative neoplasms	79.6	53.7	59.0	85.2	55.6	94.4	Recommend with improvement
Myeloma	70.4	53.7	66.7	87.0	63.9	94.4	Recommend with improvement
Neuroendocrine	74.1	46.3	67.4	87.0	69.4	94.4	Recommend with improvement
Non-small-cell lung	77.8	51.9	66.7	87.0	63.9	94.4	Recommend with improvement
Occult primary	74.1	51.9	66.0	87.0	63.9	94.4	Recommend with improvement

Continued

Table 2 Continued

Guideline	Scope and purpose Scaled per cent	Stakeholder involvement Scaled per cent	Rigour of development Scaled per cent	Clarity of presentation Scaled per cent	Applicability Scaled per cent	Editorial independence Scaled per cent	Overall assessment
Ovarian	85.2	51.9	62.5	87.0	61.1	94.4	Recommend with improvement
Pancreatic	74.1	51.9	66.7	85.2	65.3	94.4	Recommend with improvement
Penile	72.2	53.7	68.1	87.0	63.9	94.4	Recommend with improvement
Prostate	81.5	46.3	67.4	85.2	56.9	94.4	Recommend with improvement
Rectal	70.4	53.7	63.9	87.0	66.7	94.4	Recommend with improvement
Sarcoma	75.9	53.7	61.8	87.0	58.3	94.4	Recommend with improvement
Squamous cell	74.1	53.7	66.7	85.2	62.5	94.4	Recommend with improvement
Small-cell lung	70.4	42.6	50.7	85.2	48.6	94.4	Recommend with improvement
T cell	75.9	44.4	62.5	85.2	50.0	94.4	Recommend with improvement
Testicular	74.1	53.7	64.6	87.0	69.4	94.4	Recommend with improvement
Thymus	77.8	46.3	64.6	87.0	52.8	94.4	Recommend with improvement
Thyroid	63.0	53.7	64.6	87.0	68.1	94.4	Recommend with improvement
Uterine	74.1	53.7	64.6	87.0	61.1	94.4	Recommend with improvement
Vulvar	79.6	42.6	63.9	85.2	51.4	94.4	Recommend with improvement
Waldenstrom	74.1	55.6	66.0	87.0	61.1	94.4	Recommend with improvement
Average scaled domain score	73.9	48.6	62.4	84.4	57.5	94.4	

AGREE-II, Appraisal of Guidelines for Research and Evaluation.

non-profit organisation and their CPGs are developed by a team of volunteers from member institutions and no external funding is received to develop the CPGs. All guidelines have greater than 20 authors. So, the findings of this recent evaluation in *JAMA Internal Medicine* seem in line with our findings that NCCN CPGs are of comparable or higher methodological quality than other biomedical CPGs. However, the reporting quality of biomedical CPGs has been evaluated far less, owing to the fact that the RIGHT Statement is the only available tool and was published in 2017.

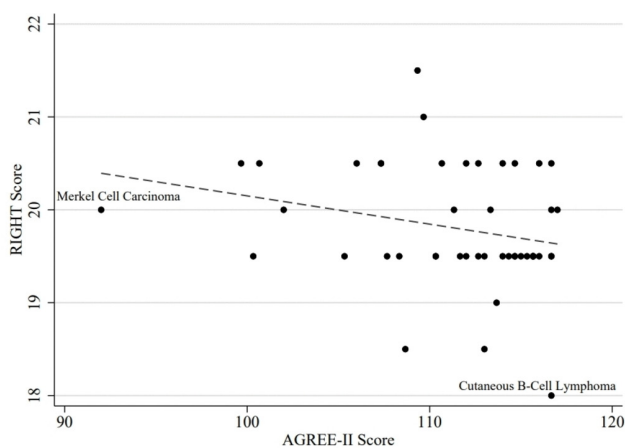


Figure 1 Scatter plot of AGREE-II raw score versus RIGHT Statement score. AGREE-II, Appraisal of Guidelines for Research and Evaluation; RIGHT, Reporting Items for practice Guidelines in HealthCare.

Only one study was identified which used the RIGHT Statement.¹⁷ This lone study evaluated 539 CPGs in traditional Chinese medicine, finding that 17 of 35 (48.6%) RIGHT Statement were reported less than 10% of the time. In comparison, our study found that only nine items were never fully reported. In an effort to provide the highest quality recommendations to physicians for the treatment of different cancers, we encourage continued improvements to the NCCN guidelines. The AGREE-II instrument⁹ was developed to assess CPG quality in six, equally essential domains ranging from describing the purpose of the CPG to the applicability of the CPG recommendations. We found that they scored well enough to continue being recommended in clinical practice, but key methodological items were not reported, thus highlighting areas where the delivery of oncology evidence can be improved. Since we assigned summary judgements related to the recommended use of NCCN CPGs in clinical practice in a continuous manner, each judgement of 'Yes, with modifications' should be interpreted continuously. Since no two CPGs were scored identically for all six domains, each judgement of 'yes, with modifications' should signal different improvements are needed in different orders of magnitude. Through applying the RIGHT Statement, which was created to be used alongside the AGREE-II instrument, we confirmed that improvements in the reporting of several key items would strengthen the impact of NCCN CPGs by increasing the clarity and comprehensiveness of the recommendations.

None of the NCCN CPGs described the process by which patient values and preferences were solicited and incorporated into the guideline recommendations, nor do they adhere to an accepted

framework for grading the quality of evidence. The primary reason for incorporating patient values and preferences into CPG recommendations is that recommendations that are aligned with patient values may be more easily adopted and implemented.^{18–20} Until recently, there were no firmly established processes for including patient values and preferences in CPG recommendations. To address this gap, the GRADE (Grading of Recommendations Assessment, Development and Evaluation) working group created the GRADE Evidence-to-Decision (EtD) framework.¹⁹ Previously, the GRADE approach has been used to assess the quality and certainty of evidence underpinning CPG recommendations. The NCCN CPGs do not currently use the GRADE approach, or any similar framework, rather they seem to rely on guideline development member assessments of the quality of evidence. The NCCN members assess the quality of evidence over certain domains, but in an effort to improve the objectivity, applicability and comparability of NCCN recommendations, we recommend adopting the GRADE approach. Concurrent adoption of the GRADE EtD framework would ensure the incorporation of patient values and preferences in all recommendations.

Additional, minor adjustments to the reporting of NCCN CPGs would improve the delivery of oncology evidence. First, stating key research questions that formed the basis for treatment recommendations in Population, Intervention, Comparator, Outcome (PICO) format would guide physicians through the purpose and scope of the guideline.^{21–23} Due to how comprehensive the NCCN CPGs are, it may be that listing all PICO format questions is not practical. Should this be the case, we recommend including a section in the CPG that clearly describes the scope, limitations and gaps in the NCCN recommendations. A second, related adjustment includes listing the outcomes that were most important when developing the CPG recommendations. For example, if efficacy outcomes are the primary basis for the recommendations, or recommending one treatment over another, physicians would benefit from that understanding.

This study has key strengths and limitations. With respect to strengths, we used two formally published and peer-reviewed tools to assess the quality of reporting and methodological rigour of NCCN guidelines. We further used three data extractors to mitigate bias in our data analysis. Each author underwent identical, comprehensive training to ensure competency prior to data extraction. With respect to limitations, our assessment of methodological quality may be limited by a lack of reporting. In other words, simply because someone was not reported as having been done, does not mean it was not done. For example, it is possible that the views of patients were sought in the formulation of the guidelines, but if this was not reported or described, we were forced to assign a low score this AGREE-II item low.

In conclusion, we simultaneously recommend the continued use of NCCN CPGs to guide oncologists in patient care and efforts to improve the weaknesses we identified in this study. Each guideline contained strengths and weaknesses, and improving the weaknesses will enhance the applicability and comparability of the recommendations. We have outlined key recommendations that would improve the completeness of reporting and increase transparency. These recommendations include the adoption of the GRADE and GRADE-EtD approach, describing key questions in PICO format, and sorting which outcomes were important when developing recommendations. We believe that adopting these recommendations will not only improve the NCCN CPGs but also oncology clinical care.

Contributors CW and MV designed the study. CW, CC and DT extracted and analysed all data. All authors wrote and approved the final version of the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

References

- Institute of Medicine Board on Health Care Services Committee on Standards for Developing Trustworthy Clinical Practice Guidelines. *Clinical practice guidelines we can trust*: National Academies Press, 2011.
- Grimshaw J, Freemantle N, Wallace S, *et al*. Developing and implementing clinical practice guidelines. *Qual Health Care* 1995;4:55–64.
- Feder G, Eccles M, Grol R, *et al*. Clinical guidelines: using clinical guidelines. *BMJ* 1999;318:728–30.
- Woolf SH, George JN. Evidence-based medicine. Interpreting studies and setting policy. *Hematol Oncol Clin North Am* 2000;14:761–84.
- Grilli R, Magrini N, Penna A, *et al*. Practice guidelines developed by specialty societies: the need for a critical appraisal. *Lancet* 2000;355:103–6.
- Benson AB, Brown E. Role of NCCN in integrating cancer clinical practice guidelines into the healthcare debate. *Am Health Drug Benefits* 2008;1:28–33.
- National Comprehensive Cancer Network. NCCN Disclosure policies and potential conflicts of interest. <https://www.nccn.org/about/disclosure.aspx> (Accessed 10 Nov 2017).
- Chen Y, Yang K, Marušić A, *et al*. A reporting tool for practice guidelines in health care: the RIGHT statement. *Ann Intern Med* 2017;166:128–32.
- Brouwers MC, Kho ME, Browman GP, *et al*. AGREE II: advancing guideline development, reporting and evaluation in health care. *CMAJ* 2010;182:E839–42.
- Wayant C, Cooper C, Turner D, *et al*. Evaluation of the NCCN guidelines using the RIGHT Statement and AGREE II instrument: a cross-sectional review. *bioRxiv* 2018.
- Hoffmann-Eßer W, Siering U, Neugebauer EA, *et al*. Guideline appraisal with AGREE II: systematic review of the current evidence on how users handle the 2 overall assessments. *PLoS One* 2017;12:e0174831.
- Wayant C, Cooper C, arcy Turner D', *et al*. RIGHT/AGREE 2 NCCN guidelines. 2018 <https://osf.io/nh46r/> (Accessed 19 Oct 2018).
- Uzeloto JS, Moseley AM, Elkins MR, *et al*. The quality of clinical practice guidelines for chronic respiratory diseases and the reliability of the AGREE II: an observational study. *Physiotherapy* 2017;103:439–45.
- Radwan M, Akbari Sari A, Rashidian A, *et al*. Appraising the methodological quality of the clinical practice guideline for diabetes mellitus using the AGREE II instrument: a methodological evaluation. *JRSM Open* 2017;8:205427041668267.
- Larenas-Linnemann DES, Antolin-Américo D, Parisi C, *et al*. National clinical practice guidelines for allergen immunotherapy: An international assessment applying AGREE-II. *Allergy* 2018;73:664–72.
- Molino C, Leite-Santos NC, Gabriel FC, *et al*. Factors associated with high-quality guidelines for the pharmacologic management of chronic diseases in primary care: a systematic review. *JAMA Intern Med* 2019 [Epub ahead of print 18 Feb 2019].
- Yun X, Yaolong C, Zhao Z, *et al*. Using the RIGHT statement to evaluate the reporting quality of clinical practice guidelines in traditional Chinese medicine. *PLoS One* 2018;13:e0207580.
- Murad MH, Montori VM, Guyatt GH. Incorporating patient preferences in evidence-based medicine. *JAMA* 2008;300:2483.
- Zhang Y, Coello PA, Brožek J, *et al*. Using patient values and preferences to inform the importance of health outcomes in practice guideline development following the GRADE approach. *Health Qual Life Outcomes* 2017;15:52.

20. Sackett DL, Rosenberg WM, Gray JA, *et al.* Evidence based medicine: what it is and what it isn't. *BMJ* 1996;312:71-2.
21. Aslam S, Emmanuel P. Formulating a researchable question: a critical step for facilitating good clinical research. *Indian J Sex Transm Dis AIDS* 2010;31:47-50.
22. Cañón M, Buitrago-Gómez Q. The research question in clinical practice: a guideline for its formulation. *Rev Colomb Psiquiatr* 2018;47:193-200.
23. Beitz JM. Writing the researchable question. *J Wound Ostomy Continence Nurs* 2006;33:122-4.