Critical Evaluation of the Quality and Recommendations of Clinical Practice Guidelines for Nasopharyngeal Carcinoma

Yu-Pei Chen, MD*; Ya-Qin Wang, MD*; Wen-Fei Li, MD*; Lei Chen, MD; Cheng Xu, MD; Tai-Xiang Lu, MD; Ai-Hua Lin, MD; Ji-Jin Yao, MD; Yang-Chan Li, MD; Ying Sun, MD, PhD; Yan-Ping Mao, MD; and Jun Ma, MD*

Abstract

Background: Given the distinct biological characteristics and regional distribution of nasopharyngeal carcinoma (NPC) compared with other head and neck cancers, and uncertainties regarding therapeutic strategies, physicians require high-quality clinical practice guidelines (CPGs) to provide transparent recommendations for NPC treatment. This study aimed to critically appraise the quality of NPC CPGs and assess the consistency of their recommendations. Methods: We identified CPGs that provided recommendations on the diagnosis and management of NPC published up to December 2015. Four investigators independently appraised CPG quality using the Appraisal of Guidelines for Research Evaluation (AGREE) II instrument. Key recommendations by CPGs were also evaluated. Results: A total of 7 CPGs were eligible for this study: 5 produced by professional organizations or governmental agencies and 2 were developed based on expert consensus. Of the 6 AGREE II domains, the applicability domain scored consistently low across CPGs (range, 13.5%–30.2%); no CPG achieved a score of >50% in all 6 domains. The scope and purpose domain (≥73.6% for 4 CPGs) and editorial independence domain (≥75.0% for 6 CPGs) scored highest. Of the 23 AGREE II items, 9 scored less than half of the points available in all 7 CPGs. The recommendations by CPGs were consistent in general; heterogeneity mainly existed among recommended therapeutic strategies. Conclusions: Variation exists in NPC CPG development processes and recommendations. Increased efforts are required to make comprehensive resources available to guide healthcare providers and enhance delivery of high-quality, evidence-based care for NPC. International collaboration is necessary to enable the development of high-quality and regionally relevant CPGs for NPC.


Because of increasing pressure to provide evidence-based medical care and increasing healthcare costs, clinical practice guidelines (CPGs) have been increasingly used over the past decade worldwide.2-4 CPGs aim to interpret and consolidate evidence from the expansive clinical research literature to offer standardized approaches to high-quality care. However, it has been suggested that the quality of many published CPGs is unsatisfactory and that substantial heterogeneity exists among guideline recommendations.2,5,6 Additionally, clinicians’ lack of adherence to CPGs may be a result of distrust in guideline development processes and recom-

From *Department of Radiation Oncology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center of Cancer Medicine; and †Department of Medical Statistics and Epidemiology, School of Public Health, Sun Yat-sen University, Guangzhou, People’s Republic of China.

*These authors contributed equally.

Submitted August 6, 2016; accepted for publication November 3, 2016.

The authors have disclosed that they have no financial interests, arrangements, affiliations, or commercial interests with the manufacturers of any products discussed in this article or their competitors.

This work was supported by grants from the National Science & Technology Pillar Program during the Twelfth Five-year Plan Period (2014BAI09810); the Health & Medical Collaborative Innovation Project of Guangzhou City, China (201400000001); the Planned Science and Technology Project of Guangdong Province (2013B020400004); and the Science and Technology Project of Guangzhou City, China (14570006).

Author contributions: Study conception and design: Chen, Chen, Ma. Acquisition of data: Chen, Li, Chen. Sun. Analysis and interpretation of data: Chen, Wang, Li, Chen, Ma. Access to raw data: Chen, Ma. Table and figure creation: Chen, Wang, Li, Chen. Analysis plan and outcome review; preparation and drafting of manuscript; final approval: All authors.

Correspondence: Jun Ma, MD, Department of Radiation Oncology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center of Cancer Medicine, 651 Dongfeng Road East, Guangzhou 510060, People’s Republic of China. E-mail: majun2@mail.sysu.edu.cn
mendations. Therefore, systematically developed guidelines providing transparent recommendations are required.

The Appraisal of Guidelines for Research & Evaluation (AGREE) instrument is a validated, generic tool for the systematic appraisal of CPG methodological development and quality. The AGREE project was established in 1998 by a group of international guideline developers and researchers; the original AGREE instrument was published in 2003 and the updated AGREE II instrument was released in 2010. AGREE has become the standard tool for evaluation and development of CPGs, with the aim of improving guideline quality and likelihood of broad endorsement.

Nasopharyngeal carcinoma (NPC) is a unique head and neck (H&N) cancer with an uneven distribution: the age-standardized incidence rate ranges from 20 to 50 per 100,000 men in southern China to 0.5 per 100,000 in mainly Caucasian populations. Despite originating from similar cell or tissue lineages, NPC is distinct to other epithelial H&N tumors. Radiotherapy (RT) is the primary and only curative modality, and combined chemoradiotherapy (chemoRT) is a pivotal treatment for locoregionally advanced NPC. However, the optimal protocols remain elusive. The value of adjuvant chemotherapy (AC) or neoadjuvant chemotherapy (NACT) in addition to concurrent chemoRT (CCRT) are uncertain; these issues need to be addressed clearly using reliable and high-quality CPGs. However, the quality and recommendations of NPC CPGs have not been specifically evaluated. Therefore, we aimed to critically appraise the quality of CPGs for NPC using the AGREE II instrument and assessed the consistency of their recommendations.

Methods

CPG Selection

CPGs were defined as systematically developed statements aiming to assist practitioner and patient decisions regarding appropriate healthcare for specific circumstances. We focused on CPGs providing recommendations on NPC diagnosis and management published in English. When >1 set of documents was produced by the same organization, the most up-to-date version was considered. We excluded guidelines that (1) focused exclusively on H&N cancer without specific recommendations for NPC; (2) focused entirely on unique techniques, such as RT planning; and (3) concentrated only on recurrent or metastatic NPC.

Two reviewers (Y.P.C. and L.C.) identified relevant studies using keyword searches of MEDLINE, ESMO, NCCN, NCI, National Guideline Clearinghouse (NGC), Cancer Care Ontario, the National Institute for Health and Care Excellence, the Scottish Intercollegiate Guidelines Network, and the Guidelines International Network. The following terms were used: “nasopharyngeal neoplasm,” “nasopharyngeal tumors,” or “nasopharyngeal cancers,” and “clinical practice guidelines,” “recommendations,” or “consensus.” The last search was performed in December 2015.

CPG Data Extraction

For each CPG, 2 reviewers (Y.P.C. and L.C.) independently collected guideline characteristics (eg, year of publication, authoring organization, country or region, level of development, funding source, whether it is distinct to H&N cancer), and key recommendations stated (eg, recommendations on diagnosis, management, and/or surveillance of NPC). The individually recorded details were compared, and any disagreement was resolved by consensus.

Quality Assessment

The quality of each CPG was assessed using the AGREE II instrument, which is composed of 23 items in 6 quality domains: (1) scope and purpose, (2) stakeholder involvement, (3) rigor of development, (4) clarity of presentation, (5) applicability, and (6) editorial independence (Table 1). Each item is scored on a 7-point Likert scale (7=strongly agree; 1=strongly disagree). Four investigators (Y.P.C., L.C., W.F.L., and Y.S.) independently appraised all selected CPGs while blinded to each other’s ratings; these investigators are all experienced in the multimodality treatment for NPC. Chance-adjusted intrerrater agreements were calculated independently for each item using Cohen’s κ statistics. Agreement was poor if κ≤0.2; fair if 0.2<κ≤0.4; moderate if 0.4<κ≤0.6; substantial if 0.6<κ≤0.8; good if κ>0.8; and perfect if κ=1. The mean score (range, 1–7) for each item was calculated from the 4 investigators’ scores. Standardized scores (range, 0%–100%) for each domain were calculated by
Table 1. Mean Item Scores for Clinical Practice Guidelines Assessed Using the AGREE-II Instrument (Range, 1–7)

<table>
<thead>
<tr>
<th>Domain</th>
<th>APHNECP (SD)</th>
<th>CHNCEP (SD)</th>
<th>ESMO (SD)</th>
<th>NCCN (SD)</th>
<th>NCI (SD)</th>
<th>NGC (SD)</th>
<th>SEOM (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain 1: scope and purpose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. The overall objective(s) of the guideline is (are) specifically described.</td>
<td>6.25 (0.50)</td>
<td>5.25 (0.96)</td>
<td>5.00 (0.82)</td>
<td>5.50 (0.58)</td>
<td>5.25 (0.50)</td>
<td>7.00 (0)</td>
<td>6.00 (1.00)</td>
</tr>
<tr>
<td>2. The health question(s) covered by the guideline is (are) specifically described.</td>
<td>6.25 (0.50)</td>
<td>4.75 (0.96)</td>
<td>5.00 (0.82)</td>
<td>5.25 (0.50)</td>
<td>5.50 (0.58)</td>
<td>6.75 (0.50)</td>
<td>4.00 (0.58)</td>
</tr>
<tr>
<td>3. The population (patients, public, etc.) to whom the guideline is meant to apply is specifically described.</td>
<td>6.00 (0.82)</td>
<td>4.25 (0.50)</td>
<td>4.50 (0.58)</td>
<td>5.50 (0.58)</td>
<td>5.75 (0.50)</td>
<td>6.75 (0.50)</td>
<td>4.00 (0.58)</td>
</tr>
<tr>
<td>Domain 2: stakeholder involvement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. The guideline development group includes individuals from all the relevant professional groups.</td>
<td>5.00 (0.82)</td>
<td>2.50 (0.58)</td>
<td>5.25 (0.96)</td>
<td>6.25 (0.50)</td>
<td>5.75 (0.50)</td>
<td>4.50 (0.58)</td>
<td>4.00 (0)</td>
</tr>
<tr>
<td>5. The views and preferences of the target population (patients, public, etc.) have been sought.</td>
<td>2.75 (0.50)</td>
<td>2.00 (0.82)</td>
<td>2.50 (0.58)</td>
<td>2.25 (0.50)</td>
<td>2.50 (0.58)</td>
<td>3.00 (0)</td>
<td>3.00 (0)</td>
</tr>
<tr>
<td>6. The target users of the guideline are clearly defined.</td>
<td>6.50 (0.58)</td>
<td>6.50 (0.58)</td>
<td>4.25 (0.50)</td>
<td>6.50 (0.58)</td>
<td>6.75 (0.50)</td>
<td>6.75 (0.50)</td>
<td>3.00 (0.58)</td>
</tr>
<tr>
<td>Domain 3: rigor of development</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Systematic methods were used to search for evidence.</td>
<td>4.25 (0.50)</td>
<td>1.50 (0.58)</td>
<td>2.50 (0.58)</td>
<td>3.25 (0.50)</td>
<td>3.25 (0.50)</td>
<td>7.00 (0)</td>
<td>3.00 (0.58)</td>
</tr>
<tr>
<td>8. The criteria for selecting the evidence are clearly described.</td>
<td>4.25 (0.50)</td>
<td>1.00 (0)</td>
<td>2.00 (0)</td>
<td>4.00 (0)</td>
<td>3.50 (0.58)</td>
<td>6.50 (0.58)</td>
<td>3.00 (0)</td>
</tr>
<tr>
<td>9. The strengths and limitations of the body of evidence are clearly described.</td>
<td>3.75 (0.50)</td>
<td>2.25 (0.50)</td>
<td>5.50 (0.58)</td>
<td>4.50 (0.58)</td>
<td>5.50 (0.58)</td>
<td>4.25 (0.50)</td>
<td>3.00 (0.58)</td>
</tr>
<tr>
<td>10. The methods for formulating the recommendations are clearly described.</td>
<td>5.75 (0.50)</td>
<td>2.00 (0.82)</td>
<td>3.75 (0.96)</td>
<td>4.00 (0)</td>
<td>4.25 (0.50)</td>
<td>6.75 (0.50)</td>
<td>3.00 (0)</td>
</tr>
<tr>
<td>11. The health benefits, side effects, and risks have been considered in formulating the recommendations.</td>
<td>4.00 (0)</td>
<td>3.50 (0.58)</td>
<td>4.25 (0.50)</td>
<td>4.50 (0.58)</td>
<td>5.00 (0)</td>
<td>3.75 (0.96)</td>
<td>3.00 (0)</td>
</tr>
<tr>
<td>12. There is an explicit link between the recommendations and the supporting evidence.</td>
<td>5.00 (0.82)</td>
<td>3.25 (0.50)</td>
<td>5.75 (0.50)</td>
<td>4.75 (0.50)</td>
<td>5.50 (0.58)</td>
<td>4.75 (0.50)</td>
<td>3.00 (0.58)</td>
</tr>
<tr>
<td>13. The guideline has been externally reviewed by experts prior to its publication.</td>
<td>3.00 (0)</td>
<td>2.00 (0.82)</td>
<td>2.50 (0.58)</td>
<td>3.50 (0.58)</td>
<td>2.75 (0.50)</td>
<td>5.50 (0.58)</td>
<td>2.00 (0)</td>
</tr>
<tr>
<td>14. A procedure for updating the guideline is provided.</td>
<td>2.00 (0)</td>
<td>1.00 (0)</td>
<td>2.50 (0.58)</td>
<td>4.25 (0.50)</td>
<td>5.25 (0.50)</td>
<td>3.25 (0.50)</td>
<td>2.00 (0)</td>
</tr>
<tr>
<td>Domain 4: clarity of presentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. The recommendations are specific and unambiguous.</td>
<td>4.75 (0.96)</td>
<td>4.00 (0.82)</td>
<td>6.50 (0.58)</td>
<td>6.50 (0.58)</td>
<td>6.00 (0.82)</td>
<td>4.25 (0.50)</td>
<td>5.00 (0.58)</td>
</tr>
<tr>
<td>16. The different options for management of the condition or health issue are clearly presented.</td>
<td>5.00 (0)</td>
<td>3.75 (0.50)</td>
<td>6.00 (0.82)</td>
<td>6.00 (0.82)</td>
<td>5.25 (0.50)</td>
<td>5.25 (0.50)</td>
<td>5.00 (0)</td>
</tr>
<tr>
<td>17. Key recommendations are easily identifiable.</td>
<td>4.50 (0.58)</td>
<td>4.25 (0.50)</td>
<td>6.25 (0.50)</td>
<td>5.75 (0.50)</td>
<td>5.75 (0.50)</td>
<td>5.50 (0.58)</td>
<td>5.00 (0.58)</td>
</tr>
<tr>
<td>Domain 5: applicability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. The guideline describes facilitators and barriers to its application.</td>
<td>2.25 (0.50)</td>
<td>3.75 (0.96)</td>
<td>2.00 (0)</td>
<td>2.00 (0)</td>
<td>1.75 (0.50)</td>
<td>2.00 (0)</td>
<td>2.00 (0)</td>
</tr>
<tr>
<td>19. The guideline provides advice and/or tools on how the recommendations can be put into practice.</td>
<td>2.00 (0)</td>
<td>2.75 (0.50)</td>
<td>2.50 (1.00)</td>
<td>2.50 (0.58)</td>
<td>2.75 (0.50)</td>
<td>3.50 (0.58)</td>
<td>2.00 (0.58)</td>
</tr>
<tr>
<td>20. The potential resource implications of applying the recommendations have been considered.</td>
<td>2.25 (0.50)</td>
<td>1.75 (0.50)</td>
<td>2.75 (0.96)</td>
<td>2.25 (0.50)</td>
<td>2.50 (0.58)</td>
<td>1.50 (0.58)</td>
<td>2.00 (0.58)</td>
</tr>
<tr>
<td>21. The guideline presents monitoring and/or auditing criteria.</td>
<td>3.50 (0.58)</td>
<td>3.00 (0)</td>
<td>1.75 (0.50)</td>
<td>2.50 (0.58)</td>
<td>2.50 (0.58)</td>
<td>2.00 (0)</td>
<td>2.00 (0.58)</td>
</tr>
<tr>
<td>Domain 6: editorial independence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22. The views of the funding body have not influenced the content of the guideline.</td>
<td>4.50 (0.58)</td>
<td>6.50 (0.58)</td>
<td>5.25 (0.50)</td>
<td>5.75 (0.50)</td>
<td>6.25 (0.50)</td>
<td>6.50 (0.58)</td>
<td>3.00 (0.58)</td>
</tr>
<tr>
<td>23. Competing interests of guideline development group members have been recorded and addressed.</td>
<td>6.50 (0.58)</td>
<td>4.75 (0.50)</td>
<td>6.5 (0.58)</td>
<td>6.75 (0.50)</td>
<td>6.00 (0)</td>
<td>5.50 (0.58)</td>
<td>7.00 (0)</td>
</tr>
</tbody>
</table>

Abbreviations: AGREE, Appraisal of Guidelines for Research & Evaluation; APHNECP, Asia Pacific Head and Neck Cancer expert panel; CHNCEP, Chinese Head and Neck Cancer expert panel; NGC, National Guideline Clearinghouse; SEOM, Spanish Society of Medical Oncology.
summing the 4 investigators’ scores and standardizing these as a percentage of the possible maximum score for that domain as:

\[
\frac{\text{score obtained} - \text{minimum score possible}}{\text{maximum score possible} - \text{minimum score possible}} \times 100\%
\]

The standardized scores for each domain and mean values for the 23 items were compared. Key recommendations stated were also compared in a descriptive manner. All statistical analyses were performed using STATA software, Version 12 (STATA Corp, College Point, TX).

**Results**

**Characteristics of Selected CPGs**

Eligible CPGs (Figure 1) were from 7 organizations: Asia Pacific H&N Cancer expert panel (APHNCEP), Chinese H&N Cancer expert panel (CHNCEP), ESMO, NCCN, NCI, NGC, and Spanish Society of Medical Oncology (SEOM); their characteristics are summarized in Table 2. All guidelines were published between 2012 and 2015. Two were developed in multiple countries: APHNCEP was approved in the Asia and Pacific area and ESMO in Europe. The remaining CPGs were produced in China, the United States, or Spain. A total of 5 guidelines were produced by professional organizations or governmental agencies, and funded by specialty societies or government; 2 were produced by expert consensus, receiving funding from industry. Four guidelines provided distinct recommendations for NPC compared with other H&N cancers.

**CPG Appraisal**

The domain scores (%) for each CPG, as assessed using the AGREE II instrument, are shown in Figure 2 and Table 3. Of the 6 domains, applicability (domain 5) scored consistently low across CPGs (range, 13.5%–30.2%); no CPG scored >50% in all 6 domains. Scope and purpose (domain 1; ≥73.6% for 4 CPGs) and editorial independence scored highest (domain 6; ≥75.0% for 6 CPGs). With the exception of domain 5, a total of 4 NPC CPGs scored >50% in the remaining 5 domains: APHNCEP (range, 50.0%–86.1%), NCCN (range, 51.6%–87.5%), NCI (range, 56.3%–85.4%), and NGC (range, 62.5%–97.2%). NGC scored highest on domains 1 and 3 (97.2% and 70.3%), and NCCN and NCI scored highest on domain 2 (both, 66.7%). NCCN also ranked first on domain 6 (87.5%), ESMO scored highest on domain 4 (87.5%), and CHNCEP scored highest on domain 5 (30.2%).

Table 1 presents the mean scores and standard deviations for the 23 items for each CPG. A substantial, good, or perfect agreement was observed for all items. On averaging the item scores for each of the 7 CPGs, 9 items scored fewer than half of the points available (<3.5); the most frequently poorly reported items were items 5, 7, 8, 13, 14, and 18 to 21 (Table 3). Unsurprisingly, the mean scores for all CPGs were low for items in domain 5 (items 18–21), indicating that applicability did not attract enough attention and was performed poorly (Table 3). The mean scores for items 5 (range, 2.00–3.00) and 13 (range, 2.00–5.50) were low, indicating that the guidelines seldom provide the views or preferences of the target population, or were externally reviewed by experts before publication. The remaining items with low scores (items 7, 8, and 14) were in domain 3 (rigor of development) and were covering evidence searching, selection methods, and procedures for updating the guideline.

**Key Recommendations of Guidelines**

Supplemental eTable 1 (available with this article at JNCCN.org) presents the key recommendations.
stated in all included CPGs. All CPGs provided recommendations on diagnosis, management, and surveillance of NPC, although the APHNCEP\textsuperscript{18} and NCI\textsuperscript{25} did not provide detailed surveillance recommendations. Considering diagnostic procedures, all CPGs advocated biopsy of the primary lesion. However, although NCCN\textsuperscript{24} and SEOM\textsuperscript{27} recommend neck node biopsy (eg, fine-needle aspiration), ESMO\textsuperscript{20} does not support this procedure and suggests it may reduce probability of cure. Nasopharynx, skull base, and neck imaging are recommended by all guidelines; CHNCEP,\textsuperscript{19} ESMO,\textsuperscript{20} and NCI\textsuperscript{25} clearly state a preference for MRI, whereas ESMO,\textsuperscript{20} NCI,\textsuperscript{25} NGC,\textsuperscript{26} and SEOM\textsuperscript{27} recommend PET/CT as the preferred method for imaging distant metastases. Except those by NGC,\textsuperscript{26} all guidelines recommend consideration of Epstein-Barr virus (EBV) DNA testing at diagnosis.

Regarding nonmetastatic/recurrent NPC, all guidelines advised RT alone for stage I. However, the recommended treatments for stage II NPC varied: APHNCEP,\textsuperscript{18} ESMO,\textsuperscript{20} and NGC\textsuperscript{26} recommend CCRT, whereas NCCN\textsuperscript{24} and NCI\textsuperscript{25} advocate CCRT followed by AC (category 2A for NCCN). In addition, NCCN\textsuperscript{24} lists CCRT alone (category 2B) or NACT plus CCRT (category 3) as choices, and NCI\textsuperscript{25} recommends RT alone for stage II NPC. SEOM\textsuperscript{27} recommends RT alone, whereas CHNCEP\textsuperscript{19} recommends RT alone for N0 and RT with or without concurrent chemotherapy for N1 NPC. For locoregionally advanced NPC, APHNCEP,\textsuperscript{18} ESMO,\textsuperscript{20} NGC,\textsuperscript{26} and SEOM\textsuperscript{27} recommend CCRT alone. NCCN\textsuperscript{24} recommendations are the same as those for stage II NPC. NCI\textsuperscript{25} advocates altered fractionation RT as an alternative, and both APHNCEP\textsuperscript{18} and CHNCEP\textsuperscript{19} focus on additional cetuximab. The agents recommended were relatively consistent for concurrent chemotherapy and AC, but inconsistent for NACT. For metastatic NPC, platinum-based combination chemotherapy was the most common basic therapeutic scheme, whereas individualized combinations of surgery, RT, and chemotherapy were recommended for recurrent NPC. CHNCEP,\textsuperscript{18} NCCN,\textsuperscript{24} NGC,\textsuperscript{26} and SEOM\textsuperscript{27} clearly define the interval and duration of surveillance. Only APHNCEP,\textsuperscript{18} CHNCEP,\textsuperscript{19} NCCN,\textsuperscript{24} and NCI\textsuperscript{25} consider monitoring of EBV DNA after treatment.

### Discussion

This study sought to assess the quality and consistency of recommendations in CPGs involving the diagnosis and management of NPC, to assist clinicians when selecting appropriate guidelines. Using the AGREE II instrument, CPGs achieved consistently low scores in applicability (domain 5). The guidelines developed by APHNCEP, NCCN, NCI, and NGC achieved scores >50% in the remaining 5 domains. Of the 23 AGREE II items, 9 were inadequately reported. Additionally, heterogeneity in treatment recommendations exists.

Developers of CPGs often fail to meet widely endorsed standards devised to ensure high-quality guidelines.\textsuperscript{29,30} Although oncology guidelines appear...
to be of better quality than those of other diseases in certain domains of guidelines. Reames et al.29 found the vast majority of oncology CPGs failed to adhere to the Institute of Medicine’s standards for trustworthy guideline development. Moreover, significant variation in CPG development processes and recommendations were reported for specific oncology subspecialties.11,32 Much work remains to make oncology guidelines as methodologically sound and evidence-based as possible. This analysis shows that current CPGs for NPC also vary in methodological quality and have inconsistent recommendations. Considering the specific biological and regional characteristics of NPC compared with other H&N cancers, it is important to improve the quality and consistency of NPC CPGs to guide high-quality, evidence-based cancer care.

AGREE II enables the systematic evaluation of CPG quality and has been validated and endorsed by WHO as an accepted standard for guideline development.33,34 Based on the rationale that a high methodological quality reflects CPG integrity, reproducibility, and transparency, we also assessed guideline development methods. Methodological quality was optimal for “scope and purpose” and “editorial independence,” (domains 1 and 6, respectively) whereas “applicability” (domain 5) received the lowest scores. In terms of individual items, applicability domain items (items 18–22) were of low quality, as expected. Additionally, most guidelines lacked explicit statements on whether the views and preferences of the target population were sought (item 5), whether the guideline had been externally reviewed by experts before publication (item 13), and the procedure for updating the guideline (item 14). Search methods and selection criteria (items 7 and 8) in for rigor of development (domain 3) were also inadequately reported. Lack of applicability may indicate that, despite investment of vast resources during development, little consideration is given to translation of CPG recommendations into practice.

For a guideline to be effective, it should take resource implications into account, describe potential facilitators and barriers, and provide clear advice on implementation. More effort is needed to improve conceptualization of implementation during development. Because patients may not have the same views as CPG developers on the outcomes they care most about (e.g., severe side effects vs survival benefits), it is important to consider the views and preferences of patients. Similarly, opinions from external experts are also important to ensure validity. The low scores for these 2 items may be explained by the difficulties in or lack of resources available for involving patients or external experts in guideline development. Certain items in the rigor of development domain (domain 3; items

![Figure 2. Comparison of the quality of the clinical practice guidelines using the AGREE II instrument.](image-url)

Abbreviations: APHNCEP, Asia Pacific Head and Neck Cancer expert panel; CHNCEP, Chinese Head and Neck Cancer expert panel; NGC, National Guideline Clearinghouse; SEOM, Spanish Society of Medical Oncology.)
7, 8, and 13) may not attract enough attention; these methodological aspects are often adequately performed, despite not being reported.\textsuperscript{35} Overall, the high-scoring items may correspond to aspects of development that are generally accepted and respected by various organizations. It is important for future projects, training, and other initiatives for CPG development to target the aspects corresponding to the low-scoring items to improve the overall methodology and quality of NPC guidelines.

Despite the differences in the rigor of development domain scores, the recommendations of the CPGs were generally consistent in terms of diagnosis and follow-up for NPC. All guidelines advocated biopsy of the primary site for diagnosis. However, only 3 guidelines (NCCN, ESMO, SEOM) provided recommendations on neck biopsy, and no firm recommendations were made. EBV DNA testing is of clinical significance for its valid effects on diagnosis, risk stratification, prognostication, and relapse supervision for NPC.\textsuperscript{36} Only APHNECP, ESMO, NCCN, and NCI recommend both pretreatment and post-treatment EBV DNA testing. For surveillance of NPC after treatment, NCCN provides relatively detailed recommendations.

Major distinctions across CPGs are mainly related to therapeutic strategies of combined chemorT. RT alone and CCRT with or without AC were the main choices for stage II NPC. Although the phase III trial by Chen et al\textsuperscript{37} found that CCRT provides a considerable survival benefit in stage II NPC, these patients underwent conventional 2-dimensional RT. As the advent of intensity-modulated RT (IMRT) has improved locoregional control and resulted in improved treatment outcomes, the benefit offered by additional chemotherapy is speculated to be smaller in patients treated with IMRT.\textsuperscript{38,39} This may explain why some CPGs recommended RT alone or CCRT without AC for stage II NPC. For locoregionally advanced NPC, CCRT is the basic therapeutic regimen; short-term evaluation of the phase III trial by Chen et al\textsuperscript{40} found that additional AC did not significantly improve survival. However, NCCN still prefers CCRT plus AC for locoregionally advanced NPC, whereas APHNECP, ESMO, NGC, and SEOM prefer CCRT alone. ESMO indicates that the efficacy of additional AC remains uncertain and that potential benefits and side effects should be considered.\textsuperscript{20} Moreover, the efficacy of additional NACT followed by CCRT has not been confirmed, and was recommended in some cases based on individualized factors (APHNECP), in patients with problematic RT planning (ESMO), or in highly symptomatic patients (SEOM); however, NACT should not negatively affect the optimal administration of CCRT (ESMO) in any case.\textsuperscript{18,20,28} The value of additional AC and NACT require further research. Therapeutic approaches for metastatic/recurrent NPC were inconclusive, and combinations of RT, chemotherapy, and/or surgery were advised based on the specific condition of each patient. Generally speaking, both NCCN and ESMO do well in their recommendations regarding the diagnosis and staging, and follow-up for NPC. In terms of treatment for NPC, the optimal combined chemorT strategies remain inconclusive, and inconsistency among CPGs may result from variations in processes for selecting evidence, and methods of interpreting data and formulating recommendations.\textsuperscript{41}

Use of AGREE II to evaluate methodologic quality depends on complete reporting of the guide-line development process.\textsuperscript{41} We evaluated the methodologic quality of the CPG publications, which are
Quality and Recommendations of NPC CPGs

There are several limitations in this study. First, we relied on materials reported in published CPGs; reporting quality could affect these findings, which potentially places reporting quality under scrutiny. Secondly, AGREE II was used to evaluate CPGs as a whole, and is not specific for individual recommendations within guidelines. However, a global appraisal of the guideline development process may reflect the validity of the recommendations to a certain extent. Finally, only CPGs written in English were included; guidelines in other languages may play an important role in guiding clinical practice in local regions.

Because NPC is relatively rare and has an extremely unbalanced geographic distribution, it is necessary to promote international collaboration for guideline development and create CPGs with regional features. For example, guideline development organizations could share interpretation of evidence, while accounting for the organizational or cultural diversity of organizations and/or countries to provide insight into the challenges in different areas to adjust guideline recommendations accordingly. Through such efforts, development of high-quality, regionally relevant CPGs for NPC will be achieved.

Conclusions

Increased efforts are required to make comprehensive resources available to guide healthcare providers and enhance delivery of high-quality, evidence-based care for NPC. Guideline developers should evaluate both methodological processes and the resulting end product to improve quality.

References


