Comparison of international breast cancer guidelines: are we globally consistent? Cancer Guideline AGREEmnt

S.E. Hogeveen,* D. Han,* S. Trudeau–Tavara,† J. Buck,† C.B. Brezden–Masley MD PhD,* M.L. Quan MD Msc,‡ and C.E. Simmons MD MSc*

Conclusions
Our review demonstrated consistency in quality and content for breast cancer practice guidelines published by various organizations. Future guidelines developed by these organizations should focus on how to implement and measure uptake of a guideline.

KEY WORDS
Breast cancer, practice guidelines, AGREE instrument, quality of care

1. INTRODUCTION
Breast cancer is a common disease that affects roughly 1 in 9 women across North America and the United Kingdom. As a result, an incredible amount of research has been dedicated to breast cancer care and management in the diagnostic, surgical, adjuvant, and metastatic settings. To enable clinicians to keep up to date with the literature and to guide treatment decisions, published research is often systematically reviewed, appraised, summarized, and presented as evidence-based clinical practice guidelines that are used globally to enhance health care.

Several studies have shown that the use of guidelines in the practice of clinical oncology can improve the care process, the quality of clinical decisions, and ultimately, patient outcomes. Such guidelines can provide recommendations that not only help clinicians to make treatment decisions, but also confirm the appropriateness of treatment policies currently in place. Clinical practice guidelines can also be directly beneficial to patients, empowering them to make informed health care choices. However, uptake and implementation of a guideline often depends on that guideline’s overall quality.

Several internationally recognized oncologic organizations publish guidelines in clinical oncology. Three...
of the main organizations are Cancer Care Ontario (CCO, Canada), the American Society of Clinical Oncology (ASCO, United States), and the National Institute for Health and Clinical Excellence (NICE, United Kingdom).

In general, clinical oncology practice guidelines are felt to be universal, because the research and evidence supporting them is globally available. The key recommendations of such guidelines should therefore be similar throughout North America and in the United Kingdom. However, the key recommendations are not the only guideline factor that can be appraised. The process by which guidelines are created, their clarity and presentation, and their applicability to the target audience are all key factors that can be assessed and compared. It has been shown that, when guidelines are poorly presented or difficult to follow, clinicians and others involved in patient care do not make use of them. It therefore seems that the quality of a guideline might affect knowledge translation of the key recommendations to the key stakeholders.

The quality of guidelines can be measured and quantified by applying the Appraisal of Guidelines for Research and Evaluation (AGREE) instrument, which is a validated assessment tool developed to provide a framework for assessment and monitoring of clinical practice guidelines. The 23 key items in the AGREE instrument are divided into 6 domains. Those domains are Scope and Purpose, Stakeholder Involvement, Rigour of Development, Clarity and Presentation, Applicability, and Editorial Independence. Each item in a domain is scored, and the item scores are compiled into an overall score for the relevant domain. In May 2009, a newer version of the AGREE instrument, with minor modifications to increase clarity, was released. Overall, the domains and the total number of items are the same in the original and the more recent AGREE instruments. The scoring methods are also consistent for both versions.

Assessments of quality across guidelines have been performed and reported in many areas of clinical medicine, including schizophrenia, osteoarthritis, low-back pain, and psoriasis. Those assessments used the original AGREE instrument to compare the quality of clinical practice guidelines in the field of medicine being examined. In general, all the studies found that guideline recommendations were consistent, but observed significant variability in the quality of guidelines across domains. It has been said that deficiencies in certain domains, specifically Stakeholder Involvement and Applicability, have the potential to undermine feasibility and the uptake of key recommendations by clinicians.

To date, oncology guidelines have not been similarly assessed for quality. The fact that multiple international organizations are publishing oncology guidelines presents an opportunity to determine whether the quality of those guidelines is consistent globally, or whether there are differences between organizations. Investigating differences between organizations is important, because if the various organizations make conflicting or unclear recommendations, physicians may be confused about which guideline to follow.

Here, we present an evaluation of breast cancer guidelines with similar themes from CCO, ASCO, and NICE. We compare and comment on guideline quality in various domains and correlate guideline content.

2. METHODS

To select guidelines for comparison, we searched for the most recently published clinical practice guidelines from CCO, ASCO, and NICE. All breast cancer guidelines from the target organizations were cross-referenced to identify guidelines on the same theme from all three organizations (Figure 1). Two large NICE guidelines address care options and provide guidance throughout the pathway of care for early-stage and for metastatic breast cancer. Because those guidelines contain multiple recommendations, they were subdivided into chapters by theme, with each chapter counting as a separate guideline.

The quality of each guideline was then independently evaluated by 4 different reviewers (including medical oncologists and trainees), using the AGREE instrument. The AGREE instrument requires at least 2 reviewers, but to increase reliability, 4 were chosen...
(as is recommended)\(^9\). Each domain in the AGREE instrument was individually scored; the scores for the individual item in each domain were then totalled and standardized as a percentage of the maximum possible score for that domain. The formula used to compute the standardized score, as published in the AGREE instrument\(^9,10\), is this:

\[
\text{Score obtained} - \text{minimum possible score} / \text{Maximum possible score} - \text{minimum possible score}.
\]

The scores were then collated and reviewed. Descriptive comparisons of the scores between organizations, within organizations, and across guidelines from different organizations were subsequently developed.

### 3. RESULTS

By cross-referencing all breast cancer guidelines from ASCO, CCO, and NICE, we identified three common themes (Figure 1):

- Use of adjuvant aromatase inhibitors (AI)
- Staging investigations and follow-up (SFU)
- Bisphosphonate (BP) use for metastatic breast cancer

Table I lists the selected guidelines.

#### 3.1 Comparison of Content Between Organizations

The key recommendations presented by each organization are consistent in overall message. The AI guidelines from ASCO, CCO, and NICE all recommended either 5 years of AI therapy or sequential therapy with tamoxifen (2–3 years, or 5 years) followed by an AI (2–3 years, or 5 years)\(^15-17\). The key recommendations in the SFU guidelines from ASCO and NICE all included annual follow-up mammography\(^18,19\). The SFU guideline from CCO focused more on postoperative staging than on follow-up\(^20\). The BP guidelines showed a consensus between the three organizations, in that all recommended that a BP should be offered to women with bone metastases (either oral clodronate, intravenous pamidronate, or intravenous zolendronic acid) to prevent skeletal-related events and to reduce bone pain\(^21-23\).

#### 3.2 Comparison of Quality Between Organizations

The quality of the evaluated guidelines showed significant variability, with the AGREE scores for each domain ranging from 8% to 100% for the various targeted organizations. Table II shows the variability. Overall, NICE had some of the highest scores across all domains in all guidelines, and ASCO had some of the lowest scores across all domains in all guidelines.

Evaluating quality across the organizations by guideline allowed for more direct comparisons to be drawn between the organizations with respect to quality. With respect to the AI guidelines, CCO had the highest scores and NICE had the lowest scores in the domains of Scope and Purpose, Rigour of Development, and Editorial Independence. Cancer Care Ontario also scored highest in the domain of Clarity and Presentation. Overall, the domain of Applicability had the lowest scores and the most variability across organizations: ASCO, 8%; CCO, 33%; NICE, 72% [Figure 2(A)]. Quality scores between the organizations were somewhat more consistent in the SFU guidelines, but significant variability was seen in the domain of Stakeholder Involvement: ASCO, 44%; CCO, 52%; NICE, 77% [Figure 2(B)]. Across four domains, NICE scored the highest. The ASCO scores were the lowest in all domains except for the domain of Editorial Independence, in which ASCO ranked evenly with NICE.

#### Table I

<table>
<thead>
<tr>
<th>Organization</th>
<th>Subject</th>
<th>Guideline title(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASCO</td>
<td>AI</td>
<td>American Society of Clinical Oncology Technology Assessment on the Use of Aromatase Inhibitors As Adjuvant Therapy for Postmenopausal Women with Hormone Receptor–Positive Breast Cancer: Status Report 2004 (16)</td>
</tr>
<tr>
<td>SFU</td>
<td></td>
<td>American Society of Clinical Oncology 2006 Update of the Breast Cancer Follow-Up and Management Guidelines in the Adjuvant Setting (19)</td>
</tr>
<tr>
<td>BP</td>
<td></td>
<td>American Society of Clinical Oncology 2003 Update on the Role of Bisphosphonates and Bone Health Issues in Women with Breast Cancer (22)</td>
</tr>
<tr>
<td>CCO</td>
<td>AI</td>
<td>The Role of Aromatase Inhibitors in Adjuvant Therapy for Postmenopausal Women with Hormone Receptor–Positive Breast Cancer (15)</td>
</tr>
<tr>
<td>SFU</td>
<td></td>
<td>Baseline Staging Tests in Primary Breast Cancer (20)</td>
</tr>
<tr>
<td>BP</td>
<td></td>
<td>Use of Bisphosphonates in Women with Breast Cancer (21)</td>
</tr>
<tr>
<td>NICE</td>
<td>AI</td>
<td>Hormonal Therapies for the Adjuvant Treatment of Early Oestrogen-Receptor-Positive Breast Cancer (17)</td>
</tr>
<tr>
<td>SFU</td>
<td></td>
<td>Follow-Up (18)</td>
</tr>
<tr>
<td>BP</td>
<td></td>
<td>Bone Metastases (23)</td>
</tr>
</tbody>
</table>

\(\text{a}\) The titles of the breast cancer guidelines that were found to have common themes are listed by organization. AI = use of adjuvant aromatase inhibitors; SFU = staging investigations and follow-up; BP = bisphosphonate use.
Finally, for guidelines on bp use in breast cancer, the consistency of quality scores once again improved. The asco scores were lowest or on par with the other organizations for all domains except for the domain of Applicability, for which asco had the highest score. The domain of Applicability showed the greatest variability in scores (asco, 75%; cco, 33%; nice, 64%). Scores were fairly even across all organizations for the domains of Rigour of Development (asco, 77%; cco, 77%; nice, 83%) and Clarity and Presentation (asco, 79%; cco, 88%, nice, 79%; Figure 2(c)). The scores for the latter domain were the least variable across organizations.

### 3.3 Quality of Guidelines Within Organizations

Overall, as can be seen in Table II, the quality of the guidelines varied significantly within the target organizations. This variability is seen despite consistency in authorship and protocol within each organization. The lowest scores were obtained in the domains of Applicability and Editorial Independence. The highest scores were obtained in the areas of Scope and Purpose and Clarity and Presentation.

#### 3.3.1 asco

The scores for the three guidelines published by asco varied greatly in all domains except for that of Stakeholder Involvement, with the most variability seen in the Applicability domain. The ai guideline received the lowest score seen in any domain across all themes: 8% for Applicability. The bp guideline received the highest scores for 4 of the 6 domains. The ai guideline had the lowest scores for 3 of the 6 domains [Figure 3(A)].

#### 3.3.2 cco

The cco guidelines for all three themes scored in the 80% range for Purpose and Scope. However, the scores for Editorial Independence varied from 29% to 100%, with 29% being the lowest score received by cco for any guideline. For 2 of the 3 guidelines, scores in the domain of Applicability were quite low [33% for the ai and bp guidelines, 58% for the sfu guideline; Figure 3(B)].

#### 3.3.3 nice

Overall, nice received consistently high scores. Its lowest score was 50% for the domain of Editorial Independence in the ai guideline. Overall, scores were lower for the ai theme than for the sfu theme. Scores across all three themes were fairly even for the domain of Applicability, with a low average of 67% [Figure 3(C)].

### 4. DISCUSSION

Our study demonstrates that, although the key recommendations offered by various oncology guideline-producing organizations are consistent, variability in the quality of breast cancer guidelines internationally is observed both within and between organizations. The formula published in the agree instrument to calculate domain scores is designed to account for inter-rater variability; the findings should therefore not be affected by differences between the reviewers. Nevertheless, the scores assigned by the 4 reviewers in the present study were reasonably consistent. Moreover, no single organization, domain, or theme was consistently scored significantly higher than the others. However, it is difficult to state that any single

---

**TABLE II** agree scores for each theme and organization

<table>
<thead>
<tr>
<th>agree domain</th>
<th>Score (%)</th>
<th>Score (%)</th>
<th>Score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Use of adjuvant aromatase inhibitors</td>
<td>Staging investigations and follow-up</td>
<td>Use of bisphosphonates</td>
</tr>
<tr>
<td></td>
<td>ASCO</td>
<td>CCO</td>
<td>NICE</td>
</tr>
<tr>
<td>Scope and purpose</td>
<td>80</td>
<td>89</td>
<td>64</td>
</tr>
<tr>
<td>Stakeholder involvement</td>
<td>46</td>
<td>52</td>
<td>60</td>
</tr>
<tr>
<td>Rigor of development</td>
<td>65</td>
<td>77</td>
<td>56</td>
</tr>
<tr>
<td>Clarity and presentation</td>
<td>62</td>
<td>85</td>
<td>67</td>
</tr>
<tr>
<td>Applicability</td>
<td>8</td>
<td>33</td>
<td>72</td>
</tr>
<tr>
<td>Editorial independence</td>
<td>67</td>
<td>100</td>
<td>50</td>
</tr>
</tbody>
</table>

*a* Each guideline was scored by domain, and then organized by theme and by organization. The scores range from 8% to 100%. Variability in scores is seen both between organizations and within organizations.

asco = American Society of Clinical Oncology; cco = Cancer Care Ontario; nice = U.K. National Institute for Health and Clinical Excellence.
(A) Guidelines on the use of adjuvant aromatase inhibitors (ai). Scores for the guidelines on use of ai were lowest and showed the most variability across organizations in the domain of Applicability. Guidelines from Cancer Care Ontario (cco) had the highest scores in the domains of Scope and Purpose, Rigour of Development, and Editorial Independence. The U.K. National Institute for Health and Clinical Excellence (nice) had the lowest scores in those areas. (B) Guidelines for staging investigations and follow-up. Scores for these guidelines were slightly more consistent between organizations. The U.K. National Institute for Health and Clinical Excellence scored highest most often; the American Society of Clinical Oncology (asco) scored lowest in every domain except the domain of Editorial Independence. (C) Guidelines on bisphosphonate use in metastatic breast cancer. The scores for these guidelines were the most consistent. Once again, the domain of Applicability showed the greatest variability, but not necessarily the lowest scores, which were seen in the domain of Editorial Independence.

**Figure 2** (A) Breast cancer guidelines from the American Society of Clinical Oncology. The scores for these guidelines varied most in the domain of Applicability and least in the domain of Stakeholder Involvement. The lowest scores for all guidelines from any organization are seen here, including an 8% in the domain of Applicability for the guideline on the use of adjuvant aromatase inhibitors. (B) Breast cancer guidelines from Cancer Care Ontario. The scores for these guidelines varied considerably in the domain of Editorial Independence. The lowest score was 29% in that domain for the guideline on bisphosphonate use. (C) Breast cancer guidelines from the U.K. National Institute for Health and Clinical Excellence. The Institute consistently scored higher overall. The scores were generally lower for guidelines on the use of adjuvant aromatase inhibitors than for guidelines on staging investigations and follow-up. ai = use of adjuvant aromatase inhibitors; sfu = staging investigations and follow-up; bp = bisphosphonate use.
organization or guideline achieved the overall highest or lowest score, because scores are not totalled across domains.

Recently, a newer version of the AGREE instrument—AGREE II—was developed. We compared the AGREE II instrument with the original AGREE instrument and determined that the scores obtained in the present study using AGREE would not be significantly altered if AGREE II had been used. Scores might have subtly differed had the present study been conducted using the updated AGREE II, but the differences would not have been large enough to change the stated conclusions, and the variability found in the quality of breast cancer guidelines internationally would be maintained.

The variability in quality within organizations is surprising, because the same people and processes are often used to create the guidelines for each theme. In fact, guidelines from CCO and NICE cite the same protocols for different themes and also show some consistency in authorship. The observed variability might be explained if the methods used to perform systematic reviews and to structure the guidelines were not transferable across themes, but this lack of transferability has not been established and may need to be explored further. For now, it certainly seems that these organizations should be better able to utilize a uniform outline or format to generate guidelines with improved within-organization consistency of quality.

The variability in quality between organizations can be attributed to different authors, different target audiences (that is, different nations), and different processes for guideline development. Keeping in mind that guidelines often help to shape policies for the funding of health care and that the resources available for funding cancer care vary globally, it is not unexpected that the strength of the recommendations and the presentation of the guidelines would vary internationally. In addition, there may be differences and discrepancies in the value placed on various components of a guideline, or even on the value of one theme compared with another between organizations and countries, that is ultimately reflected in the resources invested in the development process. Variability in cultures, value systems, and available resources all contribute to variability in quality between organizations.

Given that scores in the Applicability domain include some of the lowest across the organizations, there is room for improvement in that area, especially in the North American organizations. One component of Applicability is whether the guidelines are clear and can easily be applied in a clinical setting. When guidelines from various organizations are unclear or conflict, clinicians may become confused and frustrated. As a result, they may not apply the guidelines in their own practice. The procedures for updating the guidelines are also assessed with the AGREE instrument. Although new guidelines are regularly being developed, the process of updating existing guidelines in the face of new evidence needs improvement. That process is relevant to Applicability, even though it falls within the domain of Rigour of Development, because physicians may not make use of guidelines if they feel that the recommendations therein are out of date. Future research should evaluate the best ways of disseminating the guidelines to clinicians and to those involved in patient care. If the information is not reaching the target audience, the quality, consistency, and timing of guidelines will not make a difference. What is reassuring, however, is the consistency of key recommendations from each of the organizations targeted in the present study. Although quality varied, it seems that clinicians internationally are receiving the same overall guidance for the care of breast cancer patients.

5. CONCLUSIONS

The key recommendations in breast cancer guidelines published by ASCO, NICE, and CCO are consistent. There is, however, variability in guideline quality between and within organizations internationally. In the future, research should concentrate on exploring the most effective methods of disseminating and evaluating implementation of clinical practice guidelines. Ensuring the dissemination of high-quality guidelines through clear implementation and update plans will further improve the quality of breast cancer care globally.

6. CONFLICT OF INTEREST DISCLOSURES

The authors have no financial conflicts of interest to disclose.

7. REFERENCES


**Correspondence to:** Christine Simmons, Medical Oncology, St. Michael’s Hospital, 30 Bond Street, Toronto, Ontario M5B 1W8.

**E-mail:** simmonsch@smh.ca

* St. Michael’s Hospital, Toronto, ON.
† Sunnybrook Health Sciences Centre, Toronto, ON.
‡ Foothills Medical Centre, Calgary, AB.