BMC Health Services Research

Research article

Quality and methods of developing practice guidelines

Hugh Cruse, Magdalena Winiarek, Jan Marshburn, Otavio Clark and Benjamin Djulbegovic*

Address: H Lee Moffitt Cancer Center and Research institute, the University of South Florida, Tampa, FL, USA

E-mail: Hugh Cruse - crusehh@moffitt.usf.edu; Magdalena Winiarek - winiarmb@moffitt.usf.edu; Jan Marshburn - marshbmj@moffitt.usf.edu; Otavio Clark - clarkoa@moffitt.usf.edu; Benjamin Djulbegovic* - djulbebm@moffitt.usf.edu

*Corresponding author

Abstract

Background It is not known whether there are differences in the quality and recommendations between evidence-based (EB) and consensus-based (CB) guidelines. We used breast cancer guidelines as a case study to assess for these differences.

Methods Five different instruments to evaluate the quality of guidelines were identified by a literature search. We also searched MEDLINE and the Internet to locate 8 breast cancer guidelines. These guidelines were classified in three categories: evidence based, consensus based and consensus based with no explicit consideration of evidence (CB-EB). Each guideline was evaluated by three of the authors using each of the instruments. For each guideline we assessed the agreement among 14 decision points which were selected from the NCCN (National Cancer Comprehensive Network) guidelines algorithm. For each decision point we recorded the level of the quality of the information used to support it. A regression analysis was performed to assess if the percentage of high quality evidence used in the guidelines development was related to the overall quality of the guidelines.

Results Three guidelines were classified as EB, three as CB-EB and two as CB. The EB guidelines scored better than CB, with the CB-EB scoring in the middle among all instruments for guidelines quality assessment. No major disagreement in recommendations was detected among the guidelines regardless of the method used for development, but the EB guidelines had a better agreement with the benchmark guideline for any decision point. When the source of evidence used to support decision were of high quality, we found a higher level of full agreement among the guidelines' recommendations. Up to 94% of variation in the quality score among guidelines could be explained by the quality of evidence used for guidelines development.

Conclusion EB guidelines have a better quality than CB guidelines and CB-EB guidelines. Explicit use of high quality evidence can lead to a better agreement among recommendations. However, no major disagreement among guidelines was noted regardless of the method for their development.

Background

The objective of guidelines development is to assist physicians and patients in making optimal health care decisions, which in turn should improve the quality of clinical practice [1].

Different methods are used to develop guidelines. Some are developed by a consensus of experts while others also use a formal way to appraise the literature and create evidence-based (EB) guidelines. In general, evidence-based guidelines are considered to provide better recommendations for practice than consensus-based guidelines but are time consuming and expensive to create [2,3]. This belief that EB guidelines are superior to other types of guideline is based on our normative views of methods for guidelines development [4] and not on empirical comparison of practice recommendations using different methods for development of guidelines. To date no formal evaluation has been performed to detect if there are differences in the quality and recommendations between evidence-based and consensus-based (CB) guidelines.

If guidelines developed by using consensus or evidence-based methods have the same quality and agree in the recommendations, then obviously resources spent on the laborious and time-consuming process of locating and appraising evidence can be used elsewhere. Otherwise, if evidence based guidelines have a better quality and their recommendations differ from those guidelines produced by consensus, then creation of evidence based guidelines may become the only acceptable method of guideline development.

In this paper, we explore if there are differences in the quality and recommendations between EB and CB guidelines.

Methods

To enable meaningful comparison, multiple recommendations produced by a given guideline method should be available. This objective is best met by focusing on the guidelines that comprehensively attempt to guide clinicians in the management of one disorder. Since breast cancer is an important disease and various organizations have produced guidelines using different methods [5–7], we conducted a comparison study of comprehensive breast cancer guidelines. We assessed both the differences in the quality as measured by using different quality instruments assessment and the level of agreement among guidelines according to the method of development.

1. Identification and assessment of instruments for measurement of the quality of guidelines

Table 1: Interobserver agreement of instruments for assessment of the guidelines quality

Instrument	Guidelines with K > 0.4 among all four evaluators
Cluzeau (8)	ACCC (19), CMA (7)
Grilli (9)	ACCC (19), NHMRC (16), MPS (18)
Sanders(11)	None
Petrie (SIGN) (10)	NCCN (6), ACCC (19), SIGN (17), ICSI (15), MPS (18), SSO (5)
Shaneyfelt 1999 (12)	ACCC (19), CMA (7), NHMRC (16), ICSI (15), MPS (18), SSO (5)

Acronyms: ACCC – Association of Community Cancer Centers; CMA-Canadian Medical Association; ICSI – Institute for Clinical Systems Improvement; MPS – Multi Professional Societies; NCCN – National Comprehensive Cancer Network; NHMRC – The National Health and Medical Research Council; SIGN – Scottish Intercollegiate Guidelines Network; SSO – Society of Surgical Oncology.

Since there is no uniformly accepted instrument for evaluation of the quality of guidelines, we first performed a comprehensive literature search to identify published tools for assessment of clinical practice guideline quality. We searched MEDLINE (1996–2000) using the keywords: guidelines, practice guidelines, quality, "weights and measures", "scale", psychometrics, reproducibility. Any article considered relevant to evaluate quality of guidelines was retrieved. The list of references of each article was also scanned. After an assessment of 14 papers by four of us, four instruments to assess the quality of guidelines were identified [8–12]. An additional instrument (SIGN) was identified through Evidence-based Health Discussion Group (Table 1). For additional details on the instruments for evaluation of guidelines readers are referred to the Appendix (see 1Additional file). To assess their reliability and reproducibility, we applied all identified instruments to each guideline (see below). We calculated the coefficient of agreement (kappa) among evaluators for each guideline [13]. A good interobserver agreement was considered if kappa value exceeded 0.4 [13]. In our evaluation, two instruments [10,12] had a kappa interobserver agreement K > 0.4 among all investigators in 6 of 8 guidelines (Table 1). When it comes to evaluation of the quality of breast cancer guidelines these instruments [10,12] performed better than others and probably can be recommended for future use.

2. Identification and classification of breast cancer guidelines

A literature search was conducted for published breast cancer guidelines using MEDLINE for the years 1996 – April 2000. The following keywords were used in combi-

nation: Guidelines, Practice Guidelines, recommendations, breast neoplasms. An Internet search was also performed, using the method described by Sanders et al [11]. 131 articles were retrieved, and reviewed for their content. We considered any article that fit the definition of the National Library of Medicine for practice guidelines: directions or principles presenting current or future rules of policy for the health care practitioner to assist him in patient care decisions regarding diagnosis, therapy, or related clinical circumstances [14]. Eight papers referred to breast cancer guidelines [5–7],[15–19] and were selected for the analysis.

Each guideline was classified as CB, when there was no consideration about the quality of evidence used to make practice recommendations; as EB, when there was an explicit consideration of the quality of evidence in the development of guidelines; or as consensus based with no explicit consideration of evidence (CB-EB) when there were considerations about the evidence, but not in explicit manner. From these eight guidelines, three were classified as EB [17,7,16] three as CB-EB [15,18,6] and two as CB [19,5] (Table 2).

Table 2: Classification of Breast Cancer Guidelines according to the method of development.

Evidence Based	Consensus Based with no explicit consideration of evidence	Consensus Based
CMA (7)	NCCN (6)	ACCC(19)
SIGN (17)	ICSI (15)	SSO (5)
NHMRC(16)	MPS(18)	

Acronyms: see footnote in Table 1

3. Evaluation of guidelines

Each guideline was evaluated independently by three of us using each of the instruments. All discordances were resolved by a consensus meeting. Each guideline was scored according to the instructions of each instrument. The quality and rank was determined by the quotient of items scored positively by the total items scored for each instrument.

4. Evaluation of agreement among guidelines

Using instruments to evaluate practice guidelines yields conclusions regarding normative aspects of the guidelines development [1], but does not necessarily mean that recommendations provided by guidelines using different methods will produce different management advice to our patients. To assess if recommendations among various guidelines differ, we need to determine the level of agreement among guidelines for each specific decision point.

Since NCCN (National Comprehensive Cancer Network) guidelines [6] were presented in explicit, algorithmic format, we used this one to identify the decision points for matched comparison with other guidelines. These guidelines have been developed by the leading 18 cancer institutions in the US and have been constantly updated and re-evaluated. They have also been developed to closely mimic clinical practice. Therefore, we feel that selection of decision points based on the NCCN guidelines were appropriate. We identified fourteen decision points in the management of stage I and II breast cancer that were linked to specific recommendations in the other guidelines for our comparison. Comparison of recommendations for advanced stages of breast cancer has not been performed since there was only one guideline that included it [6].

Table 3: Quality of breast cancer guidelines

	Instruments for assessment of guidelines quality					
	Cluzeau (8)	Grilly (9)	Sanders (11)	Shaneyfelt (12)	Petrie (SIGN) (10)	Average Score
SIGN(17)	0.94	1.00	1.00	0.84	0.82	0.92
NHMRC(16)	0.77	1.00	1.00	0.80	0.75	0.86
CMA(7)	0.74	1.00	0.93	0.88	0.65	0.84
ICSI(15)	0.77	0.67	1.00	0.64	0.55	0.72
NCCN(6)	0.66	0.67	0.86	0.60	0.44	0.64
MPS(18)	0.37	0.67	0.80	0.40	0.37	0.52
ACCC(19)	0.36	0.00	0.60	0.40	0.32	0.34
SSO(5)	0.18	0.00	0.20	0.20	0.18	0.15

Acronyms: see footnote in Table 1

Table 4: Level of agreement between NCCN guideline and other breast cancer guidelines.

Guideline	Agreement	Partial Agreement	Statistical significance (p)	N. of decision points with high quality evidence
SIGN(17)	5	9	.285	13
NHMRC(16)	8	6	.593	13
CMA(7)	9	5	.285	9
ICSI(15)	4	10	.109	6
NCCN(6)	-	-	-	3
MPS(18)	0	14	<0.0001	0
ACCC(19)	9	5	.285	0
SSO(5)	0	14	<0.0001	0

Acronyms: see footnote table 1

Subsequently, four of us evaluated each of these decision points in each guideline examining level of agreement among various guidelines. Since matching between recommendations in the guidelines that were presented in non-algorithmic format was poor, we decided to use NCCN guidelines as a benchmark. We classified agreement of each guideline with the NCCN guidelines as having full agreement, partial agreement and disagreement. It was considered that guidelines agree with the NCCN if the management recommendation was the same; the guidelines were considered to disagree if they provided different recommendations. A partial agreement was judged to exist if the guideline recommended the same management but in a broadly defined sense and not in explicit, clear manner.

Each of these decision points was also classified as supported by high quality evidence or not. High quality evidence was considered to be based on randomized trials (RCT) or systematic reviews (SR)/meta-analysis (MA). If the quality evidence was not based on RCT or SR/MA or was not stated, it was classified as low quality evidence.

Subsequently, we performed a regression analysis to assess the contribution of the quality of evidence to the total score obtained by each instrument for the evaluation of the guidelines quality. Independent variable was the proportion of decisions supported by high quality evidence while dependent variable was score obtained by each instrument. A regression analysis was performed after it has assessed that the distribution of the variables was normal by Wilks-Shapiro test.

Results
Evaluation of the quality of guidelines
The results of the quality of each guideline according to each instrument are shown in Table 3. Overall, EB guidelines had higher scores than CB, and the CB-EB category ranked in the middle (Fig 1). As expected, the instruments for the evaluation of quality are based on the number of desired built-in normative features of good guidelines development, as initially recommended by Institute of Medicine[1]. This is further confirmed by the evaluation of the contribution of the quality of evidence to the final quality score: the regression analysis performed showed that the quality of guidelines, as measured by these instruments, is a function of the percentage of high quality evidence that each guideline contains. This suggests that evidence plays a major role in the composition of the quality scales. If the quality of evidence is poor, paying attention to other quality domains in the development of guidelines will not result in higher quality scores. Fig 2 illustrates a relationship between the quality of evidence and the total quality score using the two instruments that achieved best agreement among evaluators [10,12]. It is quite remarkable to note that up to >94% variation in the score could be explained by the quality of evidence alone.

Evaluation of agreement among guidelines
The agreement among each guideline for the 14 decision points is shown in Table 4. We obtained no major disagreements among guidelines, but the EB guidelines had a better agreement with the decision points in any situation than CB-guidelines and CB-EB guidelines. The fact that no major disagreements were seen regardless the method of development can probably be explained by the vagueness of recommendations by CB guidelines. As shown in Table 4, the number of decision points supported by high quality evidence is highest in the EB guidelines and zero in CB guidelines. The use of high quality evidence was significantly associated with a higher level of concordance among the decision points. When the source of evidence was of good quality (RCT or SR), we had 18 full agreements and 23 partial agreements (Chi square = 0.610, degrees of freedom = 1, p = 0.435). When the source of
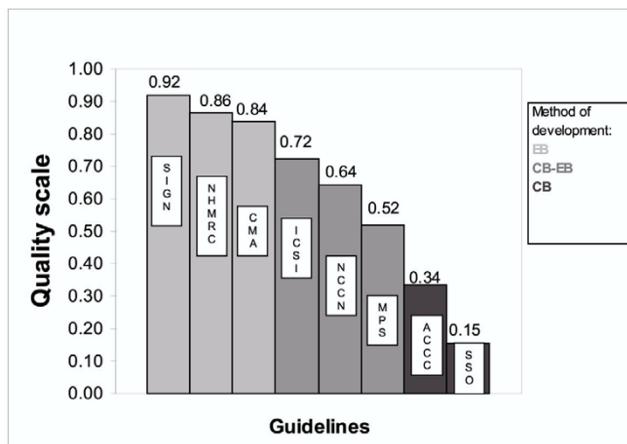
Figure 1
Average score of each guideline according to the method of development Acronyms and abbreviations: ACCC – Association of Community Cancer Centers; CMA-Canadian Medical Association; ICSI – Institute for Clinical Systems Improvement; MPS – Multi Professional Societies; NCCN – National Comprehensive Cancer Network; NHMRC – The National Health and Medical Research Council; SIGN – Scottish Intercollegiate Guidelines Network; SSO – Society of Surgical Oncology. EB: evidence-based guidelines; CB: consensus-based guidelines EB-CB: consensus-based guidelines with no explicit considerations of evidence

evidence was not stated or was of lower quality, we had 17 full agreements and 40 partial agreements (Chi-Square 9.281, Degrees of freedom 2, p= 0.002). This means that recommendations based on high quality evidence may lead to less disagreement and potentially less practice variation.

Discussion
Guidelines have been increasingly used in medical decision-making. Different methods have been used in guideline development. Does it matter how guidelines were produced? Most authors believe that it matters very much [3] and that guidelines produced using evidence-based methods are superior to other methodologies of development [2,4,9]. However, empirical investigations to assess if guidelines produced by different methods have different quality and result in different recommendations have not been performed. Here, we report such a study.

Using formal instruments for evaluation of the quality of guidelines we found that EB-guidelines had substantially higher score than CB-guidelines or guideline that considered evidence in a less formal way (CB-EB). As discussed above (see Results), this is not a surprising result, since the instruments for the guidelines evaluation measure the quality based on the number of desired normative charac-

teristics in a particular guideline. Since appraisal of evidence is considered inherently important for the development of a good guideline, one would then expect that the guidelines that pay more attention to its evidence basis (i.e., those that are evidence-based) would receive higher quality score than other types of the guidelines (i.e. guidelines developed solely by a consensus process) (see Fig 1). This is also evident in our finding that variation in the total quality score can be up to 94% explained by the quality of evidence (see Fig 2).

Not all instruments for evaluation of guidelines performed equally well. Only two of the instruments available to address the quality of guidelines had a good level of agreement among evaluators (k > 0.4) in most of guidelines. This result raises concern about the reproducibility of results using the other instruments reported in the literature. In general, a few studies have been done to evaluate reproducibility of the instruments for assessment of the guidelines quality. Any future study attempting to address the quality of guidelines should take this finding into account.

A more interesting question is to assess if the recommendations among guidelines produced by different methods actually differ. We found no instance of total disagreement among guidelines regardless of the method of development. We also found that EB and CB-EB guidelines had more points of agreement with our benchmark guidelines (NCCN)[6] than guidelines developed using exclusively consensus method. We also found that when high-quality evidence existed in the literature (see Results) less disagreement was found among various guidelines. This is not completely surprising because formulation of guidelines does not happen in a vacuum. Most guideline developers are experts in the field who have knowledge of the literature. When evidence is unequivocal, less disagreement may be expected. Consequently, less practice variation may be found when high-quality evidence exists.

Conclusions
In conclusion, EB guidelines have a better quality than CB guidelines as measured by the quality assessment instruments used in this study. The explicit use of high quality evidence is desirable and can lead to a better agreement among recommendations. However, no major disagreement among guidelines was noted regardless of the method for their development.

Competing Interest
none declared

Acknowledgements
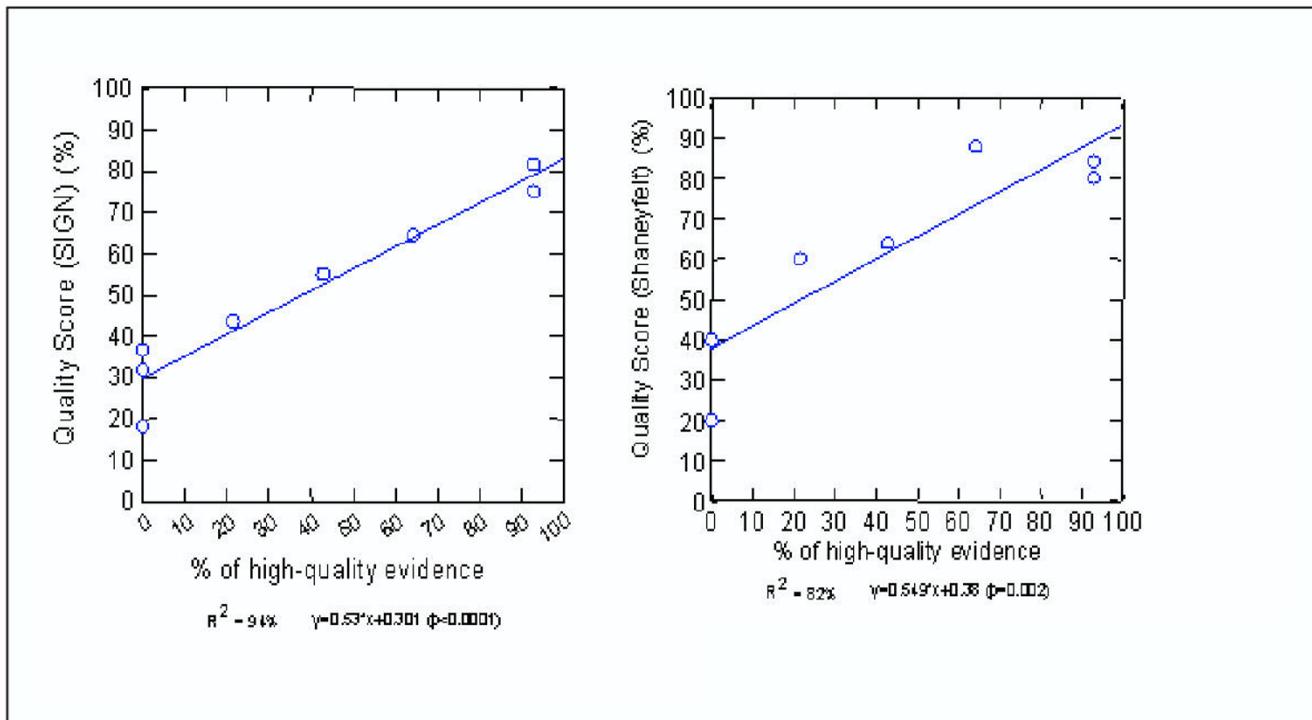
Figure 2
A relationship between quality of evidence and total guideline quality score. Note that up to 94% of variation in the quality score can be explained by the quality of evidence.

Additional material

Additional file
The file contains bibliographic details and description of the published instruments for evaluation of the quality of practice guidelines
Click here for file
[http://www.biomedcentral.com/content/supplementary/1472-6963-2-1-S1.doc]

References

1. **Institute of Medicine. Guidelines for clinical practice: from development to use.** *Washigton DC: National Academic Press;* 1992
2. Woolf SH: **Evidence-based medicine and practice guidelines: an overview.** *Cancer Control* 2000, **7**:362-7
3. Miller J, Petrie J: **Development of practice guidelines.** *Lancet* 2000, **355**:82-3
4. Eddy DM: **Clinical decision making: from theory to practice. Practice policies-guidelines for methods.** *JAMA* 1990, **263**:1839-41
5. Morrow M, Bland KI, Foster R: **Breast cancer surgical practice guidelines. Society of Surgical Oncology practice guidelines.** *Oncology (Huntingt)* 1997, **11**:877-81
6. **Update: NCCN practice guidelines for the treatment of breast cancer. National Comprehensive Cancer Network.** *Oncology (Huntingt)* 1999, **13**:187-212
7. **The Steering Committee on Clinical Practice Guidelines for the Care and Treatment of Breast Cancer.** *CMAJ* 1998, **158**:S1-2
8. Cluzeau F, Littlejohns P, Grimshaw J, Feder G: **Appraisal Instrument for Clinical Guidelines.** *London: St. George's Hospital Medical School;* 1997Available from: St. George's Hospital Medical School web site [http://www.sghms.ac.uk/depts/phs/hceu/clinguid.htm] . Accessed 11 June 2001.
9. Grilli R, Magrini N, Penna A, Mura G, Liberati A: **Practice guidelines developed by specialty societies: the need for a critical appraisal.** *Lancet* 2000, **355**:103-6
10. Petrie J, Barnwell E, Grimshaw J: **Criteria for Appraisal for National Use – Scottish Intercollegiate Guidelines Network (SIGN).** *SIGN Publication Number 39, 1995. Edinburgh: Scottish Intercollegiate Guidelines Network (SIGN);* 1995Available from SIGN web site [http://www.sign.ac.uk/guidelines/fulltext/50/index.html] (Version 2001). Accessed 11 June 2001.
11. Sanders GD, Nease RF, Owens DK: **Design and pilot evaluation of a system to develop computer-based site-specific practice guidelines from decision models.** *Med Decis Making* 2000, **20**:145-59
12. Shaneyfelt TM, Mayo-Smith MF, Rothwangl J: **Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature.** *JAMA* 1999, **281**:1900-5
13. Landis J, Koch G: **A measurement of observer agreement for categorical data.** *Biometrics* 1977, **33**:159-174
14. **NLM PubMed resources page.** 2001Available via internet [http://www.ncbi.nlm.nih.gov/entrez/meshbrowser.cgi?term=Practice+Guidelines&retrievestring=&mbdetail=n] Accessed 11 June 2001.
15. **ICSI Institute for Clinical Systems Improvement. ICSI Health Care Guideline: Breast Cancer Treatment.** *ISCI,* 2000Available from: ICSI web site [http://www.icsi.org/guidelst.htm] . Accessed 11 June 2001
16. **NHMRC-AU. NHMRC National Breast Cancer Centre – Clinical Practice Guidelines For The Management of Early Breast Cancer 1999.** *NHMRC-AU* 1999Available from: NHMRC-

AU web site [http://www.health.gov.au/nhmrc/advice/pdf/early-brs.pdf] (Version 2000). Accessed 11 June 2001.

17. **SIGN – Scottish Intercollegiate Guidelines Network. Breast Cancer in Women – A National Clinical Guideline. SIGN Publication Number 29,1998.** *Edinburgh: Scottish Intercollegiate Guidelines Network (SIGN);* 1998Available from SIGN web site [http://www.sign.ac.uk/pdf/sign29.pdf] . Accessed 11 June 2001.

18. Winchester DP, Cox JD: **Standards for diagnosis and management of invasive breast carcinoma. American College of Radiology. American College of Surgeons. College of American Pathologists. Society of Surgical Oncology.** *CA Cancer J Clin* 1998, **48**:83-107

19. **ACCC – Association of Community Cancer Centers. Oncology Patient Management Guidelines. Breast Carcinoma version 3.0.** *ACCC* 1999