Assessment of Malaysian Clinical Practice Guidelines

B. Rugayah M.P.H.¹, M.D. Noormah M.Sc¹, M.M. Mohamed Ph.D*², S.F.K. Shahnaz³

¹Malaysian Health Technology Assessment Section, Ministry of Health Malaysia.

²University Teknologi Mara, Puncak Alam Campus, Selangor.

³Hospital Tengku Ampuan Rahimah

* Corresponding author

Malaysian Journal of Pharmacy, Vol 1 Issue 10, 2012, 1-14

Abstract

The Ministry of Health Malaysia (MOH) coordinates the development of Clinical Practice Guidelines (CPGs) in Malaysia, in collaboration with the Academy of Medicine of Malaysia (AMM). This study assessed the methodological quality of 29 Malaysian national CPGs which were developed since 2000 to 2003 using Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument. The study showed high score for only domains on Scope & Purpose as well as Clarity & Presentation (68%, 75% respectively).

Keywords: clinical practice guidelines, AGREE instrument, appraisal of guidelines, Malaysian CPGs.

Introduction

Assessment of the national Malaysian Clinical Practice Guidelines (AMCPG) using Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument was proposed to study the methodological quality in the development of Clinical Practice Guidelines (CPGs) in Malaysia.

The Ministry of Health (MOH) coordinates the development of evidence-based CPGs in Malaysia. This was a collaborative effort between Ministry of Health Malaysia (MOH) and Academy of Medicine Malaysia (AMM). The goal of the CPG development was to increase quality in the delivery of health care services based on clinical evidence with scientific rigor.

This (AMCPG) Project was found to be worthy to be conducted in Malaysia, in view of the fact that Malaysia had developed several CPGs. An evaluation

of these developed CPGs could provide more information in improving the development of evidence-based CPGs in Malaysia. It is believed that evidence-based CPGs can help improve the delivery of health care, although proof to such claim has not been consistently demonstrated.

Evaluation of the CPG can be categorized into three levels^{3, 4, 5} namely: 1) Examination of the process of guideline development, dissemination and implementation; 2) Measurement of the extent of implementation of the guideline and 3) Assessment of guidelines effect on patient outcomes and health care utilization. Another way of evaluation was to classify into two levels, namely quality of the CPG and later its effectiveness. Good development methodology using current best evidence will determine the quality of the CPG. A "good quality guideline" is the one that ultimately leads to improve patient outcome. However, the

quality of a guideline is indirectly measured by assessing in whatever degree guideline producers minimized potential biases that could occur in the development process and affect validity of its recommendations^{6, 7, and 8}. Wrong recommendations affect the health professionals' credibility on guidelines, and consequently, limit their adoption.

In 1999, Shaneyfelt et al. assessed quality of CPG published in Medline between 1985 and 1997 by using systematically developed instrument. The majority of 279 assessed guidelines did not meet the pre-established methodological standards, being rigour of recommendations as one of the most deficiently reported^{6, 9}. Similar results were reported by Cluzeau et al.^{6, 10}, Grilli et al.^{6,11} and Graham et al.^{6,12} in 1999, 2000 and 2001 respectively. In 2003 the AGREE collaboration (currently the AGREE Research Trust) published the results of the first international project aimed at developing and validating a generic instrument for guidelines assessment^{7,8}. This instrument has been translated to different languages and extending its use throughout the world. In recent years, several studies showed methodological deficiencies of using the AGREE instrument in guideline development^{6,13-15}

In Malaysia, although many different institutions are interested in CPG development, there is no information about the quality of the guidelines produced. The purpose of this research was to describe trends in guidelines production in Malaysia and to assess their quality by using the AGREE instrument.

Materials and methods

A cross-sectional study was undertaken in 2004 to describe guidelines

production in Malaysia between years 2000-2003. Documents were considered as CPG if: 1) they included explicit recommendations targeted to health professional or health providers decision-making in managing diseases or condition, 2) the scope included related to screening and primary prevention, and/or diagnosis, and/or treatment and/or secondary prevention and/or rehabilitation; 3) they contained description of participants or responsible institutions and bibliographic references; 4) they were produced and diffused in the period of study (January 2000 to December 2003) and could be freely accessed. The exclusion criteria were: 1) guidelines targeted to patients (patients 'guidelines) and/or exclusively oriented to health services organization and not to clinical decision-making for managing diseases or conditions; 2) guidelines for which it was not possible to determine if a systematic process was applied in their development such as documents that lacked an explanation of the guideline development methodology that had been used or documents diffused as brief reports which only contained a set of recommendations or documents referred to as guidelines, but were undertaken by only one author without any reference to the methodology applied); 3) guidelines whose year of development could not be established as it was not stated and last but not least 4) guidelines that were not produced by a Malaysian institution (adapted guidelines were included only when the adaptation process was explicitly explained).

All guidelines registered by the Health Technology Assessment units (HTA) in MOH and AMM between January 2000 and December 2003 were selected for this study. The original published CPGs or a photocopy of the original CPGs were retrieved from the HTA unit, the chairman of the guideline developers

group/or downloaded from the Ministry of Health Malaysia website.

Quality guideline assessment was performed using the AGREE instrument. This instrument was the instrument of choice as it covers practically all the relevant dimensions of the evidence-based guideline development process. In addition, it has been internationally validated. The AGREE has fewer items and uses a numerical scale that facilitates the analysis^{8, 16, 17}.

A total of four appraisers were invited to participate voluntarily in the assessment phase. To be considered eligible, professionals should have had at least one of the following criteria: a) previous clinical epidemiology background; and b) knowledge on guidelines development. The professionals who accepted the invitation and fulfilled the eligibility criteria were trained in the use of the AGREE instrument. A learning program was developed in two stages: I. Self-reading of the tool-kit: all participants were provided with the English version of the AGREE instrument, the English version of the Training Manual. II. Pilot assessment - one CPG was assessed independently by all professionals.

All of the 29 copies of the CPGs retrieved were given to each appraiser to be appraised within one month. A data collection form designed on an Excel sheet, accompanied by a user-guide on the AGREE instrument were given to each appraiser. Results of assessments were returned to the researcher team by mail. No assessor received any honorarium.

AGREE consists of 23 key items organized in six domains. Each domain is intended to capture a separate dimension of guideline quality. Domain 1: Scope and purpose (items 1-3) is

concerned with the overall aim of the guideline, the specific clinical questions and the target patient population. Domain 2: Stakeholder involvement (items 4-7) focuses on the extent to which the guideline represents the views of its intended users. Domain 3: Rigor of development (items 8-14) relates to the process used to gather and synthesize the evidence, the methods to formulate the recommendations and to update them. Domain 4: Clarity and presentation (items 15-18) deals with the language and format of the guideline. Domain 5: Applicability (items 19-21) pertains to the likely organizational, behavioral and cost implications of applying the guideline. Domain 6: Editorial independence (items 22-23) is concerned with the independence of the recommendations and acknowledgement of possible conflict of interest from the guideline development group.

Each item is rated on a 4-point scale ranging from 4 'Strongly Agree' to 1 'strongly Disagree', with two mid points: 3 'Agree' and 2 'Disagree'. The scale measures the extent to which a criterion (item) has been fulfilled. 'Strongly Agree' means that the appraiser was confident that the criterion has been fully met, and if the appraiser was confident that the criterion has not been fulfilled at all or if there is no information available then he/she should answer 'Strongly Disagree'. If the appraiser was unsure that a criterion had been fulfilled, for example because the information was unclear or because only some of the recommendations fulfill the criterion, then he/she should answer 'Agree' or 'Disagree', depending on the extent to which he/she thought the issue had been addressed.

According to the AGREE Collaboration the domain scores of each CPG were individually considered. Scores of individual items in each domain were

summed and standardized as a percentage of the maximum possible score for that domain, taking into account the number of appraisers. Domain scores can be calculated by summing up all the scores of the individual items in a domain and by standardizing the total as a percentage of the maximum possible score for that domain⁶.

The internal consistency of each domain was evaluated using Cronbach's alpha. The Reliability between appraisers was determined for each question and each domain of the AGREE. Intraclass correlation coefficients (ICC) were calculated within each pair of appraisers and across the pool of appraisers. ICCs and Cronbach's alpha values above 0.75 were considered to represent good reliability while values at 0.40–0.75 were considered moderate and value of <0.40 was of poor reliability.

Results

A total of 29 documents were retrieved either from HTA unit or chairman of the guideline developers group or from the MOH or AMM websites. All the 29 CPGs were published locally. The financial sponsor for all these CPGs was mainly MOH. There were no pharmaceutical drug companies influencing our researches. Those developing the CPGs consist of a mixture of professionals mainly from the universities and private professional bodies like the AMM. The development process usually took about 1 to 2 years.

All the 29 documents fulfilled the inclusion criteria. All the 29 CPGs were assessed by 4 assessors. On the item Scope and purpose, only thirteen guidelines (13/29) covered diagnosis, nine guidelines (9/29) covered management, four guidelines (4/29)

covered treatment, one guideline (1/29) covered prevention and one (1/29) guideline covered screening.

CPG production was found to increase from year 2001 to year 2003 (Figure 1). Ministry of Health was the principal CPG producer during this period of time. A CPG should be strongly recommended if it was rated high (3 or 4) on the majority of items and most domain scores were above 60% indicating the CPG had a high overall quality. A CPG should be recommended with provisos or alterations if it was rated high (3 or 4) or low (1 or 2) on a similar of items and most domain scores were between 30% and 60% indicating the CPG had a moderate overall quality. A CPG should not be recommended if it was rated low (1 or 2) on the majority of items and most domain scores were below 30% indicating the CPG had a low overall quality.

Domains corresponding to *Clarity and Presentation* (overall score was 75%) and *Scope and Purpose* (overall score was 68%) (Figure 2) were high. The majority of the CPG assessed received moderate scores in nearly all domains such as *Editorial Independence* (the overall score was 56%), *Stakeholder Involvement* (the overall score was 50%) and for *Rigor of Development* (the overall score was 48%). However for *Applicability* (the overall score was 26%) the score was low. In comparison to the results of the other domains, *Clarity and Presentation* was the best scored and *applicability* was the worst scored aspect of the 29 CPGs.

There was no statistically significant difference observed in the standardized domain scores corresponding to *Applicability* and *Editorial Independence*. Statistically significant

Figure 1: Total number of CPG produced per year

differences were observed among scores corresponding to the *Scope and Purpose*, *Clarity and Presentation*, *Stakeholder Involvement* and *Rigor of Development* (Figure 3 and table 2).

Since the distribution of the items scored were skewed, the median value was used which can be a good way to determine an approximate average. Analysis by item showed median values lower than 3 in 13 of the 23 items of the AGREE instrument: 2 items received the lowest possible score (1) (table 1).

The Malaysian CPGs did show significant improvement from 2000 to 2003 for Scope & Purpose; Rigor & Development, Stakeholder involvement and Applicability using AGREE (Fig. 4) However Clarity and presentation and Editorial independence showed lowering in values from 2002 to 2003.

Inter-rater reliability is an estimation based on the correlation of scores between/among two or more raters who

rate the same item, scale, or instrument and Intraclass correlation coefficient (ICC) was used to measure the inter-rater reliability of the four appraisers. It may also be used to assess test-retest reliability. ICC may be conceptualized as the ratio of between-groups variance to total variance. ICC measured the extent to which there was agreement consistency among the appraisers. When interpreting ICC <0.4 represents poor reliability, 0.40 -0.75 represents fair to good reliability and >0.75 represents excellent reliability. For the domain, *Rigor of Development* (ICC=0.67) and *Scope & Purpose* (ICC= 0.60) the score was moderate (Table 3). The ICC for the other domains in the Malaysian CPGs was low (ICC ≤ 0.40).

Cronbach's alpha being the most common form of internal consistency reliability coefficient, was used to measure the extent to which item responses correlate with each other. Alternatively, it can be interpreted as the correlation of the observed scale with all

possible other scales measuring the same thing and using the same number of it.

For the domain, *Rigor of Development* (Cronbach's alpha = 0.68) and *Scope & Purpose* (Cronbach's alpha = 0.63), the score was moderate (Table 3). Cronbach's alpha for the other domains

in the Malaysian CPGs were low (Cronbach's alpha ≤ 0.40).

There seems to be variability of individual scores between the appraisers on items of the AGREE instrument and some of the variability may be due to differences in interpretation of several items where the instructions were broad.

Figure 2: Results of analysis of the 29 CPGs on the six AGREE instrument domains

differences were observed among scores corresponding to the *Scope and Purpose*, *Clarity and Presentation*, *Stakeholder Involvement* and *Rigor of Development* (Figure 3 and table 1).

Since the distribution of the items scored were skewed, the median value was used which can be a good way to determine an approximate average. Analysis by item showed median values lower than 3 in 13 of the 23 items of the AGREE instrument: 2 items received the lowest possible score (1) (table 2).

The Malaysian CPGs did show significant improvement from 2000 to 2003 for *Scope & Purpose*; *Rigor &*

Development, *Stakeholder involvement* and *Applicability* using AGREE (Fig. 4) However *Clarity* and *presentation* and *Editorial independence* showed lowering in values from 2002 to 2003.

Inter-rater reliability is an estimation based on the correlation of scores between/among two or more raters who rate the same item, scale, or instrument and Intraclass correlation coefficient (ICC) was used to measure the inter-rater reliability of the four appraisers. It may also be used to assess test-retest reliability. ICC may be conceptualized as the ratio of between-groups variance to total variance. ICC measured the extent to which there was agreement consistency among the appraisers. When

interpreting ICC <0.4 represents poor reliability, 0.40 - 0.75 represents fair to good reliability and >0.75 represents excellent reliability. For the domain, *Rigor of Development* (ICC=0.67) and *Scope & Purpose* (ICC= 0.60) the score

was moderate (Table 3). The ICC for the other domains in the Malaysian CPGs was low (ICC $\leq$ 0.40).

Cronbach's alpha was also used to test for internal consistency

Figure 3: Temporal evolution of the median standardized score for each AGREE instrument domain

Table 1: Comparison of the CPG quality according to independent variables

Median [Interquartile range] of standardized domain scores						
	Scope	Participation	Rigour	Clarity	Applicability	independence
2000-03 (n=29)	58% (49%)	52% (11%)	52% (33%)	79% (20%)	22% (13%)	50 (17%)
Year 2000 (n=5)	47% (11%)	40% (20%)	27% (20%)	67% (29%)	19% (10%)	50% (19%)
Year 2001 (n=10)	51% (8%)	46% (9%)	38% (16%)	67% (24%)	22% (15%)	50% (22%)
Year 2002 (N=4)	90% (40%)	54% (10%)	62% (13%)	92% (8%)	32% (29%)	56% (35%)
Year 2003 (n=10)	99% (15%)	56% (7%)	66% (8%)	82% (8%)	25% (7%)	52% (11%)
P value	0.000	0.000	0.000	0.002	0.413	0.904

reliability coefficient to measure the extent to which item responses correlate with each other. Alternatively, it can be interpreted as the correlation of the observed scale with all possible other scales measuring the same thing and using the same number of it. When interpreting Cronbach's alpha magnitudes: <0.4 represents poor reliability, 0.40 -0.75 represents fair to

good reliability and >0.75 represents excellent reliability. For the domain, *Rigor of Development* (Cronbach's alpha = 0.68) and *Scope & Purpose* (Cronbach's alpha = 0.63), the score was moderate (Table 3). Cronbach's alpha for the other domains in the Malaysian CPGs were low (Cronbach's alpha $\leq$ 0.40).

There seems to be variability of individual scores between the appraisers on items of the AGREE instrument and

some of the variability may be due to differences in interpretation of several items where the instructions were broad.

Table 2: Scores by Item of the AGREE instrument

Domain and items of the AGREE instrument	Median Value	Interquartile Range
Domain 1: Scope and Purpose		
1. The overall objectives of the guideline is (are) specifically described	3	2
2. The clinical question(s) covered by the guideline is (are) specifically described	3	2
3. The patients to whom the guideline is meant to apply are specifically described	4	1
Domain 2: Stakeholder involvement		
4. The guideline development group includes individuals from all the relevant professional groups	4	1
5. The patients' views and preferences have been sought	1	1
6. The target users of the guideline are clearly defined	3	1
7. The guideline has been piloted among end users	1	2
Domain 3: Rigour of Development		
8. The systematic methods were used to search for evidence	2	2
9. The criteria for selecting the evidence are clearly described	2	3
10. The methods used for formulating the recommendations are clearly described	2	2
11. The health benefits, side effects and risks have been considered in formulating the recommendations	3	2
12. There is an explicit link between the recommendations and the supporting evidence	3	2
13. The guideline has been externally reviewed by experts prior to its publication	2	2
14. A procedure for updating the guideline is provided	2	2.25
Domain 4 : Clarity and Presentation		
15. The recommendations are specific and unambiguous	4	1
16. The different options for management of the condition are clearly presented	4	1
17. Key recommendations are clearly identifiable	3	1
18. The guideline is supported with tools for application	3	2

Domain 5: Applicability

19. The potential organizational barriers in applying the recommendations have been discussed	2	1
20. The potential cost implications of applying the recommendations have been considered	2	1
21. The guideline presents key review criteria for monitoring and / or audit purposes	2	1

Domain 6: Editorial Independence

22. The guideline is editorially independent from the funding body	3	2
23. Conflicts of interest of guideline development members have been recorded	2	2

Figure 4: Results of 4 Reviewers based on AGREE instrument

Discussion

The results of the study showed that through the years, development of guidelines in Malaysia had progressively increased. The quality of guidelines in Malaysia was practically unknown. To our knowledge, this was the first guideline appraisal in Malaysia. From this research, the quality of the 29 Malaysian guidelines was far from ideal: scores were moderate and low in all domains.

Variability of individual scores between the appraisers on items of the AGREE instrument, was noted as evidenced by the low Cronbach's alpha and ICC. On the other hand, the Argentinean study, by María Eugenia Esandi and Zulma Ortiz et al, ICC and Cronbach's alpha for each domain were in all cases moderate or high (0.46–0.74), except for Editorial Independence which showed very low values.

First, low quality could have been the result of the absence of an explicit policy

TABLE 3: Reliability Scores of the AGREE instrument

	Single rater ICC (95% CI)	Reliability Measures Average of raters ICC (95% CI)	Cronbach alpha	P value
Scope	0.27 (0.17 to 0.38)	0.60 (0.44 to 0.71)	0.63	0.00
Participation	0.05 (0.003 to 0.10)	0.19 (- 0.12 to 0.43)	0.34	0.00
Rigour	0.20 (0.14 to 0.27)	0.67 (0.57 to 0.75)	0.68	0.00
Clarity	0.1 (0.04 to 0.18)	0.36 (0.1 to 0.55)	0.43	0.00
Applicability	0.12 (0.04 to 0.21)	0.36 (0.13 to 0.53)	0.39	0.00
Editorial Independence	-0.80 (-0.16 to 0.01)	- 0.3 (-1.3 to 0.2)	- 0.33	0.96

for guidelines production (especially development of evidence-based CPGs) and evaluation during the period under assessment. There was also no clear guidance on the integration of multiple stakeholders. Most of the guidelines did not have enough multidisciplinary representatives in the development process. In order to balance the interests, preferences and knowledge of different stakeholders whose participation in the guideline development process is required, a more integrated approach is required.

Secondly, low quality scores of the Malaysia guidelines could be explained by a slower penetration and consolidation of the evidence-based medicine concept in comparison to developed countries. Before the year 2002, the awareness of evidence-based medicine concept amongst the healthcare practitioners was still low in Malaysia. In the United States, the Consensus Development Program at the National Institute of Health developed its first guideline in 1977. In the last 30 years, all these organizations have accumulated a vast experience in guideline development, dissemination and implementation. Currently, principles of evidence-based-medicine dominate almost all of these national guideline programs. The creation of international

networks, like the Guidelines International Network (G-I-N), as well as the establishment of projects like the AGREE, have clearly contributed to the improvement and standardization of these processes in the participating countries. Contrastingly, Malaysia, did not take part in any of these activities except until recently. Diffusion and dissemination of appropriate methods for evidence-based guidelines development is limited in Malaysia. This study found that until 2003, this process was not systematized and the development of CPGs still relied heavily on the opinion of experts. The manual development of evidence-based CPGs was drafted in 2003.

Thirdly, limited accessibility to updated biomedical literature can negatively impact on the use of relevant and important evidence to support guidelines recommendations. Most of the government facilities had very limited accessibility to current biomedical literature due to the financial constrain. Even after the broad agreement on the need for systematic reviews to inform recommendations, this type of evidence was rarely referred in Malaysian guidelines. Therefore, networking activities between guideline producers should also be promoted.

Another factor that could have influenced the quality of Malaysian guidelines is the lack of economical and human resources devoted to guideline production. Since the cost of producing evidence-based guidelines is relatively high, a systematic methodology to adapt international guidelines would be an efficient way of improving not only the quantity but also their quality¹⁸. Internationally developed guidelines can be adapted to the local context, representing a considerable saving of money. However, an explicit and systematic adaptation process should be performed as guidelines' applicability and transferability can be strongly influenced by different factors, such as population needs (prevalence of disease, baseline risk status), setting (availability of resources) and other factors that modify translation of recommendations into practice¹⁹.

Although many of the Malaysian guidelines were classified as evidence-based, a thorough review of their quality utilizing the AGREE instrument led to the authors to recommend the guidelines only with provisos or alterations. Overall, almost all the guidelines performed poorly with respect to applicability. Most of the guidelines failed to address issues of barriers to implementation, monitoring criteria, and evidence of pilot testing.

On the other hand, the study on the clinical practice guidelines in Argentina (1994–2004)⁷ by María Eugenia Esandi and Zulma Ortiz et al. scored lower in the overall standardized scores. Overall standardized score for each domain were: Scope & Purpose (overall score was 39%); *Stakeholder Involvement* (overall score was 13%); Rigor & Development (overall score was 10%); Clarity and presentation (overall score was 42%); *Applicability* (overall score

was 6%); *Editorial Independence* (overall score was 0%).

One of the key factors regarding the adequacy of the guidelines pertains to the rigor of development. Many of the guidelines did not clearly delineate the literature review methodology used or the mechanism by which recommendations were formulated. This step is crucial in determining whether the recommendations were truly based on evidence or in understanding how evidence was synthesized.

As in the evidence-based decision making, patient preferences and experiences should be factored into decisions regarding clinical care, especially in diseases such as cancer in which treatments can have significant morbidity and can impact on quality of life. All guideline committees should have patient representatives and all literature reviews specifically addresses quality of life when available.

Finally, findings of this assessment highlighted the need of improving the reporting of the editorial independence of guideline producers. Practically none of the Malaysian guideline reported conflict of interests or funding sources. Lack of transparency was also reported by Papanikolaou et al. in an evaluation of 191 published guidelines: only 7 (3.7%) disclosed potential conflicts of interest²⁰. In the case of the Malaysian guidelines, omission could have been unintentional or, on the contrary, intentional (financial ties might have existed in some situations and deliberately hidden by guideline authors). However, regardless of the intent of guideline developers' actions, explicit declaration of conflict of interests at the beginning of the process is strongly recommended by most international organizations as a way of reducing the probability of biased

recommendations and increasing guidelines' credibility²¹.

There were several limitations noted in our study. It should be noted that in the AGREE instrument, the appraiser have to choose from four categories (1, 2, 3, 4) for his /her evaluation. Secondly, because we relied on materials reported in the published versions of the guidelines, our findings could be affected not only by the quality of the guidelines themselves but also by the quality of the reporting process. It was possible that in some cases, guideline developers used appropriate techniques but did not report them. We attempted to minimize this by including in our evaluation any background on supporting articles if they were available. However we feel that just as in other medical reports, documentation of methods used is important, and if explicitly stated can help determine the validity of recommendations. Thirdly, using the AGREE instrument, the inter-rater reliability was only poor to moderate. Some of the variability may be due to differences in interpretation of several items where the instructions were broad. Another potential limitation of the AGREE instrument concerns the validity of the responses to the question on the overall assessment of the guideline. Although the reviewers were instructed to consider the domain scores when making a decision about whether or not to recommend the guideline, no clear rules were established.

In Malaysia most practitioners, before the year 2003, were not clear of processes involves in evidence-based medicine approach. Most were still doing with consensus-consulting and agreeing to expert opinions.

To our knowledge this is the first time a study of this kind has been undertaken in Malaysia. Its execution was the first step in the building a network of professionals interested in improving evidence-based CPG development, dissemination and implementation in the country. Its findings have been found to be very useful in improving the methodological quality of developing CPGs in Malaysia.

Conclusion

This study was one of the firsts that systematically employed the AGREE instrument for the critical assessment of guidelines produced in Malaysia. The AGREE instrument can serve as a model to identify improvement opportunities in the guidelines development process of Malaysia. In this sense, this research shows the low quality of guidelines produced and points out areas to which training initiatives should be oriented.

A review of the current Malaysian guidelines demonstrates that many of the clinical topics of interest have been considered by at least one guideline. None covers all the necessary elements. Furthermore, although these guidelines may accurately reflect clinical practice, few adhere to the standards set forth by the AGREE instrument.

Several approaches could be used to improve the quality of guidelines. The guideline producers could become familiar with guideline development standards that have been established and make greater efforts to incorporate them into guidelines, strive to widely adopt and use them.

Acknowledgement

This research project was sponsored by the Small Research Grant, Government of Malaysia (Project Code: MRG-2004-5)

We would like to thank the following for contributing their precious idea and advice to lead us to the successful write up of this project:

- Dato' Dr. Zaki Morad, Director of Clinical Research Centre HKL

- Dr Sivalal, Unit Head of Health Technology Assessment
- Dr Jamaiah Haniff, Principal Assisstant Diector, Clinical Research Centre
- Dr Rusilawati Jawdin, Principal Assisstant Diector, Health Technology Assessment unit
- Matron Jaya Devi, Health Technology Assessment Unit
- Dr Maizun Mohd Zain, Principal Assisstant Diector, Evidence-based medicine Unit

References

1. Academy of Medicine of Malaysia accessed via the internet at <http://www.acadmed.org.my/html/index.shtml> on 18 March 2004.
2. Guidelines for Clinical Practice Guidelines. Health Technology Assessment Unit, Medical Development Division, Ministry of Health Malaysia.
3. Banks J. National Hearth, Lung, and Blood Institute marketing research study on the formatting, dissemination, and use of clinical practice guidelines: Executive Review. 1995; p 1-7.
4. Clinton JJ, Mc Cormick K, Besteman, J. Enhancing Clinical Practice: The role of practice guidelines. Amer Psychologists 1994. 49(1): p 30-33.
5. Weingarten S, Ellrodt AG. The case for intensive dissemination: Adaption of practice guidelines in the coronary care unit. Qual Rev Bull December 1992; p 449-455.
6. AGREE collaboration accessed via the internet at <http://www.agreecollaboration.org/> on 18 March 2004
7. Marfa Eugenia Esandi, Zulma Ortiz, Evelina Chapman et al, Production nad quality of clinical practice guidelines in Argentina (1994 – 2004); a cross-sectional study, Implementation Science 2008, 3:43 doi:10.1186/1748-5908-3-43.
8. The AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf Health Care*. 2003;**12**:18–23. doi: 10.1136/qhc.12.1.18.
9. Shaneyfelt T, Mayo-Smith M, Rothwangl J. Are Guidelines following guidelines? The methodological quality of Clinical Practice Guidelines in the peer-review medical literature. *Jama*. 1999;**285**:1900–1905. doi: 10.1001/jama.281.20.1900.
10. Cluzeau F, Littlejohns P, Grimshaw J, Feder G, Moran S. Development and application of a generic methodology to assess the quality of clinical guidelines. *International Journal for Quality in Health Care*. 1999;**11**:21–28. doi: 10.1093/intqhc/11.1.21.
11. Grilli R, Magrin N, Penna A, Mura G, Liberati A. Practice guidelines developed by specialty societies: the need for a critical appraisal. *Lancet*. 2000;**355**:103–106. doi: 10.1016/S0140-6736(99)02171-6.
12. Graham I, Beardall S, Carter A, Glennie J, Hebert P, Tetroe J, McAlister F, Visentin S, Anderson G. What is the quality of drug therapy in Canada? *CMAJ*. 2001;**165**:157–163.]

13. Harpole L, Kelley MJ, Schreiber G, Toloza EM, Kolimaga J, McCrory DC. Assessment of the Scope and Quality of Clinical Practice Guidelines in Lung Cancer. *CHEST*. 2003;**123**:7S–20S. doi: 10.1378/chest.123.1_suppl.7S.
14. Navarro Puerto MA, Ruiz Romero F, Reyes Domínguez A, Gutiérrez Ibarlucea I, Hermosilla Gago T, Alonso Ortiz del Río C, et al. ¿Las Guías que nos guían son fiables? Evaluación de las Guías de Práctica Clínica Españolas. *Rev Clín Esp*. 2005;**11**:533–40.
15. MacDermid JC, Brooks D, Solway S, Switzer-McIntyre S, Brosseau L, Graham ID. Reliability and validity of the AGREE instrument used by physical therapists in assessment of clinical practice guidelines. *BMC Health Services Research*. 2005;**5**:5–18. doi: 10.1186/1472-6963-5-18.
16. Rico Iturrioz R, Gutierrez-Ibarluzea I, Asua Batarrita J, Navarro Puerto MA, Reyes Domínguez A, Marín León I, Briones Pérez de la Blanca E. Valoración de escalas y criterios para la evaluación de Guías de Práctica Clínica. *Rev Esp Salud Pública*. 2004;**78**:457–467.
17. Vlayen J, Aertgeerts B, Hannes K, Sermeus W, Ramaekers D. A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. *Int J Qual Health Care*. 2005;**17**:235–242. doi: 10.1093/intqhc/mzi027.
18. Graham ID, Harrison MB, Brouwers M, Davies BL, Dunn S. Facilitating the Use of Evidence in Practice: Evaluating and Adapting Clinical Practice Guidelines for Local Use by Health Care Organizations. *JOGNN*. 2002;**31**:599–611.
19. Schünemann HJ, Fretheim A, Oxman AD. Improving the Use of Research Evidence in Guideline Development: 13. Adaptation, applicability and transferability. *Health Res Policy Syst*. 2006;**4**:25. doi: 10.1186/1478-4505-4-25.
20. Papanikolaou GN, Baltogianni MS, Contopoulos-Ioannidis DS, Haidich AB, Giannakakis IA, Ioannidis JPA. Reporting of conflicts of interest in guidelines of preventive and therapeutic interventions. *BMC Medical Research Methodology*. 2001;**1**:3. doi: 10.1186/1471-2288-1-3.
21. Boyd E, Bero L. Improving the Use of Research Evidence in Guideline Development: 4. Managing conflicts of interest. *Health Res Policy Syst*. 2006;**4**:16. doi: 10.1186/1478-4505-4-16.