

Application of the AGREE II instrument in the evaluation of a selection of South African Clinical Guidelines

K Jamaloodien

SUMMARY

Application of the AGREE II instrument in the evaluation of a selection of South African Clinical Guidelines

Guideline development processes influence the quality of clinical guidelines. The aim of this study was to use the AGREE II instrument to evaluate the variability of the quality of selected guidelines, to determine a baseline for the quality of current guidelines and determine whether guidelines demonstrated good standard practice during their development.

The AGREE II instrument was used to assess a selection of guidelines published between January 2012 and June 2013. Eleven guidelines were selected for review. Overall, guidelines scored highest in domain 1 (Scope and purpose) and 4 (Clarity of presentation); and lowest in domain 3 (Rigour of development) and 6 (Editorial independence) with the overall assessment score of three out of seven. The study demonstrated that the quality of guidelines was variable and that there are deficiencies in the guideline development process. The results from this study provide a baseline to measure the quality of future guidelines.

Key words

Clinical guidelines

Guideline development

Quality of guidelines

AGREE II instrument

University of Pretoria
Faculty of Health Sciences
School of Medicine



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Health Sciences

**Application of the AGREE II instrument in the evaluation of a selection of
South African Clinical Guidelines by**

K Jamaloodien (10430106)

Submitted in partial fulfilment of the requirement for the degree MSc Clinical
Epidemiology in the School of Health Systems and Public Health, Faculty of Health
Sciences, University of Pretoria, 2014 on 31 October 2014

Name of Supervisor: Prof P Rheeder

ABSTRACT

Background

Guideline development processes influence the quality of clinical guidelines. A review of the literature showed that although studies were done internationally to assess the quality of clinical guidelines, only one study assessed the development of guidelines in South Africa.

Objectives

The aim of this study was to use the AGREE II instrument to evaluate the variability of the quality of selected guidelines, to determine a baseline for the quality of current guidelines and determine whether guidelines reviewed demonstrated good standard practice during their development.

Methods

A selection of guidelines published in the South African Medical Journal and by the Department of Health between January 2012 and June 2013 was independently assessed by three reviewers. The AGREE II instrument, an internationally validated instrument, was used to allocate a score for the six domains and for the overall assessment of the guidelines.

Results

Eleven guidelines were selected for review; nine published in the Journal and two by the Department of Health. Overall, guidelines scored highest in domain 1 (Scope and purpose) and 4 (Clarity of presentation); and lowest in domain 3 (Rigour of development) and 6 (Editorial independence) with the overall assessment score of three out of seven.

The quality of guidelines was variable and deficiencies in the guideline development process were also demonstrated.

Conclusion

The application of the AGREE II instrument confirmed the variability of the quality of guidelines and results from this analysis provide a baseline to measure the quality of future guidelines. Furthermore, the clinical guidelines reviewed did not demonstrate good standard practice during their development.

Key words

Clinical guidelines

Guideline development

Quality of guidelines

AGREE II instrument

Contents

List of tables	6
List of figures	7
CHAPTER 1:.....	9
Introduction and Literature Review	9
1.1 Introduction	9
1.2 Clinical practice guidelines: development of assessment tools	11
1.3 History of development.....	11
1.4 Quality of clinical guidelines.....	14
1.5 Conclusion	16
CHAPTER 2:.....	17
Design and methodology	17
2.1 Summary of study methods.....	17
2.2 Aim of the study.....	17
2.3 Study design	17
2.4 Methodology.....	18
2.5 Justification for the methodology	20
2.6 Research procedures	20
2.7 Data management.....	20
2.8 AGREE II assessment criteria for each domain	21
2.9 Intradomain variation and concordance.....	24
2.10 Ethical considerations.....	24
2.11 Conclusion	25
CHAPTER 3:.....	26
Results	26
3.1 Introduction	26
3.2 Selection of guidelines for assessment.....	26
3.3 Patterns of data	27
3.4 Interguideline variation	27
3.5 Variation of domain scores between guidelines (interdomain variation) and concordance.....	28
3.6 Domain 1: Scope and purpose	32
3.7 Domain 2: Stakeholder involvement	33
3.8 Domain 3: Rigour of Development.....	34
3.9 Domain 4: Clarity of Presentation	35

3.10	Domain 5: Applicability	36
3.11	Domain 6: Editorial independence	37
3.12	Overall guideline assessment.....	38
3.13	Conclusion	39
CHAPTER 4:.....		41
<i>Discussion</i>		41
4.1	Introduction	41
4.2	Interguideline variation	41
4.3	Intradomain variation and concordance.....	42
4.4	Domain 1: Scope and purpose (moderate).....	43
4.5	Domain 2: Stakeholder involvement (low).....	43
4.6	Domain 3: Rigour of development (low).....	44
4.7	Domain 4: Clarity of Presentation (high).....	45
4.8	Domain 5: Applicability (low).....	45
4.9	Domain 6: Editorial independence (low)	45
CHAPTER 5.....		47
<i>Conclusion and recommendations</i>		47
5.1	Introduction	47
5.2	Conclusion about the research problem	47
5.3	Limitations	47
5.4	Recommendations	48
5.5	Further research.....	50
CHAPTER 6.....		51
References.....		51
ADDENDUMS.....		55
Addendum 1: Data collection tool		55
Addendum 2: Template for calculation of domain scores per guideline and level of discrepancy		59
Addendum 3: Permission letter from Ethics Committee		62

List of tables

- Table 2.1: AGREE instrument: 6 domains and 23 items
- Table 3.1: Comparison scores for each guideline across the domains
- Table 3.2: Lowest and highest guideline scores according to domain
- Table 3.3: Discrepancy levels across domains for all guidelines reviewed
- Table 3.4: Correlation between recommendation for use and concordance
- Table 4.1: Difference between mean and standard deviation per domain
- Table 5.1: Suggested recommendation to improve guideline development

List of figures

- Figure 3.1: Scope and purpose
- Figure 3.2: Stakeholder Involvement
- Figure 3.3: Rigour of development
- Figure 3.4: Clarity of Presentation
- Figure 3.5: Applicability
- Figure 3.6: Editorial Independence
- Figure 3.7: Overall assessment

Reference number and date of approval from Ethics Committee

333/2013 (29/08/2013)

Name of Journal where the article has been submitted/is proposed to be submitted.

South African Medical Journal

Acknowledgements

Prof M Blockman

Prof AG Parrish

Dr J Munsammy

Dr R de Waal

Statement of original authorship

"I declare that the dissertation, which I hereby submit for the degree MSc Clinical Epidemiology at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at another university.

CHAPTER 1: Introduction and Literature Review

1.1 Introduction

Historically, medical practitioners were deemed to be competent to practice medicine by virtue of their training, the imposition of minimal standards of competence (licensing) and experience. These competencies were supplemented by expert opinion, expressed either formally or informally. [1] However, the status of expert opinion is tenuous because of possible bias and the fact that it may not be premised on a careful assessment of research evidence. [1,2]

Since the 1970s, the explosion in medical research, and therefore literature, complicated medical practice thus creating uncertainty regarding appropriate medical interventions resulting in variations in practice. These practice variations have an impact on patient care in that some patients will receive optimal treatment and others no treatment. Yet, there are those patients who may receive treatment that may cause harm. [3]

In addition to the exponential increase in medical knowledge and the resultant variation in clinical practice, rising healthcare cost, increased burden of disease, the availability of more expensive treatment options, irrational use of medicines and inequity in access and quality of care places considerable pressure on healthcare practitioners to process and apply the available evidence into daily practice. [4]

Clinical practice guidelines (CPGs) are one of the tools that are used to meet these challenges. [4,5] The purpose of CPGs is to intentionally influence clinicians actions by making explicit recommendations. They assist healthcare practitioners to provide appropriate, cost-effective and quality care to their patients based on the latest available knowledge and evidence, thereby reducing the use of harmful or unnecessary interventions.

The development of CPGs is not new. Guidelines were initially used by public health funders and other government agencies to curb rising health costs. [5] Professional

societies also develop guidelines in an attempt to maintain professional autonomy. [6,7]

However, target users of guidelines cannot “discriminate” on the quality of the guideline. This places additional responsibility on guideline developers to ensure good quality guidelines. Therefore, the process of developing treatment guidelines should be rigorous and robust, to ensure that guidelines are valid and reliable. Internationally, there has been increased emphasis on the quality of guidelines and the evidence base used to develop recommendations. [8].

In South Africa, numerous treatment guidelines are developed by different stakeholders to provide quality care and to set the standard for practice. For example, the National Department of Health produces guidelines to guide practitioners in the public sector and various professional societies develop guidelines for the private sector.

Although the quality of guidelines developed within SADC have been evaluated for the prioritised five diseases (HIV in adults, malaria in children and adults, pre-eclampsia, diarrhoea in children and hypertension in primary care), the quality of the guidelines developed for South Africa, whether by the National Department of Health or a professional society, has not been assessed and documented.[9]

1.2 Clinical practice guidelines: development of assessment tools

Clinical practice guidelines are tools used by public health policy makers and medical funders to make access to quality care more efficient and equitable, to set the standards of care and to guide clinical practice. The US Institute of Medicine (IOM) defines clinical practice guidelines as: “*statements that include recommendations intended to optimize patient care that are informed by a systematic review of evidence and an assessment of the benefits and harms of alternate care options*”. [4]

The practice of developing guidelines as a guide for caregivers is not new. Clinical guidelines were originally based on expert consensus opinion and, in many instances, recommendations were made without an evidence base. This has serious limitations in terms of potential bias, the efficient use of available resources and the quality of care provided. [8,10]

However, the practice environment has changed due to the proliferation of scientific research, advances in medicine and the acknowledgment of the need for more rigorous reporting of study results and its implications for practice. In recent times, the emphasis has been on the systematic development of guidelines and ensuring that the guideline development process is rigorous and able to withstand scrutiny. [4,6,8,10]

1.3 History of development

Guideline developers have recognised the need to provide better quality guidelines. In 1990, a study committee appointed by the Institute of Medicine released a report in which the need for standardisation and consistency in guideline development was acknowledged. The study further highlighted the limitations associated with a systematic development of guidelines and its evaluation. In addition, no explicit method was available at the time to assess existing practice guidelines. [4]

In a second report released by the Institute of Medicine (IOM) in 1992, an “assessment instrument” to formally evaluate guidelines was developed by a second study committee. [11] The intention of this instrument was to operationalise the

attributes of a good guideline and to standardise the approach and structure for the assessment of a guideline document. The committee was of the opinion that the instrument had three potential uses:

1. As an educational tool;
2. As a self-assessment tool; and
3. As means for judging guidelines before their adoption.

Although the assessment instrument captured key attributes of good quality guidelines (sensitivity, specificity, patient responsiveness, readability, minimum obtrusiveness, feasibility, computer compatibility, appeals criteria), the committee acknowledged that more practical experience was needed before the instrument could reliably be implemented. Furthermore, the instrument was considered too complicated to implement as a generic assessment tool.

In 1993, the UK National Health Service (NHS) initiated a research programme to produce a generic appraisal tool to assess the quality of guidelines. [12] The “assessment instrument” developed by the IOM was used as a basis for developing a simplified appraisal tool. Together with a user manual to ensure the consistent interpretation of questions, the appraisal tool developed by the NHS was validated by assessing 60 UK guidelines. This tool became known as the “Cluzeau instrument”.

The predictors of guideline quality in the Cluzeau instrument were:

1. Rigour of development (attributes necessary to enhance guideline validity and reproducibility);
2. Clarity of presentation (including context and content); and
3. Implementation issues.

The appraisal instrument used a numerical scoring system (based on a Likert scale) and allowed for comparison between guidelines. The authors concluded that use of this instrument by guideline developers would result in the more accurate reflection of research evidence in guidelines.

A comparison of clinical practice guideline appraisal instruments was undertaken in 2000. [13] Of the 15 instruments identified, the Cluzeau instrument was one of two

instruments that had been validated. Subsequently, the Cluzeau instrument has been used to assess a number of clinical practice guidelines. [14,15,16,17,18]

In recognition of the need for closer collaboration, in 1998 an international task group, the AGREE (Appraisal of Guideline REsearch and Evaluation) Collaboration started to work on harmonisation and co-ordination of guideline development. [19,20].

The Collaboration published the first AGREE instrument in 2003, the purpose of which was to develop an objective tool to assess the quality of guidelines. [19] The AGREE Instrument is a 23-item tool comprising of six domains. The process by which the Instrument was developed allows for its application on international guidelines. The Instrument could be used for planning, execution and monitoring of guideline programmes as well as for comparing guidelines. The Instrument provided a common standard to improve the guideline development process and for reporting on its development.

As quality assessment is an on-going process, tools used in its measurement also require ongoing improvement. Some of the original members of the AGREE Collaboration formed the AGREE Next Steps Consortium [21]. The objectives of the Consortium were to:

- further improve the measurement properties of the instrument, including its reliability and validity;
- refine the instrument's items to better meet the needs of the intended users; and
- improve the supporting documentation (i.e., original training manual and user's guide) to facilitate the ability of users to implement the instrument with confidence.

As a result, a modified AGREE instrument was published (AGREE II) in 2009, and updated in 2013. [20,21] This Instrument is a generic tool that can be used to assess the process of development of any guideline (whether local, regional or international), original guidelines and updates thereof as well as guidelines covering any disease state. The Instrument can be used by health care providers (who would like to make an individual assessment of a guideline before incorporating into their

own practice), guideline developers (to improve their guideline development processes), policy makers and educators.

Despite its wide acceptance as an objective and generic tool to assess quality of guideline development, the AGREE II instrument has limitations. It does not assess the quality of the recommendations incorporated into the guideline based on research evidence. Nor does it assess the conclusions drawn from a critical appraisal of the evidence. [21] Although this poses a threat to the credibility of the guidelines, none of the appraisal instruments available provided for scoring the evidence base of the clinical content. [21]

1.4 Quality of clinical guidelines

The quality of the guidelines determines the potential benefit it may have for patient care. When used correctly, valid guidelines can change practice and improve patient outcomes. [22]

The variation in the quality of guidelines is a universal phenomenon. Various attempts have been made to identify common criteria which would describe quality aspects of guidelines, [12,23,24,25].

The AGREE Collaboration defines quality of guidelines as: *“the confidence that the potential biases of guideline development have been addressed adequately and that the recommendations are both internally and externally valid, and are feasible for practice”*. [20]

According to the IOM, good standard practice entails the use of approaches that are objective, scientifically valid and consistent to produce guidelines that are unbiased, scientifically valid and trustworthy. [4]

In general, clinical practice guidelines are developed in four ways, i.e.: informal consensus, formal consensus, evidence based approach and explicit approach. (26,27)

Using the informal consensus, experts make recommendations based on a subjective assessment of evidence. Inherent in this process is the risk of bias. To mitigate against a risk, more formal consensus methods have been developed. In these methods, panels of experts includes individuals with varying skill sets, including clinicians, methodologists, etc.

In order to improve the reliability and validity of guidelines, efforts were made to formulate evidence based guidelines. In this setting, panels would issue recommendations that reflected the weight of accumulated evidence. The limitation of these guidelines was that the focus of the recommended guideline was on a single outcome. The explicit approach uses the evidence based approach by systematically estimating the effects of interventions on all important health outcomes.

Although explicit, evidence based guidelines are the ideal, they are expensive to develop and are time consuming. In a review of the methodology for guideline development published in peer review journals, it was shown that guideline development did not adhere to methodological standards. [28]

The greatest challenge for guideline developers is the incorporation of evidence into practice. Strategies that improve efficiencies have been the subject of a manual published by Rosenfeld et al. [29]

To improve efficiencies in making evidence based guidelines more accessible and to ensure dissemination, implementation and use of these guidelines, the United States government established the Guideline Clearing House in 1999. [30] This National Guideline Clearinghouse (NGC), is a web based database of evidence-based clinical practice guidelines, available to all stakeholders providing and funding healthcare. Only guideline that meet specific criteria are cleared through the NGC.

Internationally, surveys have been undertaken describe the methodological quality of guidelines developed in individual countries. [15,16,17] Although an evaluation of the quality of guidelines produced in the SADC has been undertaken, no evaluation of the quality of clinical guidelines produced specifically by the National Department of Health or Clinical Societies in South Africa has been undertaken.

In recognition of the need to assess the quality of guidelines, in May 2014, the South African Medical Journal announced the appointment of a new editorial sub-committee to adjudicate guidelines and recommendations to be published in the Journal. [32] This sub-committee will use the AGREE II instrument to assess the rigour and validity of guidelines published henceforth.

1.5 Conclusion

In South Africa, numerous clinical guidelines are produced to guide clinical practice. By assessing the quality of guidelines available, recommendations can be made for a structured and rigorous development methodology to improve the quality of future guidelines. Furthermore, this study will form a baseline for the appraisal of available guidelines in the South African clinical environment.

CHAPTER 2:

Design and methodology

2.1 Summary of study methods

The quality of clinical practice guidelines varies. The AGREE II instrument was developed to address this issue of guideline quality variation. It is a generic, internationally validated tool that can be used to evaluate the quality of guidelines developed for any disease area.

This is an observational study using a systematic method of critically appraising the quality of selected guidelines. The AGREE II instrument was used as the assessment tool to critically appraise the quality of selected guidelines published in South Africa between January 2012 and June 2013. Eleven guidelines were appraised, nine published in the South African Medical Journal and two by the National Department of Health.

2.2 Aim of the study

To use the AGREE II instrument to evaluate the variability of the quality of selected guidelines, to determine a baseline for the quality of current guidelines and determine whether guidelines reviewed demonstrated good standard practice during their development.

The results of this study will represent a baseline for the appraisal of available guidelines in the South African clinical environment in the future. It is envisaged that such a baseline will stimulate the clinical guideline and health technology assessment fraternities to pay closer attention to the methodology used to develop clinical practice guidelines and how these are presented.

2.3 Study design

This is an observational study using an internationally validated quality assessment tool, to systematically perform critical appraisals of a selection of clinical practice guidelines.

2.4 Methodology

A selection clinical practice guidelines published in South Africa between January 2012 and June 2013 were critically appraised using an internationally validated quality assessment tool called the AGREE II instrument. Guidelines published in the South African Medical Journal were selected for evaluation because the Journal has the highest impact factor than any other peer reviewed journal in South Africa. In addition, guidelines published by the National department of Health within this time will also be selected for appraisal.

Data was collected using AGREE II instrument. The Instrument is a tool used to assess the methodological rigour and transparency with which a guideline is developed. It has been validated for this purpose by international guideline developers and researchers known as the AGREE Collaboration.

In the AGREEII instrument, it is recommended that between two and four appraisers rate each guideline to ensure that results are reliable. Therefore, each guideline was independently appraised by the three reviewers who have previous experience with using the AGREE II instrument to appraise guidelines. The reviewers had skills and knowledge in various fields of medicine, including:

- Evidence based medicine
- Health economics/pharmaco-economics
- Bioethics
- Public health
- Clinical pharmacology
- Clinical pharmacy
- Rational prescribing
- Pharmacovigilance

They were: Prof M Blockman, Prof AG Parrish, Dr J Munsammy, Dr R de Waal and Ms K Jamaloodien.

Scores were allocated on a 7-point rating scale (1–strongly disagree to 7–strongly agree) for each of the items in the six domains. The six domains were each scored independently and were calculated by:

- summing up all the scores of the individual items in a domain, then
- standardising the total as a percentage of the maximum possible score for that domain. The standardised domain score is given by:

$$\text{standardised domain score} = \frac{(\text{obtained score} - \text{minimum possible score})}{(\text{maximum possible score} - \text{minimum possible score})}$$

The minimum score was 0% (all appraisers score items as 1–strongly disagree). The maximum score was 100%, where all appraisers strongly agree. However, a score was allocated for the overall assessment of the quality of CPGs. Criteria and considerations included in the instrument provide a guide to allocate scores. There is no single aggregate quality score, nor minimum domain scores to differentiate the quality of guidelines.

Agreement between reviewers (concordance) was measured in terms of the levels of discrepancy:

- low discrepancy: number of domains with a standard deviation from the mean < 1.5, or
- medium discrepancy: number of domains with a standard deviation from the mean ≥ 1.5 but < 2, or
- high discrepancy: number of domains with a standard deviation from the mean ≥ 2 .

Discordant ratings among reviewers was determined as follows:

- individual item scores of reviewers were ≥ 1.5 standard deviations from the mean in 3 of 5 domains, i.e. guidelines where there was a medium level of discrepancy in 3 of 5 domains, or
- individual scores of reviewers ≥ 2 standard deviations from the mean in 1 of the 5 domains. i.e. guidelines where there was a high level of discrepancy in 1 of 5 domains.

According to the AGREE Trust, guidelines should be compared across domains. By determining the difference between the mean and standard deviation per domain across guidelines, one can determine the variation in the quality of guidelines.

The standard deviation per domain was calculated as follows:

- per item, the standard deviation for the scores allocated according to the number of reviewers was calculated.
- the average of the standard deviations per items (calculated above) was the domain standard deviation

Microsoft Excel was used to perform the calculations.

Variability was rated as follows:

	Difference between the mean and standard deviation
Low variability	0 to 10%
Medium variability	11 to 50%
High variability	> 50%

2.5 Justification for the methodology

The AGREE II instrument is an internationally validated instrument aimed at assessing quality of practice guidelines across the health spectrum; provide guidance on guideline development; and to guide the kind of information that should be reported in guidelines. It consists of 23 items evaluating quality aspects of guidelines in six domains. [13]

2.6 Research procedures

A search was conducted of the South African Medical Journal for guidelines published between 01 January 2012 and 31 June 2013. For the same period, the Department of Health's website was searched for published guidelines.

2.7 Data management

Reviewers were provided with a template for scoring guidelines. The scores were then collated in an Excel spreadsheet. Analysis was done in Excel.

2.8 AGREE II assessment criteria for each domain

Domain 1: Scope and Purpose

In this domain, a detailed description of the overall aim of the guideline, the health questions covered by the guideline as well as the population to which the guideline pertains should be provided.

Domain 2: Stakeholder Involvement

The composition, discipline and expertise of members of the guideline development group should be described. The description should also include individuals involved in reviewing the evidence and formulating the recommendations but excludes external reviewers. Incorporation of the views of the target population should also be described, e.g. formal consultations or interviews with patients. In addition, target users of the guideline should be clearly denoted.

Domain 3: Rigour of Development

In this domain, the comprehensive search strategy used to gather unbiased evidence, the search terms and sources of evidence used, as well as the timeframes within which the evidence was sourced should be described. In addition, explicit inclusion and exclusion criteria for identified evidence should be stated. The results of the critical appraisal of the evidence should also be provided. There should be a clear relationship between the evidence used and the recommendations made.

How recommendations are formulated should also be described, including how areas of disagreement are resolved. The health consequences and outcomes of the recommendations should also be provided.

External review of the guideline prior to publication is necessary and should be undertaken by reviewers not involved in the guideline development process. In addition, how guidelines are to be updated should also be included.

Domain 4: Clarity of Presentation

The recommendations should answer the health questions of the guideline, be easy to find and should be specific to the situation and the population to which applies. A

summary of recommendations in the form of summaries, flowcharts, etc. should also be provided.

Domain 5: Applicability

Strategies for dissemination and implementation of the guideline should also be included. Strategies may include summary documents, educational tools, patient leaflets, etc. In addition, the resource implication for the implementation of the guideline should also be described.

Domain 6: Editorial independence

This domain deals with undue bias due to competing interest of funding bodies. The guideline should include an explicit statement that the final recommendations were not influenced by the interest or views of the funding body. Where members of the guideline development declare a conflict of interest, the way that this is managed should also be described.

Overall Assessment

The overall quality of the guideline is rated after assessment is done using the 23-items evaluating the quality aspects. The assessor should also provide a recommendation whether the guideline should be recommended for use.

Table 2.1: AGREE instrument: 6 domains and 23 items

Domains	Criteria
Domain 1: Scope and Purpose	Overall objective(s) of guideline is (are) specifically described
	Health questions(s) covered by guideline is (are) specifically described
	Population (patients, public etc.) to who guideline is meant to apply to is specifically described

Domains	Criteria
Domain 2: Stakeholder Involvement	Guideline development group included individuals from all relevant professional groups
	Views and preferences of the target population (patients, public, etc.) have been sought
	Target users of the guideline are clearly defined
Domain 3: Rigour of Development	Systematic methods were used to search for evidence
	Criteria for selecting evidence are clearly described
	Strengths and limitations of the body of evidence are clearly described
	Methods for formulating the recommendations are clearly described
	Health benefits, side effects, and risks have been considered in formulating the recommendations
	There is an explicit link between the recommendations and the supporting evidence
	The guideline has been externally reviewed by experts prior to its publication
	A procedure for updating the guideline is provided
Domain 4: Clarity of Presentation	Recommendations are specific and unambiguous
	Different options for management if the condition or health issue are clearly presented
	Key recommendations are easily identifiable

Domains	Criteria
Domain 5: Applicability	Guideline describes facilitators and barriers to its application
	Guideline provides advice and/or tools on how recommendations can be put into practice
	The potential resource implications of applying the recommendations have been considered
	Guideline presents monitoring and/or auditing criteria
Domain 6: Editorial independence	The views of the funding body have not influenced the content of the guideline
	Competing interests of guideline development group members have been recorded and addressed.
Overall guideline assessment	Overall quality
	I would recommend this guideline for use

2.9 Intradomain variation and concordance

The AGREE Trust recommends that guidelines should be compared across domains. Variation in the quality of guidelines were determined by measuring the difference between the mean and standard deviation per domain across guidelines. Variability was rated as follows:

	Difference between the mean and standard deviation
Low variability	0 to 10%
Medium variability	11 to 50%
High variability	> 50%

2.10 Ethical considerations

None. However, the corresponding author of each guideline was contacted to inform them of the study.

2.11 Conclusion

The results of this study will represent a baseline for the appraisal of the quality of available guidelines in the South African clinical environment in the future. It is envisaged that such a baseline will stimulate the clinical guideline and health technology assessment fraternities to pay closer attention to the methodology used to develop clinical practice guidelines and how these are presented.

CHAPTER 3: Results

3.1 Introduction

The guidelines reviewed were from diverse disciplines. Scores for each guideline are summarised in table 3.1 below. Scores for the six domains were rated as a percentage and the score for overall assessment was rated out of seven. The scores from the review is a good indicator of reviewers' assessment of overall guideline quality.

3.2 Selection of guidelines for assessment

Eleven guidelines published between January 2012 and June 2013 was selected for review; nine guidelines were published in the South African Medical Journal and two were published by the Department of Health.

The guidelines selected for review:

- Guideline for the management of acute asthma in children: 2013 update. Part 3: March 2013 (AACh) [33]
- Guideline for the management of acute asthma in adults: 2013 update (AAd) [34]
- Guideline for the prevention, screening and treatment of retinopathy of prematurity (ROP) [35]
- South African guideline for the management of Chronic Hepatitis B: 2013 (HepB) [36]
- Chronic rhinitis in South Africa: Update 2013 (ChRh) [37]
- South African guidelines for the management of Gaucher disease, 2011 (Gaucher) [38]
- Venous thromboembolism: Prophylactic and therapeutic practice guideline (VTE) [39]
- Paediatric Anticoagulation Guidelines (PAG) [40]
- South African Dyslipidaemia Guideline Consensus Statement (Dyslip) [41]
- National Contraception Clinical Guidelines (NatCont) [42]
- The South African Antiretroviral Treatment Guidelines (NatHIV) [43]

3.3 Patterns of data

Table 3.1: Comparison scores for each guideline across the domains

Guideline	Domain 1: Scope and Purpose	Domain 2: Stakeholder Involvement	Domain 3: Rigour of Development	Domain 4: Clarity of Presentation	Domain 5: Applicability	Domain 6: Editorial Independence	Mean score/ guideline
AACH	63%	37%	37%	89%	18%	39%	47%
AAAd	57%	22%	4%	59%	18%	17%	30%
ROP	91%	50%	23%	74%	57%	6%	24%
HepB	37%	11%	7%	59%	13%	0%	19%
ChRh	33%	2%6	3%	39%	0%	14%	34%
Gaucher	61%	37%	8%	52%	13%	8%	30%
VTE	6%	17%	6%	67%	11%	0%	21%
AntiCoag	46%	9%	8%	37%	7%	0%	54%
Dyslip	63%	20%	7%	83%	32%	0%	35%
NatCont	74%	46%	35%	98%	57%	14%	50%
NatHIV	98%	6%	5%	93%	8%	0%	18%
Mean score/ domain	57%	29%	13%	68%	21%	9%	47%
SD	26	15	13	21	19	12	

AACH- Guideline for the management of acute asthma in children; **AAAd** - Guideline for the management of acute asthma in adults: 2013 update; **ROP** - Guideline for the prevention, screening and treatment of retinopathy of prematurity; **HepB** - South African guideline for the management of Chronic Hepatitis B: 2013; **ChRh** - Chronic rhinitis in South Africa: Update 2013; **Gaucher** - South African guidelines for the management of Gaucher disease; **VTE** - Venous thromboembolism: Prophylactic and therapeutic practice guideline; **PAG** - Paediatric Anticoagulation Guidelines; **Dyslip** - South African Dyslipidaemia Guideline Consensus Statement; **NatCont** - National Contraception Clinical Guidelines and **NatHIV** - The South African Antiretroviral Treatment Guidelines.

The results in table 3.1, which is a summary of scores for each domain within guidelines (interguideline variation) and across the domains (interdomain variation).

3.4 Interguideline variation

There is no single aggregated quality score for guidelines reviewed with the AGREE II instrument.

Nevertheless, by determining the mean of scores per guideline, considerable variations were revealed. The guideline with the highest mean score was the National Contraceptive Guideline (50%), with the National HIV Guideline scoring the lowest (18%).

3.5 Variation of domain scores between guidelines (interdomain variation) and concordance

There was great variation of scoring across domains within a guideline.

Table 3.2: Lowest and highest guideline scores according to domain

Domain 1: Scope and Purpose	Lowest score: Venous Thromboembolism
	<i>Highest score: National HIV Guideline</i>
Domain 2: Stakeholder Involvement	Lowest score: National HIV Guideline
	<i>Highest score: Retinopathy of Prematurity</i>
Domain 3: Rigour of Development	Lowest score: Chronic Rhinitis Guideline
	<i>Highest score: Paediatric Anticoagulation Guidelines</i>
Domain 4: Clarity of Presentation	Lowest score: <i>Paediatric Anticoagulation Guidelines</i>
	<i>Highest score: National Contraception Guideline</i>
Domain 5: Applicability	Lowest score: Chronic Rhinitis Guideline
	<i>Highest score: National Contraception Guideline</i>
Domain 6: Editorial Independence	Lowest score: Chronic Hepatitis B Guideline Venous Thromboembolism Paediatric Anticoagulation Guidelines National HIV Guideline South African Dyslipidaemia Guideline
	<i>Highest score: Acute Asthma in Paediatrics</i>

Table 3.2 demonstrates that some guidelines scoring the highest in one domain, did not score particularly well in other domains. For example, although the National HIV Guidelines scored the highest in terms of domain 1 (Scope and purpose), this guideline scored the lowest in two of the six domains, i.e. domains 2 (Stakeholder involvement) and 6 (Editorial independence).

Notably, five guidelines had the same lowest score for domain 6 (Editorial independence). These were:

- Chronic Hepatitis B Guideline
- Venous Thromboembolism

- Paediatric Anticoagulation Guidelines
- National HIV Guideline
- South African Dyslipidaemia Guideline

No single guideline scored consistently high in every domain. However, one guideline scored the highest in two of the six domains, i.e. the National Contraception Guideline.

The mean score for five of the six domains was less than 60%. The domain with the highest mean score was domain 4 (Clarity of presentation), while domain 6 (Editorial independence) scored the lowest.

Concordance, or agreement between reviewers, was measured in terms of the levels of discrepancy:

- low discrepancy: number of domains with a standard deviation from the mean < 1.5 , or
- medium discrepancy: number of domains with a standard deviation from the mean ≥ 1.5 but < 2 , or
- high discrepancy: number of domains with a standard deviation from the mean ≥ 2 .

Discordant ratings among reviewers were determined as follows:

- individual item scores of reviewers are ≥ 1.5 standard deviations from the mean in 3 of 5 domains, i.e. guidelines where there was a medium level of discrepancy in 3 of 5 domains, or
- individual scores of reviewers ≥ 2 standard deviations from the mean in 1 of the 5 domains. i.e. guidelines where there was a high level of discrepancy in 1 of 5 domains.

Table 3.3: Discrepancy levels across domains for all guidelines reviewed

	Domain 1: Scope and Purpose	Domain 2: Stakeholder Involvement	Domain 3: Rigour of Development	Domain 4: Clarity of Presentation	Domain 5: Applicability	Domain 6: Editorial Independence	Overall Assessment
AACH	1.60	1.07	1.31	1.15	0.77	1.08	0.58
Discrepancy Level	MEDIUM	LOW	LOW	LOW	LOW	LOW	LOW
AAd	1.77	2.31	0.29	1.53	0.96	1.32	1.15
Discrepancy Level	MEDIUM	HIGH	LOW	MEDIUM	LOW	LOW	LOW
ROP	0.38	1.54	0.91	0.96	1.97	0.58	0.58
Discrepancy Level	LOW	MEDIUM	LOW	LOW	MEDIUM	MEDIUM	LOW
HepB	1.85	0.70	0.43	2.60	0.92	0.00	0.58
Discrepancy Level	MEDIUM	LOW	LOW	HIGH	LOW	LOW	LOW
ChRh	1.79	1.75	0.22	2.16	0.00	0.58	1.15
Discrepancy Level	MEDIUM	MEDIUM	LOW	HIGH	LOW	LOW	LOW
Gaucher	2.65	1.42	0.58	1.61	1.01	0.50	1.73
Discrepancy Level	HIGH	LOW	LOW	MEDIUM	LOW	LOW	MEDIUM
VTE	0.58	0.89	0.39	2.65	0.79	0.00	1.15
Discrepancy Level	LOW	LOW	LOW	HIGH	LOW	LOW	LOW
AntiCoag	1.79	0.77	0.51	2.23	0.72	0.00	1.53
Discrepancy Level	MEDIUM	LOW	LOW	HIGH	LOW	LOW	MEDIUM
Dyslip	2.60	1.09	0.62	1.08	1.26	0.00	0.58
Discrepancy Level	HIGH	LOW	LOW	LOW	LOW	LOW	LOW
NatCont	1.44	1.41	0.55	0.19	0.79	0.58	1.00
Discrepancy Level	LOW	LOW	LOW	LOW	LOW	LOW	LOW
NatHIV	0.19	0.33	0.07	0.58	0.87	0.00	0.00
	LOW	LOW	LOW	LOW	LOW	LOW	LOW

No concordance, based on the levels of discrepancy, was found in seven of the 11 guidelines:

- Guideline for the management of acute asthma in adults: 2013 update
- Guideline for the prevention, screening and treatment of retinopathy of prematurity
- South African guideline for the management of Chronic Hepatitis B: 2013
- Chronic rhinitis in South Africa: Update 2013
- Venous thromboembolism: Prophylactic and therapeutic practice guideline
- Paediatric Anticoagulation Guidelines
- South African Dyslipidaemia Guideline Consensus Statement

The four guidelines where concordance was achieved are:

- Guideline for the management of acute asthma in children: 2013 update. Part 3: March 2013
- South African guidelines for the management of Gauchers disease, 2011
- National Contraception Clinical Guidelines
- The South African Antiretroviral Treatment Guidelines

Table 3.4: Correlation between recommendation for use and concordance

	Recommendation	Concordance
ChRh	No	No
AntiCoag	No	No
HepB	No	No
Gaucher	No	Yes
VTE	No	No
NatHIV	Yes	Yes
AACH	Yes	Yes
NatCont	Yes	Yes
AAd	Yes with modifications	No
Dyslip	Yes with modifications	No
ROP	Yes with modifications	No

Recommendation = Yes; Yes with modifications or No.

There was some correlation between concordance among reviewers and the recommendation to use the guideline in three of the four guidelines that were recommended for use. A crude estimation (in Excel) of the correlation between the

recommendation for use and concordance yielded a value of 0.77. However, the sample size is too small to make a realistic deductions of the relationship between recommendation for use and concordance.

3.6 Domain 1: Scope and purpose

The specific criteria covered in this domain are descriptions of:

- the overall objective of the guideline;
- the health questions covered by guideline; and
- population (patients, public etc.) to whom the guideline is meant to apply

As demonstrated in figure 3.1 below, guidelines generally scored well in this domain with the majority of guidelines (seven of the eleven) scoring above the mean score of 57%. The two guidelines which scored particularly well in this domain were the National HIV Guideline and Retinopathy of Prematurity.

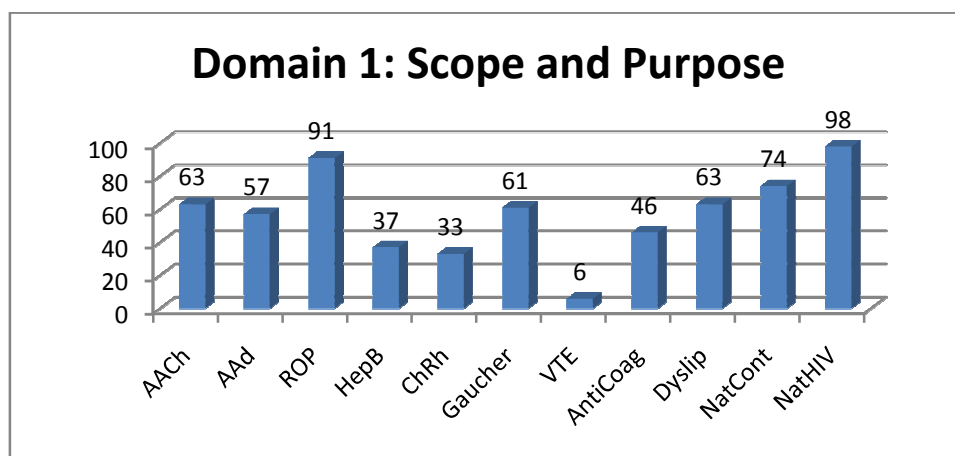


Figure 3.1: Scope and purpose

However, four guidelines scored below the mean score (57%) suggesting that that these guideline developers require more explicit information about how to describe the scope and purpose of the guideline and the potential health outcomes.

The intradomain variability was marked as evidenced by the:

- standard deviation of 26, and
- range: 6% to 98%

3.7 Domain 2: Stakeholder involvement

Guidelines should meet the needs of both the users as well as the target population to whom it would apply.

In this domain, the specific criteria are:

- the inclusion of individuals from all relevant professional groups in the guideline development group;
- obtaining the views and preferences of the target population (patients, public, etc.); and
- a clear definition of the target users of the guideline.

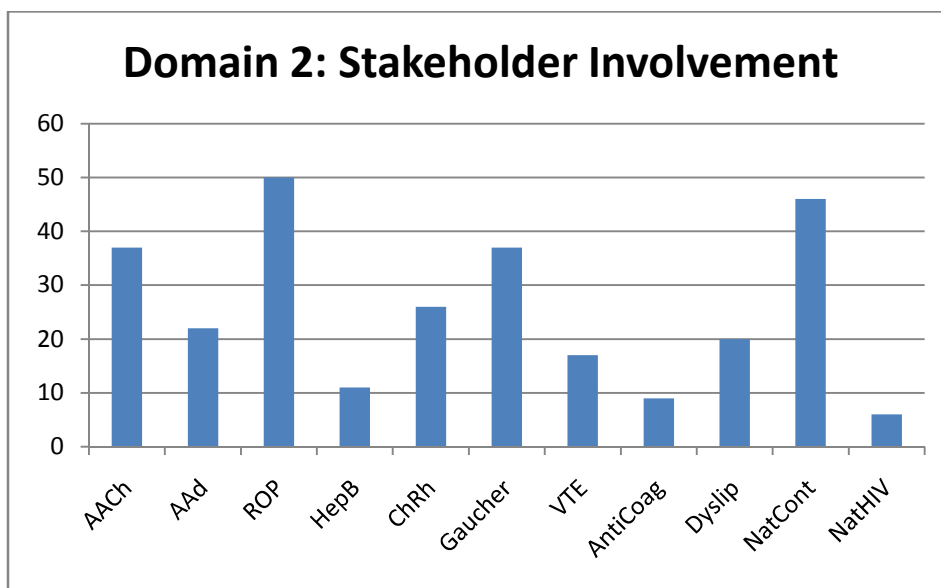


Figure 3.2: Stakeholder Involvement

In addition to the details of each individual in the guideline development, the role of each member should also be clarified. None of the guidelines assessed included this information. Furthermore, none of the guidelines provided explicit information about the intended audience for the guideline nor was a description of how the guideline should be used provided. Only one guideline sought to include the views of the target patient population (Chronic Rhinitis Guideline).

As shown in figure 3.2, only five of the guideline reviewed scored above the mean of 26% and none of the guidelines score higher than 50%.

3.8 Domain 3: Rigour of Development

Domain 3 assesses the strengths of the processes used to obtain and use evidence and how recommendations are made. The following criteria are used to assess the rigour of development:

- Use of systematic methods to search for evidence.
- A description of methods and criteria for selecting the evidence.
- A description of the strengths and limitations of the body of evidence.
- A description of methods for formulating the recommendations.
- Consideration of the health benefits, side effects, and risks in formulating recommendations.
- Linking the recommendations to the supporting evidence.
- External peer review of the guideline by experts prior to its publication.
- Provision of a procedure to update the guideline.

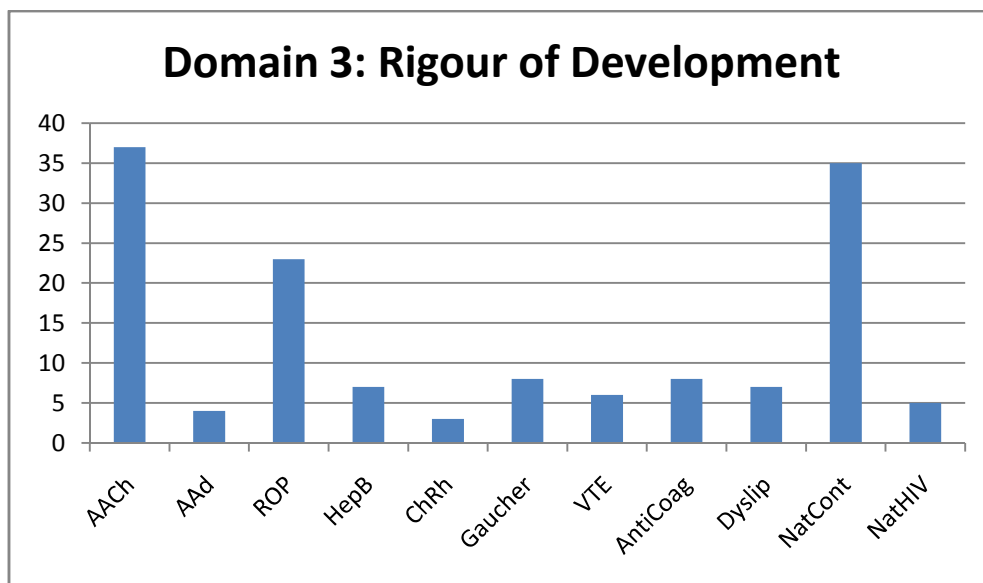


Figure 3.3: Rigour of development

In this domain, there was marked intradomain variation among guidelines as evidenced by the fact that while some guidelines scored well with regard to evidence synthesis, other guidelines scored well in terms of peer review of the guideline and consideration benefits and risks when formulating recommendations.

Although the National Contraceptive Guideline scored highest in this domain (depicted in figure 3.3), as with other guidelines assessed, the methods used to search for evidence was not included in the guideline. Only one guideline (Paediatric Asthma Guideline) provided some details of the terms used in the search strategy to find the evidence. However, the full search strategy was not provided and only a Pubmed search was conducted.

The Paediatric Asthma Guideline also scored relatively higher than the other guidelines on the criteria relating to describing methods and criteria for selecting the evidence and the strengths and limitations of the body of evidence. Furthermore, in this guideline, it was possible to identify the link between the evidence used and the recommendations made.

However, the National Contraception Guideline scored highest in this domain because there was a clear description of the health benefits, side-effects and risks that were considered when formulating the recommendation and the external peer review of the guideline before it was published.

As a domain, Rigour of development scored very poorly with a mean score of 13%. While some attempt was made to incorporate evidence based medicine principles into guideline development, there is an obvious need for more intensive integration of evidence into the guideline development process.

3.9 Domain 4: Clarity of Presentation

The criteria for this domain are:

- specific and unambiguous recommendations;
- the different options for management of the condition are clearly presented; and
- ease of identifying key recommendations.

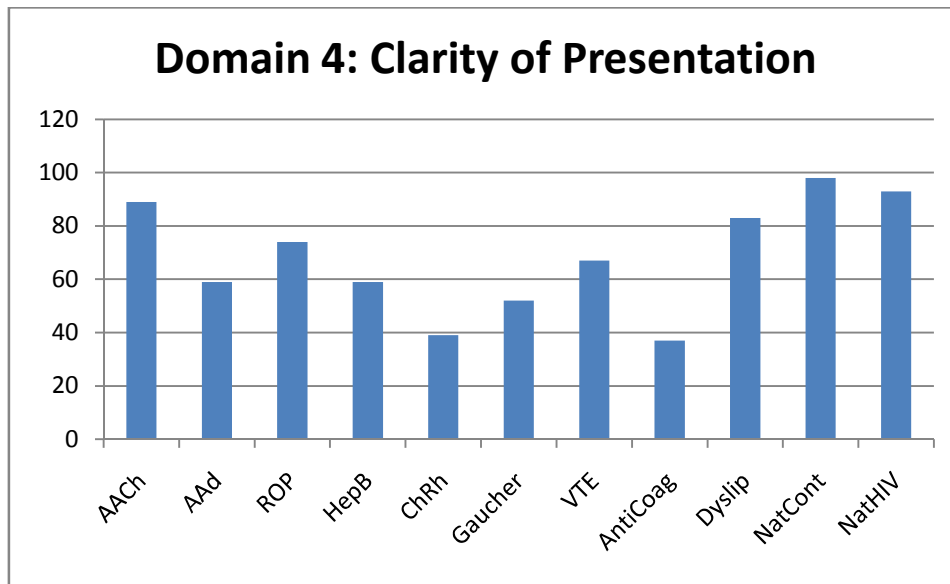


Figure 3.4: Clarity of Presentation

This domain had an average score of 68%, making it the domain with the highest score. However, only five of the 11 guidelines assessed achieved a score above the mean.

Notably, the National Contraception Guideline met all three quality criteria in this domain and therefore scored the highest (see figure 3.4). Conversely, the National ART Guideline did not because the presentation was confusing and there were conflicting recommendations within the guideline.

3.10 Domain 5: Applicability

Applicability refers to the resources and strategies for implementing the guideline.

The quality criteria in this domain are:

- a description of the facilitators and barriers to application of the guideline;
- provision of advice and/or tools on the practical implementation of recommendations;
- consideration of resource implications of the practical implementation of recommendations; and
- inclusion of monitoring and/or auditing criteria.

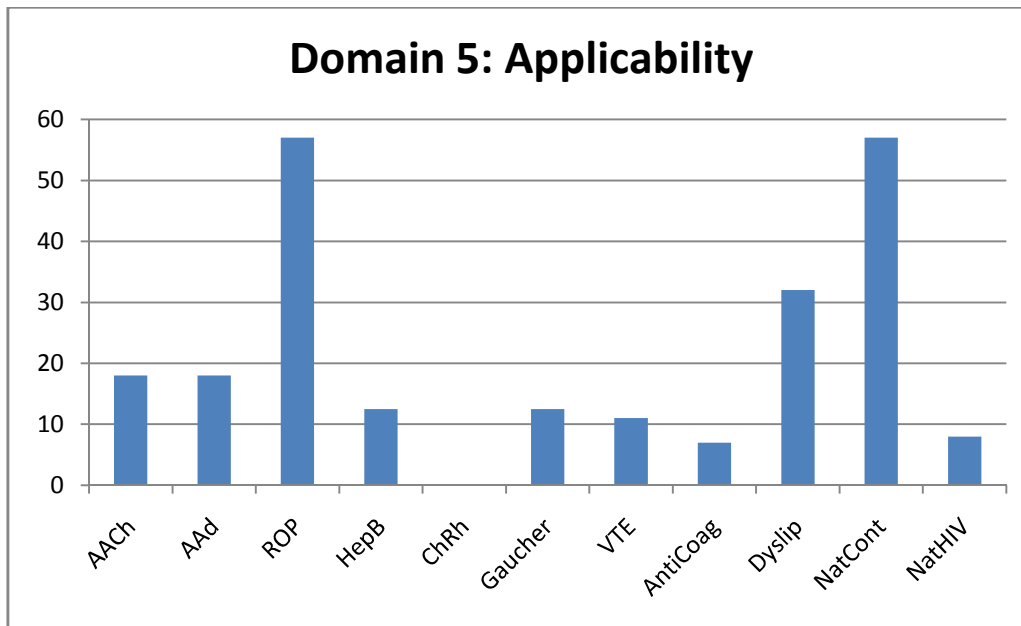


Figure 3.5: Applicability

As indicated in figure 3.5, the Retinopathy of Prematurity Guideline and the National Contraception Guideline scored equally high this domain (57%), far above the mean score of 19%. All quality items in this domain were addressed.

In the Retinopathy of Prematurity (ROP) Guideline tools were provided to assist the clinician to diagnose and manage the condition. For example, equipment necessary for ROP screening was provided in an appendix to the guideline. In the National Contraception Guideline, a reasonable attempt was also made to consider all the quality items in this domain.

3.11 Domain 6: Editorial Independence

In this domain, there are two items which addresses the issue of Editorial independence. These are:

- the influence of the funding body on the content of the guideline; and
- recording and dealing with competing interests of the guideline development group.

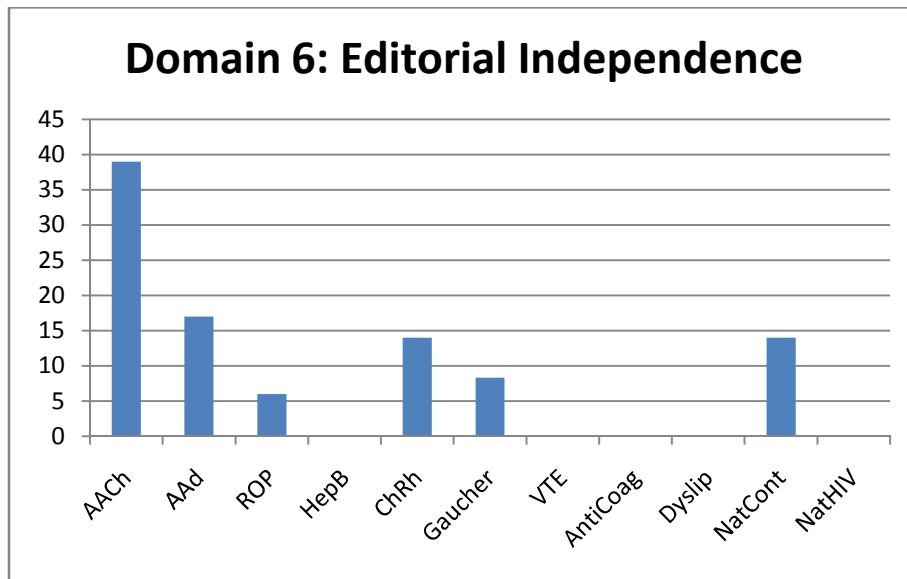


Figure 3.6: Editorial Independence

As can be seen in figure 3.6, almost half of the guidelines assessed (five of the eleven) scored very poorly in this domain with a score of 0%. In addition, the average score for this domain was 9%. The guideline scoring the highest in this domain was the Paediatric Asthma Guidelines. Although some other guidelines included a conflict of interest statement, the paediatric asthma guidelines recorded the interests of each member of the guideline development group. In the Chronic Rhinitis Guideline, although it was stated that the group meetings were sponsored, no statement was included that the funder did not influence the content of the guideline.

3.12 Overall guideline assessment

The overall assessment is a subjective measure because the reviewer must use judgement regarding the quality of the guideline. A rating (on a 7-point scale) is given after assessment is done using the 23-items evaluating the quality aspects. The assessor is also requested to provide a recommendation regarding whether the guideline should be recommended for use.

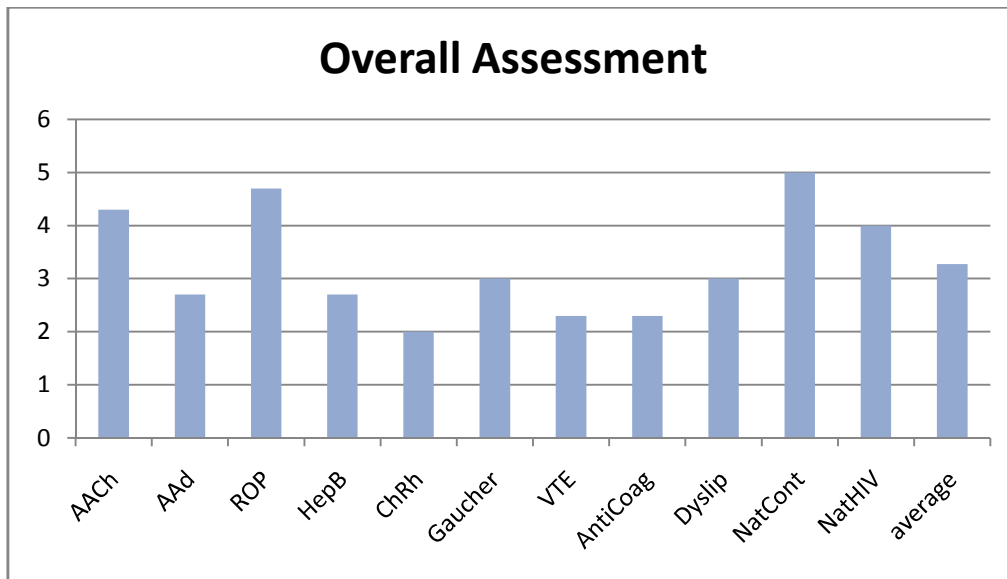


Figure 3.7: Overall assessment

The National Contraception Guideline scored the highest and was recommended for use. The ROP guideline received the second highest score but was recommended for use with modifications. Furthermore, the Dyslipidaemia Guideline scored below the mean (of three) but was recommended for use with modifications.

The mean score in terms of overall assessment of guidelines was three. Of the 11 guidelines assessed:

- Three were recommended for use, all of which scored above the mean score of three.
- Three were recommended for use with modifications, two guidelines with a equal to the mean score and one above.
- Five were not recommended for use, two of which scored equal to the mean score and three below the mean score.

3.13 Conclusion

There was great variation in scores across the six domains. Some guidelines scoring the highest in one domain, did not score particularly well in other domains. For example, although the National HIV Guidelines scored the highest in terms of domain 1 (Scope and purpose), this guideline scored the lowest in 2 of the 6 domains, i.e. domains 2 (Stakeholder involvement) and 6 (Editorial independence).

Concordance between reviewers was limited. Non-concordance between reviewers was found in seven of the 11 guidelines reviewed. This is an indication that the quality of review by reviewers were varied.

In terms of overall assessment, the scores indicate that the quality of guidelines assessed for this study varied. The mean score was 3. Of the 11 guidelines assessed:

- Three were recommended for use, all of which scored above the mean score of three.
- Three were recommended for use with modifications, two guidelines with a equal to the mean score and one above.
- Five were not recommended for use, two of which scored equal to the mean score and three below the mean score.

CHAPTER 4:

Discussion

4.1 Introduction

Clinical practice guidelines are valuable tools to provide quality, equitable care efficiently and effectively. Variability in the guideline development processes influence the quality of clinical guidelines produced and there is no agreed standard used in South Africa to guide the practice of guideline development.²

A selection of guidelines were critically appraised using the AGREE II instrument to evaluate the variability of the quality of guidelines, to determine a baseline for the quality of current guidelines and determine whether guidelines reviewed demonstrated good standard practice during their development. Scores were allocated on a 7-point rating scale (1–strongly disagree to 7–strongly agree) for each of the items in the six domains and for the overall assessment of the guidelines.

The higher the score, the better the guideline, However, the AGREE Consortium has not set minimum domain scores or patterns of scores to differentiate between low quality and high quality guidelines. [20] In some studies, mean scores were considered to be:

- acceptable if mean score is above 60%
- moderate if mean scores were between 30% and 60%, and
- low if mean score is below 30%.

Furthermore, although the methodology used to develop a guideline may be robust, or the guideline receives a high score when appraised using the AGREE II instrument, this does not automatically translate into acceptance of the guideline for implementation or improved quality of care for patients. [44]

4.2 Interguideline variation

The aggregation of domain scores into one score is not recommended. Instead, guideline comparison should take place across domains. For purpose of this dissertation, a mean score per guideline was determined for comparison.

The guideline with the highest mean score was the National Contraceptive Guideline, with the National HIV Guideline scoring the lowest. This is crude reflection of the overall quality of the guideline when compared across the six domains. As this type of analysis using the AGREE II instrument is not encouraged by the AGREE Trust, no further analysis and discussion in this regard will be pursued.

4.3 Intradomain variation and concordance

According to the AGREE Trust, guidelines should be compared across domains. By determining the difference between the mean and standard deviation per domain across guidelines, one can determine the variation in the quality of guidelines.

Table 4.1: Difference between mean and standard deviation per domain

Guideline	Domain 1: Scope and Purpose	Domain 2: Stakeholder Involvement	Domain 3: Rigour of Development	Domain 4: Clarity of Presentation	Domain 5: Applicability	Domain 6: Editorial Independence
Mean score	57%	29%	13%	68%	21%	9%
SD	26%	15%	13%	21%	19%	12%
Mean-SD	31	14	0	47	2	-3

Table 4.1 demonstrates the variation in the quality of guidelines per domain. The greatest variation was observed in the following domains:

- Domain 4: Clarity of presentation (Mean - SD = 47%)
- Domain 1: Scope and Purpose (Mean - SD = 31%)
- Domain 2: Stakeholder involvement (Mean - SD = 14%)

Domains not demonstrating significant variability in quality (i.e. low variability between domains) are:

- Domain 3: Rigour of development (Mean - SD = 0%)
- Domain 5: Applicability (Mean - SD = 2%)
- Domain 6: Editorial independence (Mean - SD = -3%)

While the variability in the quality of guidelines per domain has been demonstrated, this is not a statement of the quality of the guidelines in that domain. So, for instance, while the variability in the quality of guidelines in domain 3: Rigour of development is

very small, the overall quality of guidelines reviewed in this domain is poor with a mean score of 13%. The higher the mean score, the better the quality of comparative guidelines for that domain.

With regard to concordance, i.e. agreement between reviewers, none was observed in seven of the 11 guidelines. This lack of concordance introduces bias in the result observed. The impact of the bias can be mitigated by increasing the number of appraisers reviewing the guideline.

The lack of concordance could be explained by the differences in the clinical skills and knowledge base of the reviewers, two pharmacists and three clinicians. The lack of concordance also suggests that reviewers proficiency in the application of the tool may need to be increased through re-training. In addition, the level of discrepancy can be reduced by increasing the number of reviewers. However, the extent to which these factors impact on concordance was not the object of this study.

4.4 Domain 1: Scope and purpose (moderate)

This section deals with the expected outcome of the guideline on society, patients and individuals. Guidelines should provide explicit information about the aim, elements of management to be addressed (e.g. screening, diagnosis, etc and the patient population to which the guideline should apply (e.g. adults or children).

The average domain score was 57% with scores ranging from 6% to 98%. The difference between the mean and the standard deviation for this domain was the highest when compared to other domains. This suggest that while some guideline developers may have an understanding of what the expected outcomes of the guidelines should be, other guideline developers have demonstrated lack of awareness in this regard.

4.5 Domain 2: Stakeholder involvement (low)

The mean score for this domain was 29%, with a range of 6% to 50%. Furthermore, the difference between the mean and the standard deviation for this domain was 15% thus suggesting that communication about the guideline being developed and incorporation of the possible contributions from stakeholders is poor. The rareness of

a conditions and the dearth of robust evidence to formulate strong recommendations increase clinicians reliance on "expert opinion", e.g. Gauchers' Disease. Nevertheless, even in these instances, methods to incorporate value judgements are available as long as the methods used to incorporate values (such as consensus seeking) are documented.

There is controversy as to whether recommendations based on consensus constitute an evidence base for guideline recommendations in the absence of robust evidence. [45,46]

4.6 Domain 3: Rigour of development (low)

In this domain, the scores were as follows:

- The mean score was 13%.
- Range: for this domain was 3% to 37%
- Difference between the mean and the standard deviation was 0%.

These scores clearly demonstrate that for this domain, there is no variability of quality in this domain, the overall quality is low.

Rigour of development has been shown to be lacking in all guidelines evaluated. Despite the fact that principles of evidence based was introduced decades ago, its incorporation into the guideline development process has still not advanced. This finding is cause for concern as the quality of evidence used to formulate recommendation has a direct impact on the quality of the guideline developed.

The low scores may be explained by a lack resources and/or skill and knowledge how to perform literature searches.[46] Furthermore, it could be argued that guideline developers are unaware of the need to report how evidence was sourced and incorporated into guidelines. In this setting, the use of the AGREE II instrument is a valuable resource for creating the framework for reporting how guidelines were developed.

4.7 Domain 4: Clarity of Presentation (high)

Presentation allows for better understanding of guideline recommendations. Clearly presented guidelines that are easy to follow will also influence whether the guideline is likely to be accepted and implemented.

This is the only domain in the study where the mean is above what is considered good quality, i.e. mean = 68%. However, the range was very wide, 39% to 98% and the difference between the mean and the standard deviation = 47.

Increased understanding by guideline developers regarding the need for specific and unambiguous recommendation is required. In addition, many guidelines do not provide all the evidence options for the management of a condition, instead focusing just on medical management. Furthermore, key recommendations are not easily identified.

4.8 Domain 5: Applicability (low)

Few guidelines provide the necessary tools to ensure proper implementation of guidelines or consider the risks and benefits of recommendations made. The mean score of 19% for this domain indicates a gap in the knowledge of guideline developers regarding the implications of implementing recommendations in practice and the availability of resources to do so.

The inability of guideline developers to provide guidance on how guidelines should be implemented and the resources needed to do so has an impact on guideline implementation, and ultimately on patient outcomes. This is of particular importance when new technologies are introduced in the health system.

4.9 Domain 6: Editorial independence (low)

All guidelines scored consistently low in this domain, with a mean domain score of 9%. The mean domain score is the lowest across all the domains and has an impact on the credibility of the guideline developed. The mean domain score may be explained by lack of knowledge and declaration of information about funding sources.

In recognition of the risk of conflicts of interest poses to the validity of guidelines, strategies to limit bias have been introduced by numerous guideline developers. [47,48]. These include the declaration of conflict of interest policies, multidisciplinary panels for guideline development, using systematic procedures to review, evaluate evidence and linking evidence to recommendations as well as peer review from outside reviewers. [48]

CHAPTER 5

Conclusion and recommendations

5.1 Introduction

The results of this study suggest that guideline development in South Africa is not robust. The quality of the guidelines assessed show great variability as no structured approach for guideline development in South Africa is available.

5.2 Conclusion about the research problem

The aim of this study was to use the AGREE II instrument to evaluate the variability of the quality of selected guidelines, to determine a baseline for the quality of current guidelines and determine whether guidelines reviewed demonstrated good standard practice during their development.

The application of the AGREE II instrument successfully demonstrated the variability in the quality of guidelines evaluated. The 23-items used to evaluate quality aspects can be used to improve the methodology to develop guidelines in the future. However, the study failed to determine a baseline for the quality of current guidelines. The study demonstrates that current practice of guideline development in South Africa is poor.

5.3 Limitations

Limitations of the study include:

- Limited number of reviewers available to appraise guidelines using the AGREE II instrument.
- Limited concordance between reviewers as demonstrated by the number of guidelines recommended for review.
- The small number of guidelines assessed in this study is probably insufficient to provide a meaningful baseline for all guidelines produced in the country.
- The impact of assessor interest and knowledge in a subject matter on scores assigned to guidelines was not determined.

A shortcoming of the AGREE II instrument is that, although the rigour of guideline development is assessed, it does not assess the quality of the recommendations incorporated into the guideline based on research evidence. Nor does it assess the conclusions drawn from a critical appraisal of the evidence. [20] This poses a threat to the credibility of the guidelines. However, other tools are available to ensure that evidence based recommendation are incorporated into guidelines.

5.4 Recommendations

People involved in guideline development are often content experts and do not have sufficient expertise in developing guidelines. Therefore, guideline developers will require intensive training on all domains included in the AGREE II instrument. In addition, appraisers of clinical guidelines will also require training in order to increase the pool of trainers with the necessary skill and experience to review guidelines. This will have a positive impact on concordance. In the various domains, the following is recommended:

Table 5.1: Suggested recommendation to improve guideline development

Domain 1: Scope and Purpose	<ul style="list-style-type: none"> • Information provided should be more concise and presented in a summarised way. • The population to whom the guideline is meant to apply should be more clearly defined. • The health questions covered by the guideline should be more concise and ranked according to the priority health needs of patients.
Domain 2: Stakeholder Involvement	<ul style="list-style-type: none"> • Adopt a multidisciplinary team approach. • Involve patients in the guideline process in order to determine their preferences.

Domain 3: Rigour of Development	<ul style="list-style-type: none"> • Intensive training on evidence based medicine principles. • Skills training regarding systematic review of evidence. • Develop tools to reflect EBM decision making process such as development of presentation slides, publish EBM reviews undertaken in peer review journals.
Domain 4: Clarity of Presentation	<ul style="list-style-type: none"> • Limit the length of guidelines developed. • Provide summaries of most important recommendations in a succinct format.
Domain 5: Applicability	<ul style="list-style-type: none"> • Create awareness during the guideline development process of the potential impact of recommendations made on the health system. • Tools for implementation should be provided in the guideline. • Monitoring and evaluation of guideline development should be provided as part of the implementation plan
Domain 6: Editorial Independence	<ul style="list-style-type: none"> • Training regarding conflict of interest. • Development and implement conflict of interest policies. • Compulsory, public declaration of conflicts of interest of all panel members.

Target users of guideline are not usually aware of the quality aspects of guidelines. This places additional responsibility on guideline developers to ensure that good quality guidelines are developed.

Awareness regarding what constitutes good standard practice by guideline developers is needed. This lack of awareness is evidenced by the variation in the quality of the guidelines assessed in this study.

Furthermore, a structured approach to guideline development is needed for South Africa. Once a structured approach is adopted, it is likely that variability in the guideline development processes and, ultimately, guideline quality will be minimised.

5.5 Further research

Further research is needed to determine the methodology on how to create awareness about good standard practice in guideline development processes. In addition, research will be required to determine how to incorporate quality aspects of guideline development into current processes.

Research is needed on what the factors are that could possibly influence concordance.

CHAPTER 6

References

1. Weisz G, Cambrosio A, Keating P, Knaapen L, Schlich T, Tournay VJ. The Emergence of Clinical Practice Guidelines. *The Milbank Q* 2007; 85(4): 691-727.
2. Isaacs D, Fitzgerald D. Seven Alternatives to Evidence Based Medicine. *BMJ* 1999; 319: 1618.
3. Australian Commission on Safety and Quality in Health Care (ACSQHC). *Medical Practice Variation: Background Paper*. Sydney: ACSQHC, 2013.
4. Institute of Medicine, Committee on Clinical Practice Guidelines. *Clinical practice guidelines we can trust*. Washington, DC: National Academies Press; 2011.
5. Woolf S, Schünemann HJ, Grimshaw JM, Shekelle P. Developing clinical practice guidelines: types of evidence and outcomes; values and economics, synthesis, grading, and presentation and deriving recommendations. *Implement Sci* 2012; 7:61. <http://www.implementationscience.com/content/7/1/61> (accessed 20 October 2014).
6. Fowkes FG, Roberts CJ. Introducing guidelines into clinical practice. *Eff Health Care* 1984; 1(16): 313-323.
7. Day P, Klein R, Miller F. *A comparative UK-US study of guidelines*. The Nuffield Trust. Report number: 4, 1998.
8. Steinbrook R. Guidance for Guidelines. *Lancet* 2007; 356(4): 331-333.
9. Kredt T, Gerritsen A, van Heerden J, Conway S, Siegfried N. Clinical practice guidelines within the Southern African development community: a descriptive study of the quality of guideline development and concordance with best evidence for five priority diseases. *Health Res Policy and Syst* 2012; 10(1): 1-13.
10. 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(9198): 103-106.
11. Institute of Medicine, Committee on Clinical Practice Guidelines. *Guidelines for clinical practice: From development to use*. Washington, DC: National Academies Press; 1992.
12. Cluzeau FA, Littlejohns P, Grimshaw JM. Appraising clinical guidelines: towards a "Which" guide for purchasers. *Quality in Health Care* 1994; 3(3): 121-122.

13. Cluzeau FA, Littlejohns P, Grimshaw JM, Feder G, Moran SE. Development and application of a generic methodology to assess the quality of clinical guidelines. *Int J Qual Health* 1999; 11(1): 21-28.
14. Graham ID, Calder LA, Hébert PC, Carter AO, Tetroe JM. A comparison of clinical practice guideline appraisal instruments. *Int J Technol Assess Health Care* 2000; 16(4): 1024-1038.
15. Graham ID, Beardall S, Carter AO, Glennie J, Hébert PC, Tetroe JM, et al. What is the quality of Clinical Practice Guidelines in Canada? *Can. Med. Assoc. J.* 2001; 165(2): 157-163.
16. Burgers JS, Fervers B, Haugh M, Brouwers M, Bowman G, Philip T, Cluzeau FA. International Assessment of Clinical Practice Guidelines in Oncology Using the Appraisal Guidelines and Research and Evaluation Instrument. *J Clin Oncol* 2004; 22(10): 2000-2007.
17. Cranney A, Waldeger L, Graham ID, Man-Son-Hing M, Byszewski A, Ooi DS. Systematic assessment of the quality of osteoporosis guidelines. *BMC Musculoskeletal Disorders* 2002; 3(20): .
18. Grol R, Cluzeau FA, Burgers JS. Clinical practice guidelines: towards better quality guidelines and increased international collaboration. *Brit J Cancer* 2003; (Suppl 1): S4-S8.
19. The AGREE Collaborative Group. Guideline development in Europe: an international comparison. *Int J Technol Assess* 2000; 16(4): 1039-1049.
20. The AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf in Health Care* 2003; 12(1): 18-23.
21. The AGREE Next Steps Consortium. *AGREE II Instrument*.
http://www.agreetrust.org/wp-content/uploads/2013/10/AGREE-II-Users-Manual-and-23-item-Instrument_2009_UPDATE_2013.pdf (accessed 19 October 2014).
22. 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 C* 2005; 17(3): 235-242.
23. McAlister M, van Diepen S, Padwal RS, Johnson JA and Mujamdar SR. How evidence-based are the recommendations in evidence-based guidelines. *PLoS* 2007; 4(8): e250. www.plosmedicine.org (accessed 20 October 2014).

24. Grimshaw J, Russel IT. Achieving health gain through clinical guidelines II: Ensuring guidelines change medical practice. *Qual in Health Care* 1994; 3(1): 45-52.
25. Shiffman RN, Shekelle P, Overhage JM, Slutsky J, Grimshaw J, Deshpande AM. Standardized Reporting of Clinical Practice Guidelines: A Proposal from the Conference on Guideline Standardization. *Ann Intern Med* 2003; 139(6): 493-498.
26. Burgers J, Grol R, Klazinga NS, Mäkelä M, Zaat J. Towards evidence-based clinical practice: an international survey of 18 clinical guideline programmes. *Int J Qual Health* 2003; 15(1): 31-45.
27. Berg AO, Atkins D, Tierney W. Clinical Practice Guidelines in Practice and Education. *J Gen Intern Med* 1997; 12(Supplement 2): S25-S33.
28. Shaneyfelt TM, Mayo-Smith MF, Rothwang IJ. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA* 1999; 281(20): 1900-1905
29. Rosenfeld RM, Shiffman RN. Clinical practice guideline development manual: A quality-driven approach for translating evidence into action. *Otolaryngol Head Neck Surg* 2009; 140(6): 1-43.
30. Agency for Health Care Research and Quality. *Evaluation of the National Guideline Clearinghouse™ (NGC)*. Department of Health and Human Services' Agency for Healthcare Research and Quality. Report number: AHRQ Contract No. 4203, 2011.
31. Harbour R, Miller J. A new system for grading recommendations in evidence. *BMJ* 2001; 323(7308): 334-336.
32. The Editor. *Review of Guidelines: Use of AGREE II instrument and appointment of new editorial sub-committee*.
<http://www.samj.org.za/index.php/samj/announcement/view/7> (accessed 26 October 2014).
33. Guideline for the management of acute asthma in children: 2013 update Part 3: March 2013. *S Afr Med J* 2013;103(3):199-207. DOI:10.7196/SAMJ.6658
34. Guideline for the management of acute asthma in adults: 2013 update. *S Afr Med J* 2013;103(3):189-198. DOI:10.7196/SAMJ.6526
35. Guideline for the prevention, screening and treatment of retinopathy of prematurity (ROP) *S Afr Med J* 2013;103(2):116-125. DOI:10.7196/SAMJ.6305

36. South African guideline for the management of chronic hepatitis B: 2013. *S Afr Med J* 2013;103(5):335-349. DOI:10.7196/SAMJ.6452
37. Chronic rhinitis in South Africa: Update 2013. *S Afr Med J* 2013;103(6):419-422. DOI:10.1796/SAMJ.6972
38. South African guidelines for the management of Gaucher disease, 2011. *S Afr Med J* 2012;102(8):697-702. DOI:10.7196/SAMJ.5439
39. Venous thromboembolism: Prophylactic and therapeutic practice guideline. *S Afr Med J* 2013;103(4):260-267. DOI:10.7196/SAMJ.6706
40. Paediatric Anticoagulation Guidelines. *S Afr J Med* 2012;171-175.
41. South African Dyslipidaemia Guideline Consensus Statement. *S Afr Med J* 2012;102:177-188.
42. National Contraception Clinical Guidelines. 2012, from http://www.doh.gov.za/list.php?pageNum_rsList=2&totalRows_rsList=96&type=Policy%20Documents
43. The South African Antiretroviral Treatment Guidelines. 2013 from http://www.doh.gov.za/list.php?pageNum_rsList=2&totalRows_rsList=96&type=Policy%20Documents (accessed May 2013)
44. Spuls PI, Nast A. Evaluation of and Perspectives on Guidelines: What Is Important?. *Journal of Investigative Dermatology* 2010; 130: 2348-2349. Ollenschläger G, Marshall C, Qureshi S, Rosenbrand K, Burgers J, Mäkelä M, Slutsky J. Improving the quality of health care: using international collaboration to inform guideline programmes by founding the Guidelines International Network (GIN). *Qual Safety in Health Care* 2002; 13(6): 455-460.
45. Alonso-Coello P, Irfan A, Sola` I, Gich I, Delgado-Noguera M, Rigau D, et al. The quality of clinical practice guidelines over the last two decades: a systematic review of guideline appraisal studies. *Qual Saf in Health Care* 2010; 19(6): e56
46. Harbour R, Miller J for the Scottish Intercollegiate Guidelines Network Grading Review Group. *BMJ* 2001;323:334-6.
47. Boyd EA and Bero LA. Improving the use of research evidence in guideline development: 4. Managing conflicts of interests. *Health Res Policy Syst* 2006, 4:16 doi:10.1186/1478-4505-4-16.
48. IOM (Institute of Medicine). Conflict of Interest in Medical Research, Education, and Practice. Washington, DC: The National Academies Press, 2009.

ADDENDUMS

Addendum 1: Data collection tool

Template for scoring AGEE II scoring sheet:

Guideline:

Reviewer:

Domains	Criteria	1 Strongly disagree	2	3	4	5	6	7 Strongly agree
Domain 1: Scope and Purpose	The overall objective(s) of guideline is (are) specifically described							
	The health questions(s) covered by guideline is (are) specifically described							
	Population (patients, public etc.) to whom guideline is meant to apply to is specifically described							
Domain 2: Stakeholder Involvement	The guideline development group includes individuals from all relevant professional groups							
	The views and preferences of the target population (patients, public, etc.) have been sought							
	Target users of the guideline are clearly defined							

Domains	Criteria	1 Strongly disagree	2	3	4	5	6	7 Strongly agree
Domain 3: Rigour of Development	Systematic methods were used to search for evidence							
	The criteria for selecting evidence are clearly described							
	The strengths and limitations of the body of evidence are clearly described							
	The methods for formulating the recommendations are clearly described							
	The health benefits, side effects, and risks have been considered in formulating the recommendations							
	There is an explicit link between the recommendations and the supporting evidence							
	The guideline has been externally reviewed by experts prior to its publication							
	A procedure for updating the guideline is provided							

Domains	Criteria	1 Strongly disagree	2	3	4	5	6	7 Strongly agree
Domain 4: Clarity of Presentation	The recommendations are specific and unambiguous							
	The different options for management of the condition or health issue are clearly presented							
	Key recommendations are easily identifiable							
Domain 5: Applicability	The guideline describes facilitators and barriers to its application							
	The guideline provides advice and/or tools on how recommendations can be put into practice							
	The potential resource implications of applying the recommendations have been considered							
	The guideline presents monitoring and/or auditing criteria							
Domain 6: Editorial independence	The views of the funding body have not influenced the content of the guideline							
	The competing interests of guideline development group members have been recorded and addressed.							

Overall guideline assessment

	1 Lowest Possible Quality	2	3	4	5	6	7 Highest Possible Quality
Overall quality							
I would recommend this guideline for use	Yes		Yes, with modifications			No	

Notes:

Addendum 2: Template for calculation of domain scores per guideline and level of discrepancy

Seven-point AGREE II Score Calculator

You must fill in ALL of the Question ratings from an appraiser for the Domain score to be accurate. *Note: Please use the AGREE II User's Manual for full instructions.*

Total # of Appraisers	Appraiser				
	1	2	3	4	
0					
Domain 1 – Scope and Purpose					
Q1 - The overall objective(s) of the guideline is (are) specifically described.					0
Q2 - The health question(s) covered by the guideline is (are) specifically described.					0
Q3 - The population (patients, public, etc.) to whom the guideline is meant to apply is specifically described.					0
	Caution: Empty Cells	Caution: Empty Cells	Caution: Empty Cells	Caution: Empty Cells	0
Domain 1 Score for 0 Appraiser(s):					
Domain 2 – Stakeholder Involvement					
Q4 - The guideline development group includes individuals from all relevant professional groups.					0
Q5 - The views and preferences of the target population (patients, public, etc.) have been sought.					0
Q6 - The target users of the guideline are clearly defined.					0
	Caution: Empty Cells	Caution: Empty Cells	Caution: Empty Cells	Caution: Empty Cells	0
Domain 2 Score for 0 Appraiser(s):					
Domain 3 – Rigour of Development					
Q7 - Systematic methods were used to search for evidence.					0
Q8 - The criteria for selecting the evidence are clearly described.					0
Q9 - The strengths and limitations of the body of evidence are clearly described.					0
Q10 - The methods for formulating the recommendations are clearly described.					0
Q11 - The health benefits, side effects, and risks have been considered in formulating the recommendations.					0
Q12 - There is an explicit link between the recommendations and the supporting evidence.					0
Q13 - The guideline has been externally reviewed by experts prior to its publication.					0
Q14 - A procedure for updating the guideline is provided.					0
	Caution: Empty Cells	Caution: Empty Cells	Caution: Empty Cells	Caution: Empty Cells	0
Domain 3 Score for 0 Appraiser(s):					

Domain 4 – Clarity of Presentation				
Q15 - The recommendations are specific and unambiguous.				0
Q16 - The different options for management of the condition or health issue are clearly presented.				0
Q17 - Key recommendations are easily identifiable				0
	Caution: Empty Cells	Caution: Empty Cells	Caution: Empty Cells	Caution: Empty Cells
Domain 4 Score for 0 Appraiser(s):				
Domain 5 – Applicability				
Q18 - The guideline describes facilitators and barriers to its application.				0
Q19 - The guideline provides advice and/or tools on how the recommendations can be put into practice.				0
Q20 - The potential resource implications of applying the recommendations have been considered.				0
Q21 - The guideline presents monitoring and/or auditing criteria.				0
	Caution: Empty Cells	Caution: Empty Cells	Caution: Empty Cells	Caution: Empty Cells
Domain 5 Score for 0 Appraiser(s):				
Domain 6 – Editorial Independence				
Q22 - The views of the funding body have not influenced the content of the guideline.				0
Q23 - Competing interests of guideline development group members have been recorded and addressed.				0
	Caution: Empty Cells	Caution: Empty Cells	Caution: Empty Cells	Caution: Empty Cells
Domain 6 Score for 0 Appraiser(s):				
Overall Guideline Assessment				
1. Rate the overall quality of this guideline. <i>Scoring: 1(Least Quality) - 7(Highest Quality)</i>				
2. I would recommend this guideline for use. <i>Scoring: "Yes"; "Yes, with modifications"; "No"</i>				

DATA AUDIT

of Domains with SD that are **≥ 1.5 and < 2**
 SD (OS: Outlying Score, first level severity) **0**

of Domains with SD that are **≥ 2** SD
 (OS2: Outlying Score, 2nd level severity) **0**

Decision Rule:

Of Domains 1-5 and the Overall Assessment
OS ≥ 3 or **OS2** ≥ 1

No action required

Average Standard Deviation of Items by Domain

Domain	Standard Deviation	Discrepancy Level
1	#DIV/0!	#DIV/0!
2	#DIV/0!	#DIV/0!
3	#DIV/0!	#DIV/0!
4	#DIV/0!	#DIV/0!
5	#DIV/0!	#DIV/0!
Overall Guideline Assessment	#DIV/0!	#DIV/0!

ADDENDUM 3

Addendum 3: Permission letter from Ethics Committee