

# AGREE-II Appraisal of a clinical algorithm for hypotonia assessment

Pragashnie Govender

University of KwaZulu-Natal (Westville Campus), School of Health Sciences, Private Bag X54001, Durban, 4000.

## Abstract

**Objective:** The objective of this study was to systematically appraise the quality of an evidenced-based clinical algorithm for the clinical assessment of hypotonia in children.

**Design:** The Appraisal of Guidelines for Research and Evaluation (AGREE) II tool with 23 items and six domains was used. The study was located in South Africa. Ten appraisers, who were recruited based on specific selection criteria, completed the assessment.

**Results:** Nine appraisers recommended the EBCA without any modification. Scope and purpose (94%), stakeholder involvement (91%) and editorial independence (99%) were rated the highest with the lower scoring domains being clarity of presentation (85%) and applicability (86%) due to clarity required in areas of resource implications and auditing and monitoring criteria. Inter-rater reliability was strong (ICC 0.7) amongst the appraisers in this study.

**Conclusion:** This is the first independent assessment of the methodological rigour and transparency of a clinical algorithm using the AGREE-II instrument. Determining the quality of the EBCA for practice is essential as this would ultimately aid clinicians towards more accurate clinical assessment of hypotonia which would inevitably impact outcomes and management of the child presenting with this symptom. Whilst the AGREE-II provided initial feedback on the methodological rigour of development, understanding that the AGREE-II instrument evaluates the guideline development process and not the content is also essential in order to consider the next stage which would be to consider clinicians feedback on the clinical utility of this EBCA.

**Keywords:** AGREE-II; clinical algorithm; hypotonia, low muscle tone; paediatrics.

**DOI:** <https://dx.doi.org/10.4314/ahs.v18i3.38>

**Cite as:** Govender P. *AGREE-II Appraisal of a clinical algorithm for hypotonia assessment*. *Afri Health Sci*. 2018;18(3): 790-798. <https://dx.doi.org/10.4314/ahs.v18i3.38>

## Introduction

In recent years, the clinical assessment of hypotonia has re-emerged as contentious, given the wide range of diagnoses that have hypotonia as an underlying symptom<sup>1,2,3</sup>. The initial clinical evaluation is essential in the diagnostic

process that ensues and in determining appropriate management<sup>4</sup>. In an attempt to address this contention expressed in the scientific literature, the author engaged in a systematic process towards the development of an evidenced-based clinical algorithm (EBCA)<sup>5</sup>. This process is however incomplete without an appraisal of the quality of this process of development.

Care pathways, algorithms and practice guidelines in clinical research have developed as useful methods in standardising and guiding patient care<sup>6,7,8,9</sup> and are promulgated to encourage high quality care<sup>10</sup>. EBCA's are tools that show promise as evidence is coded into specific rules and actions that facilitate delivery of appropriate care to the relevant recipients of the care<sup>9</sup>. However, although these EBCA's and guidelines have a significant role in health-care practice, their development process and their basis

### Corresponding author:

Pragashnie Govender,  
School of Health Sciences  
University of KwaZulu-Natal (Westville Campus)  
Private Bag X54001, Durban, 4000.  
Telephone: +27312608258 (direct)  
+27312607821 (administrator)  
Mobile: +27716056931  
Fax: +27312607227  
Email: [naidoopg@ukzn.ac.za](mailto:naidoopg@ukzn.ac.za)

of evidence has been subject to criticism<sup>11</sup>. Given that these processes and tools have the potential to influence the healthcare of many individuals, the method of their development and assessment should be open to scrutiny<sup>11</sup>.

In order to assess the quality of the developed EBCA 5, the AGREE Collaboration's Appraisal of Guidelines for Research and Evaluation, version two (AGREE-II) instrument was identified as suitable<sup>12</sup>. Quality of guidelines as defined by the collaboration<sup>12</sup> is viewed 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". The AGREE II is generic, with application to guidelines across the health care continuum including screening, diagnosis and interventions. The application however has never been reported on the assessment of an evidenced-based algorithm.

The author acknowledges that only once gaps in the scientific evidence and its delivery have been addressed, then only can the issues around implementation and barriers be identified and overcome with the developed EBCA 5. In an attempt, to determine this quality of the EBCA that was developed for the clinical assessment of hypotonia, an evaluation process with the use of the AGREE II instrument<sup>12</sup> was initiated, the findings of which are described in this paper.

## Methods

The AGREE-II Instrument<sup>12</sup> was used to assess the quality of the clinical algorithm with respect to the methodological rigor in development. A sample of clinicians working in the field of paediatrics in disciplines of occupational therapy, physiotherapy and paediatrics in addition to policy-makers, and guideline developers working in these three fields were recruited for participation as appraisers. The study was located in South Africa. The AGREE-II developers recommend that a guideline be assessed by at least two appraisers and preferably four as this will increase the reliability of the assessment<sup>12</sup>. In this study, however a larger sample of ten participants were selected in order to increase the reliability of the assessment.

Quality in the AGREE-II instrument<sup>12</sup> is assessed across six domains, namely, scope and purpose, stakeholder involvement, rigor of development, clarity of presentation, applicability, and editorial independence, comprising a total of 23 items and two overall assessments. Each item is rated on a 7-point scale, either from 1 (strongly disagree) to 7 (strongly agree). Data from the AGREE-II is expressed as calculated percentage scores. Domain and overall rating scores are calculated by a sum of the individual item scores for each appraiser (obtained scores) minus the minimum possible score minimum possible score per item (1, strongly disagree)  $\times$  n items  $\times$  n appraisers. The total is expressed as a percentage of the maximum possible score maximum possible score per item (7, strongly agree)  $\times$  n items  $\times$  n appraisers minus the minimum possible score, according to the formula<sup>12</sup>:

$$\frac{\text{Obtained score} - \text{Minimum possible score}}{\text{Maximum possible score} - \text{Minimum possible score}}$$

Each appraiser received a pack of documents that included (i) a demographic questionnaire, (ii) the evidenced-based clinical algorithm, (iii) a technical report that accompanied the algorithm, (iv) a copy of the AGREE-II Manual, (v) a copy of the AGREE-II Scoring Sheets and (vi) a quick reference guide (explaining the domains and cross referencing it with the AGREE-II item and the technical Guideline item). The appraisers were given a maximum of three weeks to complete the appraisal process independently. Appraisers did not communicate or confer with each other during the appraisal process. To preserve confidentiality, a research assistant collated the scores which were uploaded onto the My AGREE Plus site ([www.agreetrust.org](http://www.agreetrust.org)) for electronic collation. Qualitative comments were transposed onto a transcription sheet for coding. Each of the items and the two overall rating items were assessed for their applicability to the clinical algorithm. A score of 1 was given when relevant information was very poorly reported or not provided or thought to be not applicable. Scores from 2 to 6 were assigned when the reporting did not meet the full criteria or considerations for an item, with scores increasing as more criteria and considerations were met. A score of 7 was given when the quality of reporting was exceptional and all criteria and considerations were met in full for an item. The overall assessment of the guideline was also assessed

and appraisers were requested to state if they would recommend the guideline, recommend it with modifications or not recommend it. Inter-rater reliability was calculated using an intraclass correlation coefficient (two way random mixed model) with SPSS version 23™ (IBM Corp, Armonk, NY, USA). The ICC provides a scalar measure of agreement or concordance between raters. The single measure of the ICC is an index for the reliability of the ratings for one, typical, single rater whilst the average measure is an index for the reliability of different raters averaged together<sup>13,14</sup>. ICC can thus be interpreted as follows, 0-0.2 (poor agreement); 0.3-0.4 (fair agreement); 0.5-0.6 (moderate agreement); 0.7-0.8 (strong agreement); and >0.8 (almost perfect agreement).

## Results

Demographic data of the sample of appraisers in this study are presented in Figure 1. The age and experience of the appraisers were varied with nine of the ten holding a master's degree. The majority of appraisers positioned themselves as guideline developers and researchers, with experience in policy and guideline development ranging from less than five years to up to 11-15 years. Half of the appraisers considered themselves novice developers/appraisers whilst the other half considered them as being between a novice and expert in guideline development or appraisal. Six participants worked in academic settings, one worked only in a clinical setting, with three working in both an academic and clinical settings.

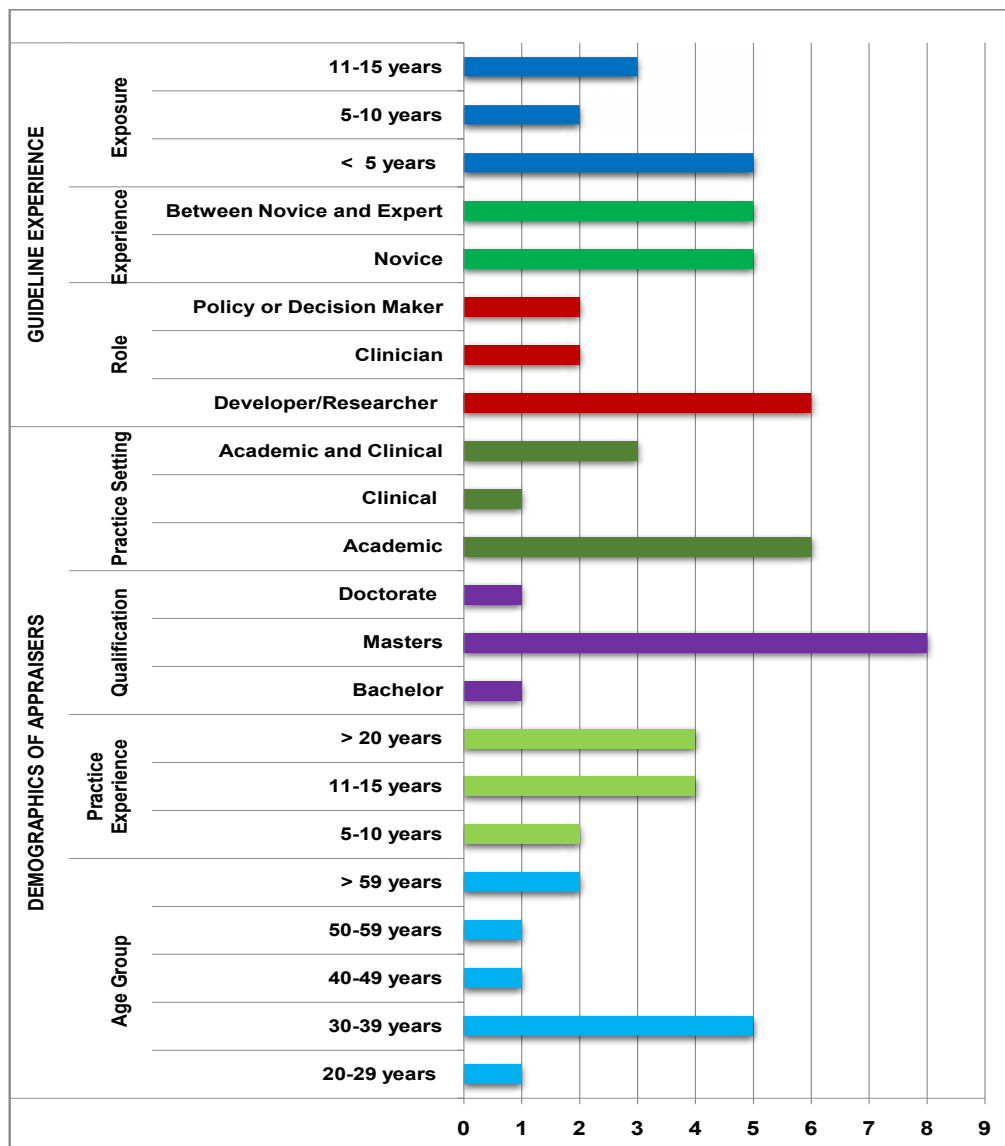


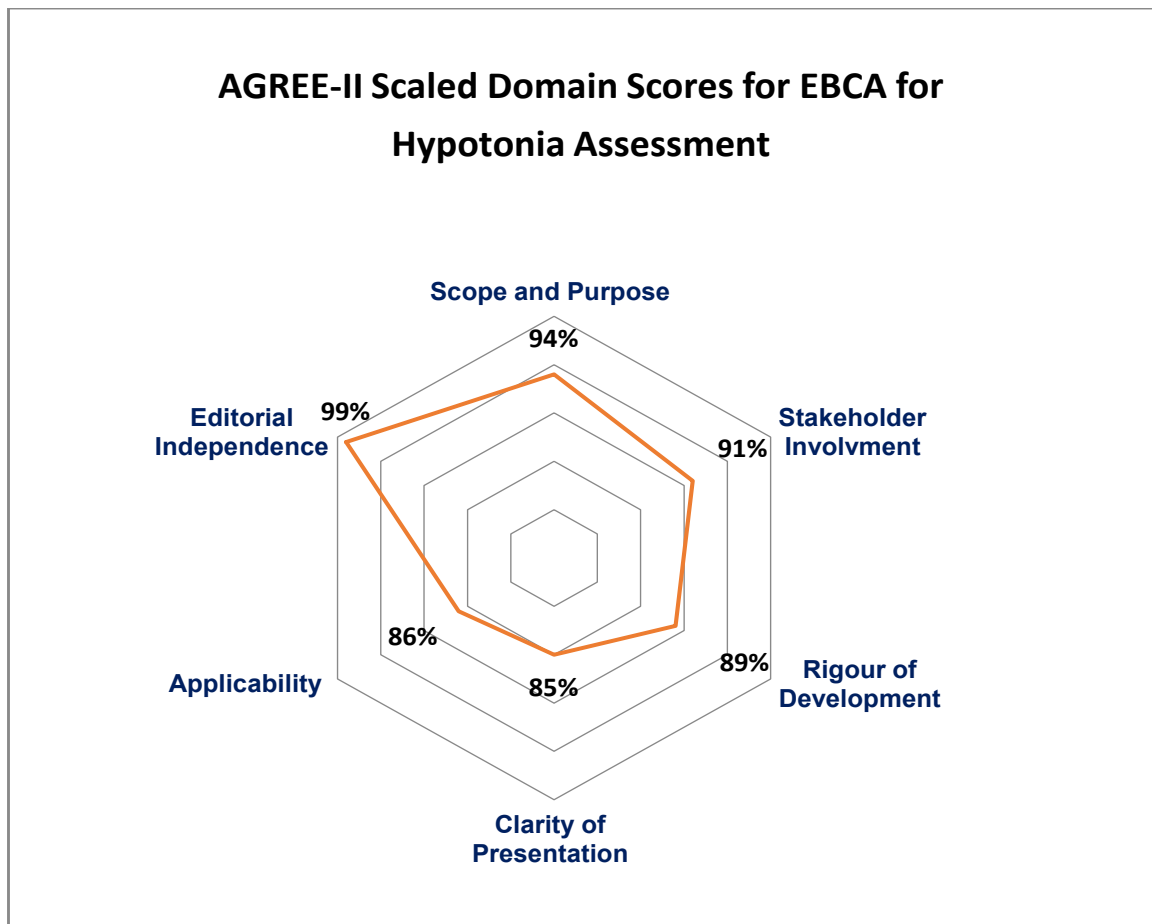
Figure 1: Demographics of Appraisers in this Study (n=10)

Of the ten appraisers that completed the AGREE-II instrument assessment, nine of them indicated that they would recommend the EBCA for use (without modification) with one appraiser indicating that they would

recommend the EBCA following modifications. Overall assessment of the EBCA revealed a score of 91%. The mean domain scores and ranges for all appraisers are reflected in Table 1 with domain percentage scores that are also graphically highlighted in Figure 2.

**Table 1:** The AGREE-II domains against which the EBCA was assessed (mean scores, range and standard deviations)

Domain	Domain Description	No. of Items (Max Score)	Domain Score Mean (Range)	AGREE-II Scaled Domain Score (%)
1	Scope and purpose is concerned with the overall aim of the guideline, the specific health questions and the target populations	3 (21)	20 (17, 21)	94%
2	Stakeholder involvement focuses on the extent to which the guideline was developed by the appropriate stakeholders and represents the views of its intended users	3 (21)	19 (15,21)	91%
3	Rigour of development relates to the process used to gather and synthesise the evidence, as well as the methods to formulate the recommendations and update them	8 (56)	51 (45,56)	89%
4	Clarity of Presentation deals with the language, structure and format of the guideline	3 (21)	18 (11,21)	85%
5	Applicability pertains to the likely barriers and facilitators to implementation, strategies to improve uptake and resource implications of applying the guidelines	4 (28)	22 (15,28)	86%
6	Editorial independence is concerned with the formulation of recommendations not being unduly biased with competing interests	2 (14)	14 (13,14)	99%



**Figure 2:** AGREE-II Scaled Domain Scores for the EBCA

The mean scores for each of the 23 items are reflected in Table 2.

Qualitative comments by appraisers indicated that there was the evidence of systematic methods and a clearly described Delphi process. There was also the open invitation to adapt the EBCA as new evidence emerges. Some appraisers felt that an expert review has been done, but may change as new evidence emerges. Others indicated that barriers and strengths including facilitators were clear. Moreover, recommendations for review as new evi-

dence becomes available had been made with evidence of changes from the original prototype presented.

However, from an interrogation of individual item scores (Table 2), and comments from the appraisers, there is the need for greater discussion on the options for management (mean 5.6), resource implications (mean 4.8) and monitoring and auditing data (mean 4.3) which reflected mean scores lower than the other 20 items and standard deviations ranging from 1.9 to 2.9 on these seven scale items.

**Table 2:** Mean scores of individual AGREE-II items against which the EBCA was assessed

Domain	Description of Domain Items	Overall rating of each item Scores could range from 1 to 7
		Item Mean (SD)
<b>Scope and Purpose</b>	The overall objective(s) of the guideline is (are) specifically described	6.7 (0.48)
	The health question(s) covered by the guideline is (are) specifically described	6.4 (0.97)
	The population (patients) to which the guideline is meant to apply is specifically described	6.9 (0.32)
<b>Stakeholder Involvement</b>	The guideline development group includes professionals from all professional groups	6.6 (0.70)
	The views and preferences of the target population have been sought	6.2 (1.14)
	The target users of the guideline are clearly defined	6.6 (0.97)
<b>Rigour of Development</b>	Systematic Methods were used to search for evidence	7.0 (0)
	The criteria for selecting the evidence are clearly described	6.5 (0.85)
	The strengths/limitations of the body of evidence are clearly described	6.2 (1.23)
	The methods for formulating the recommendations are clearly described	5.9 (0.99)
	The health benefits and risks have been considered in formulating the recommendations	6.0 (1.05)
	There is an explicit link between the recommendations and the supporting evidence	6.6 (0.70)
	The guideline has been externally reviewed by experts prior to its publication	6.4 (0.70)
A procedure for updating the guideline is provided	6.3 (1.06)	
<b>Clarity of Presentation</b>	The recommendations are specific and unambiguous	6.0 (0.94)
	The different options for management of the condition or health issue are clearly presented	5.6 (1.90)
	Key recommendations are identifiable	6.2 (0.79)
<b>Applicability</b>	The guideline describes facilitators and barriers to its application	6.5 (0.71)
	The guideline provides advice and/or tools on how recommendations can be put into practice	6.4 (0.84)
	The potential resource implications of applying the recommendations have been considered	4.8 (2.02)
	The guideline presents the monitoring or auditing criteria	4.3 (2.91)
<b>Editorial Independence</b>	The views of the funding body have not influences the content of the guideline	6.9 (0.32)
	Competing Interests have been recorded and addressed	7.0 (0)

Table 3 reports the ICC. The ICC gives a composite of intra-observer and inter-observer variability 14 thus two coefficients with their respective 95% confidence interval are indicated. The ICC will be high when there is little

variation between the scores given to each item by the appraisers. A strong agreement(0.714) was present across all appraisers in this study (mean average reflected in Table 3)

**Table 3** Intraclass Correlation Coefficient for Inter-rater reliability

	Intraclass Correlation <sup>b</sup>	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.200 <sup>a</sup>	.091	.381	3.493	22	198	.000
Average Measures	.714	.501	.860	3.493	22	198	.000

Two-way random effects model where both people effects and measures effects are random.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type C intraclass correlation coefficients using a consistency definition. The between-measure variance is excluded from the denominator variance.

Comments on the overall EBCA and accompanying technical report included the following:

*“User friendly, although appears overwhelming initially, upon reading is logical and provides good guidelines for both the inexperienced and experienced clinician.”* (Appraiser 1)

*“This is a very interesting and useful guideline that could be used for a number of other evaluations. User friendly, but a lot to take in at the beginning but great once you get used to the process”* (Appraiser 2)

*“The guideline is clear and ready for use however improvement may be needed on instructions to use the algorithms”.* (Appraiser 3)

*“The tool will certainly add value to clinical practice”* (Appraiser 5)

*“This is an extremely well developed algorithm with a thorough evidence-based explanation of its development and evolution. The one page “flow diagram” is certainly user-friendly and ready for use”.* (Appraiser 6)

*“The researcher succinctly guides the review describing and summarising each domain. It is easy to apply the AGREE-II in assessing her algorithm. Due to the paucity in this field I believe there will be immense value in the implementation”* (Appraiser 7)

*“An easy to read, easy to follow, comprehensive guideline that has taken all key areas into consideration”* (Appraiser 8)

## Discussion

In this study, the AGREE-II instrument was used to evaluate the quality of an evidenced-based clinical algorithm (EBCA) for the assessment of hypotonia. The overall ratings of the appraisers in this study generally indicate a high quality guideline. Together with the domain scores, the qualitative comments provided by the appraisers serve to assist the author in ensuring that the EBCA and accompanying technical report is ready for use in the clinical setting. Amid the six domains evaluated, the rated strengths of the EBCA appear to lie in all six domains with higher scores in domains of scope and purpose, stakeholder involvement and editorial independence. This thus reflects the transparency with which the guideline had been developed in addition to involvement of relevant stakeholders and the delineation of the scope and purpose.

Given that the AGREE II Consortium 12 has not set minimum domain scores or patterns of scores across domains to differentiate between high quality and poor quality guidelines, the scores of the AGREE-II evaluation require careful interpretation. Thus in order to best represent the findings of this evaluation, the author included item and domain mean scores and standard deviations in order to demonstrate where the strengths lay, as well as items that showed variability, in addition to calculating inter-rater reliability. Authors who have used the AGREE-II tool have reported findings using means, medians, standard deviations for domain and item scores and weighted kappa, ICC and cronbachs alpha as reli-

ability measures<sup>15,16,17,18</sup>. Brouwers et al<sup>19,20</sup> in their studies recommend that the number of appraisers required in reaching a level of inter-rater reliability of 0.7 ranged from two to five appraisers across domains. In this study ten appraisers were sampled in order to increase the reliability and validity of the findings.

This is the first independent assessment of the methodological rigour and transparency of a clinical algorithm as opposed to a clinical practice guideline using the AGREE-II instrument. The majority of appraisers have recommended the EBCA for clinical use, however there has been three items that are worth interrogating in order to improve the overall applicability and stakeholder uptake of the EBCA. Firstly, the technical report and algorithm is intended mainly for assessment and includes outputs for referrals for management. The appraisers who scored the item for options of management lower commented on the need for the assessment algorithm to include more detail on management. Given that the underlying condition may be across a spectrum of diagnosis, specific management would not be possible within this particular EBCA. A secondary process following this initial assessment would be required in order to first establish the diagnosis and then further management plans.

These issues raised by the appraisers however, highlights the need for the delineation and boundaries of this EBCA to be made explicit in the technical report to avoid confusion of what is offered. Secondly, whilst the implications of applying recommendations have been considered and has been previously reported<sup>3</sup> there was variability in the appraisers evaluations of this item although details were provided in the technical report. Surprisingly, there were no comments on this item by the appraisers. This may be partially due to some of the appraisers being convinced that this has been addressed with others differing with no recommendations for improvement. The author will thus consider resource implications as a point of discussion prior to stakeholder uptake given that there is variability as to whether this has been adequately addressed. Lastly, monitoring and auditing criteria were excluded from the technical report, but some appraiser evaluated this aspect. Comments include the fact that the annexures and evidence tables were difficult to follow whilst others indicated that the changes made were described explicitly, based on stakeholder feedback, as well as by the inclu-

sion of criteria that characterises hypotonia in the EBCA. Notwithstanding these comments and score, the author identifies the need for the inclusion of operational definitions of how each of the criteria should be measured (although this is part of the basic training of the professionals used in this study) and inferred in the EBCA. This may aid in greater uptake of the EBCA in everyday clinical practice.

## **Conclusion**

Users of clinical algorithms, care pathways and clinical practice guidelines and associated documents need to be assured that they are evidence-based. The appraisals in this study suggest that the process of guideline development and quality of reporting for an evidenced-based clinical algorithm are robust. However, successful implementation will have to be carefully considered. Whilst the AGREE-II provided initial feedback on the methodological rigour of development, understanding that the AGREE-II instrument evaluates the guideline development process and not the content is also essential in order to consider the next stage which would be to consider clinicians feedback on the clinical utility of this EBCA. Together with evaluating clinical utility and applicability, the author may augment this evaluation process with the administration of the AGREE II-GRS Instrument (5-item)<sup>12</sup> to assess how well the guideline is reported by the actual end-users of this EBCA and technical report. This may be appropriate especially when time and resources are limited for busy clinicians. The evaluation process reported in this paper can thus be seen as the first step towards determining areas of strength and potential areas for improvement that will ultimately aid in revisions and adjustments for maximal clinical benefit. As part of the implementation and knowledge exchange responsibilities, the EBCA is in the process of publication<sup>5</sup>. Whilst this may be considered effective from a clinical perspective, it does not take into account factors that may affect the adoption of the EBCA by the end-users, and as such these processes mapping the way forward may seek to address these challenges identified.

## **Acknowledgements**

The author would like to acknowledge Prof RWE Joubert for the supervision of this study; Mrs D Mahlangu for research assistance; and Mr S Govender for statistical support.

## **Funding**



This work was supported by the Medical Research Council of South Africa in terms of the National Health Scholars Programme from funds provided for this purpose by the National Department of Health and the National Public Health Enhancement Fund; the National Research Foundation of South Africa via the Thuthuka Programme grant number TTK13070720688; the University of KwaZulu-Natal; and the Office of the U.S. Global AIDS Coordinator and the U.S. Department of Health and Human Services, National Institutes of Health (NIH OAR and NIH ORWH) Grant number R24TW008863. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the government.

## References

- Hartley L, Ranjan R. Evaluation of the floppy infant. *Paediatrics and Child Health*. 2015; 25:498-504.
- Lisi EC, Cohn RD. Genetic evaluation of the pediatric patient with hypotonia: perspective from a hypotonia specialty clinic and review of the literature. *Developmental Medicine & Child Neurology*. 2011;53:586-99 PubMed.
- Soucy EA, Wessel LE, Gao F, Albers AC, Gutmann DH, Dunn CM. A Pilot Study for Evaluation of Hypotonia in Children With Neurofibromatosis Type 1. *Journal of Child Neurology*. 2015;30:382-5.
- Bodensteiner JB. The evaluation of the hypotonic infant. *Seminars in Pediatric Neurology* 2008;15:10-20.
- Govender P, Joubert RWE. Evidence-based clinical algorithm for hypotonia assessment: To pardon the errs. *Occupational Therapy International*. Volume 2018. Article ID 8967572, 7 pages. <https://doi.org/10.1155/2018/8967572>
- Miller VS, Delgado M, Iannaccone ST. Neonatal hypotonia. *Seminars in Neurology*. 1993; 13: 73-83.
- Sox HC, Stewart WF. Algorithms, clinical practice guidelines, and standardized clinical assessment and management plans: Evidence-based patient management standards in evolution. *Academic Medicine*. 2015;90:129-32 PubMed .
- Farias M, Friedman KG, Lock JE, Rathod RH. Gathering and learning from relevant clinical data: A new framework. *Academic Medicine*. 2015;90:143-8 PubMed .
- Gaddis GM, Greenwald P, Huckson S. Toward improved implementation of evidence-based clinical algorithms: clinical practice guidelines, clinical decision rules, and clinical pathways. *Acad Emerg Med*. 2007;14:1015-22

PubMed .

- Lavelle J, Schast A, Keren R. Standardizing Care Processes and Improving Quality Using Pathways and Continuous Quality Improvement. *Current Treatment Options in Pediatrics*. 2015;1:347-58.
- Hollon D, Miller IJ, Robinson E. Criteria for evaluating treatment guidelines. *American Psychologist*. 2002;57(12):1052-9.
- AGREE Next Steps Consortium. The AGREE II instrument Electronic version. Retrieved March. 2009;21:2012.
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*. 1979;86:420.
- McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological methods*. 1996;1:30.
- White PE, Shee AW, Finch CF. Independent appraiser assessment of the quality, methodological rigour and transparency of the development of the 2008 International consensus statement on concussion in sport. *British Journal of Sports Medicine*. 2014; 48:130-4.
- Tudor KI, Kozina PN, Marušić A. Methodological rigour and transparency of clinical practice guidelines developed by neurology professional societies in Croatia. *PloS one*. 2013;19;8: PubMed e69877:1-5.
- Don-Wauchope AC, Sevenpiper JL, Hill SA, Iorio A. Applicability of the AGREE II Instrument in Evaluating the Development Process and Quality of Current National Academy of Clinical Biochemistry Guidelines. *Clinical Chemistry*. 2012;58:1-13 PubMed .
- Smith CA, Toupin-April K, Jutai JW, Duffy CM, Rahman P, Cavallo S, Brosseau L. A Systematic Critical Appraisal of Clinical Practice Guidelines in Juvenile Idiopathic Arthritis Using the Appraisal of Guidelines for Research and Evaluation II (AGREE II) Instrument. *PloS one*. 2015;10:e0137180:1-22.
- Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, Fervers B, Graham ID, Hanna SE, Makarski J. Development of the AGREE II, part 1: performance, usefulness and areas for improvement. *Canadian Medical Association Journal*. 2010;182:1045-52.
- Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, Fervers B, Graham ID, Hanna SE, Makarski J. Development of the AGREE II, part 2: assessment of validity of items and tools to support application. *Canadian Medical Association Journal*. 2010; 182:E472-8.