

Critical Appraisal of the National Institute for Health and Care Excellence (NICE) Guidelines for Spine Disorders using the Appraisal of Guidelines for Research and Evaluation II Instrument (AGREE II)

Ning Liang^{1,2*}, Sizhan Wu^{2#}, Simon Roberts³, Navnit Makaram⁴, James Reeves Mbori Ngwayi¹ and Daniel Porter^{1,2}

¹Department of Clinical Medicine, Tsinghua University, China

²Department of Orthopedic Surgery, First Hospital of Tsinghua University, China

³ECAT Department, University of Edinburgh, United Kingdom

⁴Department of Trauma and Orthopaedics, Royal Infirmary of Edinburgh, United Kingdom

Abstract

Introduction: Disorders of the spine (as defined by the musculoskeletal structures surrounding the spinal neural elements) require evidence based, approach to their care. This evaluation used the Appraisal of Guidelines for Research and Evaluation (AGREE) II instrument to evaluate the methodological quality of evidence based guidelines on spine disorders published by The National Institute for Health and Care Excellence (NICE).

Materials and methods: We systematically searched clinical guidelines on spine disorders published by NICE until December 2019. Four appraisers across three international centers independently evaluated the quality of eligible clinical guidelines using AGREE II. Mean AGREE II scores for each domain were calculated. In higher quality domains scores for individual items were analysed. The guidelines were grouped according to type and year of publication. Comparative statistics and intraclass correlation (ICC) calculations were performed.

Results: A total of 37 guidelines published by NICE on spine disorders were identified. Mean scores for all six domains were as follows: Scope and Purpose (73.2%), Stakeholder Involvement (63.9%), Rigour of Development (68.1%), Clarity of Presentation (73.6%), Applicability (53.2%) and Editorial Independence (64.5%). The mean score for overall quality of all NICE spinal related guidelines was 68.8% (95% CI: 62.3-75.3). Interventional Procedure Guidelines were evaluated as possessing significantly lower overall quality than other types ($p=0.007$). Overall quality was significantly associated with year of publication ($r_s=0.476$, $p=0.0029$). Evaluator ICC for each guideline ranged from 0.39 to 0.95.

Conclusion: NICE guidelines on spine disorders demonstrated acceptable or good quality across most domains. Despite deficiencies in the applicability domain, their quality has improved over time. We recommend use of NICE guidelines for assessment and treatment of spine disorders.

Keywords: AGREE II • Clinical Practice Guideline • NICE • Spine

Background

The World Health Organization reported in 2021 that musculoskeletal diseases are the main cause of global disability with approximately 1.71 billion people affected worldwide [1]. In 2019 the prevalence of low back pain was estimated as 568 million and neck pain 223 million [2]. An aging population will result in increasing numbers of people with these and other common spine disorders [3].

Clinical practice guidelines can provide health care providers with decision making recommendations from an evidence base [4]. Around the world, many institutions, organisations or groups have formulated and issued clinical practice guidelines. The National Institute for Health and Care

Excellence (NICE), was established as a special health authority in 1999 and a non-departmental public body in 2013, performing statutory functions in the United Kingdom with the aim of 'improving health and wellbeing by putting science and evidence at the heart of health and care decision making' [5]. To date, more than 1750 different guidelines have been formulated under several headings, with spine disorders contributing significantly to the burden of disease, NICE guidance in this clinical area encompasses several guideline types, including: Clinical Guidelines (abbreviation CG) which were succeeded by NICE guidelines (NG) in 2015 these guidelines review the evidence across broad health care topics; Interventional Procedure Guidance (IPG) which review the efficacy and safety of procedures; Technology Appraisal Guidance (TG) which review clinical and cost effectiveness of new treatments; Medical Technology Guidance (MTG) which review new medical technologies for adoption in the UK National Health Service for multiple clinical conditions [5]. The effectiveness of a clinical practice guideline is dependent on its inherent quality. During guideline development the World Health Organization (WHO) recommended that "Prior to submission for clearance, the AGREE-II appraisal instrument should be used to check whether the guideline meets international quality standards and reporting criteria" [6]. In 2010 AGREE II was developed and formulated by an international development and research team, based on quantitative scores to evaluate the quality of the guideline [7]. Since then NICE has undertaken internal audit to ensure that the processes and methods for guideline development are based on internationally accepted criteria of quality, as detailed in the AGREE II instrument [8]. However internal quality assurance processes may not reflect the evaluation of external auditors. Furthermore, many guidelines

Address for Correspondence: Ning Liang, Department of Clinical Medicine, Tsinghua University, China, Email: liangning1984@hotmail.com

Note: #Co-first authorship

Copyright: ©2022 Liang Ning et al. This is an open-access article distributed under the terms of the creative commons attribution license which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited

Received: 01-July-2022, Manuscript No. jnd-22-69502; **Editor assigned:** 04-July-2022, PreQC No. P-69502 (PQ); **Reviewed:** 18-July-2022; QC No. Q-69502; **Revised:** 25-July-2022, Manuscript No. R-69502; **Published:** 01-August-2022, DOI: 10.4172/2329-6895.10.7.504

were developed prior to the adoption of these standards use this style when you need to begin a new paragraph.

The AGREE II tool has been used independently to evaluate several NICE guidelines, involving urological and endocrine system disorders [9-12], but there are currently no studies using the AGREE II tool to evaluate NICE guidelines for musculoskeletal or neurosurgical conditions. The purpose of this study is to use the AGREE II tool to assess quality of NICE guidelines for spine disorders.

Material and Methods

Search keyword methodology

Our goal was to identify guidelines related to spine disorders, defined as disorders of the musculoskeletal structures surrounding the spinal neural elements. Two researchers independently performed keyword searches for spine disorders on the official website of International Classification of Diseases 11th Revision (ICD-11) (<https://icd.who.int/browse11/l-m/en>). Search terms identified via dictionary linkage are shown in Electronic Supplementary Material S1. Search keywords retrieval process followed the PRISMA flow algorithm [13]. Inclusion and exclusion criteria for categories of disorders searched in the official website of ICD-11 are shown in Electronic Supplementary Material S2 and S3 respectively. In our study, a total of 28 search keywords for spine disorders were formed from ICD-11 and are shown in Electronic Supplementary Material S4. Use this for the first paragraph in a section, or to continue after an extract.

Guideline identification

The search for spine guidelines was carried out by two researchers (first author and second author) independently using the NICE website (<https://www.nice.org.uk/>) up to 31 December 2019. The guideline search used

keywords obtained in the aforementioned process. Both keyword and manual searches were performed.

Inclusion criteria were:

1. Literature related to spine disorders
2. Literature meets the guidelines standard of the National Guidelines Clearinghouse [14].

Exclusion criteria were:

1. Disorders secondary or metastatic without causing spinal cord compression
2. Systemic diseases including sites other than the spine
3. Physiological or pathological abnormalities of spinal cord, spinal neural structures, and vertebral vascular conditions caused by disorders not primarily of the spine. The guidelines retrieval process followed the PRISMA flow algorithm [13].

AGREE II instrument

The AGREE II assessment system is an internationally validated tool for assessing guideline quality, including 23 main items in 6 domains and 2 overall assessment items (Table 1). Each domain addresses an aspect of guideline quality, namely: "scope and purpose", "stakeholder involvement", "rigour of development", "clarity of presentation", "applicability", and "editorial independence". The 23 field items and one of the overall assessment items are graded on a seven point Likert scale. Item scores range from 1 (no information or very poor quality) to 7 (all conditions met and of excellent quality) [7].

Table 1. NICE guidelines on spine disorders.

Title of Guidelines	Published Year	Reference Number
Automated percutaneous mechanical lumbar discectomy	2005	IPG141
Balloon kyphoplasty for vertebral compression fractures	2006	IPG166
Direct C1 lateral mass screw for cervical spine stabilisation	2005	IPG146
Endoscopic laser foraminoplasty	2003	IPG31
Epiduroscopic lumbar discectomy through the sacral hiatus for sciatica	2016	IPG570
Functional electrical stimulation for drop foot of central neurological origin	2009	IPG278
Golimumab for treating non-radiographic axial spondyloarthritis	2018	TA497
iFuse for treating chronic sacroiliac joint pain	2018	MTG39
Insertion of an annular disc implant at lumbar discectomy	2014	IPG506
Interspinous distraction procedures for lumbar spinal stenosis causing neurogenic claudication	2010	IPG365
Intramuscular diaphragm stimulation for ventilator-dependent chronic respiratory failure caused by high spinal cord injuries	2017	IPG594
Lateral interbody fusion in the lumbar spine for low back pain	2017	IPG574
Low back pain and sciatica in over 16s: assessment and management	2016	NG59
Metastatic spinal cord compression in adults: risk assessment, diagnosis and management	2008	CG75
Minimally invasive sacroiliac joint fusion surgery for chronic sacroiliac pain	2017	IPG578
Nerve transfer to partially restore upper limb function in tetraplegia	2018	IPG610
Non-rigid stabilisation techniques for the treatment of low back pain	2010	IPG366
Percutaneous coblation of the intervertebral disc for low back pain and sciatica	2016	IPG543
Percutaneous electrothermal treatment of the intervertebral disc annulus for low back pain and sciatica	2016	IPG544
Percutaneous endoscopic laser cervical discectomy	2009	IPG303
Percutaneous endoscopic laser thoracic discectomy	2004	IPG61
Percutaneous insertion of craniocaudal expandable implants for vertebral compression fracture	2016	IPG568

Percutaneous interlaminar endoscopic lumbar discectomy for sciatica	2016	IPG555
Percutaneous intradiscal laser ablation in the lumbar spine	2010	IPG357
Percutaneous intradiscal radiofrequency treatment of the intervertebral disc nucleus for low back pain	2016	IPG545
Percutaneous transforaminal endoscopic lumbar discectomy for sciatica	2016	IPG556
Percutaneous vertebroplasty	2003	IPG12
Percutaneous vertebroplasty and percutaneous balloon kyphoplasty for treating osteoporotic vertebral compression fractures	2013	TA279
Peripheral nerve-field stimulation for chronic low back pain	2013	IPG451
Prosthetic intervertebral disc replacement in the cervical spine	2010	IPG341
Prosthetic intervertebral disc replacement in the lumbar spine	2009	IPG306
Spinal injury: assessment and initial management	2016	NG41
Spondyloarthritis in over 16s: diagnosis and management	2017	NG65
Therapeutic endoscopic division of epidural adhesions	2010	IPG333
Therapeutic percutaneous image-guided aspiration of spinal cysts	2007	IPG223
TNF-alpha inhibitors for ankylosing spondylitis and non-radiographic axial spondyloarthritis	2016	TA383
Transaxial interbody lumbosacral fusion for severe chronic low back pain	2018	IPG620

Ethics statement

This study is an evaluation of existing literature without human subjects; hence it is not subject to ethics committee evaluation.

Evaluation of the guidelines

Each guideline was assessed by a panel of four appraisers. All appraisers were familiar with the AGREE II instrument having used it before to evaluate clinical guidelines and completed the online training tools recommended by AGREE II (www.agreetrust.org) [7]. No communication between appraisers occurred during the rating process. Data analysis was performed after completing the evaluation of all NICE spine related guidelines.

Statistical analysis

The score for each domain was obtained by the sum of all scores of the individual items in a domain and then standardized as follows: (obtained score - minimum possible score)/(maximum possible score - minimum possible score) [7]. Mean values and 95% confidence intervals (CI) for all raters were calculated. Although domain scores can be used to compare different guidelines and to help determine whether guidelines should be recommended, the AGREE II tool does not set a minimum domain score, nor does it define the boundary criteria for identifying the quality of the guidelines. These decisions are made by the user. According to convention in existing research reports, the domain score criteria we used were: <40% very low quality, 40%–59% low quality, 60%–79% acceptable quality, ≥ 80% good quality [15,16].

Scores obtained for individual items within domains were calculated using the same method as for domain scores. Since the purpose of the AGREE scale is to emphasise and encourage best practice, we took the view that

domains which failed to reach 'acceptable' quality threshold should not be the focus of detailed critique. Consequently, although all domain scores are presented, individual item scores are only displayed and discussed for domains which reach at least 'acceptable' quality.

For overall guideline assessment, appraiser scores for item 1 of the 'overall assessment' section were used to derive a score for each guideline by the same method as for domain scores.

Statistical analysis of the data was performed using Statistical Package for the Social Science (SPSS Inc, Chicago, Illinois, USA) version 22.0 and Stata (Version 16.1, StataCorp LP, College Station, TX) software programs. Correlation between overall score and guideline publication date was calculated using Spearman's test. Inter rater reliability of domain scores was assessed using intra-class correlation coefficient (ICC). ICC values less than 0.40, between 0.40 and 0.59, between 0.60 and 0.74, and greater than 0.75 were indicative of poor, moderate, good, and excellent reliability, respectively [17]. Mann-Whitney U test was used to investigate the quality differences between guideline types. The level of statistical significance was set at $p < 0.05$.

Results

A total of 37 guidelines fulfilled the inclusion criteria, including 29 IPG, 1 MTG, 4 NG or CG and 3 TA (Table 1). The scores of each domain and overall scores in the 37 guidelines after evaluation by AGREE II criteria are shown in Table 2. All four CG/NG guidelines were categorized as having 'acceptable' quality in all domains. Seven (7/29) IPG guidelines and one (1/3) TA guideline were categorized as having 'acceptable' quality in all domains. The overall scores in 26 NICE spine related guidelines were above the "acceptable" level. The mean overall score of all NICE spine related guidelines was 68.8% (95% CI: 62.3%–75.3%).

Table 2. AGREE II domain scores and overall score for each NICE spine disorder guideline.

Number of Guidelines	Domain 1 Scope and purpose (%)	Domain 2 Stakeholder involvement (%)	Domain 3 Rigour of development (%)	is Domain 4 Clarity of presentation (%)	Domain 5 Applicability (%)	Domain 6 Editorial independence (%)	Overall score (%)
IPG 31	47.2	55.6	57.8	50.0	36.5	52.1	41.7
IPG 141	54.2	54.2	60.9	65.3	30.2	47.9	62.5
IPG 146	55.6	54.2	60.4	55.6	41.7	52.1	58.3
IPG 166	65.3	54.2	60.4	62.5	44.8	50.0	58.3
IPG 278	66.7	51.4	58.9	55.6	40.6	47.9	54.2
IPG 365	61.1	48.6	60.4	51.4	52.1	54.2	54.2

IPG 506	62.5	47.2	61.5	55.6	28.1	52.1	54.2
IPG 570	56.9	52.8	60.9	56.9	40.6	56.3	54.2
IPG 574	59.7	45.8	63.0	51.4	44.8	56.3	54.2
IPG 594	65.3	45.8	62.0	54.2	34.4	56.3	58.3
MTG 39	70.8	59.7	45.8	69.4	47.9	62.5	66.7
NG 59	83.3	76.4	64.6	84.7	67.7	58.3	79.2
TA 497	77.8	59.7	45.3	75.0	63.5	52.1	62.5
CG 75	93.1	88.9	86.4	94.4	82.3	62.5	87.5
IPG 578	79.2	76.4	77.1	86.1	60.4	77.1	79.2
IPG 610	77.8	79.2	76.0	81.9	63.5	77.1	79.2
IPG 366	79.2	75.0	78.1	83.3	50.0	72.9	70.8
IPG 543	81.9	75.0	77.1	84.7	61.5	70.8	79.2
IPG 544	72.2	70.8	76.0	80.6	57.3	70.8	75.0
IPG 303	75.0	73.6	78.6	84.7	64.6	72.9	79.2
IPG 61	75.0	66.7	73.4	83.3	58.3	77.1	75.0
IPG 568	79.2	73.6	78.1	77.8	59.4	85.4	79.2
IPG 555	80.6	76.4	77.6	84.7	61.5	85.4	79.2
IPG 357	83.3	73.6	80.2	77.8	59.4	66.7	75.0
IPG 545	81.9	72.2	81.3	81.9	60.4	79.2	79.2
IPG 556	76.4	79.2	81.3	80.6	60.4	85.4	79.2
IPG 12	61.1	50.0	54.7	77.8	42.7	47.9	58.3
TA 279	86.1	65.3	62.0	76.4	79.2	81.3	83.3
IPG 451	81.9	68.1	74.5	81.9	59.4	79.2	70.8
IPG 341	65.3	54.2	67.2	70.8	41.7	50.0	66.7
IPG 306	69.4	65.3	66.7	70.8	43.8	58.3	66.7
NG 41	87.5	61.1	72.4	86.1	60.4	68.8	75.0
NG 65	94.4	79.2	87.5	90.3	75.0	72.9	83.3
IPG 333	77.8	52.8	66.1	75.0	39.6	58.3	62.5
IPG 223	69.4	50.0	61.5	70.8	36.5	54.2	54.2
TA 383	86.1	81.9	59.9	83.3	77.1	75.0	87.5
IPG 620	69.4	48.6	64.1	69.4	41.7	60.4	62.5
Mean	73.2	63.9	68.1	73.6	53.2	64.5	68.8
(95%CI)	(67.0 ~ 79.5)	(57.0 ~ 70.7)	(62.3 ~ 73.9)	(66.6 ~ 80.5)	(45.4 ~ 61.0)	(57.7 ~ 71.3)	(62.3 ~ 75.3)

Mean domain and item scores are shown in Table 3. Mean domain score is highest for 'Clarity of Presentation' (73.6%), followed by 'Scope and Purpose' (73.2%), 'Rigour of Development' (68.1%), 'Editorial Independence' (64.5%) and 'Stakeholder Involvement' (63.9%). Mean domain score for 'Applicability' (53.2%) is the lowest.

Table 3. AGREE II instrument mean scores and 95% confidence intervals (CI) of domains and items 1-17 and 22-23 for NICE spine disorder guidelines.

Domain and Domain item	Mean(95%CI)
Scope and Purpose	73.2(67.0 ~ 79.5)
1. The overall objective(s) of the guideline is (are) specifically described.	76.4(73.1 ~ 79.6)
2. The health question(s) covered by the guideline is (are) specifically described.	73.9(70.1 ~ 77.6)
3. The population (patients, public, etc.) to whom the guideline is meant to apply is specifically described.	69.5(65.2 ~ 73.8)
Stakeholder involvement	63.9(57.0 ~ 70.7)
4. The guideline development group includes individuals from all the relevant professional groups.	70.4(66.9 ~ 73.9)
5. The views and preferences of the target population (patients, public, etc.) have been sought.	59.0(53.8 ~ 64.2)
6. The target users of the guideline are clearly defined.	62.2(57.7 ~ 66.6)
Rigour of development	68.1(62.3 ~ 73.9)
7. Systematic methods were used to search for evidence.	69.9(65.3 ~ 74.6)
8. The criteria for selecting the evidence are clearly described.	74.9(71.1 ~ 78.7)
9. The strengths and limitations of the body of evidence are clearly described.	74.5(70.6 ~ 78.5)
10. The methods for formulating the recommendations are clearly described.	62.8(59.8 ~ 65.8)

11. The health benefits, side effects, and risks have been considered in formulating the recommendations.	77.8(75.0 ~ 80.6)
12. There is an explicit link between the recommendations and the supporting evidence.	67.2(63.6 ~ 70.8)
13. The guideline has been externally reviewed by experts prior to its publication.	56.6(52.9 ~ 60.4)
14. A procedure for updating the guideline is provided.	60.7(54.8 ~ 66.6)
Clarity of presentation	73.6(66.6 ~ 80.5)
15. The recommendations are specific and unambiguous.	76.5(72.4 ~ 80.6)
16. The different options for management of the condition or health issue are clearly presented.	61.3(56.5 ~ 66.0)
17. Key recommendations are easily identifiable.	83.0(79.1 ~ 86.9)
Applicability	53.2(45.4 ~ 61.0)
18. The guideline describes facilitators and barriers to its application.	
19. The guideline provides advice and/or tools on how the recommendations can be put into practice.	
20. The potential resource implications of applying the recommendations have been considered.	
21. The guideline presents monitoring and/or auditing criteria.	
Editorial independence	64.5(57.7 ~ 71.3)
22. The view of the funding body have not influenced the content of the guideline.	62.5(58.9 ~ 66.1)
23. Competing interests of guideline development group members have been recorded and addressed.	66.4(61.9 ~ 71.0)

Five domains exceeded the threshold for 'acceptable' quality (mean score of 60%): 'Scope and Purpose', 'Stakeholder involvement', 'Rigour of development', 'Clarity of presentation', and 'Editorial Independence'. The quality evaluation of two items in these five domains fell below this 60% threshold for acceptability; item 5 'the views and preferences of the target population (patients, public, etc.) have been sought' and item 13 'the guideline has been externally reviewed by experts prior to its publication' (Table 3) (Figure 1).

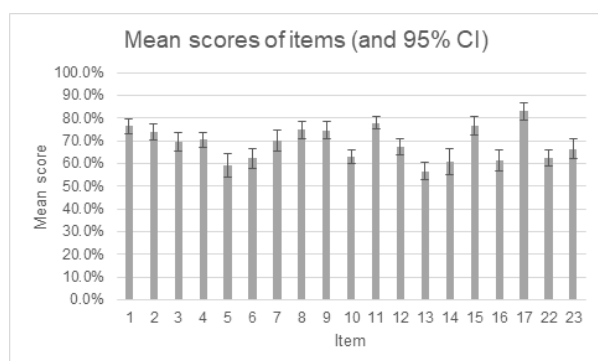


Figure 1. A 34 year old male patient before surgical intervention (own material) Clinical presentation with a mass-like lesion caused swelling and deformity in the right cheek and buccal region.

Mean domain score for IPG and non-IPG documents are 64.1 (95% CI: 58.0~70.1) and 73.4 (95% CI: 67.3~79.4) respectively ($z=-2.085$, $p=0.037$) (Table 4). Significant differences in domain scores for domains 1, 2, 4 and 5 were also found (Table 4). Non-IPGs also have higher overall scores (78.1) than IPGs (66.2) ($z=-2.687$, $p=0.007$) (Table 4) (Figure 2).

Table 4. Individual domain scores mean and overall scores in Interventional Procedure Guidance (IPG) and non-IPG documents.

	IPG	Non-IPG	p- value
Domain 1	70.0 (66.5 ~ 73.6)	84.9 (79.9 ~ 89.9)	0.001

Domain 2	61.7 (57.5 ~ 66.0)	71.5 (64.1 ~ 79.0)	0.037
Domain 3	68.8 (65.8 ~ 71.9)	65.5 (55.1 ~ 75.9)	0.567
Domain 4	71.1 (66.7 ~ 75.5)	82.5 (77.0 ~ 87.9)	0.024
Domain 5	48.8 (44.9 ~ 52.8)	69.1 (61.7 ~ 76.6)	<0.001
Domain 6	63.9 (59.3 ~ 68.6)	66.7 (60.5 ~ 72.9)	0.448
Mean Domain Score	64.1 (58.0 ~ 70.1)	73.4 (67.3 ~ 79.4)	0.037
Overall Score	66.2 (62.3 ~ 70.2)	78.1 (72.0 ~ 84.2)	0.007

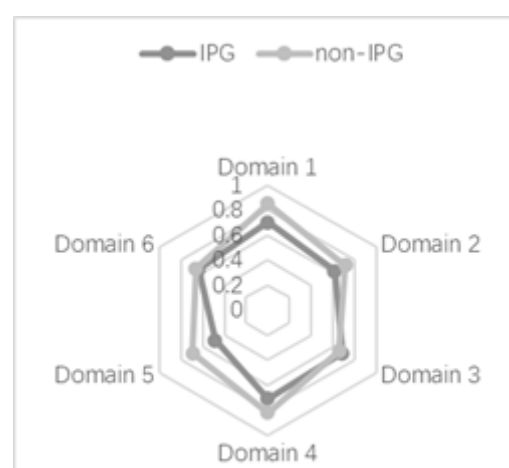


Figure 2. Radar maps comparing the average domain scores of NICE spine disorder guidelines between Interventional Procedure Guidance (IPG) and non-IPG documents.

The overall scores in 26 NICE spine related guidelines were above the "acceptable" level. The mean overall score of all NICE spine related guidelines was 68.8% (95% CI: 62.3%~75.3%) (Table 2). Intraclass correlation (ICC) values for each NICE guideline ranged from 0.393 to

0.953 (Table 5).

Table 5. Intraclass correlation (ICC) (and 95% confidence interval) for each NICE spine-disorder guideline.

	IPG31	IPG141	IPG146	IPG166	IPG278	IPG365	IPG506	IPG570	IPG574	IPG594	MTG39	NG59	TA497
ICC	0.695	0.657	0.73	0.733	0.688	0.672	0.687	0.671	0.679	0.694	0.393	0.623	0.498
-95%CI	0.291	0.260	0.306	0.313	0.260	0.244	0.270	0.245	0.237	0.272	0.054	0.220	0.128
+95%CI	0.941	0.931	0.950	0.951	0.940	0.936	0.939	0.935	0.938	0.941	0.836	0.922	0.881
	CG75	IPG578	IPG610	IPG366	IPG543	IPG544	IPG303	IPG61	IPG568	IPG555	IPG357	IPG545	IPG556
ICC	0.95	0.946	0.947	0.943	0.932	0.941	0.916	0.931	0.944	0.935	0.929	0.942	0.932
-95%CI	0.842	0.830	0.829	0.825	0.790	0.815	0.728	0.735	0.827	0.782	0.784	0.820	0.783
+95%CI	0.992	0.991	0.991	0.991	0.989	0.990	0.986	0.989	0.991	0.990	0.935	0.991	0.989
	IPG12	TA279	IPG451	IPG341	IPG306	NG41	NG65	IPG333	IPG223	TA383	IPG620		
ICC	0.795	0.761	0.953	0.731	0.712	0.639	0.892	0.761	0.706	0.709	0.71		
-95%CI	0.364	0.430	0.852	0.329	0.287	0.240	0.566	0.370	0.264	0.357	0.262		
+95%CI	0.965	0.956	0.992	0.950	0.946	0.926	0.983	0.957	0.945	0.943	0.946		

Discussion

The purpose of NICE at its inception was 'to create consistent guidelines and end rationing of treatment by postcode across the UK [18]. Ours is the first study to use the AGREE II instrument to assess the quality of NICE guidelines for spine related disorders. Up to end of 2019, we identified 37 guidelines which fulfilled our inclusion criteria as related to spine disorders. This represented approximately 3% of the total guideline cohort in the NICE library at that time point.

Reliability

Intraclass correlation (ICC) of overall guideline scores in our study ranged from 0.39 to 0.95. 35 out of 37 guideline evaluations (94.6%) were categorized as exhibiting good or excellent inter rater reliability.

Domains reaching acceptability threshold

The highest mean score was for the domain "clarity of presentation", with mean scores of a further four domains ('scope and purpose', 'rigour of development', 'editorial independence' and 'stakeholder involvement') also exceeding the 60% threshold. Evidence based recommendations in NICE guidelines include that they should be 'developed by independent committees, including professionals and lay members, and consulted on by stakeholders [19]. Since the first guideline was released in 1999, NICE has published over 1750 guidelines, and technology appraisal guidance alone exceeds five hundred in number [18]. After more than 20 years of development, NICE has accumulated considerable experience in their formulation and publication [20]. However, items 5 and 13 failed to reach acceptability threshold in any of the guidelines; finding patients willing and able to provide input into guideline development has proved difficult, for example several guidelines report that 'NICE's Patient and Public Involvement Programme were unable to gather patient commentary for procedures under evaluation'. NICE recognises the need to support patients, nursing staff, and the public to participate in the development and formulation of the guidelines, and has taken measures to increase the participation of these personnel, such as establishing a Public Involvement Programme and Citizens Council project [21]. High quality guidelines should be externally reviewed by experts prior to their publication; however it appeared to some assessors that there may have been a lack of engagement from key stakeholders identified as important contributors by NICE in the consultation phase of guideline development.

Domains not reaching acceptability threshold

Although mean domain 'Applicability' scored below 60%, more than one-third (14/37) of the guidelines scored this domain at 'acceptable' level or above. All NG/CGs scored over 60%. All guidelines contained accessible documents to assist doctors putting 'guidance into practice'. However in many cases, assessors may have found these documents non guideline specific. Four previous NICE guidelines evaluated using AGREE II resulted in a wide range of scores in this domain, with some authors agreeing that applicability represented their weakest domain (scores of 48 and 56 [11-12]), whereas others rated it highly (scores of 82 and 100 [9,10]). The applicability of a guideline is key to its success but may be dependent on heterogenous structural factors within the National Health Service systems, and therefore requires independent consideration.

Overall guideline assessment

Regarding overall evaluation, the AGREE II manual does not describe how to perform quantitative scoring [22]. Previous studies have applied domain score calculation methods to calculate mean scores for the item 'rate the overall quality of this guideline' without reference to the item 'recommendations of the guideline for use' [23,24]. In others, assessors have scored this based on the average rating given to the six domains [9,25]. In our study we gave no instructions to reviewers about providing overall recommendations. The mean overall score for all NICE spine disorder guidelines exceeded the 60% threshold for acceptability and a majority of assessors recommended every guideline for use.

Overall guideline scores were significantly lower in IPGs and this was observed across several domains. IPGs are of more limited scope than other guidelines, especially in particular CG/NGs in which supporting evidence can run to several thousand pages.

Overall guideline evaluation scores correlated significantly with year of publication, suggesting a dynamic process of continuous guideline quality improvement.

Limitations

Our study has several limitations. First, AGREE II remains a subjective evaluation tool. Second, the method for calculating a consensus derived overall guideline score using AGREE II is established neither by developer instructions nor by precedence in the literature. Consequently, the method

we chose may be considered arbitrary (although consistent with the calculation method of AGREE II domain and item scores). Although inter rater reliability was good or excellent in 95% of evaluations, the lack of agreement between assessors in a small number may weaken reliability of the method. We attempted to overcome this by fulfilling training and assessor number recommendations beforehand.

Conclusion

Our consensus is that the NICE spine disorder guidelines should be recommended for clinical practice as they demonstrate either acceptable or good overall quality. Evidenced ongoing quality improvement over time continues to be reassuring.

Disclosure of Interest

The authors report there are no competing interests to declare.

Statement of Using Databases

The International Classification of Diseases 11th Revision database was publicly available during the study period. The authors confirm that all methods were carried out in accordance with relevant guidelines and regulations.

Data Availability Statement

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Funding Details

None.

Acknowledgement

None

References

- World Health Organization. "Musculoskeletal conditions." Accessed August 2021.
- Cieza, A, Causey K and Kamenov K. "Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: A systematic analysis for the global burden of disease study 2019." *Lancet* 396(2021):2006-2017.
- Hartvigsen, J, Hancock MJ and Kongsted A. "What low back pain is and why we need to pay attention." *Lancet* 391(2018):2356-2367.
- Woolf, SH, Grol R and Hutchinson A. "Clinical guidelines: Potential benefits, limitations, and harms of clinical guidelines." *BMJ* 318(1999):527-530.
- National Institute for Health and Care Excellence. Accessed August 2021.
- WHO handbook for guideline development, 2nd Edition. Accessed August 2021.
- AGREE II-AGREE Enterprise website. Accessed August 2021.
- Developing NICE guidelines: The manual. Accessed 2021 Oct.
- Chua, ME, Mendoza J and See M. "A critical review of recent clinical practice guidelines on the diagnosis and treatment of non-neurogenic male lower urinary tract symptoms." *Can Urol Assoc J* 9(2015):E463-E470.
- Li-Zhen, L, Yun X and Xiao-Dong Z. "Evaluation of guidelines on the screening and diagnosis of gestational diabetes mellitus: Systematic review." *BMJ Open* 9(2019):e023014.
- Enikeev, D, Misrai V and Rijo E. "EAU, AUA and NICE Guidelines on surgical and minimally invasive treatment of benign prostate hyperplasia: A critical appraisal of the guidelines using the AGREE-II tool." *Urol Int* 9(2021):1-10.
- Perez-Panero, AJ, Ruiz-Munoz M and Cuesta-Vargas AI. "Prevention, assessment, diagnosis and management of diabetic foot based on clinical practice guidelines: A systematic review." *Med Balt* 98(2019):e16877.
- Moher, D, Shamseer L and Clarke M. "Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement." *Syst Rev* 4(2015):1.
- Agency for Healthcare Research and Quality, Rockville, MD. NGC and NQMC Inclusion Criteria. Accessed August 2021.
- Doniselli, FM, Zanardo M and Manfre L. "A critical appraisal of the quality of low back pain practice guidelines using the AGREE II tool and comparison with previous evaluations: A EuroAIM initiative." *Eur Spine J* 27(2018):2781-2790.
- Sekercioglu, N, Al-Khalifah R and Ewusie JE. "A critical appraisal of chronic kidney disease mineral and bone disorders clinical practice guidelines using the AGREE II instrument." *Int Urol Nephrol* 49(2017):273-284.
- Cichetti, DV. "Guidelines, criteria and rules of thumb for evaluating normed and standardized assessment instruments in psychology." *Psychol Assessment* 6(1994):284-290.
- History of NICE. Accessed August 2021.
- NICE-The National Institute for Health and Care Excellence. Accessed August 2021.
- NICE: 20 years of evidence-based decision making. Accessed August 2021.
- Patient and public involvement policy. Accessed 2021 Oct.
- Yaman, ME, Gudeloglu A and Senturk S. "A critical appraisal of the north american spine society guidelines with the appraisal of guidelines for research and evaluation II instrument." *Spine J* 15(2015):777-781.
- Sabharwal, S, Patel NK, Gauher S. "High methodologic quality but poor applicability: Assessment of the AAOS guidelines using the AGREE II instrument." *Clin Orthop Relat Res* 472(2014):1982-1988.
- Gavrilidis, P, Roberts KJ and Askari A. "Evaluation of the current guidelines for resection of hepatocellular carcinoma using the appraisal of guidelines for research and evaluation II instrument." *J Hepatol* 67(2017):991-998.
- Hatakeyama, Y, Seto K and Amin R. "The structure of the quality of clinical practice guidelines with the items and overall assessment in AGREE II: A regression analysis." *BMC Health Serv Res* 19(2019):788.

How to cite this article: Liang, Ning, Sizhan W, Simon R and Navnit M, et al. "Critical Appraisal of the National Institute for Health and Care Excellence (NICE) Guidelines for Spine Disorders using the Appraisal of Guidelines for Research and Evaluation II Instrument (AGREE II)" *J Neural Disord* 10(2022):504.