



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

Fabio Martino Doniselli¹ · Moreno Zanardo²  · Luigi Manfrè³ · Giacomo Davide Edoardo Papini⁴ · Alex Rovira⁵ · Francesco Sardanelli^{4,6} · Luca Maria Sconfienza^{6,7} · Estanislao Arana^{8,9}

Received: 23 July 2018 / Accepted: 9 September 2018 / Published online: 15 September 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Purpose To assess the methodologic quality of guidelines for the management of low back pain (LBP) and compare their recommendations.

Methods No ethics committee approval was needed for this systematic review. In March 2017, a systematic search was performed using MEDLINE, EMBASE, National Guideline Clearinghouse, and National Institute for Health and Clinical Excellence to find practice guidelines of assessment and management of LBP. The evaluation of guidelines quality was performed independently by four authors using the AGREE II tool, and the results were compared with previous appraisals performed in 2004 and 2009.

Results Of 114 retrieved guidelines, eight were appraised. All except one reached the level of “acceptable” in overall result, with two of them reaching the highest scores. Only two guidelines reached a level of “acceptable” in every domain; the others had at least one domain with low scores. The guidelines had the higher scores (range = 63–94%) on “Scope and purpose” and “Clarity of presentation” (47–89%). “Stakeholder Involvement” has the highest variability between the guidelines results (40–96%). “Rigor of Development” reached an intermediate mean result (34–90%), “Applicability” (42–70%), and “Editorial Independence” (38–85%). Only three guidelines had a radiologist among authors and reached higher scores compared to guidelines without a radiologist among the authors. Compared to previous assessments, low-level guidelines were 53% in 2004, 36% in 2009, and 13% in 2017.

Conclusions Considering all guidelines, only one had a “low” overall score, while half of them were rated as of “high” quality. Future guidelines might take this into account to improve clinical applicability.

Keywords Systematic review · Guidelines · AGREE II · Low back pain · Lumbar pain

Introduction

Low back pain (LBP) is a widespread and well-documented health problem, although often underestimated in terms of prevalence among general population. Some statistics report that about two-thirds of adults suffer from LBP at some point in their life, being second only to upper respiratory tract

diseases among reasons why to refer to a physician [1]; it is the greatest contributor to years lived in disability throughout much of the world and the first cause of everyday life activity limitation as well as absence from work [2]. Much less frequently, LBP is the presenting symptom of a more serious disease, such as cancer or infections. While some causes of LBP can undergo surgery, most of these symptoms are related to a nonspecific disease, self-limiting but recurrent, resulting from a sum of various medical conditions of different severity that could be very difficult to define. About 90% of patients presenting in primary care have stopped consulting the physician within 3 months [3], as symptoms tend to rapidly improve regardless of the treatment provided, if any, probably reflecting the nonspecific benefic effects of seeking and receiving care [4]. Thus, the most important

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00586-018-5763-1>) contains supplementary material, which is available to authorized users.

✉ Moreno Zanardo
moreno.zanardo@unimi.it

Extended author information available on the last page of the article

diagnostic task is to distinguish the minority of patients with serious cause of pain and/or neurologic impairment from those with a benign although disabling disease [5]. However, a large amount of lumbar spine imaging investigations is requested even in those patients who do not present with symptoms or risk factors suggesting a severe underlying condition (the so-called red flags). This approach has been largely demonstrated not to improve clinical outcome, while exposing patients to unnecessary potential radiation harms and exponentially increases costs [6, 7].

In the last decades, many different diagnostic guidelines for LBP have been proposed, and the evidence-based medicine approach, defined in the 1990s, is currently changing radiology practice to optimize patient care [8]. Relevant medical societies or specialized working groups issued several guidelines, providing clinical suggestions about diagnosis and management of patients experiencing LBP, based on the highest level of evidence. This is typically obtained with the evaluation of systematically searched primary studies [9, 10]. However, these guidelines may widely vary in terms of quality and it is important to evaluate the methodology used to develop guidelines to be confident with their recommendations [11, 12].

A few guidelines evaluation can be found in the literature, the most recent dated 2010 [13–15]. The authors used the AGREE tool to evaluate the quality of the great amount of clinical indications, treatment option, and diagnostic instruments for LBP. In 2009, a revised version of AGREE [16] was released. This is more clinician oriented and easy to use, and the comparison with previous AGREE results is still possible [17].

The European Network for the Assessment of Imaging in Medicine (EuroAIM) is a working group of the European Institute for Biomedical Imaging Research (<http://www.eibir.org/scientific-activities/joint-initiatives/euroaim/>), aimed to assess and increase the evidence for the use of imaging in clinical practice. In the last years, the EuroAIM focused on the evaluation of guidelines related to different aspects of imaging. So far, guidelines on musculoskeletal ultrasound [18] and dual-energy X-ray absorptiometry in osteoporosis [19] have been developed in collaboration with the European Society of Musculoskeletal Radiology.

The objective of this study is to update the assessment of the methodological quality of LBP guidelines published after 2010 using the AGREE II tool.

Materials and methods

Search strategy and eligibility criteria

No ethics committee approval was needed for this systematic review. In March 2017, a systematic search of the literature

was performed by using MEDLINE (PubMed; www.pubmed.gov), EMBASE (Embase; www.embase.com), Google (Google; www.google.com), National Guideline Clearinghouse (www.guideline.gov; key word: low back pain), and National Institute for Health and Clinical Excellence (NICE) (www.nice.org.uk; key word: low back pain) for searching practice guidelines of assessment and management of the LBP.

A controlled vocabulary (medical subject headings in PubMed and EMBASE thesaurus keywords in EMBASE) was used. Keywords included “guidelines” OR “practice guidelines” AND “low back pain” and their expansions.

The search key was: (('practice guideline'/exp OR 'clinical practice guidelines' OR 'guidelines' OR 'guidelines as topic' OR 'practice guideline' OR 'practice guidelines' OR 'practice guidelines as topic') AND ('low back pain'/exp OR 'acute low back pain' OR 'back pain, low' OR 'chronic low back pain' OR 'loin pain' OR 'low back pain' OR 'low backache' OR 'low backpain' OR 'lowback pain' OR 'lower back pain' OR 'lumbago' OR 'lumbal pain' OR 'lumbal syndrome' OR 'lumbalgia' OR 'lumbalgia' OR 'lumbar pain' OR 'lumbar spine syndrome' OR 'lumbar syndrome' OR 'lumbodinia' OR 'lumbosacral pain' OR 'lumbosacral root syndrome' OR 'lumbosacroiliac strain' OR 'pain, low back' OR 'pain, lumbosacral' OR 'strain, lumbosacroiliac')).

Only guidelines published after 2009 were included in this systematic review, because guidelines published before 2001 were already included and assessed in a previous guidelines evaluation [13], while the guidelines published after 2001 and before 2009 were included and assessed by Bouwmeester et al. [14].

We excluded from the results of our search those papers that were not primarily focused on LBP, such as national/international guidelines in which LBP was briefly mentioned in the context of a more comprehensive disease evaluation.

Exclusion criteria were:

- posters, oral presentations, animal studies, reviews;
- guidelines not issued by national and international medical societies;
- full manuscript not available in English.

The initial selection of eligible articles was performed by two independent readers (F.M.D. and M.Z., with 3 and 2 years of experience in evaluation of practice guidelines, respectively) basing on titles and abstracts only. After downloading eligible articles, the full text was reviewed for a thorough assessment. Finally, references of included articles were manually searched for further eligible studies.

Previous guidelines evaluation

We searched for previous guidelines evaluations using AGREE instrument (both first and second versions) to compare historical progression of guideline quality. We chose randomly two guidelines previously analyzed to be evaluated by our appraisers, to compare our results to other works to check the reproducibility of AGREE instrument.

Guidelines evaluations

The evaluation of guidelines quality was performed using the AGREE II tool through the official Web site dedicated online platform (<http://www.agreetrust.org/>).

The AGREE II tool consists of 23 different items organized in six domains (Fig. 1): domain 1 (items 1–3) called “Scope and Purpose”; domain 2 (items 4–6) “Stakeholder Involvement”; domain 3 (items 7–14) “Rigor of Development”; domain 4 (items 15–17) “Clarity of presentation”; domain 5 (items 18–21) “Applicability”; domain 6 (items 21–22) “Editorial Independence.” These six domains are followed by two additional items (“Overall Assessment”), which includes “the rating of the overall quality of the

guideline and whether the guideline would be recommended for use in practice.”

Four independent appraisers (E.A., F.M.D., L.M. and G.D.E.P.) with variables experience (2–20 years) in scientific research scored each guideline. All appraisers were previously trained to use AGREE II rating system by means of the user manual, which was available on the online platform; in addition, appraisers were asked to complete two online training tools specifically developed to assist users in effectively applying the tool (<http://www.agreetrust.org/resource-centre/agree-ii-training-tools/>). According to instruction tool, each item was rated on a 7-point scale ranging from 1 (strongly disagree, which means that no relevant information is provided) to 7 (strongly agree, which means that the quality of reporting is exceptional). Final domain scores were calculated by summing up all item scores within the domain and by scaling the total as a percentage of the maximum possible score for that domain (<http://www.agreetrust.org/about-the-agree-enterprise/introduction-to-agree-ii/scoring-the-agree-ii/>). According to AGREE tool, each appraiser could leave a comment on each field of evaluation, to help elucidate his/her personal judgment.

Domain 1: Scope and purpose	
Item 1	The overall objective(s) of the guideline is (are) specifically described
Item 2	The health question(s) covered by the guideline is (are) specifically described
Item 3	The population (patients, public, etc.) to whom the guideline is meant to apply is specifically described.
Domain 2: Stakeholder involvement	
Item 4	The guideline development group includes individuals from all the relevant professional groups.
Item 5	The views and preferences of the target population (patients, public, etc.) have been sought.
Item 6	The target users of the guideline are clearly defined.
Domain 3: Rigor of development	
Item 7	Systematic methods were used to search for evidence.
Item 8	The criteria for selecting the evidence are clearly described.
Item 9	The strengths and limitations of the body of evidence are clearly described.
Item 10	The methods for formulating the recommendations are clearly described.
Item 11	The health benefits, side effects, and risks have been considered in formulating the recommendations.
Item 12	There is an explicit link between the recommendations and the supporting evidence.
Item 13	The guideline has been externally reviewed by experts prior to its publication.
Item 14	A procedure for updating the guideline is provided.
Domain 4: Clarity of presentation	
Item 15	The recommendations are specific and unambiguous.
Item 16	The different options for management of the condition or health issue are clearly presented.
Item 17	Key recommendations are easily identifiable.
Domain 5: Applicability	
Item 18	The guideline describes facilitators and barriers to its application.
Item 19	The guideline provides advice and/or tools on how the recommendations can be put into practice.
Item 20	The potential resource implications of applying the recommendations have been considered.
Item 21	The guideline presents monitoring and/or auditing criteria.
Domain 6: Editorial independence	
Item 22	The views of the funding body have not influenced the content of the guideline.
Item 23	Competing interests of guideline development group members have been recorded and addressed.

Fig. 1 AGREE II domains descriptions and items synthesis

Data analysis

For analysis purposes, the evaluations performed by the four appraisers were averaged and the average of each domain is reported in the results. Agreement between appraisers' scores was calculated using the intraclass correlation coefficient (ICC), defined as follows: <0.20, poor; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, good; 0.81–1.00, very good. As for previous studies, the overall quality of each guideline was evaluated using a threshold of 60% for the final score of each domain [20]. High quality was defined when 5 or more domains scored > 60%, average quality when 3 or 4 domains scored > 60%, and low quality when ≤ 2 domains scored > 60%. In addition, the total score (expressed as mean \pm standard deviation, SD) of guidelines and domains was calculated. Domain scores were categorized as good ($\geq 80\%$), acceptable (60–79%), low (40–59%), or very low (< 40%), similar to a previous similar paper [20]. Data collection, extraction, and scoring were performed by a fifth independent reviewer (M.Z.) with 3 years of experience in scientific research, using a Microsoft Excel[®] 2016 spreadsheet. ICC calculations were performed using the SPSS software (version 24, IBM, Armonk, NY, USA).

Results

Our strategy of the literature search identified 97 studies, eight of them met our inclusion criteria [21–28]. Exclusion of studies was mainly: not a clinical practice guideline, review of clinical practice guidelines, not focused on both treatment and diagnosis of LBP. The flowchart of the literature search is shown in Fig. 2.

We found two previous works [13, 14] which used AGREE instrument to evaluate LBP guidelines. We randomly chose two guidelines from the paper by Bouwmeester et al. [14] to be evaluated by our appraisers, who were blinded to this decision. We considered guidelines published in the last 10 years not already evaluated by previous works [13, 14]. At the end, eight guidelines were included and are reported in Table 1.

Figure 3 summarizes the total score for each domain as well as the final judgment of overall quality. We considered “Core” the number of the pages of the guideline text, while “Total” the number of pages including eventual background, appendix or appendices, references, or other supplementary materials. Of the eight evaluated guidelines, all except one [26] reached the acceptance level in overall result (at least 60%) with two of them [25, 27] reaching the highest scores (at least 80%).

Considering all the domains, only two guidelines [23, 25] reached an acceptable score in every domain; all the others had at least one domain with low score.

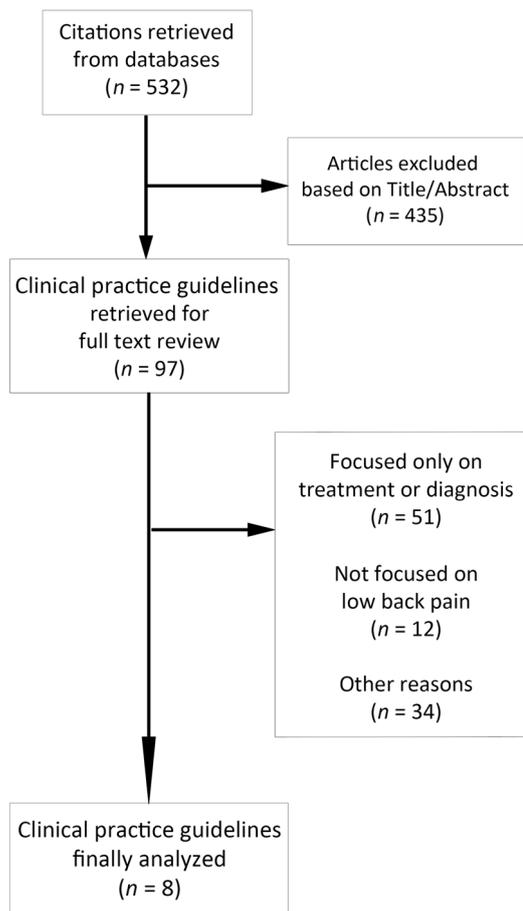


Fig. 2 Study flow for the systematic review

Acceptability in domain 1 (“Scope and Purpose”) was obtained by all the guidelines, with scores ranging from the lowest 63% (acceptable level) [26] to the highest of 94% [24, 28] (good level). This domain was the best in terms of overall results considering all the evaluated guidelines (86% of mean evaluation considering all the guidelines).

Domain 2 (“Stakeholder Involvement”) has the highest variability between the guidelines results (SD = 21%), ranging from the lowest level of Allegri et al. guideline [26] (40%) and Chiodo et al. guideline [21] (44%) to the highest results (good level) of the NICE guideline [25] (96%) and the Danish Health Authority guideline [27] (88%).

Domain 3 (“Rigor of Development”) reached an intermediate mean result (72%) with a slightly higher deviation standard (17%) when considering all the guidelines. The lowest scores were obtained by Allegri et al. guideline [26] (very low, 34%), while the highest were the Danish Health Authority guideline [26] (good, 90%) and the American College of Physicians guideline [28] (good, 83%).

Domain 4 (“Clarity of Presentation”) had a high mean quality score (83%), with 14% SD. All evaluated guideline

Table 1 Characteristics of the analyzed guidelines

Ref	Organization or author	Year	Country	Guideline title
21	University of Michigan	2011	USA	Acute low back pain guideline
22	Orthopedic section of the American Physical Therapy Association	2012	USA	Clinical practice guidelines linked to the international classification of functioning, disability, and health
23	Institute for clinical systems improvement (ICSI)	2012	USA	Adult acute and subacute low back pain
24	Institute of Health Economics (IHE)	2015	Canada	Low back pain—clinical practice guidelines
25	National Institute for Health and Care Excellence (NICE)	2016	UK	Low back pain and sciatica in over 16s: assessment and management
26	Allegri et al.	2016	Italy	Mechanisms of low back pain: a guide for diagnosis and therapy
27	Danish Health Authority (DHA)	2017	Denmark	National Clinical Guidelines for non-surgical treatment of patients with recent onset low back pain or lumbar radiculopathy
28	American College of Physicians (ACP)	2017	USA	Noninvasive treatments for acute, subacute, and chronic low back pain: a clinical practice guideline from the American College of Physicians

Fig. 3 Scores obtained by all guidelines for each domain. OA stands for overall assessment

ref	Pages		Authors (Radiologists)	Domain 1	Domain 2	Domain 3	Domain 4	Domain 5	Domain 6	OA	QUALITY
	Core	Total									
21	16		8 (0)	85%	44%	63%	89%	50%	63%	75%	average
22	11	81	8 (0)	78%	61%	79%	88%	47%	38%	67%	average
23	25	92	14 (1)	93%	83%	74%	88%	65%	83%	79%	high
24	4	59	not specified	94%	72%	79%	89%	57%	71%	79%	high
25	18	>1000	20 (1)	92%	96%	71%	86%	70%	77%	83%	high
26	11		9 (0)	63%	40%	34%	47%	42%	63%	46%	low
27	16		22 (1)	89%	88%	90%	88%	48%	71%	92%	high
28	11	29	4 (0)	94%	57%	83%	85%	42%	85%	79%	average
MEAN				86%	68%	72%	83%	53%	69%	75%	
DS				11%	21%	17%	14%	10%	15%	14%	

scores were above 80% (good level) with Chiodo et al. guideline [21] and the Institute of Health Economics guideline having the highest score (89%); only Allegri et al. [26] were below the 80% threshold with a score of 47% (low).

Domain 5 (“Applicability”) was characterized by the lowest scores for most of the guidelines and the lowest variability between the scores (DS = 10%), with a mean of 53% in overall analysis. Only two guidelines [23, 25] were above the threshold of acceptability, and the NICE guideline [25] obtained the best result (70%).

Domain 6 (“Editorial Independence”) reached an intermediate mean result (69%) when considering the whole guidelines group. The lowest scores were obtained by the orthopedic section of the American Physical Therapy Association guideline [22] (very low, 38%), while the highest were the American College of Physicians guideline [28] (good, 85%) and the Institute for Clinical Systems Improvement guideline [23] (good, 83%).

Interestingly, when evaluating the subspecialty composition of the authors of each guideline, only three [23, 25, 27] had a radiological contribute; moreover, these guidelines have the highest score (92%, 83%, 79%) compared to the group without a radiologist.

Comparison with previous studies was possible only for the papers by Van Tulder et al. [13] and Bouwmeester et al. [14] where percentage scores were reported for every domain in every study. Although overall score as reported in our Table 2 as “OA” followed a different formulation, these data were not given in previous evaluations. To make our results comparable, we measured total score as mean of the domains and we assessed quality as described before (high, average, low). Results are listed in Table 3, and guidelines previously evaluated by Van Tulder et al. [13] and by Bouwmeester et al. [14] were reported [29–59].

To assess inter-study reproducibility, we randomly chose two guidelines previously evaluated by Bouwmeester et al. [14]. The National Health Committee of

Table 2 Guidelines evaluated by Van Tulder [13] are shown in bold, by Bouwmeester [14] in italics, and present evaluation in bold italics

	References	Domains						Mean	Quality
		Scope and purpose	Stakeholder involvement	Rigor of development	Clarity and presentation	Applicability	Editorial independence		
Van Tulder et al. (Ref [13])	29	100	42	48	92	22	67	62	A
	30	100	83	86	83	33	33	70	A
	31	78	42	48	75	33	33	52	L
	32	67	42	38	63	33	33	46	L
	33	67	50	33	67	33	33	47	L
	34	56	50	57	67	33	33	49	L
	35	67	50	38	63	33	33	47	L
	36	100	33	48	92	11	17	50	L
	37	67	33	10	67	0	0	30	L
	38	100	50	86	83	44	33	66	A
	39	89	25	67	67	0	17	44	A
	40	67	42	62	75	44	33	54	A
	41	78	42	62	75	44	33	56	A
	42	89	83	86	83	33	33	68	A
	43	44	50	38	42	0	0	29	L
44	78	50	33	100	33	0	49	L	
45	89	75	76	83	33	50	68	A	
Bouwmeester et al. (Ref. [14])	<i>46</i>	<i>61</i>	<i>75</i>	<i>76</i>	<i>100</i>	<i>61</i>	<i>33</i>	<i>68</i>	H
	<i>47</i>	<i>89</i>	<i>50</i>	<i>93</i>	<i>100</i>	<i>28</i>	<i>58</i>	<i>70</i>	A
	<i>48</i>	<i>61</i>	<i>79</i>	<i>98</i>	<i>92</i>	<i>56</i>	<i>50</i>	<i>73</i>	A
	<i>49</i>	<i>61</i>	<i>46</i>	<i>74</i>	<i>75</i>	<i>44</i>	<i>83</i>	<i>64</i>	A
	<i>50</i>	<i>61</i>	<i>50</i>	<i>81</i>	<i>67</i>	<i>72</i>	<i>92</i>	<i>71</i>	H
	<i>51</i>	<i>61</i>	<i>58</i>	<i>88</i>	<i>100</i>	<i>44</i>	<i>42</i>	<i>66</i>	A
	<i>52</i>	<i>56</i>	<i>50</i>	<i>50</i>	<i>88</i>	<i>89</i>	<i>42</i>	<i>63</i>	L
	<i>53</i>	<i>89</i>	<i>71</i>	<i>95</i>	<i>96</i>	<i>39</i>	<i>100</i>	<i>82</i>	H
	<i>54</i>	<i>56</i>	<i>83</i>	<i>48</i>	<i>96</i>	<i>33</i>	<i>0</i>	<i>53</i>	L
	<i>55</i>	<i>39</i>	<i>46</i>	<i>62</i>	<i>88</i>	<i>44</i>	<i>17</i>	<i>49</i>	L
	<i>56</i>	<i>56</i>	<i>38</i>	<i>88</i>	<i>96</i>	<i>56</i>	<i>83</i>	<i>70</i>	A
	<i>57</i>	<i>89</i>	<i>46</i>	<i>95</i>	<i>92</i>	<i>56</i>	<i>83</i>	<i>77</i>	A
	<i>58</i>	<i>67</i>	<i>54</i>	<i>48</i>	<i>46</i>	<i>11</i>	<i>8</i>	<i>39</i>	L
	<i>59</i>	<i>22</i>	<i>25</i>	<i>43</i>	<i>92</i>	<i>22</i>	<i>25</i>	<i>38</i>	L
Present evaluation	<i>21</i>	<i>85</i>	<i>44</i>	<i>63</i>	<i>89</i>	<i>50</i>	<i>63</i>	<i>66</i>	A
	<i>22</i>	<i>93</i>	<i>83</i>	<i>74</i>	<i>88</i>	<i>65</i>	<i>83</i>	<i>81</i>	H
	<i>23</i>	<i>78</i>	<i>61</i>	<i>79</i>	<i>88</i>	<i>47</i>	<i>38</i>	<i>65</i>	A
	<i>24</i>	<i>94</i>	<i>72</i>	<i>79</i>	<i>89</i>	<i>57</i>	<i>71</i>	<i>77</i>	H
	<i>25</i>	<i>92</i>	<i>96</i>	<i>71</i>	<i>86</i>	<i>70</i>	<i>77</i>	<i>82</i>	H
	<i>26</i>	<i>63</i>	<i>40</i>	<i>34</i>	<i>47</i>	<i>42</i>	<i>63</i>	<i>48</i>	L
	<i>27</i>	<i>89</i>	<i>88</i>	<i>90</i>	<i>88</i>	<i>48</i>	<i>71</i>	<i>79</i>	H
	<i>28</i>	<i>94</i>	<i>57</i>	<i>83</i>	<i>85</i>	<i>42</i>	<i>85</i>	<i>74</i>	A

The quality was indicated as *H* high, *A* average, and *L* low

New Zealand guideline [51] and Chou et al. guidelines [57] were evaluated by our four appraisers and their results were compared with Bouwmeester et al. [14]. The comparison of the previous studies [13, 14] with the current work is given in Table 3. When assessing quality level as

previously done, low-level guidelines represented 53% in 2004, 36% in 2009, and only 13% in 2017, while high-quality guidelines, not present at all in 2004, represent the 50% of evaluated guidelines in 2017.

Table 3 Comparison of two guidelines (Refs. [51, 57]) evaluated by previous group (Bouwmeester [14]) and this work (see column *present*)

Evaluated by	Ref. [51]		Ref. [57]	
	Ref. [14]	Present	Ref. [14]	Present
<i>Domains</i>				
1. Scope and purpose	61	65	89	92
2. Stakeholder involvement	58	53	46	71
3. Rigor of development	88	73	95	79
4. Clarity and presentation	100	81	92	93
5. Applicability	44	68	56	34
6. Editorial independence	42	54	83	75
Mean	66	66	77	74
Quality	Average	Average	Average	High

Discussion

LBP remains a major health problem worldwide, with direct implications on daily life and thus affecting everyday human activities. This condition has deep relationship with other pathologies such as depression, anxiety, and sleep disturbances [60], but is still not well managed. As found by Rego et al. [61], up to 40% of patients exhibited no indications to imaging. Up to about 59% of lumbar spine MRI examinations requested among different social and economic settings have been deemed as inappropriate, depending on the evaluation method [62–65]. Of all lumbar and thoracic spine CT examinations in young patients, up to 77% were judged as inappropriate; therefore, such examinations are responsible of an unjustified cost for individuals or health systems and unnecessary radiation exposure [66].

Our main finding is that overall quality of guidelines has improved since last revision [13, 14], as only one was rated below acceptable quality, while the others ranged from acceptable to high quality.

The highest scores were found in domain 1 (“Scope and purpose”) and in domain 4 (“Clarity of presentation”). This can be probably due to the clinical address of the guidelines, particularly for domain 1, as LBP patients are traditionally well described throughout a large number of signs and symptoms (an example is the “red flags” always cited in the literature). In this sense, LBP guidelines are mainly addressed to clinicians. High scores in domain 4 are also closely connected to the clinical focus: If guidelines should be easily consulted, simplicity and clarity are certainly two crucial requisites.

Domain 2 (“Stakeholder involvement”) was characterized by the highest variability between the guidelines. This

could be related to the variability in the composition of the authors’ specializations: Only three guidelines had a radiologist among authors (Table 2). The role of radiologists in LBP decision-making remains negligible, as shown in the composition of the authors for each assessed guideline. As demonstrated by the results, the involvement of radiologists among the authors increases every single domain of evaluation. We did not compare previous studies’ evaluators’ specialization to check if the presence of radiologists determined a better quality also in previous guidelines. We may argue that involvement of radiologists in this kind of guidelines may improve clinical applicability and patient management.

Domain 3 (“Rigor of development”) and domain 6 (“Editorial independence”) are on average worse than the other domains, but it has to be noted how improvement occurred since the first evaluation [13]. This is particularly true for domain 6 and we believe that AGREE instrument can play a major role in guiding and helping guideline writing.

Lowest scores were found on domain 5 (“Applicability”) and may be related to the wide targets regarding population and outcomes. One of them is imaging indications, where radiologists may find several issues, especially regarding appropriateness. Another issue is related to the different healthcare systems in the world: Managing and evaluating recommendations imply to relate the results to the different way LBP patients are assisted (private healthcare or public health systems) and how much resources are dedicated to that.

Considering all guidelines in our evaluation, only one had a “low” overall score, while half of them were rated as “high” quality. Two of the guidelines, as reported in Table 3, were previously compared, and nearly perfect agreement was found with the scores of Van Bouwmeester et al. [14]. With this evidence for robustness of the AGREE tool, we therefore compared all the previously evaluated guidelines in Van Tulder et al. [13] and Bouwmeester et al. [14]. As shown in Fig. 4, since the adoption of AGREE standard, the quality of guidelines has pushed forward from the beginning. Differently from Bouwmeester et al. [14] (who found that “adding more than 3 appraisers did not consistently improve reliability”), we followed the recommendations by MacDermid et al. [67] (who recommend in the AGREE II Manual “... that each guideline be assessed by at least 2 appraisers, and preferably 4, as this will increase the reliability of the assessment”). Indeed, we used four independent appraisers from different countries to improve not only the reliability of the assessment, but also the quality of the evaluations with a critical judgment of the guidelines.

Some limitations should be taken into account. First, the AGREE II tool does not assess the consistency of guidelines with the published evidence and does not evaluate the clinical content of guidelines. These two limitations are common

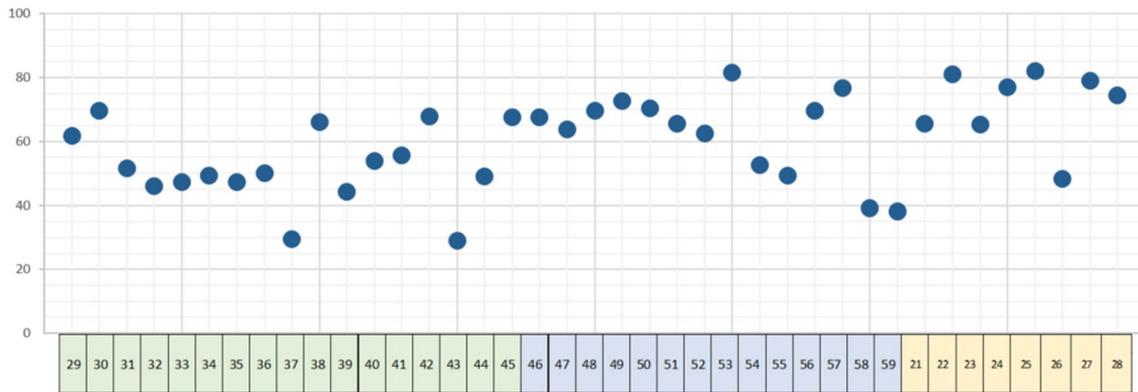


Fig. 4 Graphic representation of the overall assessment score (y-axes) obtained by all guidelines evaluated by Van Tulder [13], by Bouwmeester [14] and by the present evaluation (on x-axes)

to many similar tools. Then, the experience of the reviewers was quite variable. However, having averaged all evaluations has certainly tempered this drawback.

In conclusion, our analysis revealed that half of the assessed guidelines on LBP have high quality. The domain “Applicability” was the most critical in terms of overall score, probably due to the wide range of physicians to whom these guidelines are addressed. Future guidelines might take this into account to improve clinical applicability.

Acknowledgements This work has been conducted within the framework of the Network for Assessment of Imaging in Medicine (EuroAIM), research platform of the European Institute for Biomedical Research under the umbrella of the European Society of Radiology (<http://www.eibir.org/scientific-activities/jointinitiatives/euroaim/>). Francesco Sardanelli is chair of the EuroAIM initiative, Luca Maria Sconfienza is the coordinator of the guideline evaluation project, while the other authors are members of the EuroAIM working group.

Compliance with ethical standards

Conflicts of interest Authors have no conflicts of interest to disclose related to the present paper.

References

- Deyo RA, Weinstein JN (2001) Low back pain. *N Engl J Med* 344:363–370. <https://doi.org/10.1056/NEJM200102013440508>
- Hoy D, March L, Brooks P et al (2014) The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. *Ann Rheum Dis* 73:968–974. <https://doi.org/10.1136/annrheumdis-2013-204428>
- Croft PR, Macfarlane GJ, Papageorgiou AC et al (1998) Outcome of low back pain in general practice: a prospective study. *BMJ* 316:1356–1359. <https://doi.org/10.1136/bmj.316.7141.1356>
- Artus M, van der Windt D, Jordan KP et al (2014) The clinical course of low back pain: a meta-analysis comparing outcomes in randomised clinical trials (RCTs) and observational studies. *BMC Musculoskelet Disord* 15:68. <https://doi.org/10.1186/1471-2474-15-68>
- Breen A (2017) Low back pain: identifying sub-groups, clinical prediction rules and measuring results. *Complement Ther Clin Pract*. <https://doi.org/10.1016/j.ctcp.2017.07.005>
- Chou R, Deyo RA, Jarvik JG (2012) Appropriate use of lumbar imaging for evaluation of low back pain. *Radiol Clin N Am* 50:569–585. <https://doi.org/10.1016/j.rcl.2012.04.005>
- Andersen JC (2011) Is immediate imaging important in managing low back pain? *J Athl Train* 46:99–102. <https://doi.org/10.4085/1062-6050-46.1.99>
- Lavelle LP, Dunne RM, Carroll AG et al (2015) Evidence-based practice of radiology. *Radiographics* 35:1802–1813. <https://doi.org/10.1148/rg.2015150027>
- Egger M, Smith GD, Altman DG (2008) *Systematic reviews in health care: meta-analysis in context*. BMJ Books, London. ISBN 978-0-727-91488-0
- Sardanelli F, Bashir H, Berzaczky D et al (2014) The role of imaging specialists as authors of systematic reviews on diagnostic and interventional imaging and its impact on scientific quality: report from the EuroAIM evidence-based radiology working group. *Radiology* 272:533–540. <https://doi.org/10.1148/radiol.14131730>
- Shaneyfelt TM, Mayo-Smith MF, Rothwangl J (1999) Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA* 281:1900–1905. <https://doi.org/10.1001/jama.281.20.1900>
- Grilli R, Magrini N, Penna A et al (2000) Practice guidelines developed by specialty societies: the need for a critical appraisal. *Lancet* 355:103–106. [https://doi.org/10.1016/S0140-6736\(99\)02171-6](https://doi.org/10.1016/S0140-6736(99)02171-6)
- van Tulder MW, Tuut M, Pennick V et al (2004) Quality of primary care guidelines for acute low back pain. *Spine* 29:E357–E362. <https://doi.org/10.1097/01.brs.0000137056.64166.51>
- Bouwmeester W, van Enst A, van Tulder M (2009) Quality of low back pain guidelines improved. *Spine* 34:2562–2567. <https://doi.org/10.1097/brs.0b013e3181b4d50d>
- Dagenais S, Tricco AC, Haldeman S (2010) Synthesis of recommendations for the assessment and management of low back pain from recent clinical practice guidelines. *Spine J* 10:514–529
- Brouwers MC, Kho ME, Browman GP et al (2010) AGREE II: Advancing guideline development, reporting and evaluation in healthcare. *CMAJ* 182:E839–E842. <https://doi.org/10.1016/j.spinee.20>

17. AGREE Next Steps Consortium (2017) The AGREE II Instrument [Electronic version]. <http://www.agreetrust.org>. Accessed 13 June 2018
18. Messina C, Bignotti B, Tagliafico A et al (2017) A critical appraisal of the quality of adult musculoskeletal ultrasound guidelines using the AGREE II tool: an EuroAIM initiative. *Insights Imaging* 8:491–497. <https://doi.org/10.1007/s13244-017-0563-4>
19. Messina C, Bignotti B, Bazzocchi A et al (2017) A critical appraisal of the quality of adult dual-energy X-ray absorptiometry guidelines in osteoporosis using the AGREE II tool: an EuroAIM initiative. *Insights Imaging* 8:311–317. <https://doi.org/10.1007/s13244-017-0553-6>
20. Sekercioglu N, Al-Khalifah R, Ewusie JE et al (2017) A critical appraisal of chronic kidney disease mineral and bone disorders clinical practice guidelines using the AGREE II instrument. *Int Urol Nephrol* 49:273–284. <https://doi.org/10.1007/s11255-016-1436-3>
21. Chiodo AE, Alvarez DJ, Graziano GP et al (2010) Acute low back pain. Clinical alignment and performance excellence. Updated 2010. <http://www.med.umich.edu/1info/FHP/practiceguides/back/back.pdf>. Accessed 13 June 2018
22. Delitto A, George SZ, Van Dillen LR et al (2012) Low back pain. *J Orthop Sports Phys Ther* 42:A1–A57. <https://doi.org/10.2519/jospt.2012.0301>
23. Goertz M, Thorson D, Bonsell J et al (2012) Adult acute and subacute low back pain. Institute for Clinical Systems Improvement, Bloomington
24. Toward Optimized Practice (TOP) Low Back Pain Working Group (2015) Evidence-informed primary care management of low back pain: clinical practice guideline. Edmonton, AB: Toward optimized practice. <http://www.topalbertadoctors.org/cpgs/885801>. Accessed 13 June 2018
25. The National Guideline Centre (UK) (2016) Low back pain and sciatica in over 16 s: assessment and management. 2016, NICE, London. <http://www.nice.org.uk/guidance/ng59>. Accessed 13 June 2018
26. Allegri M, Montella S, Salici F et al (2016) Mechanisms of low back pain: a guide for diagnosis and therapy. Version 2. F1000Res. <https://f1000research.com/articles/5-1530/v2#article-reports>. Accessed 13 June 2018
27. Stochkendahl MJ, Kjaer P, Hartvigsen J et al (2018) National clinical guidelines for non-surgical treatment of patients with recent onset low back pain or lumbar radiculopathy. *Eur Spine J* 27:60. <https://doi.org/10.1007/s00586-017-5099-2>
28. Qaseem A, Wilt TJ, McLean RM et al (2017) Noninvasive treatments for acute, subacute, and chronic low back pain: a clinical practice guideline from the American College of Physicians. *Ann Internal Med* 166:514–530. <https://doi.org/10.7326/M16-2367>
29. Spitzer WO, LeBlanc FE, Dupuis M et al (1987) Scientific approach to the assessment and management of activity-related spinal disorders. *Spine* 12:S1–S59
30. Bigos S, Bowyer O, Braen G (1994) Acute low back problems in adults. Clinical practice guideline no. 14. AHCPR Publication No. 95-0642. Rockville, MD: Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services
31. Faas A, Chavannes AW, Koes BW et al (1996) NHG-Standard ‘lage-rugpijn’. *Huisarts Wet* 39:18–31
32. Borkan JM, Reis S, Werner S et al (1996) Guidelines for the treatment of low back pain in primary care. *Harefuah* 130:145–151
33. National Health Committee. National Advisory Committee on Health and Disability, Accident Rehabilitation and Compensation Insurance Corporation (1997) New Zealand acute low back pain guide. Wellington, New Zealand. <https://www.healthnavigator.org.nz/media/1006/nz-acute-low-back-pain-guide-acc.pdf>. Accessed 13 June 2018
34. Kendall NAS, Linton SJ, Main CJ (1997) Guide to assessing psychosocial yellow flags in acute low back pain: risk factors for long-term disability and work loss. Accident Rehabilitation and Compensation Insurance Corporation of New Zealand and the National Health Committee, Wellington, New Zealand
35. Malmivaara A, Kotilainen E, Laasonen E et al (1999) Clinical practice guidelines of the Finnish medical association duodecim. Diseases of the Low Back, Finland
36. Keel P, Weber M, Roux E et al (1998) Kreuzschmerzen: Hintergründe, Prävention, Behandlung. *Verbindung der Schweizer rzte*, Bern
37. Veterans Health Administration (2001) Low back pain or sciatica in the primary care setting. Department of veterans affairs, office of quality and performance, Washington, DC. <https://www.healthquality.va.gov/guidelines/Pain/lbp/VADoDLBPCPG092917.pdf>. Accessed 13 June 2018
38. Agency for Health Care Administration and the Florida Department of Health (1999) University of Florida patients with low back pain or injury: medical practice guidelines
39. Nachemson AL, Jonsson E (2000) Neck and back pain: the scientific evidence of causes, diagnosis, and treatment. Lippincott Williams and Wilkins, Philadelphia
40. Danish Institute for Health Technology Assessment (2000) Low back pain: frequency, management and prevention from a health technology perspective. National Board of Health, Copenhagen, Denmark. http://www.chiro.org/LINKS/GUIDELINES/FULL/Low_Back_Pain_Frequency_Management.shtml. Accessed 13 June 2018
41. Arzneimittelkommission der deutschen rzeschaft (2000) Empfehlungen zur Therapie von Kreuzschmerzen. Koln, Germany
42. Abenheim L, Rossignol M, Valat JP et al (2000) The role of activity in the therapeutic management of back pain. Report of the international Paris task force on back pain. *Spine* 25:1S–33S
43. Hutchinson A, Waddell G, Feder G et al (1996) Clinical guidelines for the management of acute low back pain. Royal College of General Practitioners, London. http://www.chiro.org/LINKS/GUIDELINES/FULL/Royal_College/index.html
44. Bekkering GE, Hendriks HJM, Koes BW et al (2001) KNGF-richtlijn lage rugpijn. *Ned Tijdschr Fysiother* 111:1–24
45. Institute for Clinical Systems Improvement (2001) Health care guideline: adult low back pain. Institute for Clinical Systems Improvement, Bloomington
46. Philadelphia panel (2001) Philadelphia panel evidence-based clinical practice guidelines on selected rehabilitation interventions for low back pain. *Phys Ther* 81:1641–1674
47. The Dutch Institute for Healthcare Improvement (2003) Clinical guideline for non-specific low back pain. *Ned Tijdschr Fys* 113:1–24 (in Dutch)
48. Australian Acute Musculoskeletal Pain Guidelines Group (2003) Management of acute musculoskeletal pain (National Library of Australia Cataloguing-in-Publication). Brisbane, Australia: Australian Academic Press. <http://www.sif-fisioterapia.it/wp-content/uploads/2014/12/Acute-Musculoscheletral-Pain-Australia-2003.pdf>. Accessed 13 June 2018
49. Airaksinen O, Brox JI, Cedraschi C et al (2006) European guidelines for the management of chronic non-specific low back pain. *Eur Spine J* 15:S192–S300. <https://doi.org/10.1007/s00586-006-1072-1>
50. Van Tulder MW, Becker A, Bekkering T et al (2006) European guidelines for the management of acute low back pain in primary care. *Eur Spine J* 15(Suppl 2):S169–S191. <https://doi.org/10.1007/s00586-006-1071-2>
51. National Health Committee (2004) New Zealand acute low back pain guide. National Advisory Committee on Health and Disability, Accident Rehabilitation and Compensation Insurance Corporation, Wellington, New Zealand. <https://www.healthnavi>

- gator.org.nz/media/1006/nz-acute-low-back-pain-guide-acc.pdf. Accessed 13 June 2018
52. Institute for Clinical Systems Improvement (ICSI) (2005) Adult low back pain. ICSI, Bloomington
 53. Bekkering GE, Hendriks HJM, Koes BW et al (2003) National practice guideline for the physiotherapeutic management of patients with low back pain. *Physiotherapy* 89:82–96. [https://doi.org/10.1016/S0031-9406\(05\)60579-2](https://doi.org/10.1016/S0031-9406(05)60579-2)
 54. Chavannes AW, Mens JMA, Koes BW et al (2005) Dutch general practice guideline for non-specific low back pain. *Huisarts Wet* 48:113–123 (in Dutch)
 55. Negrini S, Giovannoni S, Minozzi S et al (2006) Diagnostic therapeutic flow-charts for low back pain patients: the Italian clinical guidelines. *Euro Medicophys* 42:151–170
 56. Drug Committee of the German Medical Society (2007) Recommendations for treatment of low back pain. Koln, Germany (in German)
 57. Chou R, Qaseem A, Snow V, Clinical Efficacy Assessment Subcommittee of the American College of Physicians, American College of Physicians, American Pain Society Low Back Pain Guidelines Panel et al (2007) Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. *Ann Intern Med* 147:478–491. <https://doi.org/10.7326/0003-4819-147-7-200710020-00006>
 58. Friedrich M, Lika R (2007) Evidenz- und konsensusbasierte österreichische Leitlinien für das Management akuter und chronischer unspezifischer Kreuzschmerzen. *Wien Klin Wochenschr* 119:189–197
 59. Rossignol M, Arsenault B, Dionne C et al (2007) Clinic on Low-back pain in interdisciplinary practice (CLIP) guidelines. <http://www.santpub-mtl.qc.ca/clip>. Accessed 13 June 2018
 60. Stubbs B, Koyanagi A, Thompson T et al (2016) The epidemiology of back pain and its relationship with depression, psychosis, anxiety, sleep disturbances, and stress sensitivity: data from 43 low- and middle-income countries. *Gen Hosp Psychiatry* 43:63–70. <https://doi.org/10.1016/j.genhosppsych.2016.09.008>
 61. Rego MH, Nagiah S (2016) Over-imaging in uncomplicated low back pain: a 12-month audit of a general medical unit. *Intern Med J* 46:1437–1439. <https://doi.org/10.1111/imj.13279>
 62. Jame SZ, Sari AA, Majdzadeh R et al (2014) The extent of inappropriate use of magnetic resonance imaging in low back pain and its contributory factors. *Int J Prev Med* 5:1029–1036
 63. Gidwani R, Sinnott P, Avoundjian T et al (2016) Inappropriate ordering of lumbar spine magnetic resonance imaging: are providers Choosing Wisely? *Am J Manag Care* 22:e68–e76
 64. Emery DJ, Shojania KG, Forster AJ et al (2013) Overuse of magnetic resonance imaging. *JAMA Intern Med* 173:823–825. <https://doi.org/10.1001/jamainternmed.2013.3804>
 65. Avoundjian T, Gidwani R, Yao D et al (2016) Evaluating two measures of lumbar spine MRI overuse: administrative data versus chart review. *J Am Coll Radiol* 13:1057–1066. <https://doi.org/10.1016/j.jacr.2016.04.013>
 66. Oikarinen H, Meriläinen S, Pääkkö E et al (2009) Unjustified CT examinations in young patients. *Eur Radiol* 19:1161–1165. <https://doi.org/10.1007/s00330-008-1256-7>
 67. MacDermid JC, Brooks D, Solway S et al (2005) Reliability and validity of the AGREE instrument used by physical therapists in assessment of clinical practice guidelines. *BMC Health Serv Res* 5:18. <https://doi.org/10.1186/1472-6963-5-18>

Affiliations

Fabio Martino Doniselli¹ · Moreno Zanardo²  · Luigi Manfrè³ · Giacomo Davide Edoardo Papini⁴ · Alex Rovira⁵ · Francesco Sardanelli^{4,6} · Luca Maria Sconfienza^{6,7} · Estanislao Arana^{8,9}

¹ Postgraduation School in Radiodiagnostics, Università degli Studi di Milano, Milan, Italy

² PhD Course in Integrative Biomedical Research, Department of Biomedical Sciences for Health, Università degli Studi di Milano, Via Mangiagalli, 31, 20133 Milan, Italy

³ Department of Neurosurgery/Interventional Radiology, Minimal Invasive Spine Therapy, Institute of Oncology in Mediterranean (IOM), Viagrande, Catania, Italy

⁴ Unit of Radiology, IRCCS Policlinico San Donato, San Donato Milanese, Milan, Italy

⁵ Section of Neuroradiology and Magnetic Resonance Unit, Department of Radiology (IDI), Vall d’Hebron Institut de Recerca, Hospital Universitari Vall d’Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain

⁶ Dipartimento di Scienze Biomediche per la Salute, Università degli Studi di Milano, Via Pascal 36, 20100 Milano, Italy

⁷ Unità Operativa di Radiologia Diagnostica ed Interventistica, IRCCS Istituto Ortopedico Galeazzi, Milano, Italy

⁸ Spanish Back Pain Research Network, Kovacs Foundation, Paseo de Mallorca 36, 07012 Palma de Mallorca, Spain

⁹ Department of Radiology, Valencian Oncology Institute Foundation, C/Professor Beltrán Bágüena, 19, 46009 Valencia, Spain