Joint Bone Spine 79 (2012) 176–185

Available online at ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com/en

Original article

An updated overview of clinical guidelines for chronic low back pain management in primary care

Paolo Pillastrini[a,1,*], Ivan Gardenghi[a], Francesca Bonetti[a], Francesco Capra[a], Andrew Guccione[b], Raffaele Mugnai[a], Francesco S. Violante[a]

[a] *Occupational Medicine Unit, Department of Internal Medicine, Geriatrics and Nephrology, Alma Mater Studiorum–University of Bologna, S. Orsola-Malpighi Hospital, via P. Palagi 9, 40138 Bologna, Italy*
[b] *Department of Rehabilitation Science, College of Health and Human Services, George Mason University, 4400 University Drive, Fairfax, VA, 22030 USA*

ARTICLE INFO

Article history:
Accepted 23 March 2011
Available online 12 May 2011

Keywords:
Chronic low back pain
Clinical guidelines
Quality assessment
Diagnosis
Treatment

ABSTRACT

Objectives: In the past decade many countries around the world have produced clinical practice guidelines to assist practitioners in providing a care that is aligned with the best evidence. The aim of this study was to present and compare the most established evidence-based recommendations for the management of chronic nonspecific low back pain in primary care derived from current high-quality international guidelines.
Methods: Guidelines published or updated since 2002 were selected by searching PubMed, CINAHL, EMBASE, guidelines databases, and the World Wide Web. The methodological quality of the guidelines was assessed by three authors independently, using the Appraisal of Guidelines for Research and Evaluation (AGREE) Instrument. Guideline recommendations were synthesized into diagnostic and therapeutic approaches that were supported by strong, moderate or weak evidence.
Results: Thirteen guidelines were included. In general, the quality was satisfactory. Guidelines had highest scores on clarity and presentation and scope and purpose domains, and lowest scores on applicability. There was a strong consensus among all the guidelines particularly regarding the use of diagnostic triage and the assessment of prognostic factors. Consistent therapeutic recommendations were information, exercise therapy, multidisciplinary treatment, and combined physical and psychological interventions.
Conclusion: Compared to previous assessments, the average quality of the guidelines dealing with chronic low back pain has improved. Furthermore, all guidelines are increasingly aligning in providing therapeutic recommendations that are clearly differentiated from those formulated for acute pain. However, there is still a need for improving quality and generating new evidence for this particular condition.

1. Introduction

Low back pain (LBP) is a common problem affecting both genders and most age groups such that about one in four adults seeks care in a six-month period. LBP has substantial direct and indirect costs to the person, workplace and society [1]. Although most episodes of LBP appear self-limiting [2], recurrence with a variable course is common [3], with 10–15% of cases leading to chronic pain [4]. In order to decrease this burden, the use of interventions with demonstrated effectiveness is essential [5]. Clinical practice guidelines (CPGs) can be powerful tools for promoting evidence-based practice (EBP), as they integrate research findings in order to support decision-making. Following the publication of the first LBP guideline in 1987 by the Quebec Task Force, which also highlighted the absence of high-quality evidence [6], there has been a steady worldwide interest on this subject, culminating with the publication of specific CPGs in many countries over the past few years.

Nevertheless, previous reviews [7–9] reported disappointing results with regard to the methodological quality of guidelines assessed using the Appraisal of Guidelines for Research and Evaluation (AGREE) Instrument [10]. Furthermore, although many guideline recommendations were similar with respect to diagnosis and therapeutic interventions especially for acute LBP, researchers repeatedly indicated the need to place additional emphasis on differentiating acute from chronic LBP (CLBP) and providing more consistent recommendations for the management of this distinct condition [9]. Recently, separate reviews by Koes et al. [11] and Dagenais et al. [12] structured their findings to provide comprehensive guidance to clinicians. However some important mono- and multidisciplinary guidelines dealing with the management of

* Corresponding author. Tel.: +39 051 63 62 496; fax: +39 051 63 62 609.
E-mail address: paolo.pillastrini@unibo.it (P. Pillastrini).
[1] *Website*: www.unibo.it.

1297-319X/$ – see front matter © 2011 Société française de rhumatologie. Published by Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.jbspin.2011.03.019

CLBP were not included. The purpose of the present study was to assess the literature and rate the methodological quality of currently available guidelines for the management of nonspecific CLBP in primary care using the validated AGREE tool, and to provide a specific, updated and evidence-based overview of the most important clinical recommendations regarding the management of this particular condition.

2. Methods

2.1. Data sources

CPGs were identified using specific search strategies in various sources:

- MEDLINE and PubMed, CINAHL, EMBASE (from 2002 to December 2010). The search included combinations of the following keywords (MeSH terms): *low back pain* plus *guideline* or *practice guideline* or *clinical practice guideline* and the same combination using the plural form "guidelines".
- Guideline databases (up to December 2010), including the National Guideline Clearinghouse, Canadian Medical Association InfoBase, Guidelines International Network, National Institute for Clinical Excellence, National Library for Health guidelines database and Scottish Intercollegiate Guidelines Network. The search term used was *low back pain*.
- World Wide Web (up to December 2010), by means of the Google and Google Scholar browsers, using the term *low back pain guideline* and the same combination using the plural form "guidelines".
- Additional guidelines were identified by manually searching the reference lists of retrieved guidelines, reports and review articles [13].

2.2. Selection of guidelines

Only guidelines published or updated since 2002 were considered. Only the most recent CPG was included when multiple versions were available. Additionally, guidelines had to meet the following criteria for inclusion in the study: (a) addressed the clinical management of nonspecific CLBP in primary care; (b) published by a professional group; (c) available in English, Italian or German languages/versions; (d) stated recommendations for therapeutic interventions explicitly. Guidelines were excluded on the basis of the following criteria: (a) addressed only the management of acute LBP, occupational-related LBP, secondary care of LBP or prevention of LBP; (b) developed by one individual or one regional health care centre/hospital; (c) copied or summarized another included guideline; (d) comprised a single report or article on guideline evaluation/implementation; (e) provided a narrative review without evidence-based recommendations or limited its objectives to teaching. Additional information (other publication or website) was only taken into account when the guideline explicitly referred to it.

2.3. Quality assessment

All guidelines were reviewed independently by three authors (I.G., F.B., R.M.) and scored for methodological quality according to the AGREE instrument [10], which has been shown to be reliable when used by physiotherapists to assess the quality of clinical guidelines relevant to physical therapy practice [14]. This tool consists of 23 items organized in six domains (Table 1) so that each domain is intended to capture a separate dimension of guideline quality. Each item is rated on a four-point scale ranging from 4 = "Strongly Agree" to 1 = "Strongly Disagree," with two mid points: 3 = "Agree" and 2 = "Disagree." The scale measures the extent to which a criterion (item) has been fulfilled. Discrepancies between the scores of the three reviewers were resolved in a consensus meeting only when there was a difference in positive and negative assessment (e.g., scoring 1 or 2 vs. 3 or 4). Domain scores were calculated by summing the scores of all the individual items in a domain and then dividing the difference between the obtained score and the minimum possible score by the difference between the maximum possible and minimum possible score.

2.4. Guideline assessment

There is no validated measure for distinguishing among "excellent," "good," "fair", and "poor" guidelines. Therefore, we adapted the novel approach used by the Institute of Health Economics to fill this gap [15]. The TOP research team used a modification of the AGREE tool to reduce the ambiguity and subjectivity associated with item scoring and enable the differentiation of good from poor quality guidelines [16]. We similarly identified ten "essential" criteria (AGREE items) for categorizing guidelines on the basis of their quality. In addition to the seven items already considered "essential" by the Evidence-based Medicine Working Group [17], we included all three items from AGREE's Scope and purpose domain. Guidelines were therefore rated on how well they described their scope and purpose (items 1-3), how their methods excluded bias by examining the search strategy used (item 8), how the recommendations were formulated and presented (item 10 and 15), whether the recommendations were directly linked to the evidence (item 12), the external review process (item 13), and whether funding sources (item 22) and conflicts of interest of developers (item 23) were reported. The average quality rating score for these items was derived by calculating the arithmetic mean of the scores given on each item by the three reviewers. This mean score was then categorized as follows: Excellent - average score of 36 to 40, Good - average score of 31 to 35, Moderate - average score of 21 to 30, Poor - average score of 0 to 20.

This nonlinear four-level scale was devised *a posteriori* to allow a clear distinction between the guidelines included in the study and to obtain a correlation with the overall assessment categories provided by the AGREE instrument (strongly recommend, recommend, would not recommend, not recommend or unsure).

2.5. Diagnostic and therapeutic recommendations

Finally, we selected only the most established recommendations that had been included in those guidelines we assessed as "excellent" by our criteria. The methods for grading the levels of evidence and the strength of recommendations were highly variable across all the examined guidelines so that making comparisons was highly problematic [18]. We decided to extract these diagnostic and therapeutic recommendations and report them maintaining the original, albeit different, grading systems used by the authors of each guideline.

To provide greater utility to the reader, we formulated a measure, "clinical inference", to adjudicate the strength of similar recommendations drawn from different guidelines and summarize our own opinions about each kind of recommendation. Each "clinical inference" was derived by considering three evidence elements: (1) the strength of each recommendation as judged and reported by each guideline development group (GDG), (2) the levels of evidence supporting each recommendation as ranked in each guideline and (3) the number of studies concerning each recommendation [19].

Although no validated tool was available to assist this procedure, and the results were simply based on group discussion of all relevant criteria, they conformed to the following scheme for implementing a recommendation based on the strength of the

Table 1
Domains and Items of the Appraisal of Guidelines Research and Evaluation (AGREE) Instrument.

AGREE Domains	AGREE Items
Scope and purpose	1) The overall objective of the guideline is specifically described 2) The clinical question covered by the guideline is specifically described 3) The patients to whom the guideline is meant to apply are specifically described
Stakeholder involvement	4) The guideline development group includes individuals from all the relevant professional groups 5) The patients' views and preferences have been sought 6) The target users of the guideline are clearly defined 7) The guideline has been piloted among target users
Rigor of development	8) Systematic methods were used to search for evidence 9) The criteria for selecting the evidence are clearly described 10) The methods used for formulating the recommendations are clearly described 11) The health benefits, side effects and risks have been considered in formulating the recommendations 12) There is an explicit link between the recommendations and the supporting evidence 13) The guideline has been externally reviewed by experts prior to its publication 14) A procedure for updating the guideline is provided
Clarity and presentation	15) The recommendations are specific and unambiguous 16) The different options for the management of the condition are clearly presented 17) Key recommendations are easily identifiable 18) The guideline is supported with tools for application
Applicability	19) The potential organizational barriers in applying the recommendations have been discussed 20) The potential cost implications of applying the recommendations have been considered 21) The guideline presents key review criteria for monitoring and/or audit purposes
Editorial independence	22) The guideline is editorially independent from the funding body 23) Conflicts of interest of guideline development members have been recorded

aggregated evidence for an intervention: Do - recommendations with strong supporting evidence, Might do - recommendations with moderate supporting evidence, Don't do - recommendations with strong evidence against intervention, Don't know - recommendations with limited or inconclusive evidence.

3. Results

3.1. Selection of guidelines

Our search strategy identified 34 guidelines, of which 21 were potentially relevant but were excluded for different reasons (Fig. 1). Ultimately, 13 guidelines for primary care management of CLBP were included, listed below indicating country and year of publication:

- Institute for Clinical Systems Improvement (ICSI); United States, 2008 [20].

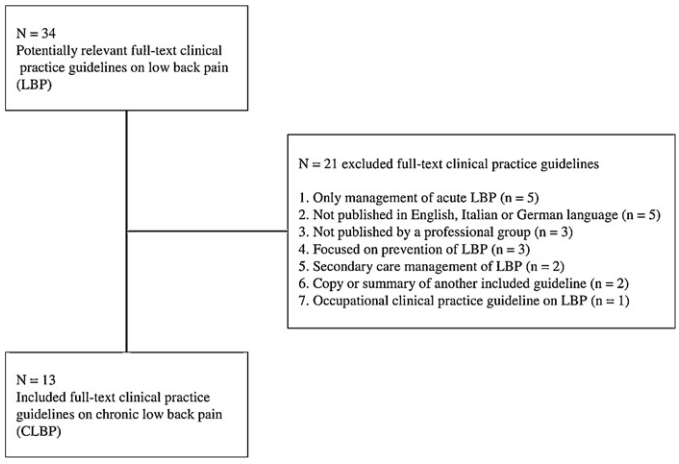

Fig. 1. Study Flow: Summary of inclusion/exclusion selection process of clinical practice guidelines.

- American College of Physicians (ACP); American Pain Society Low Back Pain Guidelines Panel (APS); United States, 2007 [21].
- Institute of Health Economics, Toward Optimized Practice (TOP) Program; Canada, 2009 [15].
- Clinic on Low-Back Pain in Interdisciplinary Practice (CLIP) Guidelines; Canada, 2007 [22].
- New South Wales Therapeutic Assessment Group (NSW TAG); Australia, 2002 [23].
- COST B13 Working Group on Guidelines for Chronic Low Back Pain in Primary Care; Europe, 2004 [24].
- The Care and Research Institute (IRCCS) Don Carlo Gnocchi Foundation, ONLUS; Italy, 2006 [25].
- Royal Dutch Society for Physiotherapy (KNGF); physiotherapy guidelines; the Netherlands, 2003 [26].
- The National Collaborating Centre for Primary Care (NCCPC); Royal College of General Practitioners (RCGP); United Kingdom, 2009 [27].
- The Chartered Society of Physiotherapy (CSP); United Kingdom, 2006 [28].
- Centre of Excellence for Orthopaedic Pain Management Speising (CEOPS); Austria, 2007 [29].
- Drug Committee of the German Medical Society (AKDA); Germany, 2007 [30].
- The German College of General Practitioners and Family Physicians (DEGAM); Germany, 2003 [31].

3.2. Quality assessment

In general, the quality of many guidelines was satisfactory (Table 2). The domain that received the lowest mean score (less than 50% of the maximum possible score) was "Applicability", with 49.54%. On the other hand, the best domains were "Clarity and presentation" and "Scope and purpose" (82.69% and 81.23% respectively). It was also verified that three guidelines [15,27,28] achieved high results in each domain with an overall average quality above 80%. Specifically, in the domain of clarity and presentation, most guidelines presented unambiguous recommendations and well listed management options. Moreover, key recommendations were easily identifiable and all the guidelines were supported by some

Table 2
Domain Scores of the Guidelines Expressed as Percentages.

Guidelines	References	AGREE Domains						Average Quality of Guidelines
		Scope and purpose	Stakeholder involvement	Rigor of development	Clarity and presentation	Applicability	Editorial independence	
TOP (Canada 1)	[15]	**100**[a]	**92**	**98**	94	**89**	**100**	95.50
NCCPC, RCGP (UK 1)	[27]	**100**	89	84	**100**	81	67	86.83
CSP (UK 2)	[28]	**100**	75	90	97	81	50	82.17
KNGF (Netherlands)	[26]	89	75	86	92	33	**100**	79.17
ACP, APS (US 2)	[21]	89	50	90	92	56	83	76.67
ICSI (US 1)	[20]	**100**	64	56	89	56	83	74.67
DEGAM (Germany 2)	[31]	96	61	89	58	67	67	73.00
AKDA (Germany 1)	[30]	67	47	86	83	52	67	67.00
COST B13 (Europe)	[24]	67	56	73	78	44	83	66.83
IRCCS (Italy)	[25]	85	69	67	89	48	33	65.17
CEOPS (Austria)	[29]	67	61	60	50	7	33	46.33
NSW TAG (Australia)	[23]	74	58	32	67	19	17	44.50
CLIP (Canada 2)	[22]	22	25	37	86	11	17	33.00
Domain Mean Scores		81.23	63.23	72.92	82.69	49.54	61.54	

US: United States; UK: United Kingdom; 1: first guideline; 2: second guideline.
[a] Bolded numbers identify the highest scoring guidelines in each domain. Guidelines are listed in descending order according to their average quality rating score.

tools for application, except for one [29]. Similarly, all guidelines except one [22] clearly described their scope and purpose, even if many of them did not get the maximum score for this domain due to missing descriptions of clinical questions and of the target patient population [22–26,29,30]. Most guidelines scored well also on the domain of rigor of development with only three of them [20,22,23] showing several low-moderate scores. However, the description of an external review process by experts before publication was frequently absent. In addition, some guidelines did not provide a procedure for their update [22–24,29].

The involvement of stakeholders in the development of guidelines was moderate. The working group often included individuals from all the relevant disciplines and the target users were almost always mentioned. On the other hand, most guidelines did not perform a pilot test among target users [20–25,29], and many of them did not take into account patients' views and preferences [20,22,24,30] or lacked a description on how they had been included. "Editorial independence" was not wholly described, and ratings were generally poor, although the domain mean score was not the worst (61.54%). Only two [15,26] of the 13 guidelines clearly stated their independence from the funding body; however, most of them described possible conflicts of interests of their members [15,20,21,24,26–31]. Finally, scores were lowest on the domain of applicability. This was particularly evident on item 21, which concerns the presentation of key review criteria for monitoring and/or audit. Potential organizational barriers and cost implications were largely often ignored by the authors of guidelines, with only two [15,27] taking these issues into account.

3.3. High-quality guidelines and their recommendations

According to our ten criteria for categorizing guidelines, we judged five [15,21,26–28] of the 13 guidelines included in the study as "excellent." Of these CPGs, three of them [15,26,28] were not included in the reviews of Koes et al. [11] or Dagenais et al. [12]. One CPG [27] was included only in Dagenais et al. study, while another [21] could be found in both. All of these guidelines provide recommendations on the diagnosis and treatment of CLBP, providing a comprehensive picture of the management of this condition in primary care.

For each guideline, Table 3 reports the levels of evidence and the recommendation grading systems used by their authors. Finally the seven most important diagnostic recommendations with strong supporting evidence (Do) are summarized in Table 4, while Table 5 reports findings from CPGs pertinent to the management of CLBP. The most important therapeutic recommendations included are categorized by: general behaviour, pharmacologic therapy, conservative nonpharmacologic therapy, invasive procedures, referral for surgery.

4. Discussion

Currently, many guidelines have directed their attention to CLBP, providing specific recommendations for this particular condition, which still remains a very important clinical challenge in medicine [32]. After the evaluation of 13 international CPGs on the management of nonspecific CLBP using the validated AGREE instrument, this study presents an overview of the most important diagnostic and therapeutic recommendations found within five high-quality guidelines [15,21,26–28].

4.1. Guidelines evaluation

Similarly to a recent review of Bouwmeester et al. [13], the quality of CPGs has improved over time compared to the disappointing overall quality assessed by previous reviews [8,9]. These results are almost certainly related to the fact that AGREE is internationally more known and widely accepted at present than in 2004 when it had just been published. Some GDGs indeed explicitly mention the use of this instrument to improve the methodological quality of their guidelines [15,28,29]. However, although the direction taken seems to be the right one, some AGREE components still need special attention. Consistent improvement in applicability and editorial independence are needed since these domains are essential for effective guideline implementation. Furthermore, there must be assurance that recommendations have not been biased by third parties. Guideline authors should specifically describe the development of the CPG, particularly the affiliations of contributors and external reviewers. Finally, CPGs should be pilot tested among target users to further validate a GDG's conclusions and increase practitioners' awareness [33].

4.2. From guidelines to clinical practice

While methodological high-quality guidelines are becoming increasingly available, implementation of their recommendations into concrete daily practice remains a thorny task [34]. Although the reason for this difficulty could be initially attributed

Table 3
Evidence Scoring and Recommendation Grading Systems of the five High-quality Guidelines.

Guidelines	Levels of Evidence	Recommendation grading and basis
American College of Physicians (ACP); American Pain Society Low Back Pain Guidelines Panel (APS); United States, 2007 [21]	Good = Evidence includes consistent results from well-designed, well-conducted studies in representative populations that directly assess effects on health outcomes (at least 2 consistent, higher-quality trials)	A = The panel strongly recommends that clinicians consider offering the intervention to eligible patients. The panel found good evidence that the intervention improves health outcomes and concludes that benefits substantially outweigh harms
	Fair = Evidence is sufficient to determine effects on health outcomes, but the strength of the evidence is limited by the number, quality, size, or consistency of included studies; generalizability to routine practice; or indirect nature of the evidence on health outcomes (at least 1 higher-quality trial of sufficient sample size; 2 or more higher-quality trials with some inconsistency; at least 2 consistent, lower-quality trials, or multiple consistent observational studies with no significant methodological flaws)	B = The panel recommends that clinicians consider offering the intervention to eligible patients. The panel found at least fair evidence that the intervention improves health outcomes and concludes that benefits moderately outweigh harms, or that benefits are small but there are no significant harms, costs, or burdens associated with the intervention
	Poor = Evidence is insufficient to assess effects on health outcomes because of limited number or power of studies, large and unexplained inconsistency between higher-quality trials, important flaws in trial design or conduct, gaps in the chain of evidence, or lack of information on important health outcomes	C = The panel makes no recommendation for or against the intervention. The panel found at least fair evidence that the intervention can improve health outcomes, but concludes that benefits only slightly outweigh harms, or the balance of benefits and harms is too close to justify a general recommendation
		D = The panel recommends against offering the intervention. The panel found at least fair evidence that the intervention is ineffective or that harms outweigh benefits
		I = The panel found insufficient evidence to recommend for or against the intervention. Evidence that the intervention is effective is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined
Institute of Health Economics, Toward Optimized Practice (TOP) Program; Canada, 2009 [15]	Systematic review (SR) = as cited by the "seed" guideline(s) or identified from supplementary literature search required by the Ambassador Guideline Development Group (GDG)	Do = The GDG accepted the original recommendation, which provided a prescriptive direction to perform the action or used the term "effective" to describe it
	Randomized controlled trial (RCT) = as cited by the "seed" guideline	Not Recommended = The GDG accepted the original recommendation, which provided a prescriptive direction "not" to perform the action; used the term "ineffective" to describe it; or stated that the evidence does "not support" it
	Nonrandomized trial (NRT) = in the form of nonsystematic/narrative review, nonrandomized comparative study, and case series study, as cited by the "seed" guideline	Do Not Know = The GDG accepted the original recommendation, which did not recommend for or against the action or stated that there was "no evidence", "insufficient or conflicting evidence", or "no good evidence" to support its use
	Guideline (G) = as cited by the "seed" guideline	The GDG supplemented a recommendation or created a new one, based on their collective professional opinion
	Expert opinion (EO) = after examining the individual studies cited by the "seed" guideline(s) or additional SRs on low back pain, the original recommendation was rejected and a new one was drafted based on the collective expert opinion of the Ambassador GDG	
	When no evidence was provided by the "seed" guideline in support of the recommendation, the supporting evidence for that recommendation was labelled as expert opinion	
Royal Dutch Society for Physiotherapy (KNGF); physiotherapy guidelines; the Netherlands, 2003 [26]	Only evidence from systematic reviews and meta-analysis was considered	Strong = Consistent findings in several high-quality randomized controlled trials (RCTs)
		Moderate = Consistent findings in one high-quality RCT and one or more low-quality RCTs Limited or contradictory = One RCT (high or low quality) or inconsistent findings in several RCTs None = No RCTs available

Table 3 (*Continued*)

Guidelines	Levels of Evidence	Recommendation grading and basis
The National Collaborating Centre for Primary Care (NCCPC); Royal College of General Practitioners (RCGP); United Kingdom, 2009 [27]	1++ = High-quality meta-analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias	No specific grading reported (recommendations based on GDG's weighting of the evidence and expert consensus)
	1+ = Well-conducted meta-analyses, systematic reviews of RCTs, or RCTs with a low risk of bias	
	1− = Meta-analyses, systematic reviews of RCTs, or RCTs with a high risk of bias	
	2++ = High-quality systematic reviews of case-control or cohort studies. High-quality case-control or cohort studies with a very low risk of confounding, bias or chance and a high probability that the relationship is causal	
	2+ = Well-conducted case-control or cohort studies with a low risk of confounding, bias or chance and a moderate probability that the relationship is causal	
	2− = case-control or cohort studies with a high risk of confounding, bias, or chance and a significant risk that the relationship is not causal	
	3 = Nonanalytical studies (for example, case reports, case series)	
	4 = Expert opinion, formal consensus	
The Chartered Society of Physiotherapy (CSP); United Kingdom, 2006 [28]	Ia = Evidence obtained from a systematic review of RCTs	A = At least one RCT of overall higher quality and consistency addressing the specific recommendation (evidence levels Ia and Ib)
	Ib = Evidence obtained from at least one RCT	B = Well-conducted clinical studies but not RCTs on the topic of the recommendation (evidence levels IIa, IIb and III)
	IIa = Evidence obtained from at least one well-designed controlled clinical trial without randomisation or a poor quality RCT	C = Evidence from the nominal group technique or other expert committee reports This indicates that directly applicable clinical studies or higher quality are absent (evidence level IV)
	IIb = Evidence obtained from at least one other type of well-designed quasi-experimental study.	D = Recommended good practice based on the clinical experience of the GDG
	III = Evidence obtained from well-designed nonexperimental descriptive studies, such as comparative studies, correlation studies and case studies	
	IV = Evidence obtained from expert committee reports or opinions and/or clinical experience of respected authorities e.g., from the nominal consensus process	
	Where the research evidence was either incomplete or inconsistent then consensus evidence was sought	

SR: systematic review; RCT: randomized controlled trial; NRT: nonrandomized trial; G: guideline; EO: expert opinion; GDG: guideline development group

to guidelines themselves, in reality we found that three of our five selected high-quality guidelines [15,27,28] examined in detail the issue of their implementation, identifying potential facilitators and barriers as well as strategies to manage them. However, agreed-upon quality might not necessarily be followed by acceptance and use [35]. Research is needed to study the effectiveness of these guideline implementation plans, particularly in health professions other than medicine, and demonstrate that implementing these guideline recommendations improves patient outcomes.

4.3. AGREE limitations and strengths

As others authors have reported previously [8,14,19], the AGREE instrument proved itself easy-to-use and transparent providing a useful way to score CPGs' methodological quality. However, it does not assess the clinical content of the guidelines nor the quality of evidence supporting the recommendations, which is a common deficit in all the existing appraisal tools [36]. Nevertheless, by using AGREE, guideline users can undertake preliminary evaluation of recommendations respective to overall guideline quality and then undertake further targeted evaluations of individual high-quality guidelines [19]. For this very reason, we decided not to report the recommendations of all the included guidelines, but to use the methodological quality of a guideline as the basis to explore its clinical context, considering this aspect crucial to ensuring target users are provided with a global overview of rigorous work from every point of view.

4.4. Evidence and Recommendations

Along with the importance of assessing and guaranteeing the methodological quality of a guideline, it is also fundamental for potential users to trust the validity and clinical relevance of its recommendations. In this regard, the authors of previous reviews [8,9] underscored that many guidelines did not explicitly describe how they had identified, selected, and summarized the available evidence. Bouwmeester et al. [13] concluded recently that this item has improved. Moreover our study shows that also the methods used to formulate the recommendations appear to

Table 4
The Most Established Diagnostic Recommendations with Strong Supportive Evidence.

Diagnostic Recommendations	Guidelines	Level of Evidence	Strength of Recommendations
Conduct a focused history, physical and neurologic examination to make a distinction between nonspecific LBP, radicular syndrome, and specific pathology; the so-called diagnostic triage	ACP, APS TOP KNGF NCCPC, RCGP CSP	Moderate[a] SR Not disclosed Not disclosed Not disclosed	Strong[a] Do Non-graded Non-graded Non-graded
Assess red flags, using history taking and examination, to exclude specific serious pathology	ACP, APS TOP KNGF CSP	Moderate[a] EO Not disclosed Not disclosed	Strong[a] Do Non-graded Non-graded
Assess prognostic factors (clinical, psychosocial and work related factors), directly or in case of no improvement, to assess the risk for chronic disabling LBP. These factors are known as "yellow, blue and black flags"	ACP, APS TOP KNGF CSP	Moderate[a] SR Not disclosed Not disclosed	Strong[a] Do Non-graded Non-graded
Assess severity of pain and other components related to LBP as well as associated impairments, functional limitations and problems with participation. Use validated and standardized outcome measures to accomplish this goal, including: Visual analogue scale (VAS) Quebec Back Pain Disability Scale Roland-Morris Disability Questionnaire (RDQ) Oswestry Disability Index (ODI), Quality of life e.g., the short form 36 health survey questionnaire (SF-36), and other measures of psychological status	ACP, APS KNGF CSP	Moderate[a] Not disclosed Not disclosed	Strong[a] Non-graded Non-graded
Do not routinely obtain imaging or other diagnostic tests in patients with nonspecific LBP			
Consider additional investigation, such as diagnostic imaging (Rx, CT, MRI, bone scanning, discography, facet nerve blocks), electromyography or laboratory tests, only in case of red flags, severe low back pain or when progressive neurologic deficits are present	ACP, APS TOP NCCPC, RCGP CSP	Moderate[a] NRT 1 ++ (1 RCT) Not disclosed	Strong[a] Do Non-graded Non-graded
Perform a physical and neurologic examination that includes neuromuscular (muscle strength, sensibility and deep tendon reflexes) and specific tests like the straight leg raising test to assess the presence and severity of nerve root dysfunction	ACP, APS KNGF CSP	Moderate[a] Not disclosed Not disclosed	Strong[a] Non-graded Non-graded
Evaluate patients with persistent low back pain and signs or symptoms of radiculopathy or spinal stenosis with MRI (preferred) or CT to identify potential candidates for surgery or epidural steroid injection	ACP, APS NCCPC, RCGP CSP	Moderate[a] 1 ++ (1 RCT) Not disclosed	Strong[a] Non-graded Non-graded

CT: computed tomography; MRI: magnetic resonance imaging; Rx: radiography (plain X-rays); LBP: low back pain; SR: systematic review; RCT: randomized controlled trial; NRT: nonrandomized trial; EO: expert opinion.

[a] Refer to the original guideline document [21] for the description of the Methods for Recommendation Grading and Evidence Scoring.

be generally more rigorous, more explicit and better explained, although it is sometimes unclear which recommendations are based mainly on scientific evidence and which on a common consensus.

4.4.1. Diagnostic recommendations

The content of the diagnostic recommendations has remained unchanged over the past years and is strongly similar across all the guidelines, as also was found by Koes et al. [11] and Dagenais et al. [12] in their recent reviews. All authors agree that clinicians should conduct a focused history, physical and neurologic examination, undertaking diagnostic triage to assess severity and identify the type of LBP. The assessment of red flags, prognostic factors (often indicated with the terms "yellow, blue and black flags"), severity of pain and functional impact, extreme symptom reporting, prior episodes of LBP and patients expectations are almost always strongly recommended as well. In particular, recommendations regarding the assessment of clinical, work and especially psychosocial risk factors for chronicity are more firmly evident both in the literature [37] and in current guidelines than a decade ago.

4.4.2. Therapeutic recommendations

The fact that many disciplines are involved in treating chronic patients with LBP and the current state of research likely contribute to a lower consensus regarding therapeutic recommendations. However, our study shows that conflicting recommendations about the effectiveness of various interventions have greatly decreased in recent years and all the guidelines are slowly aligning in providing them. Moreover, all the high-quality guidelines now provide therapeutic recommendations for chronic pain that are clearly differentiated from those formulated for acute and, with one exception [21], also for subacute pain. Nevertheless, explicit recommendations are still often ambiguous with respect to clear indications for implementation for a considerable proportion of these interventions.

The most established nongraded therapeutic recommendation supports taking an individual's expectations, beliefs and preferences into account when weighing treatment alternatives. Moreover, all the guidelines explicitly underline the importance of educating and providing patients with information on LBP with regard to their expected course and the possibility of effective prevention and self-care options. An important role should be placed

also on advising patients to remain active and continue with normal activities as far as possible.

Considering the broad array of conservative nonpharmacologic therapeutic interventions practiced by physiotherapists, few are consistently and widely recommended across various guidelines. First of all, there is generally strong evidence that physical activity and therapeutic exercise are effective for the management of CLBP, even if it is not clear which type and quantity is the best. A supervised, patient-specific, graded and active exercise program is nearly always recommended. Next, multidisciplinary approaches and combined physical and psychological (CPP) interventions with cognitive-behavioural therapy and exercise are particularly recommended for people who have received at least one course of less intensive treatment and have high disability and/or significant psychological distress. The same conclusion was reached in a very recent systematic review by van Middelkoop et al. [38], which identified multidisciplinary treatment, behavioural treatment and exercise therapy as the most promising interventions to be provided as conservative treatments in daily management of CLBP.

On the other hand, the recommendations regarding spinal manipulation continue to show some discrepancies probably due to the underlying conflicting evidence [11]. In some guidelines manipulation is recommended, or at least presented as a therapeutic option [21,27], but others find inconclusive evidence [15] or do not recommend it [28]. Finally, the evidence supporting physical agents and modalities (tens, biofeedback, lumbar supports, ultrasound, electro and laser therapy) is almost always limited or inconclusive. These modalities are quite clearly discouraged among all the guidelines, while traction is not recommended.

Table 5
The Most Established Therapeutic Recommendations and Clinical Inferences.

Therapeutic Recommendations	Guidelines	Level of Evidence			Strength of Recommendations			Clinical Inference (do, might do, don't do, don't know)
General behaviour								
Information, education and self-care	ACP, APS	Moderate[a]			Strong[a]			DO
	TOP	SR			Do			
	KNGF	Not disclosed			Not graded			
	NCCPC, RCGP	1- (3 RCT)			Recommend			
	CSP	Not disclosed			Not graded			
Pharmacologic therapy (only for short-term relief)								
Acetaminophen and NSAIDs	ACP, APS	Fair/Good			B			MIGHT DO
	TOP	SR			Do			
	NCCPC, RCGP	1++ (1 SR)			Recommend			
Opioids and tramadol	ACP, APS	Fair			B			MIGHT DO
	TOP	SR			Do			
	NCCPC, RCGP	1+ (3 RCT)			Recommend (for short duration use)			
Benzodiazepines	ACP, APS	Fair			B			MIGHT DO
Antiepileptic drugs	ACP, APS	Fair (for gabapentin) to poor (for topiramate)			C (gabapentin), I (topiramate)			DON'T KNOW
Aspirin	ACP, APS	Poor			I			DON'T KNOW
Muscle relaxants	ACP, APS	Poor			I			MIGHT DO
	TOP	SR			Do (in selected patients)			
Tricyclic antidepressants	ACP, APS	Good			B/C			MIGHT DO
	TOP	SR			Do			
	NCCPC, RCGP	1++ (1 SR)			Recommend			
Conservative nonpharmacologic therapy								
Psychological therapy (mainly cognitive-behavioural therapy or progressive muscle relaxation)	ACP, APS	Fair/Good			B			MIGHT DO
	TOP	SR			Do			
	KNGF	2 MA			Limited or moderate evidence of effectiveness			
	NCCPC, RCGP	1+ (2 RCT)			Not recommend (as stand-alone)			
Physical activity and therapeutic exercise								DO
Exercise programmes	ACP, APS	Good			B			
	TOP	SR			Do			
	KNGF	2 SR			Strong evidence of effectiveness			
	NCCPC, RCGP	1++ (1 SR)			Recommend			
Specific exercises		Pain	PS	Function	Pain	PS	Function	
Mobilising	CSP	IV	IV	IV	C	C	C	
Strengthening	CSP	Ib	IV	IV	A	C	C	
Aerobic	CSP	Ib	IV	Ib	A	C	A	
Unsupervised walking	CSP	No evidence	IV	IV	Not graded	C	C	
General	CSP	Ib	IV	Ib	A	C	A	
Core Stability	CSP	No evidence	No evidence	IV	Not graded	Not graded	C	
McKenzie	CSP	Ib	IV	IV	A	C	C	
Active rehabilitation	TOP	SR			Do			DO
Educational interventions (educational and education-exercise programmes like Back Schools)	ACP, APS	Fair			C			MIGHT DO
	NCCPC, RCGP	1++ (SR), 1+ (4 RCT), 1- (2 RCT)			Recommend (but not as stand-alone)			

Table 5 (*Continued*)

Therapeutic Recommendations	Guidelines	Level of Evidence			Strength of Recommendations			Clinical Inference (do, might do, don't do, don't know)
Brief individualized educational interventions	ACP, APS	Fair			B			MIGHT DO
Multidisciplinary treatment programs, interdisciplinary rehabilitation, CPP interventions	ACP, APS	Good			B			DO
	TOP	SR			Do			
	NCCPC, RCGP	1+ (2 RCT), 1- (10 RCT)			Recommend (high levels of disability and/or psychological distress)			
Manual therapy								MIGHT DO
Spinal manipulation	ACP, APS	Good			B			
	TOP	SR			Don't know			
	NCCPC, RCGP	1++ (1 RCT)			Recommend			
	CSP	Ib			A (Not recommend alone)			
Spinal mobilisation	NCCPC, RCGP	1++ (1 RCT)			Recommend			
Combined manipulation and mobilisation		Pain	PS	Function	Pain	PS	Function	
	CSP	Ib	No evidence	Ib	A	Not graded	A	
Massage	ACP, APS	Fair			B			
	TOP	SR			Do			
	KNGF	1 R			Effectiveness unclear			
	NCCPC, RCGP	1++ (1 SR and 1 RCT)			Recommend			
	CSP	No evidence			Not graded			
Traction	ACP, APS	Fair			D (continuous or intermittent traction), C (auto-traction for sciatica)			DON'T DO
	KNGF	1 SR + 1 RCT			Strong evidence of ineffectiveness			
	NCCPC, RCGP	1++ (1 SR)			Not recommend			
Lumbar supports	ACP, APS	Poor			I			DON'T KNOW
	TOP	RCT			Don't know			
	NCCPC, RCGP	1++ (1 SR)			Not recommend			
Biofeedback	ACP, APS	Poor			I			DON'T DO - DON'T KNOW
	KNGF	1 SR			Moderate evidence of ineffectiveness			
Hydrotherapy (exercise in water)	KNGF	2 RCT			Limited or moderate evidence of effectiveness			MIGHT DO
		Pain	PS	Function	Pain	PS	Function	
	CSP	IV	IV	Ib	C	C	A	
Yoga	ACP, APS	Fair (for Viniyoga) to poor (for Hatha yoga)			B (Viniyoga)			MIGHT DO
TENS	ACP, APS	Poor			I			DON'T KNOW
	TOP	SR			Do			
	KNGF	1 R			Effectiveness unclear			
	NCCPC, RCGP	1+ (2 RCT), 1- (1 RCT)			Not recommend			
Ultrasound therapy, Electrotherapy, Low-level laser therapy	ACP, APS	Poor			I			DON'T DO - DON'T KNOW
	KNGF	1 SR + 1 MA on ultrasound therapy; 1 R on electrotherapy; 1 MA + 1 SR on laser therapy			Effectiveness unclear (because they are passive interventions, they are not recommended)			
	NCCPC, RCGP	1++ (1 SR)			Not recommend			
Invasive procedures								
Acupuncture	ACP, APS	Fair			B			MIGHT DO
	TOP	SR			Do			
	NCCPC, RCGP	1++ (1 SR), 1+ (3 RCT), 1- (1 RCT)			Recommend			
Prolotherapy	TOP	SR			Do (in selected patients)			DON'T KNOW
	NCCPC, RCGP	1++ (1SR)			Not recommend			
Intra-muscular, spinal, epidural or nerve block injections								
Epidural steroid injection	TOP	SR			Do (for patients with leg pain)			MIGHT DO
Facet-joint steroid injection	NCCPC, RCGP	1+ (1 RCT in 1 SR)			Not recommend			DON'T DO - DON'T KNOW
Intra-discal steroid injection	NCCPC, RCGP	1+ (1 RCT)			Not recommend			
Referral for surgery (or other invasive procedures)								
	ACP, APS	Not disclosed			Recommend			MIGHT DO (when patients do not respond to standard noninvasive treatments)
	NCCPC, RCGP	1+ (2 SR), 1+ (1 MA), 1+ (3 RCT), 1- (1 RCT)			Recommend			

4.5. Future directions in research and guideline development

Overall, it seems that recommendations in current guidelines regarding the diagnosis and treatment of CLBP have not changed substantially compared to those included in old guidelines and scientific literature [39] about a decade ago. However, some refinements and valuable results have been obtained as the new consensus in favour of exercise therapy and against traction surely demonstrate.

For the future, it will be very important to achieve consensus on precise recommendations for manual therapy and educational interventions that have currently conflicting, limited or moder-

ate evidence of effectiveness. On the other hand, for some of the already explicitly recommended treatments, *i.e.* behavioural and exercise therapy, it will be important to describe their specific characteristic and clarify which subgroups of patients can benefit most from a particular type of intervention [28,40]. Updates or future guidelines should devote more attention to the definitions of CLBP itself (chronic, persistent, recurrent, etc.), of the interventions, their cost-effectiveness (especially psychological therapy and multidisciplinary treatment), and what constitutes a successful prognosis. Moreover, our work highlights the need to invest effort and resources on developing an internationally validated method to distinguish between "excellent" and "poor" guidelines, identifying the most reliable and available CPGs, and implementing them properly in local health care settings. Finally, a well-established system, as the promising one proposed by the GRADE Working Group [18], should be internationally adopted to grade evidence and recommendations, thus preventing confusion and allowing effective communication and comparison across guidelines and target users.

Disclosure of interest

The authors declare that they have no conflicts of interest concerning this article.

Acknowledgements

The authors would like to thank Angela Longo and Judith Herlemann for technical support, Chiara Scardoni and Cristian Mugnai for English language editing.

Appendix A. Supplementary data

Supplementary material (Table S1) associated with this article can be found at http://www.sciencedirect.com, at doi:10.1016/j.jbspin.2011.03.019.

References

[1] Kent PM, Keating JL. The epidemiology of low back pain in primary care. Chiropr Osteopat 2005;13:13.
[2] van Tulder MW, Koes BW, Bouter LM. Conservative treatment of acute and chronic nonspecific low back pain. A systematic review of randomized controlled trials of the most common interventions. Spine 1997;22:2128–56.
[3] Hestbaek L, Leboeuf-Yde C, Manniche C. Low back pain: what is the long-term course? A review of studies of general patient populations. Eur Spine J 2003;12:149–65.
[4] Quittan M. Management of back pain. Disabil Rehabil 2002;24:423–34.
[5] Sackett DL, Straus S, Richardson WS, et al. Evidence-based medicine. How to practice and teach EBM. 2nd ed. New York: Churchill Livingstone; 2000.
[6] Spitzer WO, Le Blanc FE, Dupuis M, et al. Scientific approach to the assessment and management of activity-related spinal disorders: a monograph for clinicians–report of the Quebec Task Force on Spinal Disorders. Spine 1987;12:S1–59.
[7] Koes BW, van Tulder MW, Ostelo R, et al. Clinical guidelines for the management of low back pain in primary care. An international comparison. Spine 2001;26:2504–14.
[8] van Tulder MW, Tuut M, Pennick V, et al. Quality of primary care guidelines for acute low back pain. Spine 2004;29:E357–62.
[9] Arnau JM, Vallano A, Lopez A, et al. A critical review of guidelines for low back pain treatment. Eur Spine J 2006;15:543–53.
[10] AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. Qual Saf Health Care 2003;12:18–23.
[11] Koes BW, van Tulder M, Lin CW, et al. An updated overview of clinical guidelines for the management of non-specific low back pain in primary care. Eur Spine J 2010;19:2075–94.
[12] Dagenais S, Tricco AC, Haldeman S. Synthesis of recommendations for the assessment and management of low back pain from recent clinical practice guidelines. Spine J 2010;10:514–29.
[13] Bouwmeester W, van Enst A, van Tulder MW. Quality of low back pain guidelines improved. Spine 2009;34:2562–7.
[14] MacDermid JC, Brooks D, Solway S, et al. Reliability and validity of the AGREE instrument used by physical therapists in assessment of clinical practice guidelines. BMC Health Serv Res 2005;5:18.
[15] Institute of Health Economics, Toward Optimized Practice (TOP) Program. Guidelines for the Evidence-Informed Primary Care Management of Low Back Pain. Alberta, Canada; 2009.
[16] Scott NA, Moga C, Harstall C. Making the AGREE tool more user-friendly: the feasibility of a user guide based on Boolean operators. J Eval Clin Pract 2009;15:1061–73.
[17] Hayward RSA, Wilson MC, Tunis SR, et al. Users' guides to the medical literature. VIII. How to use clinical practice guidelines. A. Are the recommendations valid? The Evidence-Based Medicine Working Group. JAMA 1995;274:570–4.
[18] GRADE. Working Group. Grading quality of evidence and strength of recommendations. BMJ 2004;328:1490–7.
[19] Holohan V, Deenadayalan Y, Grimmer K. Evidence-based physiotherapy for acute low back pain: a composite clinical algorithm synthesized from seven recent clinical guidelines. Physiother Can 2006;58:280–92.
[20] Institute for Clinical Systems Improvement (ICSI). Adult Low Back Pain. Bloomington, MN: ICSI; 2005.
[21] Chou R, Qaseem A, Snow V, et al. Clinical Efficacy Assessment Subcommittee of the American College of Physicians; American College of Physicians; American Pain Society Low Back Pain Guidelines Panel. 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 2007;147:478–91.
[22] Rossignol M, Arsenault B, Dionne C, et al. Clinic on Low-Back Pain in Interdisciplinary Practice (CLIP) Guidelines. Montréal: Direction de santé publique, Agence de la santé et des services sociaux de Montréal; 2007.
[23] New South Wales Therapeutic Assessment Group Inc (NSW TAG). Low Back Pain. NSW: Darlinghurst; 2002.
[24] Airaksinen O, Brox JI, Cedraschi C, et al. European guidelines for the management of chronic non-specific low back pain. Eur Spine J 2006;15:S192–300.
[25] Negrini S, Giovannoni S, Minozzi S, et al. Diagnostic therapeutic flow-charts for low back pain patients: the Italian clinical guidelines. Euro Medicophys 2006;42:151–70.
[26] Bekkering GE, Hendriks HJM, Koes BW, et al. National Practice Guideline for the physiotherapeutic management of patients with low back pain. Physiotherapy 2003;89:82–96.
[27] Savigny P, Kuntze S, Watson P, et al. Low back pain: early management of persistent non-specific low back pain. London, UK: National Collaborating Centre for Primary Care and Royal College of General Practitioners; 2009.
[28] Mercer C, Jackson A, Hettinga D, et al. Clinical guidelines for the physiotherapy management of persistent low back pain (LBP). London, UK: Chartered Society of Physiotherapy; 2006.
[29] Friedrich M, Likar R. Evidenz- und konsensusbasierte österreichische Leitlinien für das Management akuter und chronischer unspezifischer Kreuzschmerzen. Wien Klin Wochenschr 2007;119:189–97.
[30] Drug Committee of the German Medical Society. Recommendations for treatment of low back pain. Köln, Germany; 2007 [in German].
[31] Becker A, Chenot JF, Niebling W, et al. The German College of General Practitioners and Family Physicians (DEGAM) low back pain guideline. Omikron Publishing Düsseldorf, 2003 [in German].
[32] Chou R. Evidence-based medicine and the challenge of low back pain: where are we now? Pain Practice 2005;5:153–78.
[33] Côté AM, Durand MJ, Tousignant M, et al. Physiotherapists and use of low back pain guidelines: a qualitative study of the barriers and facilitators. J Occup Rehabil 2009;19:94–105.
[34] Chenot JF, Scherer M, Becker A, et al. Acceptance and perceived barriers of implementing a guideline for managing low back in general practice. Implement Sci 2008;3:7.
[35] Breen AC, van Tulder MW, Koes BW, et al. Mono-disciplinary or multidisciplinary back pain guidelines? How can we achieve a common message in primary care? Eur Spine J 2006;15:641–7.
[36] Vlayen J, Aertgeerts B, Hannes K, et al. A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. Int J Qual Health Care 2005;17:235–42.
[37] Heymans MW, van Buuren S, Knol DL, et al. The prognosis of chronic low back pain is determined by changes in pain and disability in the initial period. Spine J 2010;10:847–56.
[38] van Middelkoop M, Rubinstein SM, Kuijpers T, et al. A systematic review on the effectiveness of physical and rehabilitation interventions for chronic non-specific low back pain. Eur Spine J 2011;20:19–39.
[39] Furlan AD, Clarke J, Esmail R, et al. A critical review of reviews on the treatment of chronic low back pain. Spine 2001;26:E155–62.
[40] van Middelkoop M, Rubinstein SM, Verhagen AP, et al. Exercise therapy for chronic nonspecific low-back pain. Best Pract Res Clin Rheumatol 2010;24:193–204.