

Are Clinical Practice Guidelines for Low Back Pain interventions of high quality and updated? A systematic review using the AGREE II instrument

Greta Castellini

IRCCS Istituto Ortopedico Galeazzi

Valerio Iannicelli

Universita Vita Salute San Raffaele

Matteo Briguglio

IRCCS Istituto Ortopedico Galeazzi

Davide Corbetta

Universita Vita Salute San Raffaele

Luca Maria Sconfienza

IRCCS Istituto Ortopedico Galeazzi

Giuseppe Banfi

IRCCS Istituto Ortopedico Galeazzi

Silvia Gianola (✉ silvia.gianola@grupposandonato.it)

IRCCS Istituto Ortopedico Galeazzi <https://orcid.org/0000-0003-3770-0011>

Research article

Keywords: Low Back Pain, Spine, Clinical Practice Guidelines, Systematic Review, Critical Appraisal, AGREE II, GRADE, Quality of Evidence

Posted Date: September 11th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-42082/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published on October 22nd, 2020. See the published version at <https://doi.org/10.1186/s12913-020-05827-w>.

Abstract

BACKGROUND: Clinical practice guidelines (CPGs) provide recommendations for practice, but the proliferation of CPGs issued by multiple organisations in recent years has raised concern about their quality. The aim of this study was to systematically appraise CPGs quality for low back pain (LBP) interventions and to explore inter-rater reliability (IRR) between quality appraisers. The time between systematic review search and publication of CPGs was recorded.

METHODS: Electronic databases (PubMed, Embase, PEDro, TRIP), guideline organisation databases, websites, and grey literature were searched from January 2016 to January 2020 to identify GPCs on rehabilitative, pharmacological or surgical intervention for LBP management. Four independent reviewers used the Appraisal of Guidelines for Research and Evaluation II (AGREE II) tool to evaluate CPGs quality and record the year the CPGs were published and the year the search strategies were conducted.

RESULTS: A total of 21 CPGs met the inclusion criteria and were appraised. Seven (33%) were broad in scope and involved surgery, rehabilitation or pharmacological intervention. The score for each AGREE II item was: Editorial Independence (median 67%, interquartile range [IQR] 31 – 84%), Scope and Purpose (median 64%, IQR 22 – 83%), Rigour of Development (median 50%, IQR 21 – 72%), Clarity and Presentation (median 50%, IQR 28 – 79%), Stakeholder Involvement (median 36%, IQR 10 – 74%), and Applicability (median 11%, IQR 0 – 46%). The IRR between the assessors was nearly perfect (interclass correlation 0.90; 95% confidence interval 0.88 – 0.91). The median time span was 2 years (range, 1-4), however, 38% of the CPGs did not report the coverage dates for systematic searches.

CONCLUSIONS: We found methodological limitations that affect CPGs quality. A universal database is needed in which guidelines can be registered and recommendations dynamically developed through a living systematic reviews approach to ensure that guidelines are based on updated evidence.

LEVEL OF EVIDENCE: 1

REGISTRATION PROSPERO DETAILS: CRD42019127619.

Introduction

The worldwide point prevalence of Low Back Pain (LBP) is 9.4% (95% CI, 9.0-9.8) in 2010. (1) Next to the common cold, it is one of the commonest reasons why people seek their physician, with a substantial medical social and economic impact for individuals, families, and society due to its high direct and indirect costs. (2-4) Back pain is a leading cause of years lived with disability and the first cause of activity limitation and absence from work. (1) The overall burden of LBP arising from ergonomic exposures at work was estimated at 21.8 million [95% Confidence Interval (CI) 14.5–30.5] disability adjusted life years (DALYs) in 2010. (5) In response to the global burden, numerous CPGs have been issued by medical societies and working groups, providing recommendations for its diagnosis and management. (6, 7) While the principles for developing CPGs are well established, their proliferation has raised concern about quality. Published CPGs appraisals report that the quality is generally poor, though it appears to have recently improved, and that their applicability is generally low. (8, 9) Appraisals of CPGs for LBP (9-14) do not take into account the most recently published guidelines. Since CPGs provide a bridge between scientific literature and clinical decision making, their implementation in clinical practice should be based on recent evidence, and consider as much as possible a wide range of therapeutic choices. (15)

But because 1 out of 5 recommendations in clinical guidelines go out of date within 3 years, the validity of recommendations beyond 3 years is potentially questionable. (16) As a general rule, CPGs should be reviewed every 3 years after their issue. (17) The National Institute of Clinical Excellence (NICE), the benchmark in guidelines production, has stated that “A formal review of the need to update a guideline is usually undertaken by NICE 3 years after its publication”. (18) This is warranted by the time span between the year of running the systematic search strategy during guideline production and the year of publication in a systematic review. (19) This time span is further stretched because guidelines production and dissemination need to be based on systematic reviews. The use of guidelines older than 3 years would be considered unethical in clinical decision making. (20)

Moreover, existing appraisals of guidelines for LBP do not rely on a comprehensive search of the many possible therapeutic options (rehabilitative, pharmacological or surgical) for treating acute and chronic LBP. (21) The scope is an important item in the AGREE II favoring guidelines that are broad in scope rather than those focusing on a particular set of interventions for a specific condition. (22)

With this study, we critically appraised only the most recent evidence-based CPGs for LBP interventions by means of the AGREE (Appraisal of Guidelines Research and Evaluation) II instrument, the gold standard for critical appraisal of guidelines (22, 23), consistent with the assumption that time can influence CPG reliability. Also, we evaluated the inter-rater reliability of AGREE II and recorded the time span as the years between the *date* of last search and period covered by the search and guideline publication date.

Materials And Methods

The reporting of this systematic review fulfils the Preferred Reporting Items for Systematic Reviews and Meta-Analyses. (24, 25) The PRISMA checklist given is in **Supplementary Digital Content 1**. No ethics committee approval was needed. The protocol is registered in PROSPERO (CRD42019127619).

Inclusion and exclusion criteria

In line with the World Health Organization, we defined a CPG as a document containing “systematically developed evidence-based statements that assist providers, patients, policy makers and other stakeholders to make informed decisions on health care and public health policy”. (26)

Inclusion criteria were: (i) the systematic process evaluated the recommendations; (ii) the CPG was focused on rehabilitation, pharmacological or surgical therapeutic intervention for LBP management; (iii) the full text was published in the last 4 years (2016-2020). We used the most up-to-date version and its supplementary documents. No language restrictions were applied. Exclusion criteria were: (i) 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; (ii) not issued by a national or international society (e.g., designed for local use); (iii) declaration of recommendations was based exclusively on consensus statements or systematic reviews or commentary editorials related to published CPGs; (iv) focus on interventions other than therapeutic (e.g., prevention, diagnosis); (v) based on population subgroups (e.g., pregnant women), specific causes (e.g. spondyloarthritis) or mixed/generic population (e.g., musculoskeletal chronic pain).

Information sources and search strategy

We systematically searched the PubMed, Embase, PEDro, and TRIP databases using the adapted terms and keywords derived from the scoping search outlined in the search strategy. We checked guideline organisation databases (e.g., National Institute for Clinical Excellence) and guideline websites (e.g., eGuidelines). **Supplementary Digital Content 2** illustrates the search strategy. Two reviewers (SG, GC) with a solid background in clinical epidemiology ran the search strategy in March 2019 and updated the results in January 2020. Grey literature was searched using Google Scholar and reference lists were screened for further eligible CPGs.

Selection of clinical practice guidelines

Search results were uploaded to Endnote software and duplicates were removed. (27)(28) Two independent reviewers (SG, VI) screened the titles and abstracts according to the eligibility criteria. Full texts were retrieved when abstracts gave insufficient information or in case of disagreement between the two reviewers. When disagreement persisted, a third reviewer was consulted (GC). Rayyan software (<https://rayyan.qcri.org/>) was used to manage screening and selection. (29) Reasons for study exclusion are reported.

Appraisal of clinical practice guidelines

Four independent researchers (MB, GC, SG, VI) appraised each CPG using the AGREE II instrument and recorded with a self-chronometer the time taken for each assessment. The researchers received training in the use of AGREE II. They completed the AGREE II Online Training Tool (<http://www.agreetrust.org/resource-centre/agree-ii-training-tools/>) and participated in two calibration rounds with a sample of four relevant CPGs of varying quality from a previous overview of clinical guidelines for chronic LBP restricted to 2012. (30) The original AGREE tool was published in 2003 has since then been revised in an updated version. The AGREE II instrument (22) consists of 23 items organized into six quality domains: scope and purpose, stakeholder involvement, rigour of development, clarity of presentation, applicability, and editorial independence. Box 1 shown the items and domains of the AGREE II instrument. (31) Answers to items are graded on a 7-point scale from 1 (strongly disagree) to 7 (strongly agree). A standardized score (range, 0 to 100%) was calculated for each domain.

The appraisers completed the first global rating item on a 7-point scale (1=lowest possible quality, 7= highest possible quality) and the second global rating item of recommending the guidelines for use in practice, with one of three options (Yes, Yes, with modifications, and No). One author (VI) calculated the standardised domain score for each of the six domains as recommended by AGREE II. (22, 32) The general data from each CPG were collected: i) authors and year of publication; ii) ex novo, update or adoption/adolpment CPG status; iii) continent of origin; iv) organization/society/association, funding source, conflict of interest. We also extracted content information such as target population, target interventions (i.e., surgery, physical therapy, pharmaceuticals, educational / behavioural, alternative medicine), rating methods for the quality of evidence (e.g., the Grading of Recommendations Assessment, Development and Evaluation - GRADE), presence of a multidisciplinary panel (as defined by AGREE II: potential candidates for a panel group include clinicians, content experts, researchers, policy makers, clinical administrators, and funders; at least one methodology expert), and patient involvement (as defined by AGREE II: to capture patient/public views and preferences).

[box 1]Data synthesis

We used descriptive statistics to summarize the characteristics of CPGs deemed eligible for inclusion. Data are summarized as frequency number (percentage) or median and interquartile range (IQR). We calculated a quality score for each of the six domains of CPGs using the formula presented in the AGREE II User's Manual. (32) The appraisers added notes and completed the two global rating items at the end of each AGREE II assessment. The first global rating item asks appraisers to rate the overall quality of the guideline on a 7-point scale (1=lowest possible quality and 7= highest possible quality). Domain scores are calculated by summing up the appraisers' scores of the individual items in a domain and then scaling the total as a percentage of the maximum possible score for that domain, which is then automatically generated on the platform My AGREE PLUS. (33)

The second global rating item asks whether the appraiser would recommend the guideline for use in practice and to respond with one of three options (Yes, Yes, with modifications, and No).

The first global rating was adopted to formulate the agreement on the overall assessment between the four appraisers measuring the intraclass correlation coefficient (ICC) with 95% confidence interval (CI). The degree of agreement was graded according to Landis and Koch (34): slight (0.01-0.2); fair (0.21-0.4); moderate (0.41-0.6); substantial (0.61-0.8); and almost perfect (0.81-1). Statistical significance was a P value < 0.05. All tests were two-sided. (34) All data analyses were performed using STATA (StataCorp. 2017. *Stata* Statistical Software: *Release 15*. College Station, TX, USA: StataCorp LLC).

Results

Search results

The systematic search retrieved 2502 citations; additional 30 citations were retrieved from the grey literature. A total of 70 CPGs and related documents underwent full-text screening, 25 of which met the inclusion criteria. Four are awaiting assessment (**Fig. 1**). Finally, we appraised 21 CPGs using AGREE II (**Supplementary Digital Content 3**).

Characteristics of CPGs

Table 1 presents the main characteristics of the 21 CPGs: 10 (47.6%) addressed multiple interventions. Rating of evidence quality was planned in 76% of the guidelines and reported in 67%. More than half (52%) had a multidisciplinary panel and less than half (38%) reported patient involvement (**Supplementary Digital Content 4**).

Table 1. Characteristics of CPGs.

REF	Publication year	Country	Status	Topic	Publication dates of systematic search strategy
ACP(35)	2017	USA	Update	Educational / behavioural, physical therapy, pharmaceuticals	2008 - 2015
AIM(36)	2019	USA	Update	Surgery	Not reported
AOA(37)	2016	USA	Update	Physical therapy	2003 - 2014
ASIPP(38)	2019	USA	New	Pharmaceuticals	Not reported
BMA(39)	2018	Brazil	New	Educational / behavioural	Not reported
CAMM(40)	2016	China	New	Alternative medicine	Not reported
CCGI(41)	2018	Canada	New	Educational / behavioural, physical therapy	2015 - 2017
CCGPP(42)	2016	USA	Adoption ¹	Physical therapy	2009 - 2014
CPLA(43)	2018	Latin America	Adoption ²	Physical therapy, pharmaceuticals	2004 - 2014
DSA(44)	2016	Netherlands	New	Surgery	1990 - 2011
GSCI(45)	2018	International	New	Surgery	Not reported
ICSI(46)	2018	USA	Update	Educational / behavioural, physical therapy, pharmaceuticals	2000 - 2017
KCE(47)	2017	Belgium	Adoption ³	Educational / behavioural, physical therapy, pharmaceuticals, surgery	2010 - 2015
KIOM(48)	2017	Korea	New	Physical therapy, pharmaceuticals, alternative medicine	start date not reported - 2015
KSSS(49)	2017	Korea	Adoption ⁴	Educational / behavioural, physical therapy, pharmaceuticals	2000 - 2016
L&I(50)	2016	USA	Update	Surgery	Not reported
NICE(51)	2016	UK	Update	Educational / behavioural, physical therapy, pharmaceuticals, surgery	2013 - 2015
PSP(52)	2017	Poland	New	Physical therapy	Not reported
PSSS(53)	2016	Poland	New	Surgery	Not reported
TOP(54)	2017	USA	Update	Educational / behavioural, physical therapy, pharmaceuticals, surgery	2010 - 2014
VADoD(55)	2017	USA	Update	Educational / behavioural, physical therapy, pharmaceuticals	2006 - 2016

Legend

¹ Three evidence-based clinical practice guidelines for the surgical and interventional management of persistent and disabling spine disorders were selected by consensus (Globe et al. J Manipulative Physiol Ther 2008; Baker et al. Top Integr Health Care, 2012; Farabaugh RJ, J Manipulative Physiol Ther 2010).

² Eleven evidence-based clinical practice guidelines were selected according to the field of interest. Table 4 of original publication contains all references.

³ One evidence-based clinical practice guidelines was selected (de Campos, NICE 2016);

⁴ Three evidence-based clinical practice guidelines were selected according to the field of interest (Dagenais et al. Spine J 2010; Koes et al. Eur Spine J 2010; Chou et al. Ann Intern Med 200).

Note: For full details of CPG see **Supplement Digital Content 3**.

AGREE II domains assessment

Overall, the highest rating AGREE II domain was Editorial Independence (median 67%, interquartile range [IQR] 31 – 84%), followed by Scope and Purpose (median 64%, IQR 22 – 83%), Rigour of Development (median 50%, IQR 21 – 72%), Clarity and Presentation (median 50%, IQR 28 – 79%), Stakeholder Involvement (median 36.1%, IQR 10 – 74%), and Applicability (median 11%, IQR 0 – 46%). In the overall guideline assessment, the median of the overall quality item was 42% (IQR 15 – 67%) and the most frequent recommendation regarding the use of the guideline was “No” (**Table 2**).

The NICE guideline (51) had the highest quality (96%) in the area of Educational/behavioural, physical therapy, pharmaceutical interventions. The Belgian Healthcare Knowledge Centre (KCE) (83%) guideline (56) had high quality and covered the same

interventions plus surgery with a short time span (1 and 2 years, respectively) for searching evidence (**Supplementary Digital Content 4**).

Table 2. Overall domain assessment of CPGs.

Clinical Practice Guideline		Scope and Purpose	Stakeholder Involvement	Rigour of Development	Clarity of Presentation	Applicability	Editorial Independence	OVERALL	
								first global rating (Personal rating)	second global rating (I would recommend?)
NICE (51)	2016	98,61%	95,83%	90,10%	100,00%	65,63%	93,75%	95,83%	Yes
CCGI(41)	2018	90,28%	84,72%	88,54%	90,28%	73,96%	85,42%	87,50%	Yes
KCE(47)	2017	88,89%	77,78%	83,85%	69,44%	86,46%	95,83%	83,33%	Yes
ACP(35)	2017	97,22%	75,00%	76,56%	80,56%	22,92%	93,75%	75,00%	Yes
VADoD(55)	2017	84,72%	90,28%	80,21%	87,50%	41,67%	60,42%	70,83%	Yes, with mod.
ICSI(46)	2018	80,56%	72,22%	68,23%	86,11%	51,04%	79,17%	62,50%	Yes, with mod.
TOP(54)	2017	69,44%	48,61%	66,15%	77,78%	71,88%	75,00%	58,33%	No
KIOM(48)	2017	72,22%	40,28%	47,92%	59,72%	1,04%	83,33%	45,83%	No
CAMM(40)	2016	63,89%	44,44%	35,94%	63,89%	0,00%	29,17%	45,83%	No
GSCI(45)	2018	52,78%	11,11%	25,00%	70,83%	11,46%	75,00%	41,67%	No
ASIPP(38)	2019	66,67%	25,00%	57,29%	33,33%	6,25%	81,25%	41,67%	No
AOA(37)	2016	76,39%	36,11%	52,60%	25,00%	33,33%	58,33%	41,67%	No
CPLA(43)	2018	26,39%	13,89%	27,60%	27,78%	23,96%	93,75%	33,33%	No
CCGPP(42)	2016	43,06%	33,33%	53,13%	47,22%	16,67%	66,67%	29,17%	No
DSA(44)	2016	63,89%	56,94%	50,52%	50,00%	4,17%	45,83%	29,17%	No
BMA(39)	2018	22,22%	0,00%	34,90%	29,17%	0,00%	29,17%	20,83%	No
KSSS(49)	2017	15,28%	2,78%	17,71%	20,83%	0,00%	29,17%	8,33%	No
PSSS(53)	2016	20,83%	2,78%	12,50%	29,17%	0,00%	62,50%	8,33%	No
AIM(36)	2019	18,06%	8,33%	6,77%	27,78%	0,00%	0,00%	4,17%	No
PSP(52)	2017	22,22%	0,00%	1,56%	19,44%	0,00%	33,33%	4,17%	No
L&I(50)	2016	12,50%	13,89%	0,00%	29,17%	0,00%	0,00%	0,00%	No

Inter-rater reliability and time for AGREE II appraisal

Inter-rater agreement was nearly perfect (ICC 0.90; 95% CI 0.88 – 0.91). Guidelines appraisal took 42 minutes on average to complete (95% CI 35-50).

Time to publication

Overall, 38.1% of the CPGs did not report the dates of systematic search strategy, whereas less than half (47.6%) reported a median of 2 years (IQR 1 - 4) from search to publication. Only half provided a search within 1 year after publication (**Table 1**).

Discussion

Here we report the results of quality appraisal using AGREE II of the most recent CPGs for LBP interventions (published January 2016 to January 2020) that we retrieved by systematic search of electronic medical databases and guidelines websites. A key finding was the variability in the quality of the CPGs across all six AGREE II domains; the highest average scores (>60%) were recorded for *Domain 6 - Editorial Independence* and the *Domain 1 - Scope and Purpose* and the lowest (<15%) for *Domain 5 - Applicability*. The overall quality was rated low and the most frequent response for guideline recommendation was “No” (15 out of 21 CPGs).

Our findings are shared by previous appraisals of CPGs for rehabilitation (57) and other contexts (8, 58, 59) that suggest room for improvement regarding rigour of development, stakeholder involvement, and applicability (8, 58, 59). While only half of the

CPGs were noted to have acceptable rigour of development (*Domain 3 - Rigour of Development*), the variability in this domain was considerable. A low score for this domain is worrying, as it has been identified as a strong predictor of quality by the AGREE instrument. (8) Regression analysis showed a statistically significant influence of the assessment of the items in this domain on overall guideline quality. (60) The item assessing the systematic search can have great importance (i.e., "Item 7: Systematic methods were used to search for evidence") because CPGs ought to be based on recently updated evidence. However, we found that less than half did not report the time coverage of systematic search and, when reported, it ranged from 1 to 4 years before publication. Two-thirds of the CPGs in our sample adequately planned and judged the body of the evidence linked to recommendations (e.g., GRADE). However, because the application of a system for grading the evidence (i.e., GRADE) cannot always ensure inclusion of the most updated evidence within an acceptable time span, reliability should be evaluated with caution.

The validity of each recommendation, and of the CPG, is determined by the methodological quality and the transparency of its development and by the "living evidence" on which it is based. As suggested by Garcia et al., waiting more than 3 years to review a guideline is potentially too long, in which case the recommendations may be outdated by the time of guideline publication. (16) This critical issue has been addressed by the living CPGs concept, (61) which draws inspiration from the established model of living systematic reviews, where evidence is continuously updated and incorporated as soon as available in the literature through a process of continuous surveillance. (62) Accordingly, AGREE II should place importance on timing and rate CPG a high-quality score when the search is conducted within 2 years of completion of the review. (63)

Less than one third of the CPGs in this sample met the AGREE II criterion for participation of patients and their advocate (*Domain 2 - Stakeholder Involvement*). Guideline developers need to prioritize patient and stakeholder involvement starting from the early stages of CPG development. They should be actively involved as members on guideline panels and their comments and inputs included in the draft guideline. (64) Furthermore, evidence suggests that involvement of patients and stakeholders leads to the inclusion of patient-relevant topics and enhances CPG implementation. (65) Unfortunately, development and implementation are erroneously considered as separate activities. (8) In our appraisal, the poorest score was recorded for CPGs applicability (*Domain 5 - Applicability*), with results similar to other CPGs in rehabilitation (57) and other conditions. (8, 12, 66-68) CPGs can provide healthcare professionals with the necessary guidance to access the best research evidence efficiently. Nonetheless, they have little effect on changing clinical behavior.

Only half of the CPGs in our sample were rated satisfactory for adequacy of the reporting of recommendations and options for management (*Domain 4 - Clarity of Presentation*). This may be related to the purpose of AGREE II: the current version makes no distinction between quality of reporting and quality of conduct of a CPG. Despite good reporting, the methodological conduct underlying a guideline can still be weak. (69) Quality of conduct and reporting should be judged separately, just as for all other study designs. (70, 71) In systematic reviews, for instance, PRISMA and the AMSTAR assess the quality of reporting and the quality of conduct, respectively. (72)

We recorded high compliance of the CPGs with the overall aim of the guideline, the clinical question, and the target population (*Domain 1 - Scope and Purpose*). This could be explained by the focus on LBP, which is the most prevalent musculoskeletal condition for which guidelines are needed in view of the years lived with disability in most countries. (73) Lastly, we recorded high compliance of the CPGs with the reporting of sources of support (*Domain 6 - Editorial Independence*). Given the global socioeconomic burden of LBP and the need for care, CPGs must report the presence and management of conflict of interests.

Strengths and limitations

Our appraisal has several strengths. We performed an exhaustive search that included explicit eligibility criteria and independent duplicate assessment of eligibility. Four reviewers were involved in the appraisal, with a nearly perfect inter-rater reliability. While all appraisers were trained in the use of AGREE II, it should be acknowledged that the appraisers shared a similar background (methodology and rehabilitation), which may partially explain the high overall agreement. Indeed, our team included clinical experts and methodologists with experience in clinical epidemiology, including systematic reviews and CPGs. Even after receiving the same training however, guideline appraisers from different areas may still interpret the items and the

scoring system differently. (74) Furthermore, it is possible that the appraisers, basing the assessment on their own experience, paid more attention to assessing the quality of reporting than the quality of conduct and vice versa. We analysed a reliable subset of CPGs restricted to LBP in order to ensure consistency of appraisal, while avoiding discrepancies in item judgement due to different clinical contexts (e.g., AGREE II to assess CPGs in oncology differs from orthopaedics). We focused on the most recent guideline versions in order to offer stakeholders, policy makers, clinicians, and patients the latest evidence for the effectiveness of interventions. However, selecting the CPGs was a challenge, since the definition of guidelines is not universally established and the meaning of consensus and that of evidence-based CPG are sometimes confused. The rigour of methods and panel of experts have to be simultaneously considered in a CPG, but the current definition does not explicate these elements.

A possible limitation of our work is linked to characteristics of the AGREE II itself. It focuses on the quality of the development of CPGs, but this is not sufficient to ensure implementation of single clinical recommendations and improvement in health outcomes. (75) While high-quality CPGs can guarantee rigour in the production of recommendations, their implementation depends largely on how health care professionals decide whether or not to implement a single recommendation in the balance between content (strength and direction of a recommendation), clinical expertise, and resources available. The implementation of a single clinical recommendation cannot be disjointed from overall CPG quality.

Future spin for research

At the time of its publication, a CPG can already be outdated and so will not reflect the most recent evidence. Indeed, time can influence its reliability: (a) during the conduction of systematic reviews for the production of the body of the evidence needed during CPG development; (b) between finalization of a CPG and its publication. In order to avoid waste of effort and of resources due to duplication of CPGs or CPGs outdated before their time, we urge for the creation of a universal database in which guidelines can be registered and updated along the lines of registers for RCTs (e.g., WHO or clinicaltrials.gov) and systematic reviews (e.g., PROSPERO) but for CPGs. In this way, a “living and dynamic” development of recommendations can be better recognized by identifying the most recent literature. (76)

Conclusion

We found methodological limitations affecting CPG quality. Our work highlights the importance of adoption of high quality and updated CPGs to guarantee the validity of a single recommendations, notwithstanding the possibility that implementation of each single recommendation may be the result of a balanced decision between content (strength and direction of a recommendation), clinical expertise, and available resources. We call for a universal database in which guidelines can be registered and recommendations dynamically developed through a living systematic reviews approach to ensure that CPGs are based on recent evidence.

Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Competing interests: The authors declare no competing interests.

Funding: The work was supported by the Italian Ministry of Health “Linea 3 – Valutazione della qualità delle attuali linee guida in ortopedia e in riabilitazione” L3042. The funding sources had no controlling role in the study design, data collection, analysis, interpretation or report writing.

Availability of data and materials: All data generated or analysed during this study are included in this published article with all additional materials. Raw data are stored at the following link: https://osf.io/xwbu2/?view_only=d3aa81b467874b468bd1207d96df7376

Authors' Contributions:

Concept development providing idea for the research: SG, CG

Design planning the methods to generate the results: SG, CG, VI

Supervision: SG, GC

Data collection: SG, CG, VI, MB

Analysis and interpretation (statistics, evaluation and presentation of the results): VI, SG, GC, DC

Writing: VI, SG, GC

Critical review (revised manuscript for intellectual content; this does not relate to spelling and grammar checking): DC, GB, LMS

Abbreviations

A Measurement Tool to Assess systematic Reviews (AMSTAR); Appraisal of Guidelines for Research and Evaluation II (AGREE II); Confidence Interval (CI); Clinical Practice Guidelines (CPGs); disability adjusted life years (DALYs); Grading of Recommendations Assessment, Development and Evaluation (GRADE); Interquartile Range (IQR); Intraclass correlation (ICC); Low Back Pain (LBP); National Institute of Clinical Excellence (NICE); Physiotherapy Evidence Database (PEDro); Preferred Reporting Intervention for Systematic Review and Meta-analysis (PRISMA); Turning Research into Practice (TRIP); World Health Organization (WHO).

References

1. Hoy D, March L, Brooks P, Blyth F, Woolf A, Bain C, et al. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. *Annals of the rheumatic diseases*. 2014;73(6):968-74.
2. Vrbancic TS. [Low back pain—from definition to diagnosis]. *Reumatizam*. 2011;58(2):105-7.
3. Deyo RA, Phillips WR. Low back pain. A primary care challenge. *Spine*. 1996;21(24):2826-32.
4. Childs JD, Fritz JM, Flynn TW, Irrgang JJ, Johnson KK, Majkowski GR, et al. A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: a validation study. *Annals of internal medicine*. 2004;141(12):920-8.
5. Hoy D, March L, Brooks P, Woolf A, Blyth F, Vos T, et al. Measuring the global burden of low back pain. *Best practice & research Clinical rheumatology*. 2010;24(2):155-65.
6. O'Connell NE, Cook CE, Wand BM, Ward SP. Clinical guidelines for low back pain: A critical review of consensus and inconsistencies across three major guidelines. *Best practice & research Clinical rheumatology*. 2016;30(6):968-80.
7. O'Sullivan K, O'Keefe M, O'Sullivan P. NICE low back pain guidelines: opportunities and obstacles to change practice. *British journal of sports medicine*. 2017;51(22):1632-3.
8. Alonso-Coello P, Irfan A, Sola I, Gich I, Delgado-Noguera M, Rigau D, et al. The quality of clinical practice guidelines over the last two decades: a systematic review of guideline appraisal studies. *Quality & safety in health care*. 2010;19(6):e58.
9. Meroni R, Piscitelli D, Ravasio C, Vanti C, Bertozzi L, De Vito G, et al. Evidence for managing chronic low back pain in primary care: a review of recommendations from high-quality clinical practice guidelines. *Disability and rehabilitation*. 2019:1-15.
10. van Tulder MW, Tuut M, Pennick V, Bombardier C, Assendelft WJ. Quality of primary care guidelines for acute low back pain. *Spine*. 2004;29(17):E357-62.
11. Bouwmeester W, van Enst A, van Tulder M. Quality of low back pain guidelines improved. *Spine*. 2009;34(23):2562-7.

12. Doniselli FM, Zanardo M, Manfre L, Papini GDE, Rovira A, Sardanelli F, et al. 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. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society.* 2018;27(11):2781-90.
13. Dagenais S, Tricco AC, Haldeman S. Synthesis of recommendations for the assessment and management of low back pain from recent clinical practice guidelines. *The spine journal : official journal of the North American Spine Society.* 2010;10(6):514-29.
14. Ng JY, Mohiuddin U. Quality of complementary and alternative medicine recommendations in low back pain guidelines: a systematic review. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society.* 2020;29(8):1833-44.
15. Gurgel RK. Updating Clinical Practice Guidelines: How Do We Stay Current? *Otolaryngology–head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery.* 2015;153(4):488-90.
16. Martinez Garcia L, Sanabria AJ, Garcia Alvarez E, Trujillo-Martin MM, Etxeandia-Ikobaltzeta I, Kotzeva A, et al. The validity of recommendations from clinical guidelines: a survival analysis. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne.* 2014;186(16):1211-9.
17. Shekelle PG, Ortiz E, Rhodes S, Morton SC, Eccles MP, Grimshaw JM, et al. Validity of the Agency for Healthcare Research and Quality clinical practice guidelines: how quickly do guidelines become outdated? *Jama.* 2001;286(12):1461-7.
18. (NICE) TNIfHaCE. Developing NICE guidelines: the manual. Process and methods, Published: 31 October 2014, niceorguk/process/pmg20

©. 2014.

19. Yoshii A, Plaut DA, McGraw KA, Anderson MJ, Wellik KE. Analysis of the reporting of search strategies in Cochrane systematic reviews. *Journal of the Medical Library Association : JMLA.* 2009;97(1):21-9.
20. Moher D. Reporting guidelines: doing better for readers. *BMC medicine.* 2018;16(1):233.
21. Wong JJ, Cote P, Sutton DA, Randhawa K, Yu H, Varatharajan S, et al. Clinical practice guidelines for the noninvasive management of low back pain: A systematic review by the Ontario Protocol for Traffic Injury Management (OPTIMa) Collaboration. *European journal of pain.* 2017;21(2):201-16.
22. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne.* 2010;182(18):E839-42.
23. Brouwers MC, Kerkvliet K, Spithoff K, Consortium ANS. The AGREE Reporting Checklist: a tool to improve reporting of clinical practice guidelines. *Bmj.* 2016;352:i1152.
24. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. *Open medicine : a peer-reviewed, independent, open-access journal.* 2009;3(3):e123-30.
25. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *Bmj.* 2009;339:b2700.
26. WHO WHO. WHO handbook for guideline development, 2nd ed. World Health Organization. <http://www.who.int/iris/handle/10665/145714>. 2014.
27. Eapen BR. EndNote 7.0. *Indian journal of dermatology, venereology and leprology.* 2006;72(2):165-6.
28. Bramer WM, Milic J, Mast F. Reviewing retrieved references for inclusion in systematic reviews using EndNote. *Journal of the Medical Library Association : JMLA.* 2017;105(1):84-7.
29. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Systematic reviews.* 2016;5(1):210.

30. Pillastrini P, Gardenghi I, Bonetti F, Capra F, Guccione A, Mugnai R, et al. An updated overview of clinical guidelines for chronic low back pain management in primary care. *Joint bone spine*. 2012;79(2):176-85.
31. Appraisal of guidelines for research and evaluation II: AGREE II instrument [http://www.agreetrust.org/wp-content/uploads/2013/10/AGREE-II-Users-Manual-and-23-item-Instrument_2009_UPDATE_2013.pdf].
32. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. Development of the AGREE II, part 2: assessment of validity of items and tools to support application. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*. 2010;182(10):E472-8.
33. Makarski J, Brouwers MC, Enterprise A. The AGREE Enterprise: a decade of advancing clinical practice guidelines. *Implementation science : IS*. 2014;9:103.
34. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-74.
35. Qaseem A, Wilt TJ, McLean RM, Forciea MA, Clinical Guidelines Committee of the American College of P. Noninvasive Treatments for Acute, Subacute, and Chronic Low Back Pain: A Clinical Practice Guideline From the American College of Physicians. *Annals of internal medicine*. 2017;166(7):514-30.
36. AIM. AIM Specialty Health - Musculoskeletal Program - Clinical Appropriateness Guidelines for Spine Surgery. https://aimspecialtyhealthcom/guidelines/PDFs/2019/May18/AIM_Guidelines_MSK_Spine-Surgery.pdf.
37. Task Force on the Low Back Pain Clinical Practice G. American Osteopathic Association Guidelines for Osteopathic Manipulative Treatment (OMT) for Patients With Low Back Pain. *The Journal of the American Osteopathic Association*. 2016;116(8):536-49.
38. Navani A, Manchikanti L, Albers SL, Latchaw RE, Sanapati J, Kaye AD, et al. Responsible, Safe, and Effective Use of Biologics in the Management of Low Back Pain: American Society of Interventional Pain Physicians (ASIPP) Guidelines. *Pain physician*. 2019;22(1S):S1-S74.
39. Brazilian Medical A, Silvinato A, Simoes RS, Buzzini RF, Bernardo WM. Lumbar herniated disc treatment with percutaneous hydrodiscectomy. *Revista da Associacao Medica Brasileira*. 2018;64(9):778-82.
40. Zhao H, Liu B, Liu Z, Xie L, Fang Y, Zhu Y, et al. Clinical practice guidelines of using acupuncture for low back pain. . *World Journal of Acupuncture - Moxibustion* 2016;26:4 (1-13) (<https://www.sciencedirect.com/science/article/pii/S1003525717300168>).
41. Bussieres AE, Stewart G, Al-Zoubi F, Decina P, Descarreaux M, Haskett D, et al. Spinal Manipulative Therapy and Other Conservative Treatments for Low Back Pain: A Guideline From the Canadian Chiropractic Guideline Initiative. *Journal of manipulative and physiological therapeutics*. 2018;41(4):265-93.
42. Globe G, Farabaugh RJ, Hawk C, Morris CE, Baker G, Whalen WM, et al. Clinical Practice Guideline: Chiropractic Care for Low Back Pain. *Journal of manipulative and physiological therapeutics*. 2016;39(1):1-22.
43. Amescua-Garcia C, Colimon F, Guerrero C, Jreige Iskandar A, Berenguel Cook M, Bonilla P, et al. Most Relevant Neuropathic Pain Treatment and Chronic Low Back Pain Management Guidelines: A Change Pain Latin America Advisory Panel Consensus. *Pain medicine*. 2018;19(3):460-70.
44. Itz CJ, Willems PC, Zeilstra DJ, Huygen FJ, Dutch Society of A, Dutch Orthopedic A, et al. Dutch Multidisciplinary Guideline for Invasive Treatment of Pain Syndromes of the Lumbosacral Spine. *Pain practice : the official journal of World Institute of Pain*. 2016;16(1):90-110.
45. Acaroglu E, Nordin M, Randhawa K, Chou R, Cote P, Mmopelwa T, et al. The Global Spine Care Initiative: a summary of guidelines on invasive interventions for the management of persistent and disabling spinal pain in low- and middle-income communities. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society*. 2018;27(Suppl 6):870-8.
46. Thorson D, Campbell R, Massey M, Mueller B, McCathie B, Richards H, et al. Low Back Pain, Adult Acute and Subacute. Revision date: March 2018 / 16th edition. Available from: <https://www.icsi.org/wp-content/uploads/2019/01/March-2018-LBP-Interactive.pdf>.

47. Van Wambeke P, Desomer A, Ailliet L, Berquin A, Demoulin C, Depreitere B, et al. Low back pain and radicular pain: assessment and management. Good Clinical Practice (GCP) Brussels: Belgian Health Care Knowledge Centre (KCE). KCE Reports 287. D/2017/10.273/36. Available from: https://kce.fgov.be/sites/default/files/atoms/files/KCE_287_Low_back_pain_Report_2.pdf
- 2017.
48. Jun J, Cha Y, Lee J, Choi J, Choi T-Y, Park W, et al. Korean medicine clinical practice guideline for lumbar herniated intervertebral disc in adults: An evidence based approach. *European Journal of Integrative Medicine* 2017;9 (18-26). Available from: <https://www.sciencedirect.com/science/article/abs/pii/S1876382017300033>
49. Hong JY, Song KS, Cho JH, Lee JH. An Updated Overview of Low Back Pain Management in Primary Care. *Asian spine journal*. 2017;11(4):653-60.
50. Surgical Guideline for Lumbar Fusion (Arthrodesis) - Washington State Dept. of Labor & Industries (L&I). Available from: <https://www.lni.wa.gov/ClaimsIns/Files/OMD/MedTreat/LumbarfusionUpdate020216.pdf>.
51. de Campos T. Low Back Pain and Sciatica in Over 16s: Assessment and Management. National Institute for Health and Care Excellence: Clinical Guidelines. London 2016. Available from: <https://www.nice.org.uk/guidance/ng59/resources/low-back-pain-and-sciatica-in-over-16s-assessment-and-management-pdf-1837521693637>. Low Back Pain and Sciatica in Over 16s: Assessment and Management. National Institute for Health and Care Excellence: Clinical Guidelines. London 2016.
52. Kassolik k, Rajkowska-labon E, Tomasik T, Pisula-lewadowska A, Gieremek K, Andrzejewski W, et al. Recommendations of the Polish society of Physiotherapy, the Polish society of Family medicine and the college of Family Physicians in Poland in the field of physiotherapy of back pain syndromes in primary health care. . *FamMedPrimCare Rev* 2017;19(3): 323–334. doi: <https://doi.org/10.5114/fmpcr.2017.69299>.
53. Latka D, Miekisiak G, Jarmuzek P, Lachowski M, Kaczmarczyk J. Treatment of lumbar disc herniation with radiculopathy. Clinical practice guidelines endorsed by The Polish Society of Spinal Surgery. *Neurologia i neurochirurgia polska*. 2016;50(2):101-8.
54. Toward Optimized Practice (TOP) Low Back Pain Working Group. 2017 December. Evidence-informed primary care management of low back pain: Clinical practice guideline. 3rd Edition – minor revision 2017.
55. Pangarkar S. Low Back Pain Work Group. VA/DoD clinical practice guideline for diagnosis and treatment of Low Back Pain. . 2017.
56. Van Wambeke P, Desomer A, Ailliet L, Berquin A, Demoulin C, Depreitere B, et al. Low back pain and radicular pain: assessment and management. Good Clinical Practice (GCP) Brussels: Belgian Health Care Knowledge Centre (KCE). 2017. KCE Reports 287. D/2017/10.273/36.
57. Dijkers MP, Ward I, Annaswamy T, Dedrick D, Feldpausch J, Moul A, et al. Quality of Rehabilitation Clinical Practice Guidelines: An Overview Study of AGREE II Appraisals. *Archives of physical medicine and rehabilitation*. 2020;101(9):1643-55.
58. Armstrong JJ, Goldfarb AM, Instrum RS, MacDermid JC. Improvement evident but still necessary in clinical practice guideline quality: a systematic review. *J Clin Epidemiol*. 2017;81:13-21.
59. Gagliardi AR, Brouwers MC. Do guidelines offer implementation advice to target users? A systematic review of guideline applicability. *BMJ open*. 2015;5(2):e007047.
60. Hoffmann-Esser W, Siering U, Neugebauer EA, Brockhaus AC, Lampert U, Eikermann M. Guideline appraisal with AGREE II: Systematic review of the current evidence on how users handle the 2 overall assessments. *PloS one*. 2017;12(3):e0174831.
61. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. *Annals of internal medicine*. 2007;147(4):224-33.

62. Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, et al. Living systematic review: 1. Introduction-the why, what, when, and how. *J Clin Epidemiol.* 2017;91:23-30.
63. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *Bmj-Brit Med J.* 2017;358.
64. Medicine Io. *Clinical Practice Guidelines We Can Trust.* Washington, DC: The National Academies Press. <https://doi.org/10.17226/13058>. 2011.
65. Armstrong MJ, Mullins CD, Gronseth GS, Gagliardi AR. Impact of patient involvement on clinical practice guideline development: a parallel group study. *Implementation science : IS.* 2018;13(1):55.
66. Acuna SA, Huang JW, Scott AL, Micic S, Daly C, Brezden-Masley C, et al. Cancer Screening Recommendations for Solid Organ Transplant Recipients: A Systematic Review of Clinical Practice Guidelines. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons.* 2017;17(1):103-14.
67. Tong A, Chapman JR, Wong G, de Bruijn J, Craig JC. Screening and follow-up of living kidney donors: a systematic review of clinical practice guidelines. *Transplantation.* 2011;92(9):962-72.
68. Acuna-Izcaray A, Sanchez-Angarita E, Plaza V, Rodrigo G, de Oca MM, Gich I, et al. Quality assessment of asthma clinical practice guidelines: a systematic appraisal. *Chest.* 2013;144(2):390-7.
69. Jarl G, Hellstrand Tang U, Norden E, Johannesson A, Rusaw DF. Nordic clinical guidelines for orthotic treatment of osteoarthritis of the knee: A systematic review using the AGREE II instrument. *Prosthetics and orthotics international.* 2019:309364619857854.
70. Chen Y, Yang K, Marusic A, Qaseem A, Meerpohl JJ, Flottorp S, et al. A Reporting Tool for Practice Guidelines in Health Care: The RIGHT Statement. *Annals of internal medicine.* 2017;166(2):128-32.
71. Huwiler-Muntener K, Juni P, Junker C, Egger M. Quality of reporting of randomized trials as a measure of methodologic quality. *Jama.* 2002;287(21):2801-4.
72. Pussegoda K, Turner L, Garritty C, Mayhew A, Skidmore B, Stevens A, et al. Identifying approaches for assessing methodological and reporting quality of systematic reviews: a descriptive study. *Systematic reviews.* 2017;6(1):117.
73. Disease GBD, Injury I, Prevalence C. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet.* 2018;392(10159):1789-858.
74. Marciano NJ, Merlin TL, Bessen T, Street JM. To what extent are current guidelines for cutaneous melanoma follow up based on scientific evidence? *International journal of clinical practice.* 2014;68(6):761-70.
75. Watine J. Is it time to develop AGREE III? *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne.* 2019;191(43):E1198.
76. Akl EA, Meerpohl JJ, Elliott J, Kahale LA, Schunemann HJ, Living Systematic Review N. Living systematic reviews: 4. Living guideline recommendations. *J Clin Epidemiol.* 2017;91:47-53.

Figures

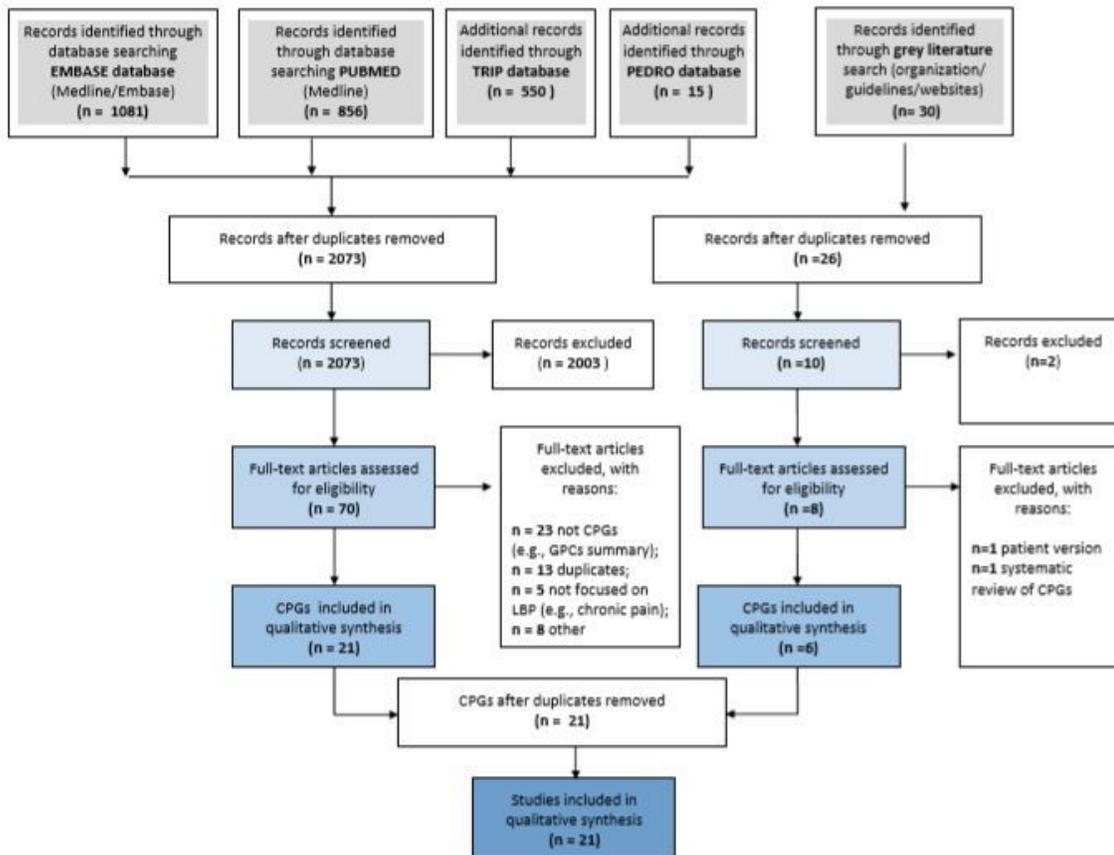


Figure 1

Flow diagram of CPG selection

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaldigitalcontent1LiteratureSearchstrategy.docx](#)
- [SupplementaryPRISMA2009checklist.doc](#)
- [Box1.docx](#)
- [Supplementaldigitalcontent2.docx](#)
- [Supplementarydigitalcontent3.xlsx](#)