CrossMark

ORIGINAL ARTICLE

Quality assessment of osteoporosis clinical practice guidelines for physical activity and safe movement: an AGREE II appraisal

James Jacob Armstrong[1] · Isabel Braganca Rodrigues[2] · Tom Wasiuta[1] ·
Joy C. MacDermid[3]

Abstract

Summary Many osteoporosis clinical practice guidelines are published, and the extent to which physical activity and safe movement is addressed varies. To better inform clinical decision-making, a quality assessment and structured analysis of recommendations was undertaken. Guideline quality varied substantially, and improvement is necessary in physical activity and safe movement recommendations.

Purpose The purpose of the present study is to survey available osteoporosis clinical practice guidelines (CPGs) containing physical activity and safe movement recommendations in order to assess the methodological quality with which they were developed. An analysis of the various physical activity and safe movement recommendations was conducted to determine variability between CPGs.

Methods An online literature search revealed 19 CPGs meeting our inclusion criteria. Three independent scorers evaluated CPG quality using the Appraisal of Guidelines for Research and Evaluation version II (AGREE II) instrument. Two separate individuals used a standard table to extract relevant recommendations.

Results Intra-reviewer AGREE II score agreement ranged from fair to good (intra-class correlation coefficient (ICC) = 0.34 to 0.65). The quality of the 19 included CPGs was variable (AGREE sub-scores: 14 to 100 %). CPGs scored higher in the "scope and purpose" and "clarity of presentation" domains. They scored the lowest in "applicability" and "editorial independence." Four CPGs were classified as high quality, ten average quality, and five low quality. Most CPGs recommended weight-bearing, muscle-strengthening, and resistance exercises. Information on exercise dosage, progression, and contraindications was often absent. Immobility and movements involving spinal flexion and/or torsion were discouraged.

Conclusions There were several high-quality CPGs; however, variability in quality and lack of specific parameters for implementation necessitates caution and critical examination by readers. CPG development groups should pay special attention to the clinical applicability of their CPGs as well as fully disclosing conflicts of interest. CPGs were in general an agreement regarding safe physical activity and safe movement recommendations. However, recommendations were often vague and the more specific recommendations were inconsistent between CPGs.

✉ James Jacob Armstrong
  jarmstrong2018@meds.uwo.ca

  Isabel Braganca Rodrigues
  rodrigib@mcmaster.ca

  Tom Wasiuta
  twasiuta2018@meds.uwo.ca

  Joy C. MacDermid
  jmacderm@uwo.ca

[1] Schulich School of Medicine and Dentistry, 1465 Richmond St., London, ON N6G 2M1, Canada

[2] McMaster University School of Rehabilitation Science, 1225 Royal York Rd., Toronto, ON M9A 4B5, Canada

[3] Hand and Upper Limb Center Clinical Research Lab, 930 Richmond St., London, ON N6A 3J4, Canada

Springer

Introduction

Osteoporosis is a disease characterized by reduced bone mass and deterioration of bone micro-architecture. Patients are usually diagnosed using bone mineral density (BMD) scan measurements at the hip and spine [1]. These bone abnormalities result in enhanced bone fragility, and as a consequence, patients with osteoporosis have a much higher fracture risk. The likelihood of developing osteoporosis is highest in North America and Europe. However, incidence rates in developing countries are predicted to rise as population longevity continues to increase in these regions [2]. The global burden of osteoporosis is substantial with an estimated 9.0 million osteoporotic fractures occurring per year [3].

Current mainstays of treatment are bisphosphonates, calcium, and vitamin D. However, there is evidence that these therapies cannot address non-BMD-related determinants of fracture risk [4, 5]. Previous work has suggested that targeting the non-BMD determinants of fracture risk may be a more effective means of treatment in some patients, or at the least somewhat helpful to most [6]. Exercise programs designed to improve balance, strength, and coordination are simple yet cost-effective interventions that may lower a patient's risk of falling and experiencing negative health outcomes [7]. In addition, certain types of exercise have been shown to improve BMD through the simulation of bone remodeling [8]. This suggests that physical activity can partially address both the BMD and non-BMD determinants of fracture risk for osteoporotic patients. However, a RCT published in the CMAJ found that a major barrier for patients with osteoporosis engaging in physical activity was the fear of potential injury [9].

To address this, education regarding safe movement can be combined with exercise programs. This has been shown to be effective at reducing fall risk and therefore the risk of associated fractures [10]. As the benefits of these interventions greatly outweigh the costs, physicians should consider employing both physical activity and safe movement education as methods of risk reduction for osteoporotic patients. In accordance with this, many nationally and internationally published osteoporosis clinical practice guidelines (CPGs) now contain recommendations regarding physical activity and/or safe movement.

Healthcare professionals rely heavily on such CPGs [11] that aim to improve quality, consistency, and effectiveness of care by applying evidence-based medicine and providing healthcare practitioners with expert summaries of the most recent evidence [12]. However, evidence exists to suggest that in general, CPG quality may be low [13] and the rigor with which CPGs follow standard development methods is unsatisfactory [14–16]. Therefore, a common, widely accepted, and standardized method to evaluate CPGs is required. The benefit is twofold: readers will know in which CPGs to place the most

trust, and CPG developers will be able to improve the quality of future publications.

To address this need, several CPG quality appraisal instruments have been developed. However, to date, there has been no evaluation of the methodological quality of osteoporosis CPGs. Our study utilized the Appraisal of Guidelines for Research and Evaluation version II (AGREE II) instrument [17] to evaluate the methodological quality of 19 osteoporosis CPGs as they pertain to physical activity and safe movement. This instrument, along with its previous versions, has previously been applied to many CPGs for different diseases [18–22]. This instrument has been validated and is considered a reliable and useful tool [23, 24]. In addition, differences in region, publication date, and quality of CPGs may result in some variability in recommendations made. Therefore, the purpose of this review is to provide an AGREE II quality assessment of currently available osteoporosis CPGs and to assess physical activity and safe movement recommendations, including a discussion of how they differ between CPGs.

Methods

Clinical practice guideline identification

CPGs were identified between June 15, 2015, and July 1, 2015, using PubMed, the Wiley Online Library, Scholars Portal, the National CPG Clearinghouse, and the International Osteoporosis Foundation's National Guideline Database. A search strategy using the keywords "osteoporosis," "national," "guidelines," "safe movement," "physical activity," and "exercise" was employed. The results of this search were further filtered to include only papers produced or commissioned by national/international professional associations or health ministries. This strategy identified 45 papers, of which 19 met the pre-defined inclusion criteria: (1) complete CPG text is available in English or as a translation (studies with only the abstract translated were excluded); (2) CPG contains recommendations regarding safe movement and/or exercise for patients with primary osteoporosis; (3) the target audience was primary or secondary healthcare providers; and (4) the most recent version of the CPG was published no later than 2004. Reasons for CPG exclusions were collected.

Quality assessment

This study employed the latest version of the AGREE II instrument [17] to evaluate each CPG meeting our inclusion criteria. According to AGREE II protocol, each CPG was scored on 23 items within 6 domains. Domain 1 (scope and purpose) is divided into 3 items: guideline objectives, health questions, and population application. Domain 2 (stakeholder involvement) is based on 3 items: guideline development

group, preferences of target population, and target users. Domain 3 (rigor of development) includes 8 items: systematic methods used to search evidence, criteria for selection, strengths and limitations of the evidence, methods for formulating the evidence, health benefits and side effects of recommendations, explicit links between recommendation and supporting evidence, expert reviewers, and updating guideline for future use. Domain 4 (clarity and presentation) introduces 3 items: recommendations are specific and unambiguous, different options for management, and key recommendations. Domain 5 (applicability) includes 4 items: facilitators and barriers, advice/tools to implement recommendations into practice, resources for implications, and auditing criteria. Domain 6 (editorial independence) is based on 2 items: editorial independence from the funding body and conflicts of interest of the guideline development members.

CPGs were scored by three independent reviewers, each of whom were trained to use the AGREE II instrument and were provided with the AGREE II user manual. The user manual defines each item and assists the user in determining a CPG's score for that item. Items were scored based on a scale ranging from 1 (absence of item) to 7 (item is reported with exceptional quality). Domain scores were calculated by summing item scores within each domain from each reviewer, then standardizing them as a percentage of the maximum possible score. Agreement between each reviewer's scores was tested using a two-way ANOVA with single-rater two-way intra-class correlation coefficients (ICCs) for each domain across all guidelines as in a previous study [25]. This method was chosen based on recommendations from Shrout and Fleiss [26]. The degree of reviewer score agreement was defined using a previously used scale: agreement for ICC <0.20, poor; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, good; 0.81–1.00, very good [25]. To ensure uniformity, reviewers were instructed not to refer to supporting documents published separately unless they were explicitly referred to within the CPG.

AGREE II protocol states that no overall score is calculated to determine if a CPG is recommended or not recommended. Instead, CPGs in this study were ranked, as in a previous study, according to the number of domains scoring >60 % [26]. High-quality CPGs have 5 or 6 domains scoring >60 %, average-quality CPGs 3 or 4, and poor-quality CPGs have 2 or fewer of their domains scoring >60 % [27].

Extraction and analysis of relevant recommendations

Two reviewers used a pre-existing standardized table to independently extract each CPG's recommendations on safe movement and physical activity. The standardized table was used to record the occurrence of specific physical activity and/or safe movement recommendations. Types of physical activity and/or safe movement were recorded as either recommended (1), recommended against (−1), or not addressed (0) by

each CPG. This way, the positive as well as the negative frequency of each recommendation could be tracked.

Once extracted, the frequency of the different recommendations made and their consistency between CPGs were analyzed. Recommendation frequency was compared with CPG publication date, and CPG AGREE quality scores.

Results

Guideline selection

Our literature search identified 45 CPGs and 19 met inclusion criteria. CPG exclusion was mostly on the basis of language; 19 CPGs were not available in English and 3 had English language abstracts or summaries only. Three CPGs were excluded based on an absence of physical activity and safe movement recommendations. Finally, one CPG was excluded due to a publication date prior to 2004. The remaining 19 CPGs met the inclusion criteria and are summarized in Table 1.

Quality assessment

Intra-reviewer item score agreement ranged from fair to good (Table 2). Dispute regarding whether an item's criteria had been fulfilled or unfulfilled never occurred. However, when disagreement did occur, it was related to the degree to which an item's criteria had been fulfilled.

Domain scores resulting from the AGREE II quality assessment are illustrated in Table 3. Collectively, the included CPGs' scores varied significantly both within and across the six domains. Overall, CPGs were strongest in the "scope and purpose" and "clarity of presentation" domains, scoring 87 % with a SD of 15 % and 84 % with a SD of 11 % respectively. The CPGs scored poorly on the remaining four domains with the worst quality domain "editorial independence" scoring 46 % with a SD of 28 %.

CPGs published by the National Osteoporosis Foundation of America, Australia, Scotland, and Malaysia met the criteria for being high quality, having at least 5 domains scoring greater than 60 %. The majority of CPGs included in this study were of average quality, meaning they had either three or four domains with scores over 60 %. The CPGs published by Greece, Asia, Ibero-america, the Middle East and North Africa, and Taiwan all had two or fewer domains that scored over 60 % and were considered low quality.

For domain 1, scope and purpose, CPGs included in this study had a mean AGREE II quality score of 87 %. Quality scores varied little between the individual CPGs for this domain with a standard deviation of 15 %. CPGs published by Malaysia, Australia, and the American Association of Endocrinologists best described their scope and purpose, while

Table 1 Intraclass Correlation Coefficients for each CPG AGREE Score

Guideline title	Guideline country of origin	Guideline organization	Accessed through	Publication date
Guidelines for diagnosing, prevention + A3:A21, and treatment of osteoporosis in Asia	Asia	None identified	Wiley Online	2006
Clinical guideline for the prevention and treatment of osteoporosis in postmenopausal women and older men	Australia	The Royal Australian College of General Practitioners (RACGP)	IOF	2010
2010 clinical practice guidelines for the diagnosis and management of osteoporosis in Canada: summary	Canada	Osteoporosis Canada	PubMed	2010
European guidance for the diagnosis and management of osteoporosis in postmenopausal women	Europe	ESCEO/IOF	PubMed	2013
Guidelines for the Diagnosis and Management of Osteoporosis in Greece	Greece	Greek National Medical Agency (GNMA)	IOF	2004
The Osteoporosis Society of Hong Kong (OSHK): 2013 OSHK guideline for clinical management of postmenopausal osteoporosis in Hong Kong	Hong Kong	Osteoporosis Society of Hong Kong (OSHK)	PubMed	2013
Ibero-american Consensus on Osteoporosis SIBOMM 2009	South America	The Sociedad Ibero-americana de Osteología y Metabolismo Mineral (SIBOMM)	Scholars Portal	2009
Clinical practice guidelines on postmenopausal osteoporosis: an executive summary and recommendations	India	Indian Menopause Society (IMS)	PubMed	2013
Japanese 2011 guidelines for prevention and treatment of osteoporosis—executive summary	Japan	Japan Osteoporosis Society (JOP)	PubMed	2012
First update of the Lebanese guidelines for osteoporosis assessment and treatment	Lebanon	Lebanese Society for Osteoporosis and Metabolic Bone Disorders (OSTEOS)	PubMed	2007
Clinical Guidance on Management of Osteoporosis 2012	Malaysia	Malaysian Osteoporosis Society (MOS)	IOF	2012
Middle East and North Africa consensus on osteoporosis	Middle East and N. Africa	Multiple	PubMed	2007
Management of osteoporosis and the prevention of fragility fractures	Scotland	Scottish Intercollegiate Guidelines Network (SIGN)	IOF	2015
Osteoporosis: MOH Clinical Practice Guidelines 3/2008	Singapore	Singapore Ministry of Health	IOF	2009
NOFSA Guideline for the Diagnosis and Management of Osteoporosis	South Africa	National Osteoporosis Foundation of South Africa (NOFSA)	IOF	2010
Taiwanese Guidelines for the Prevention and Treatment of Osteoporosis	Taiwan	Taiwanese Osteoporosis Association	PubMed	2013
Osteoporosis Clinical guideline for prevention and treatment: executive summary	UK	National Osteoporosis Guideline Group	IOF	2009
Clinician's Guide to Prevention and Treatment of Osteoporosis	USA	National Osteoporosis Foundation (NOF)	PubMed	2010
American Association of Clinical Endocrinologists medical guidelines for clinical practice for the diagnosis and treatment of postmenopausal osteoporosis: executive summary of recommendations	USA	American Association of Clinical Endocrinologists (AACE)	PubMed	2010

CPGs published by Greece, Asia and the Middle East, and North Africa failed to do so adequately, if at all.

The mean CPG quality score for domain 2, stakeholder involvement, was 59 %. Scores were marginally more variable with a standard deviation of 17 %. The Australian, Singaporean, and Scottish CPGs had the highest quality methods ensuring appropriate stakeholder involvement. The Greek and Asian CPG and the CPG from the Middle East and North Africa received the lowest quality scores in this domain and failed to adequately address this methodological category.

In domain 3, rigor of development, the mean CPG quality score was 58 %. Variation was slightly larger with a standard deviation of 19 %. Receiving scores of 96, 87, and 81 % respectively, CPGs from Australia, Canada, and Scotland were the most rigorously developed. Scoring the lowest in this category were CPGs from Greece, Asia and the Middle East, and North Africa.

Domain 4, clarity and presentation, had a mean score of 84 %. Small variation was observed with a standard deviation of 11 %. CPGs from Scotland, Australia, and South Africa

Table 2 Intraclass correlation coefficients for each CPG AGREE score

Guideline	Intraclass correlation coefficient (ICC)	Degree of agreement
Asia	0.35	Fair
Australia	0.53	Moderate
Canada	0.37	Fair
European	0.50	Moderate
Greek	0.44	Moderate
Hong Kong	0.47	Moderate
Ibero-american	0.65	Good
India	0.40	Fair
Japan	0.37	Fair
Lebanon	0.39	Fair
M. E. and N. A.	0.53	Moderate
Malaysia	0.34	Fair
Scotland	0.50	Moderate
Singapore	0.51	Moderate
South Africa	0.65	Good
Taiwan	0.59	Moderate
UK	0.53	Moderate
USA (NOF)	0.51	Moderate
USA (AACE)	0.55	Moderate

ICC <0.20, poor; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, good; 0.81–1.00, very good

scored the highest in this category, while CPGs from Lebanon, the UK, and Asia received the lowest scores.

In terms of general applicability, domain 5, CPGs had a mean AGREE quality score of 52 % and a standard deviation of 11 %, the smallest variation of any domain. CPGs from Malaysia, Scotland, and the National Osteoporosis Foundation of America scored the highest, while CPGs from the Middle East and North Africa, Greece, and Ibero-america scored the lowest in terms of applicability.

Domain 6, editorial independence, had a mean of 46 %. This domain was the most variable with a standard deviation of 28 %. CPGs with methods facilitating the greatest editorial independence were from Australia, Malaysia, and the UK while CPGs from Asia, Greece, and Ibero-america scored the lowest in this category, and their methods may have had the greatest potential for allowing a conflict of interest to affect their recommendations.

The quality of included CPGs demonstrated improvement over time. Included CPGs were published between 2004 and 2015, and those published in the first half (2004–2009) scored an average of 67, 42, 41, 74, 41, and 25 % for domains 1 through 6, respectively. CPGs with publication dates from 2010 to 2015 had average domain scores of 90, 64, 63, 89, 56, and 54 %, respectively.

Analysis of physical activity and safe movement recommendations

Physical activity and safe movement recommendations compiled from all CPGs are listed in Table 4 along with the proportion of CPGs making each specific recommendation. Safe movement recommendations were defined as instructions detailing specific movements and body positions that reduce a patient's vulnerability to fractures and falls. As per inclusion criteria, 100 % of CPGs recommended physical activity. Commonly recommended physical activities included weight-bearing exercises (low-impact ones being the most commonly recommended), muscle-strengthening exercises, resistance exercises, and balance training exercises. Safe movement recommendations were much less common than physical activity recommendations, with only 58 % of included CPGs making any type of safe movement recommendation. The most frequent safe movement recommendations advised against immobility, forward spinal flexion, and spinal torsion. Recommendations usually used broad categories of exercise types or goals instead of indicating specific exercises.

The most frequently made recommendations regarding physical activity were not substantially different between high- and average-quality CPGs. However, some differences were found in low-quality CPG recommendations. The most commonly made

Table 3 AGREE II domain scores

Country of origin	Scope and purpose (%)	Stakeholder involvement (%)	Rigor of development (%)	Clarity of presentation (%)	Applicability (%)	Editorial independence (%)	# of domains scoring >60 %
Asia	52	31	34	60	36	14	1
Australia	100	83	96	100	57	86	5
Canada	90	62	84	93	50	54	4
European	88	48	40	90	64	29	3
Greek	48	31	29	79	36	14	1
Hong Kong	90	60	45	79	34	68	4
South America	83	36	40	76	30	14	2
India	93	50	55	81	46	43	2
Japan	90	43	48	76	34	75	3
Lebanon	93	57	72	76	36	14	3
M. E. & N. A.	64	36	36	64	21	32	2
Malaysia	100	67	75	90	75	79	6
Scotland	86	93	80	100	68	57	5
Singapore	86	83	47	98	59	14	4
South Africa	98	64	61	95	52	54	4
Taiwan	95	48	37	76	34	14	2
UK	100	67	49	67	46	82	4
USA (AACE)	100	55	58	98	38	50	2
USA (NOF)	95	57	44	95	63	68	4
Mean	87	56	54	84	46	45	
SD	15	17	19	11	16	28	

recommendations were low-impact weight-bearing exercise (such as walking), muscle-strengthening exercises, and balance training exercises in high-quality CPGs; muscle-strengthening exercises, low-impact weight-bearing exercises, and resistance training in average-quality CPGs; and balance training, high- and low-impact weight-bearing, and muscle-strengthening exercises in low-quality CPGs.

Safe movement recommendations were mostly consistent between high-, average-, and low-quality CPGs. In general, high- and average-quality CPGs tended to make safe movement recommendations more often than low-quality CPGs. High and average quality CPGs made an average of 3.4 (SD = 2.7) and 3.5 (SD = 3.9) recommendations per CPG. Low-quality CPGs made an average of 0.6 (SD = 1.5) recommendations per CPG. The most frequently made recommendations were to avoid immobility as well as excessive flexion or torsion of the spine.

The number of recommendations made by CPGs has increased significantly over time. CPGs published from 2004 to 2009 recommended an average of 4.4 (SD = 2.22) different types of physical activity whereas CPGs published from 2010 to 2015 recommended an average of 7.5 (SD = 3.28; $P < 0.05$). The same held true

for safe movement recommendations; CPGs published from 2004 to 2009 made an average of 0.86 recommendations (SD = 1.46) while CPGs published from 2010 to 2015 made an average of 3.3 recommendations (SD = 3.53; $P < 0.05$).

Discussion

Our findings suggest that the overall quality of osteoporosis CPGs varies substantially and that the information on exercise and safe movement is variable and poorly defined. As CPGs play such an integral role in the provision of care [11], identifying a CPG's quality before relying on it is crucial. As of now, with the current variable quality of osteoporosis CPGs, patients worldwide may not all be receiving care based on the best, most current peer-reviewed evidence. Presently, the onus is on the healthcare provider to ensure recommendations are sourced from CPGs with the highest methodological quality.

In our evaluation, the highest quality domains were scope and purpose and clarity of presentation. These results are in accordance with previous studies of CPG quality from a wide variety of healthcare disciplines [18–20, 28]. In this study, CPGs from Australia, Malaysia, the UK, and the American Association of Endocrinologists received perfect scores in this

domain. Why certain domains consistently score well on average across many different fields of healthcare is not currently known. Perhaps it is easier to fulfill the requirements of certain domains, or perhaps current authors place a higher priority on certain domains.

The poorest scoring domains were "editorial independence" and "applicability." In a review of physician adherence to CPGs, it was suggested that as many as 38 % of physicians described CPGs as inconvenient or too difficult to use [29]. Making CPGs easy to implement is a crucial step toward increasing their rate of use. Domain 5, applicability, was the second lowest scoring domain among included CPGs. Hence, the lack of specific information about exercise and safe movement can be considered a barrier for implementation. Our results regarding this domain are in agreement with a multitude of previous studies of CPGs in other healthcare disciplines in which this domain scored the second lowest or lowest [30–33]. The Malaysian CPG we reviewed scored the highest in this domain. It included a section defining specific factors required to be in place in order to ensure that the CPG was used effectively. Most other CPGs we reviewed completely neglected to include any such information. For exercise and safe movement, it might be anticipated that clinicians or patient guides would provide dosage information, and visual aids might facilitate implementation. Perhaps the low average score in this domain is due to the CPG development group composition. Groups usually include a variety of medical experts, patient representatives, and epidemiologists; however, they often lacked anyone with a high degree of administrative, financial, economic, knowledge translation, or logistical experience. If future CPG development groups included such individuals, the group may feel more comfortable devising a strategy easing the economic and logistical burden of applying CPGs and perhaps score higher in this domain as a result.

It is unclear why editorial independence is the lowest scoring and most variable domain in this study as well as in many other studies of CPG quality [22, 33–35]. The requirements for this domain are rather straightforward: any funding received for CPG development must be disclosed as well as any potential conflicts of interest from its authors or members of the development group. It is possible that development groups assume that it is implied to the reader, or they simply do not appreciate the importance of editorial independence or its disclosure. It has been shown that conflicts of interest are almost endemic among CPG developers across many fields of healthcare [36–39]. Further evidence suggests that these conflicts may actually have an effect on the recommendations found within CPGs [40]. It seems imperative that emphasis be placed on this domain. Perhaps if new conflict of interest reporting standards or even standards for CPG development panel member composition were created in order to limit potential conflicts from the beginning, CPG quality in this

domain would improve. Since we focused on exercise and safe movement, it is unlikely that conflict of interest would affect which exercises were recommended. However, it might be that a lack of attention to these noncommercial interventions could be related to such influences.

It is important to acknowledge which AGREE II domains are most often lacking as this may aid in the development of future CPGs. We hope that with data from our study, future osteoporosis CPG development groups can concentrate on improving the applicability and editorial independence of their CPGs. In fact, the AGREE II instrument is also intended to serve as a framework for CPG development [17]. Indeed, when reviewing the included CPGs, it was noted that those which disclosed knowledge of the AGREE instrument tended to score higher than those which did not.

Overall, our CPGs showed an improvement in quality over time. Domain scores improved substantially when CPGs published before 2010 were compared with those published after. This contrasts a 2012 review of multidisciplinary CPGs which found that little improvement, if any, had occurred over the previous two decades [16]. The CPGs in our study that showed this improvement were largely published after that review was conducted, so the quality improvement is perhaps a more recent trend, or is more obvious in osteoporosis CPGs.

Physical activity recommendations found within the included CPGs were highly varied with the exception of a few generally agreed upon recommendations. The generally agreed upon recommendations of weight-bearing, muscle-strengthening, and resistance training exercises were often vaguely defined with deference often being given to physical therapists' clinical judgment. It appears that osteoporosis CPG development groups agree that some sort of physical activity should be recommended to patients. However, there is no definitive consensus as to what specific type of exercise should be done, how often, and at what intensity. Furthermore, as the number of physical activity and safe movement recommendations increases over the years, recommendation variability seems to be increasing. The trend is moving toward increasing variability and decreased specificity. CPG readers may inevitably become more confused and find it more difficult to apply these recommendations to clinical practice. This trend is likely due to a multitude of newly published studies proclaiming specific types of exercise as beneficial to osteoporotic patients [41–44]. While considerable research exists stating the value of several singular types of physical activity, little if any research comparing the benefits of different physical activities relative to one another has been conducted. Research in this direction is required in order to decrease the variability and increase the specificity of physical activity recommendations, making them easier to apply for healthcare providers.

Special attention should be given to the category of weight-bearing exercise. It was usually defined by CPGs as high (i.e.,

Table 4 Relative frequency of recommendations

Physical activity recommendation	Frequency (%)	Safe movement recommendation	Frequency (%)
Adequate physical activity/exercise	100	Avoid immobility	37
Weight-bearing exercises	95	Avoid forward flexion	32
Low impact weight bearing (walking)	78	Avoid twisting motions	26
Moderate impact weight bearing (stair climbing)	44	Avoid lateral flexion exercises	21
High impact weight bearing (jogging)	11	Avoid heavy lifting	21
Muscle-strengthening exercises	84	Promote the use of assistive devices	21
Resistances exercises	58	Avoid lifting objects over shoulders	16
Balance training exercises	58	Training on safe ADLs	16
Combination of exercise types	32	Avoid forward flexion with load on spine	11
Flexibility improvement/stretching	32	Avoid high-impact loading	11
Physiotherapy (preemptive—before fracture)	32	Avoid pushing	11
Tai Chi	26	Avoid pulling	11
Posture exercise	26	Avoid explosive movements	11
Physiotherapy (rehabilitative after fracture)	26		
Dance	16		
Endurance training	11		
Trunk/core stabilization exercises	11		
Hip stabilization exercises	5		

Percentage of CPGs containing each category of recommendation is reported

jogging), moderate (i.e., stair climbing), or low (i.e., walking) impact. CPGs with high AGREE quality scores almost unanimously recommended a low-impact exercise and advised against a high-impact exercise. Lower quality CPGs, on the other hand, while also recommending a low-impact weight-bearing exercise, frequently recommended high-impact weight-bearing activities and never cautioned against these. Most often, CPGs recommending against a high-impact weight-bearing exercise cited the ambiguous, weak, or low-quality evidence regarding its benefit, as well as safety concerns stemming from development panel members' clinical judgment. Further, it may be important to consider safe movement principles when recommending exercise, as certain activities or exercises might expose people to more at-risk postures.

The methodological quality of this study is in line with previous work. However, it is not without its limitations. While we searched several online databases, only English language CPGs were examined. If CPGs in other languages were included in this study, results may have been different. Reviewer bias also could have affected our results. For example, one reviewer mentioned that if a particular CPG exceeded pre-existing expectations, it may have received a slightly higher mark than it otherwise would have. The same concern

was also voiced for CPGs that underperformed compared to pre-existing expectations. Finally, there are limitations inherent to the AGREE II instrument. It is a tool for the assessment of methodological quality. It does not assess the clinical content. This limitation is common to all existing appraisal tools [45]. They can never fully replace a reader's critical judgment or clinical decision-making that considers the nature of the health condition and person to whom recommendations would be applied.

In conclusion, CPGs aim to provide healthcare providers with summaries of the most recent and highest quality evidence in a particular field. However, given the currently variable quality of osteoporosis CPGs, the burden of critically appraising evidence still partially lies on the healthcare provider. To our knowledge, this study is the first to evaluate the quality of osteoporosis CPGs and provide guidance as to which are developed with the highest methodological standards. It would be prudent for CPG development groups worldwide to address and report adherence to the AGREE II instrument framework. Special attention is needed to ensure CPGs are easily applicable to current clinical practice paradigms. Improving editorial independence standards or reporting standards is also necessary. Lastly, regarding the content of the examined CPGs, greater detail and specificity

is required for physical activity recommendations. More research is required in this area before CPG recommendations can improve.

Compliance with ethical standards

Conflicts of interest None.

References

1. A Report of a WHO Study Group (1994) Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: synopsis of a WHO report. World Heal Organ Tech Rep Ser 843:1–129. doi:10.1007/BF01622200

2. Genant H, Cooper C, Poor G, Reid I (1999) Interim Report and Recommendations of the World Health Organization Task-Force for Osteoporosis. Osteoporos Int 259–264. doi: 10.1007/s001980050224

3. Johnell O, Kanis JA (2004) An estimate of the worldwide prevalence, mortality and disability associated with hip fracture. Osteoporos Int 15:897–902. doi:10.1007/s00198-004-1627-0

4. McClung M, Geusens P, Miller P et al (2001) Effect of risedronate on the risk of hip fracture in elderly women. N Engl J Med 344:333–340

5. McCloskey EV, Beneton M, Charlesworth D et al (2007) Clodronate reduces the incidence of fractures in community-dwelling elderly women unselected for osteoporosis: results of a double-blind, placebo-controlled randomized study. J Bone Miner Res 22:135–141. doi:10.1359/jbmr.061008

6. Cheung AM, Detsky AS (2008) Osteoporosis and fractures: missing the bridge? J Am Med Assoc 299:1468–1470

7. Sherrington C, Whitney JC, Lord SR et al (2008) Effective exercise for the prevention of falls: a systematic review and meta-analysis. J Am Geriatr Soc 56:2234–2243. doi:10.1111/j.1532-5415.2008.02014.x

8. Kelley GA, Kelley KS, Tran ZV, et al. (2014) Exercise and bone mineral density in men: a meta-analysis. 1730–1736.

9. Carter ND, Khan KM, Mckay HA et al (2002) Community-based exercise program reduces risk factors for falls in 65- to 75-year-old women with osteoporosis: randomized controlled trial. CMAJ 167:997–1004

10. Chang JT, Morton SC, Rubenstein LZ et al (2004) Interventions for the prevention of falls in older adults: systematic review and meta-analysis of randomised clinical trials. Br Med J 328:1–7

11. Qaseem A, Snow V, Gosfield A et al (2010) Pay for performance through the lens of medical professionalism. Ann Intern Med 152:366–369. doi:10.7326/0003-4819-152-6-201003160-00006

12. Field MJ, Lohr KN (1990) Clinical practice guidelines: directions for a new program. Washington, D.C

13. Grilli R, Magrini N, Penna A et al (2000) Practice guidelines developed by specialty societies: the need for a critical appraisal. Lancet 355:103–106. doi:10.1016/S0140-6736(99)02171-6

14. Rivara FP (1999) Are guidelines following guidelines? AAP Grand Rounds 281:1900–1905. doi:10.1542/gr.2-2-15

15. Qaseem A, Forland F, Macbeth F et al (2012) Guidelines International Network: toward international standards for clinical practice guidelines. Ann Intern Med 156:525–531

16. Kung J (2012) Failure of clinical practice guidelines to meet Institute of Medicine standards: two more decades of little, if any, progress. Arch Intern Med 172:1628–1633. doi:10.1001/2013.jamainternmed.56

17. Brouwers MC, Kho ME, Browman GP et al (2010) AGREE II: advancing guideline development, reporting and evaluation in health care. Can Med Assoc J 182:E839–842. doi:10.1503/090449

18. Vasse E, Vernooij-Dassen M, Cantegreil I et al (2012) Guidelines for psychosocial interventions in dementia care: a European survey and comparison. Int J Geriatr Psychiatry 27:40–48. doi:10.1002/gps.2687

19. Van den Berg T, Engelhardt EG, Haanstra TM et al (2011) Methodology of clinical nutrition guidelines for adult cancer patients: how good are they according to AGREE criteria? J Parenter Enter Nutr 36:316–322. doi:10.1177/0148607111414027

20. Alonso-Coello P, Irfan A, Solà I et al (2010) The quality of clinical practice guidelines over the last two decades: a systematic review of guideline appraisal studies. Qual Saf Health Care 19, e58. doi:10.1136/qshc.2010.042077

21. Al-Ansary LA, Tricco AC, Adi Y et al (2013) A systematic review of recent clinical practice guidelines on the diagnosis. Assessment and management of hypertension. PLoS ONE. doi:10.1371/journal.pone.0053744

22. Nelson AE, Allen KD, Golightly YM et al (2014) A systematic review of recommendations and guidelines for the management of osteoarthritis: The Chronic Osteoarthritis Management Initiative of the U.S. Bone and Joint Initiative. Semin Arthritis Rheum 43:701–712. doi:10.1016/j.semarthrit.2013.11.012

23. MacDermid JC, Brooks D, Solway S et al (2005) Reliability and validity of the AGREE instrument used by physical therapists in assessment of clinical practice guidelines. BMC Health Serv Res 5:18. doi:10.1186/1472-6963-5-18

24. Brouwers MC, Kho ME, Browman GP et al (2010) Development of the AGREE II, part 1: performance, usefulness and areas for improvement. Cmaj 182:1045–1052. doi:10.1503/cmaj.091714

25. Lytras T, Bonovas S, Chronis C et al (2014) Occupational asthma guidelines: a systematic quality appraisal using the AGREE II instrument. Occup Environ Med 71:81–6. doi:10.1136/oemed-2013-101656

26. Shrout P, Fleiss J (1979) Intraclass correlations: uses in assessing rater reliability. Psychol Bull 86:420–428

27. Ou Y, Goldberg I, Migdal C, Lee PP (2011) A critical appraisal and comparison of the quality and recommendations of glaucoma clinical practice guidelines. Ophthalmology 118:1017–1023. doi:10.1016/j.ophtha.2011.03.038

28. Zeng L, Zhang L, Hu Z et al (2014) Systematic review of evidence-based guidelines on medication therapy for upper respiratory tract infection in children with AGREE instrument. PLoS ONE 9, E87711. doi:10.1371/journal.pone.0087711

29. Cabana MD, Rand CS, Powe NR et al (1999) Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA 282:1458–1465. doi:10.1001/jama.282.15.1458

30. Yan J, Min J, Zhou B (2013) Diagnosis of pheochromocytoma: a clinical practice guideline appraisal using AGREE II instrument. J Eval Clin Pract 19:626–632. doi:10.1111/j.1365-2753.2012.01873.x

31. Sabharwal S, Patel NK, Gauher S et al (2014) High methodologic quality but poor applicability: assessment of the AAOS guidelines using the AGREE II instrument. Clin Orthop Relat Res 472:1982–1988. doi:10.1007/s11999-014-3530-0

32. Don-Wauchope AC, Sievenpiper JL, Hill SA, Iorio A (2012) Applicability of the AGREE II instrument in evaluating the development process and quality of current National Academy of

Clinical Biochemistry guidelines. Clin Chem 58:1426–1437. doi: 10.1373/clinchem.2012.185850

33. Zhang Z, Guo J, Su G et al (2014) Evaluation of the Quality of Guidelines for Myasthenia Gravis with the AGREE II instrument. PLoS ONE 9, e111796. doi:10.1371/journal.pone.0111796

34. Holmer HK, Ogden LA, Burda BU, Norris SL (2013) Quality of clinical practice guidelines for glycemic control in type 2 diabetes mellitus. PLoS ONE 8:1–7. doi:10.1371/journal.pone.0058625

35. Cates JR, Young DN, Bowerman DS, Porter RC (2006) An independent AGREE evaluation of the occupational medicine practice guidelines. Spine J 6:72–77. doi:10.1016/j.spinee.2005.06.012

36. Norris SL, Holmer HK, Ogden LA et al (2012) Conflict of interest disclosures for clinical practice guidelines in the National Guideline Clearinghouse. PLoS ONE 7:1–8. doi:10.1371/journal.pone. 0047343

37. Norris SL, Holmer HK, Ogden LA, Burda BU (2011) Conflict of interest in clinical practice guideline development: a systematic review. PLoS ONE. doi:10.1371/journal.pone.0025153

38. Neuman J, Korenstein D, Ross JS, Keyhani S (2011) Prevalence of financial conflicts of interest among panel members producing clinical practice guidelines in Canada and United States: cross sectional study. BMJ. doi:10.1136/bmj.d5621

39. Mendelson TB, Meltzer M, Campbell EG et al (2011) Conflicts of interest in cardiovascular clinical practice guidelines. Arch Intern Med 171:577–584. doi:10.1001/archinternmed.2011.96

40. Norris SL, Burda BU, Holmer HK et al (2012) Author's specialty and conflicts of interest contribute to conflicting guidelines for screening mammography. J Clin Epidemiol 65:725–733. doi:10. 1016/j.jclinepi.2011.12.011

41. Martyn-St. James M, Carroll S (2006) High-intensity resistance training and postmenopausal bone loss: a meta-analysis. Osteoporos Int 17:1225–1240. doi:10.1007/s00198-006-0083-4

42. Hackney ME, Kantorovich S, Levin R, Earhart GM (2007) Effects of tango on functional mobility in Parkinson's disease: a preliminary study. J Neurol Phys Ther 31:173–179. doi:10.1097/NPT. 0b013e31815ce78b

43. Martyn-St James M (2008) A meta-analysis of impact exercise on postmenopausal bone loss: the case for mixed loading exercise programmes. Br J Sports Med. doi:10.1136/bjsm.2008.052704

44. De Kam D, Smulders E, Weerdesteyn V, Smits-Engelsman BCM (2009) Exercise interventions to reduce fall-related fractures and their risk factors in individuals with low bone density: a systematic review of randomized controlled trials. Osteoporos Int 20:2111–2125. doi:10.1007/s00198-009-0938-6

45. Vlayen J, Aertgeerts B, Hannes K et al (2005) A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. Int J Qual Heal Care 17:235–242. doi: 10.1093/intqhc/mzi027