


RESEARCH ARTICLE

Open Access



An interpretive review of consensus statements on clinical guideline development and their application in the field of traditional and complementary medicine

Jennifer Hunter^{1,2*} , Matthew Leach³, Lesley Braun^{1,4,5} and Alan Bensoussan¹

Abstract

Background: Despite ongoing consumer demand and an emerging scientific evidence-base for traditional and complementary medicine (T&CM), there remains a paucity of reliable information in standard clinical guidelines about their use. Often T&CM interventions are not mentioned, or the recommendations arising from these guidelines are unhelpful to end-users (i.e. patients, practitioners and policy makers). Insufficient evidence of efficacy may be a contributing factor; however, often informative recommendations could still be made by drawing on relevant information from other avenues. In light of this, the aim of this research was to review national and internationally endorsed consensus statements for clinical guideline developers, and to interpret how to apply these methods when making recommendations regarding the use of T&CM.

Method: The critical interpretive review method was used to identify and appraise relevant consensus statements published between 1995 and 2015. The statements were identified using a purposive sampling technique until data saturation was reached. The most recent edition of a statement was included in the analysis. The content, scope and themes of the statements were compared and interpreted within the context of the T&CM setting; including history, regulation, use, emerging scientific evidence-base and existing guidelines.

Results: Eight consensus statements were included in the interpretive review. Searching stopped at this stage as no new major themes were identified. The five themes relevant to the challenges of developing T&CM guidelines were: (1) framing the question; (2) the limitations of using an evidence hierarchy; (3) strategies for dealing with insufficient, high quality evidence; (4) the importance of qualifying a recommendation; and (5) the need for structured consensus development.

Conclusion: Evidence regarding safety, efficacy and cost effectiveness are not the only information required to make recommendations for clinical guidelines. Modifying factors such as burden of disease, magnitude of effect, current use, demand, equity and ease of integration should also be considered. Uptake of the recommendations arising from this review are expected to result in the development of higher quality clinical guidelines that offer greater assistance to those seeking answers about the appropriate use of T&CM.

Keywords: Evidence based medicine, Guideline, Practice guideline, Complementary medicine, Integrative medicine, Literature review

* Correspondence: drjenniferhunter@yahoo.com.au

¹NICM, University of Western Sydney, Locked Bag 1797, Penrith, NSW 2751, Australia

²Menzies Centre for Health Policy, School of Medicine, University of Sydney, Sydney, Australia

Full list of author information is available at the end of the article



Background

Traditional and complementary medicine (T&CM) refers to a conglomerate of health-related interventions and therapies not usually considered mainstream by the Western medical system. T&CM includes (but is not limited to) naturopathy, traditional Chinese medicine, Ayurvedic medicine, homeopathy, chiropractic, osteopathy, massage therapy, yoga and meditation. In such a multifarious field with divergent training requirements, different models of regulation, and myriad treatment options informed by varying (and sometimes inconsistent) evidence, it is not surprising there is considerable diversity in clinical practice [1]. The impact of these inconsistent practices on patient outcomes, patient satisfaction and professional credibility can be significant [2].

Clinical guidelines are “systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances” [3] that aim to reduce unnecessary variations in service delivery by informing a rational approach to the management of patients, as well as guiding healthcare policies. Evidence-based clinical guidelines were initially almost solely based on evidence of efficacy and safety [4–6]. The limitation of this approach was that it ignored other important considerations when developing guidelines to meet the healthcare needs of a population [7]. Increasingly, the importance of contextual information and qualifying statements about the burden of disease, economic impact, current use, patient values and preferences and equity, and the need for transparency throughout the development process have been adopted as guideline development standards [8–10]. Despite these standards, it is not uncommon for clinical guidelines and health policies regarding T&CM to only consider the evidence for safety, efficacy and cost-effectiveness, if they are considered at all [11].

The quality of clinical guidelines continues to be a matter of concern, hence the development of various guideline appraisal tools such as the AGREE II [12, 13]. In the field of T&CM, standard medical guidelines are fraught with inconsistencies and unhelpful recommendations. For example, reviews of guidelines endorsed by the UK National Institute for Health Care and Excellence (NICE) or the Scottish Intercollegiate Guidelines Network (SIGN) have found that many lacked transparency and consistency about the inclusion or exclusion of T&CM [14–16]. The conclusions drawn from the available evidence often overestimated or underestimated potential benefits. In many instances, even when one or more T&CM interventions were reviewed by the guideline developers, either no recommendations or nonspecific recommendations, such as ‘practitioners should discuss T&CM use with their patients’ or ‘more research is needed’ were made. General statements provide little guidance for clinical decision

making and could be viewed as ‘holding statements’ rather than serving any real purpose.

Given the aforementioned findings, clinical guidelines of higher quality are urgently required to guide the safe and rational use of T&CM in practice [17]. Indeed, there are many instances where specific recommendations are often needed in T&CM practice and policy. As Table 1 illustrates, the decision to appraise an intervention or otherwise in a guideline is not always (nor should it be) dependent on data from clinical trials.

Insufficient evidence about any intervention or practice poses significant challenges for guideline developers. In the case of T&CM, failure to evaluate the field or even make a recommendation when there is insufficient evidence of efficacy may simply widen the gap between what practitioners and users of T&CM are doing, and what is considered best practice. Guideline developers should always attempt to make specific, informative recommendations about the use of T&CM [18–20]. As Petitti et al. state:

“Decision makers do not have the luxury of waiting for certain evidence. Even though evidence is insufficient, the clinician must still provide advice, patients must make choices, and policy makers must establish policies” [21].

Further guidance on how to use all available information and evidence for clinical decision making will help improve the utility of clinical guidelines that consider T&CM. Debate continues in the T&CM field and more generally around the appropriate use of evidence for evaluating interventions [22–24]. The objectives of this review, however, were to identify, appraise and synthesise nationally or internationally endorsed consensus statements for clinical

Table 1 Instances where T&CM should be considered in guideline development

1. The potential for at least modest clinical benefit where there is a large burden of disease^a;
2. High quality traditional evidence about safety and effectiveness that address a continuing burden of disease;
3. Cost-effective interventions that address a continuing burden of disease^a;
4. The potential for catastrophic risks from the T&CM intervention or its alternative;
5. Unknown benefit from a T&CM intervention that is commonly used^a;
6. Inequalities of access or large unmet need;
7. High demand to integrate within conventional healthcare settings;
8. High costs relative to alternatives or vice versa;
9. Large system changes required for its introduction^a; and
10. Conflict in choices between individual and societal perspectives^a.

^aCriteria outlined by the World Health Organisation for prioritising guideline development [7]

guideline developers; and ‘interpret’ how these statements might apply to the field of T&CM, particularly in instances where there is low quality or inconsistent evidence regarding safety, efficacy or cost-effectiveness.

Method

Study design

A critical interpretive review of consensus statements for guideline developers was undertaken [25, 26]. The literature search and analysis of the consensus statements was an iterative process. A sampling frame was created where the identified consensus statements were coded and categorised into themes and subthemes on an electronic spreadsheet. The first author then summarised the findings for further discussion amongst the co-authors. Numerous iterations explored how best to categorise and interpret the themes until consensus was reached. Literature searching continued until there was data saturation (i.e. the point at which no new major themes emerged).

Literature search & sampling

The consensus statements for guideline developers were identified using a similar approach to that used in the interpretive synthesis outlined by Dixon-Woods et.al [26]. The literature was searched from 7th April 2014 through to 10th October 2015. A systematic literature search was not conducted, because unlike Schünemann et.al, this review was not a content analysis where all consensus statements are identified to formulate a comprehensive list of items from these statements [27]. Instead, purposive sampling was used to identify statements published before the end-date of the search that clearly addressed the research objective. Consensus statements and publications were first identified through the authors’ expert knowledge of the topic. This was augmented by literature searches on Google Scholar and PubMed. Database searches using various sets of search terms (e.g. guidelines*, “Practice Guidelines as Topic/standards”[Mesh], Evidence-Based Medicine/methods[Mesh]) and search functions (e.g. customizing Article types) were abandoned because the results were either too broad or too narrow. Alternate search strategies were therefore employed, such as bibliographic searching of previously published systematic reviews [27], bibliographic cluster searching [28], and the use of ‘PubMed/Similar articles’ or ‘Google Scholar/Related articles’ functions.

Inclusion & exclusion criteria

The authors defined a consensus statement as a document or similar resource (e.g. website) developed by an independent panel of experts that provided systematic guidance. In this instance, the guidance was on methodologies for formulating clinical guidelines or related health policies. Only statements endorsed by national or

international authorities and published in English were included. Consensus statements on health policy making were also included since clinical guidelines are used not only to inform clinical decision making but to inform health service delivery and public health policies. Consensus statements describing how to appraise the quality of clinical guidelines were excluded as no new themes could be identified that were not already addressed in detail, including the rationale, in the statements on guideline development. Statements published from 1995 until the end of the search date in 2015 were included. For those with multiple iterations, only the most recent edition of a statement was included in the analysis.

Data extraction & analysis

An interpretive approach was used to appraise and synthesise the information [25, 26]. This was an inductive process. As consensus statements and their related publications were identified, their content was reviewed for relevant themes applicable to the use, practice and context of T&CM (see Table 2), and the known shortcomings of existing clinical guidelines for T&CM [14–16, 29–31]. The statements were compared for similarities (*reciprocal translational analysis*) and contradictions (*refutational synthesis*). Lines-of-arguments (*synthesising arguments*) were generated by integrating the content and themes identified in the individual statements. The aim was to identify overarching themes and constructs, and then *interpret* how they apply to T&CM.

Results

Eight consensus statements for guideline developers met the inclusion criteria for in-depth review; this was the point at which data saturation was reached and no new major themes emerged. Three of the statements were international [7, 18–20, 32–63]; the remaining five were national statements from Australia [64–66], Germany [67], Scotland [68], US [21, 69–71] and UK [72–74].

The primary focus of the first seven statements (as listed in Table 3) was the development of clinical

Table 2 Contextual information about T&CM

1. T&CM often has different regulation than biomedicine because the interventions are considered relatively safe and/or there is a historical precedence of use spanning hundreds, if not thousands of years [29].
2. Clinical guidelines are often uninformative regarding the use of T&CM; and are lacking in transparency and consistency [14–16].
3. In instances of insufficient scientific evidence, information about clinical indications and safety; patient demand and preferences; or equity and costs may still be available [29].
4. The outcomes that users seek from T&CM do not always align with the outcomes evaluated in clinical trials [30].
5. The ongoing demand for T&CM suggests that factors other than scientific evidence are important and influence the decision making process of patients and practitioners [31].

Table 3 Summary of consensus statements for guideline developers

1. Guidelines for W.H.O. guidelines [7].
A concise outline of the principles for guideline development, the statement emphasises the importance of considering local circumstances such as applicability of the research findings to the population and the feasibility and opportunity costs of implementing the guideline. Annex C provides a succinct yet comprehensive checklist for appraising treatment guidelines.
2. GRADE Grading of Recommendations Assessment, Development and Evaluation [18, 19, 32–44].
Endorsed by the Cochrane Collaboration and W.H.O., GRADE is an internationally recognised system for evaluating and comparing alternative interventions. Recommendations are graded based on the quality of evidence (i.e. study design, bias, consistency of the results, and precision of the overall estimates); risks and benefits; resources used; and values and preferences. Resources include the online software GRADEproGDT for producing guidelines, Summary of Findings tables and Evidence to Decision Frameworks [45].
3. AWMF Regelwerk/Guidance Manual and Rules for Guideline Development [67].
The Association of the Scientific Medical Societies in Germany (AWMF) builds on the GRADE process. It also provides explicit guidance on how to structure consensus development, classify the strength of consensus and manage justified dissent to formulate a transparent set of recommendations. Accompanying statements included the DELBI 2.0 - Deutsche Leitlinien-Bewertungsinstrument/German Guideline Appraisal Instrument.
4. NHMRC: A guide to the development, evaluation and implementation of clinical practice guidelines [64].
The Australian National Health and Medical Council (NHMRC) guidelines along with seven accompanying handbooks were published from 1999 to 2003. Together they provide detailed information about reviewing and appraising clinical, economic and socioeconomic evidence; implementing and disseminating guidelines; and presenting the information to consumers. Guidelines on appraising evidence and grading recommendations were updated in 2009 and included a Body of Evidence Matrix [65, 66]. A 'hierarchy of evidence' constrains the grading of recommendations.
5. NICE: Developing NICE guidelines - The manual [72].
The National Institute for Health Care and Excellence (NICE) guidelines manual from the United Kingdom has evolved over time. Initially recommendations were rated according to a 'hierarchy of evidence' that reflected the quality of studies. This rating was replaced in 2006–7 with an approach that more closely aligns with GRADE. Differences to GRADE include modifications to the economic evidence profiles and the absence of 'overall summary' labels for the quality of evidence and strength of recommendations [74].
6. SIGN 50: A Guideline Developer's Handbook [68].
The Scottish Intercollegiate Guidelines Network (SIGN), first published their handbook in 1999. It was modified over the years to reflect methodological changes in appraising the evidence and grading recommendations. A hierarchy of evidence was used from 2001 to 2012 that was discontinued in line with the SIGN's adoption of the GRADE process in 2013.
7. U.S. Preventive Services Task Force (USPSTF) system [69].
The system is used to assess the quality, magnitude and confidence in the evidence about benefits and harms of an intervention and provide a graded recommendation [70]. An analytic structured causal framework of process and outcome is recommended to help guide the systematic evaluation of non-RCT study designs [71]. If the problem of insufficient evidence persists, the four "Domains of Information Pertinent to Clinical Decision making for Preventive Services" may be used [21].

Table 3 Summary of consensus statements for guideline developers (*Continued*)

8. SUPPORT Tools for evidence-informed health Policymaking (STP) [20, 46–63].
STP was written for people responsible for making decisions about health policies and programmes and for those who support these decision makers. The series addresses four broad areas: 1) Supporting evidence-informed policymaking 2) Identifying needs for research evidence in relation to three steps in policymaking processes, namely problem clarification, options framing, and implementation planning 3) Finding and assessing both systematic reviews and other types of evidence to inform these steps, and 4) Going from research evidence to decisions.

practice guidelines for the management or prevention of disease. These guidelines all used similar methodologies for systematically identifying and appraising the evidence of efficacy, safety and cost-effectiveness [7, 32, 64, 67, 68, 70, 72]. However, there were differences in the terminology and categories used to summarise the evidence and formulate recommendations. GRADE, AWMF and the NHMRC, for example, categorised the quality of the evidence and the strength of the recommendations [32, 66, 67]. Alternatively, NICE provided guidance on the wording of phrases to reflect the strength of the recommendations rather than using explicit grades or categories [72]. Both the USPSTF and SIGN included an option to make a non-specific recommendation for instances of genuine uncertainty [68, 70].

The eighth consensus statement included in this review, the SUPPORT guidelines [46] was the only statement aimed solely at evidence-informed health policy making, including decisions about healthcare services. The SUPPORT guidelines acknowledge controlled trials and systematic reviews as important, but in addition, they emphasise the value of obtaining other information and local evidence of modifying factors such as needs, values, costs and the availability of resources. Importantly, they also offer guidance on preparing and using policy briefs [53, 61].

Following a detailed analysis of the eight selected consensus statements and their related publications, five main themes emerged that were relevant to the challenges of developing T&CM recommendations and are particularly relevant when there is low quality, conflicting or inconsistent evidence. These were:

1. The importance of framing the question.
2. The limitations of an evidence hierarchy.
3. Methods for dealing with insufficient evidence.
4. Qualifying a recommendation.
5. Structured consensus development.

Framing the question

All eight statements provided guidance about clarifying at the outset of the guideline development process the

intended scope, questions, interventions and outcomes to be covered. The PICO process (Patient problem, Intervention, Comparison, Outcome) was often recommended to help formulate clinically relevant questions and patient-important outcomes were increasingly emphasised [36, 67, 68, 72].

Little guidance was provided however about methods for systematically identify potentially relevant interventions and selecting interventions for further in-depth systematic reviews. The WHO and NICE both provided guidance around choosing priority topics and interventions [7, 72]. This included interventions that were commonly used with unclear benefits and risks. The NICE 2012 edition was the only consensus statement that specifically mentioned high T&CM use by patients for managing the problem as a reason for inclusion, and the importance of searching databases relevant to T&CM evidence [73].

“The effects of complementary and alternative therapies may be addressed in the guideline if such therapies are commonly used in the clinical area of interest. If commonly used complementary and alternative therapies are not to be covered in the guideline, this should be stated clearly in the scope.” [73].

Limitations of an evidence hierarchy

As the various recommendations for developing guidelines have been updated, there has been a move away from using a ‘hierarchy of evidence’ or ‘levels of evidence’ towards the GRADE approach to making recommendations. This is due to ongoing concerns that a hierarchy can inappropriately encourage guideline developers and policy makers to directly link study design to recommendation strength, or ignore lower levels of evidence that should also be included when grading the strength of the recommendation [74]. NICE ceased using an evidence hierarchy in 2007–8, followed by SIGN in 2012; notwithstanding, the 2014 edition of SIGN still refers to levels of evidence in the “Example pages from an evidence table” [68].

The Australian National Health and Medical Research Council (NHMRC) guideline was the only included consensus statement that continued to use an evidence hierarchy as a direct constrainer on the strength of the recommendations [66]. According to these guidelines, the strongest recommendations, an ‘A’ or ‘B’, can only be made if the evidence quality is also graded as an ‘A’ or ‘B’. The NHMRC does acknowledge that questions about safety – especially for uncommon adverse events from treatments or harms from diagnostic testing – are unlikely to be answered through randomised controlled trials and in such cases, consideration of lower levels of evidence are permitted.

Dealing with insufficient evidence

GRADE and the USPSTF statements provided the most specific advice on how to manage the challenges of insufficient, high quality evidence. The SUPPORT statement most clearly emphasised that inconclusive results or lack of research should not be misinterpreted as evidence of no effect. Despite insufficient evidence about effectiveness, informed decisions can still be made about interventions that are potentially harmful or when the potential benefits are not worth the cost [20].

In situations of low quality evidence for an intervention and a lack of confidence in the effect estimates of the risks and benefits, the GRADE statements outlined five instances when a strong recommendation could still be made [18]. Table 4 is a modification of the GRADE guidelines where the original non-T&CM examples are replaced with examples pertinent to T&CM.

In the case of inconclusive or absent evidence from randomised controlled trials (RCTs) and meta-analyses, the USPSTF proposed several instances where the assemblage of non-RCT evidence would be admissible in clinical guidelines. The first is where an intervention is potentially effective, there is a large burden of disease and there is no research investigating the direct effects of the intervention on the health outcome [71]. In this instance, a Generic Analytic Framework (GAF) could be constructed to answer a sequence of key questions that form a chain of evidence about benefits and risks [71]. Recommendations can then be formulated based on indirect evidence linking the intervention to the outcome. For example, an intervention that demonstrates a reduction in the incidence rate of ischaemic heart disease (IHD) is direct evidence. When the same intervention has only demonstrated an ability to lower a person’s weight, other research must be linked to provide indirect evidence that losing weight can reduce known IHD risk factors and the likelihood of developing IHD. The safety, acceptability and costs of the intervention are also considered. The USPSTF further recognises that different types and quality of evidence will be required to link the evidence.

The second instance proposed by the USPSTF for the assemblage of non-RCT evidence is when the intervention is not amenable to being evaluated under RCT conditions [21]. Examples given included various behavioural interventions for substance abuse where either there is no appropriate control for blinding, or it is impossible to provide the treatment fidelity required for an RCT because it would eliminate the individualised, adaptive treatment approach that is needed for success.

For instances when despite using the above two suggestions there remains insufficient evidence, the USPSTF recommends structuring information around the following four domains to explicitly present data for decision makers:

Table 4 The application of the GRADE “Paradigmatic situations in which a strong recommendation may be warranted despite low or very low confidence in effect estimates” for T&CM^a

1. When low quality evidence suggests benefit in a life-threatening situation (evidence regarding harms can be low or high)

e.g. Very low quality evidence for intravenous silibinin for life-threatening amatoxin mushroom poisoning [91, 92], or intravenous ascorbate for life-threatening viral infections [93, 94]. Recommend to use if there are no alternate proven interventions.
2. When low quality evidence suggests benefit and high quality evidence suggests harm or a very high cost

e.g. Low quality evidence that intravenous ascorbate improves quality of life and reduces chemotherapy toxicity in cancer treatment, and high quality of evidence of very high costs compared to oral ascorbate [95]. If there are significant opportunity costs to the patient, recommend not to use intravenously.
3. When low quality evidence suggests equivalence of two alternatives, but high quality evidence of less harm for one of the competing alternatives

e.g. Low quality evidence of equivalence of glucosamine and non-steroidal anti-inflammatory drugs (NSAIDs) for the long-term symptomatic management of osteoarthritis; but high quality evidence of increased gastrointestinal and cardiovascular risks for NSAIDs only [96, 97]. Recommend glucosamine as first-line treatment, especially for patients with a higher risk of complications from NSAIDs.
4. When high quality evidence suggests equivalence of two alternatives and low quality evidence suggests harm in one alternative

e.g. High quality equivalence of certain proprietary extracts of St. John's Wort and selective serotonin reuptake inhibitors (SSRIs) for depression; and low quality evidence of a higher risk of suicide with SSRIs [98–100]. If these certain proprietary St. John's Wort extracts are available and affordable, then recommend as first-line treatment before an SSRI.
5. When high quality evidence suggests modest benefits and low/very low quality evidence suggests the possibility of catastrophic harm

e.g. High quality evidence of modest benefit from peri-operative use of fish oil for elective Coronary Artery Bypass Surgery; and low quality evidence for catastrophic haemorrhage [80–82]. Recommend not to cancel CAB surgery if the patient has taken fish oil pre-operatively.

^aModified from GRADE where non-TC&M examples were presented [18]

1. Burden of suffering – the incidence and prevalence of a condition; the degree of personal, family and community suffering; and the burden to families, society and health care systems.
2. Potential harm – the immediate and long-term harms to individuals and patients from delivering an intervention or service and from alternatives, including the potential harms associated with doing nothing.
3. Cost – the direct monetary costs of a service or intervention; the opportunity costs, such as the time, money and resources that would be diverted to provide an intervention with less evidence or acceptability to patients; and the costs of decommissioning the intervention should it then prove to be ineffective.
4. Current practice – the potential negative consequences (including legal) of providing a novel, less widely used service or intervention compared to those commonly in place; and the extra resources

that will be needed to change ingrained practice [21]. (In the case of T&CM, this question might also be extended to consider the consequences of removing or restricting access to commonly used interventions).

Both the USPSTF and SIGN included a category for recommending the use of interventions in the research setting only [68, 70]. The USPSTF stated that only-in-research recommendations should be reserved for promising interventions where there is the potential to cause significant harm or there are high costs [70]. The latter includes interventions where there is a large component of fixed costs that cannot be retrieved if the intervention is withdrawn [21]. Conversely, GRADE did not provide an only-in-research category. Such a recommendation is possible however, if the following three conditions are met:

1. There is genuine uncertainty from the existing evidence;
2. Further research is very likely to remove or reduce this uncertainty; and
3. The cost of further research is deemed to be good value [19].

Qualifying a recommendation

There was general consensus across all statements that an evidence-based guideline is unhelpful if it fails to provide information about modifying factors. Contextual information about the burden of disease and available interventions; generalisability and applicability to population groups; direct and indirect costs; demand, accessibility and equity; and the values and preferences of patients and providers is increasingly being used to help select interventions, identify relevant outcomes for appraising the evidence, provide information about benefits and risks, and to qualify recommendations [4–7, 72].

High quality evidence in support of these modifying factors may justify upgrading or downgrading a recommendation [18]. For example, patients may consider the most effective intervention to be unacceptable due to their personal tolerance for risk, or other personal values such as a preference for natural therapies. In the case of healthcare providers and policy makers, equity, costs and current service provision are likely to be influencing factors. An intervention with small clinical impact (effect size) that is widely used or readily available, may be preferred to an intervention with large clinical impact that is significantly more expensive or requires substantial system changes to integrate into practice. That patients or policy makers make different choices based on preferences, values and costs, are reasons why an intervention with high quality scientific evidence of efficacy may still be downgraded to a weak recommendation and vice versa [18, 19].

The NHMRC proposed a system that includes the grading of modifying factors [65, 66]. The NHMRC Evidence Matrix grades the evidence for safety, efficacy, cost-effectiveness, consistency of results, clinical impact, the generalisability of the evidence and its applicability to the Australian healthcare setting. Evidence about other important modifying factors however, such as patient and provider preferences were not included.

Only-in-research recommendations also require qualification. GRADE for example actively discouraged blanket statements recommending further scientific research [18]. Instead, such recommendations should include justification of the need for further research and detail the research questions with particular attention given to patient-important outcomes [19, 21].

Structured consensus development

All eight statements in this review emphasised that membership of a guideline development committee should represent the relevant stakeholders. AWMF 2.0 was the only statement however to recommend and outline scientifically sound formal consensus methods to promote transparency and resolve conflicts arising from differences of opinion [67]. Given the complexity of the decision-making process that necessitates sourcing and appraising all the information, non-objective personal and professional biases are likely to emerge when selecting interventions and outcomes, appraising modifying factors, and formulating recommendations. Standardised methods such as the Nominal Group Process, the Structured Consensus Conference and the Delphi Technique were recommended. The ultimate aim is to improve the transparency, quality, reproducibility and acceptability of the recommendations [67].

Discussion

This is first known review to synthesise the content and themes of national and international consensus statements for developing clinical and health policy guidelines and to interpret these through the lens of T&CM. Given the influence of the evidence-based medicine movement on clinical practice, education and health policy, it is not surprising that the majority of statements reviewed in this paper provided detailed guidance on how to systematically identify and appraise evidence of efficacy [27]. The limitations of using a didactic 'recipe book' approach when formulating recommendations was increasingly being recognised; particularly the limitations of using an evidence hierarchy and the importance of modifying factors [24, 74]. The USPSTF statements provided the clearest guidance and strategies for dealing with insufficient evidence.

Notwithstanding alternate, more pragmatic approaches to evidence appraisal such as those proposed by the

USPSTF, the paucity and heterogeneity of scientific evidence for many T&CM interventions remains a significant challenge to guideline developers. It is important not to imply that inconsistent evidence or an absence of evidence means there is evidence of no effect [20]. In these instances the general consensus was that guidelines should still attempt to make specific recommendations or at least offer some information to help guide decisions [18, 20, 21]. Table 4 lists the paradigmatic circumstances proposed by GRADE where a strong recommendation could be made despite low quality evidence [18]. Guideline developers should be mindful of these instances and not automatically default to a recommendation not to use an intervention based solely on low quality scientific evidence regarding efficacy [18, 21].

The early use of an evidence hierarchy that places the RCT and meta-analyses at the pinnacle may help explain the ad-hoc inclusion and appraisal of T&CM in clinical guidelines, especially older guidelines [16]. If higher levels of evidence are lacking and lower levels of evidence are discounted with no qualifying statements, gaps in the evidence review are likely to occur and an intervention overlooked [74]. The guidelines may then default to non-informative statements and recommendations, as was found to be the case in the reviews of UK clinical guidelines [16]. Consistent with international standards, bodies such as the Australian NHMRC should cease endorsing the use of 'levels of evidence' as a direct constrainer of ensuing recommendations and instead make greater use of qualifying statements that consider important modifying factors, including those relevant to patients and practitioners.

The USPSTF suggested a number of instances when the double-blind RCT is not the most appropriate study design [18–21]. Although the specific components of a T&CM intervention may be amenable to assessment using an RCT design, there are many instances where this is not appropriate [75]. For example, for some T&CM interventions, finding an adequate control may be difficult or impossible; and for others, treatment fidelity would be lost due to the individualised, multifaceted approach of the therapy or the complexity of the study outcomes that are multiple and holistic, with some being immediate and others delayed [76, 77].

A potential T&CM example for the assemblage of admissible non-RCT evidence is acupuncture for depression [21, 71]. Depression is an illness where there is a large burden of disease and there is growing pragmatic evidence of effectiveness, but weak or conflicting evidence from double-blind RCTs about the efficacy of acupuncture [78]. The challenge with finding a suitable control for acupuncture, as well as the individualised nature of the intervention, may explain the mixed results from efficacy (explanatory) trials compared to the more

consistent positive results from effectiveness (pragmatic) trials [79]. In cases such as this, it may even be justified to give a lower weighting to the quality score of study designs that use a non-individualised treatment protocol or an inappropriate control.

Due to the paucity of a large body of high quality evidence regarding efficacy for many T&CM interventions, a common recommendation from systematic reviews and clinical guidelines is to make a general call for further research. This is unhelpful to clinicians and patients who need immediate guidance and should only be made if the research is warranted [19, 21]. A recommendation for further research should only be made for interventions where there is true uncertainty about risks and benefits; especially if there are large direct costs or opportunity costs, or there is the potential for large benefit from wider, more equitable use [20]. For example, along with the strong recommendation not to cancel coronary artery bypass surgery if a patient has taken fish oil preoperatively (see Table 4: example 5) [80–82]; recommendations for further research are justified. Treatment duration and doses of EPA and DHA requires further clarification. There is potential for different populations to disproportionately benefit (e.g. socioeconomic, ethnic, or other groups with specific cardiovascular risk factors). Economic evaluations are also warranted since the cost of fish oil, even with only modest clinical benefit, may be cost-effective compared to the cost of surgical complications; and health inequalities are a concern since patients commonly pay 100% out-of-pocket to use fish oil.

Including modifying factors when qualifying recommendations enhances their relevance to different clinical scenarios and populations [18, 21]. The diverse and potentially conflicting information about efficacy and relevant modifying factors is particularly challenging for guideline developers. Modifying factors can be used for example to upgrade or downgrade the strength of a recommendation independent of the quality of the evidence [18, 19]. This point is particularly relevant to T&CM interventions where there is insufficient high quality scientific evidence regarding efficacy and effectiveness. In these instances, high quality evidence may still be available about the burden of disease; risks of alternate therapies; direct and indirect costs; demand, access, affordability and equity; generalisability and applicability of the intervention to specific population groups; patient and provider values and preferences; and implementation and feasibility [7–10, 27, 32, 83]. It is therefore inappropriate to limit systematic literature reviews for informing guideline development only to questions about safety, efficacy and cost-effectiveness.

To elaborate, there is high quality evidence that hormone replacement therapy (HRT) is effective for managing menopausal symptoms; however, there is also

high quality evidence about the risks of HRT. [84, 85] By contrast, there is conflicting evidence about the efficacy of Black Cohosh for managing menopausal symptoms and very low quality evidence questioning its safety [86, 87]. There is also high quality evidence that some women would prefer to use potentially less effective natural approaches to manage these non-life-threatening symptoms, of which herbs such as Black Cohosh are amongst the most popular choices [88–90]. Although many clinical guidelines qualify the recommendation to use HRT with a statement about assessing the risks and benefits of hormone use for an individual patient, most fail to make any qualifying statements about known patient preferences to use T&CM, its comparative safety, and the direct costs and opportunity costs of first trialling a potentially less efficacious intervention [90].

The inconsistencies regarding the inclusion of T&CM and recommendations made about their use in clinical guidelines calls for a more transparent and systematic approach to guideline development. Even formal methods for consensus development such as those outlined in AWMF [67] will be prone to bias if the expert committee for example, brainstorm or uses other non-systematic methods to select comparison interventions. The scope, interventions and outcomes will then likely reflect the experience and knowledge of the members of the committee, or other biases such as only considering interventions that are thought to have high quality evidence and worthy of consideration. As WHO and NICE highlighted, amongst other reasons (see Table 1) T&CM should be considered if they are commonly used in the clinical context [7, 73], irrespective of the quality of evidence about their benefits and risks [7].

Conclusion & recommendations

This interpretive review has considered, for the first time, the usefulness of directives for developing guidelines and recommendations regarding T&CM practice and policy. Like many areas of healthcare, insufficient evidence about efficacy poses significant challenges to guideline developers, which in the field of T&CM has contributed towards insufficient and inconsistent recommendations. The emerging and heterogeneous evidence-base for many T&CM interventions necessitates a range of methodologies to ensure the systematic selection of interventions and consideration of modifying factors when formulating and qualifying recommendations. In light of these issues and the high demand for T&CM, we believe guideline developers to consider T&CM from a number of perspectives when appraising the evidence, and to make clinically useful and specific recommendations regarding their use.

Specifically, guideline developers should cease endorsing an evidence hierarchy as a direct constrainer of

recommendations. Strict use of the levels of evidence runs the risk of inappropriately linking the quality of the evidence for efficacy directly to the strength of the recommendation, whilst ignoring admissible non-RCT evidence and important modifying factors. In instances of very low quality or equivocal evidence of efficacy, guideline developers must consider the paradigmatic situations where nonetheless a strong recommendation can be made. Failing this, broader contextual information is often available for T&CM even when there is low quality scientific evidence regarding efficacy. Information about modifying factors should be presented to facilitate informed decision making and improve clinical relevance. Finally, greater attention must be given to adopting a systematic and transparent approach to the entire development process, including the selection of comparative interventions and patient relevant outcomes. The uptake of these recommendations is expected to result in higher quality clinical guidelines that offer greater assistance to those seeking answers about the appropriate use of T&CM.

Funding

None.

Availability of data and materials

Not applicable.

Authors' contributions

JH conceived the study design, undertook the literature searching and wrote the draft manuscripts. ML, LB, AB reviewed and interpreted the literature, and provided content and editorial input. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Author details

¹NICM, University of Western Sydney, Locked Bag 1797, Penrith, NSW 2751, Australia. ²Menzies Centre for Health Policy, School of Medicine, University of Sydney, Sydney, Australia. ³Australian Research Centre in Complementary & Integrative Medicine, University of Technology Sydney, Sydney, Australia. ⁴Blackmores Institute, Sydney, Australia. ⁵Monash/Alfred Psychiatric Research Centre, Monash University, Melbourne, Australia.

Received: 26 July 2016 Accepted: 27 January 2017

Published online: 17 February 2017

References

- Ooi SL, Rae J, Pak SC. Implementation of evidence-based practice: A naturopath perspective. *Complement Ther Clin Pract.* 2016;22:24–8.
- Leach MJ. Clinical decision making in complementary and alternative medicine. Sydney: Churchill Livingstone; 2010.
- Field MJ, Lohr KN. Clinical practice guidelines: directions for a new program. Washington: National Academy Press; 1990.
- Truman BL, Smith-Akin CK, Hinman AR, Gebbie KM, Brownson R, Novick LF, Lawrence RS, Pappaioanou M, Fielding J, Evans Jr CA, et al. Developing the Guide to Community Preventive Services—overview and rationale. The Task Force on Community Preventive Services. *Am J Prev Med.* 2000;18(1 Suppl):18–26.
- Woolf SH, DiGiuseppi CG, Atkins D, Kamerow DB. Developing evidence-based clinical practice guidelines: lessons learned by the US Preventive Services Task Force. *Annu Rev Public Health.* 1996;17:511–38.
- Shekelle PG, Woolf SH, Eccles M, Grimshaw J. Clinical guidelines: developing guidelines. *BMJ.* 1999;318(7183):593–6.
- WHO. Guidelines for WHO guidelines. Geneva: World Health Organisation; 2003. p. 46.
- Eccles M, Mason J. How to develop cost-conscious guidelines. *Health Technol Assess.* 2001;5(16):1–69.
- Woolf S, Schunemann HJ, Eccles MP, Grimshaw JM, Shekelle P. Developing clinical practice guidelines: types of evidence and outcomes; values and economics, synthesis, grading, and presentation and deriving recommendations. *Implement Sci.* 2012;7:61.
- Edejer TT. Improving the use of research evidence in guideline development: 11. Incorporating considerations of cost-effectiveness, affordability and resource implications. *Health Res Policy Syst.* 2006;4:23.
- NHMRC. Review of the Australian Government Rebate on Natural Therapies for Private Health Insurance. Canberra: Commonwealth of Australia; 2015.
- Brouwers M, Kho ME, Browman GP, Cluzeau F, feder G, Fervers B, Hanna S, Makarski J on behalf of the AGREE Next Steps Consortium. AGREE II Instrument: Advancing guideline development, reporting and evaluation in healthcare. The AGREE Research Trust, May 2009, UPDATE: September 2013. [<http://www.agreertrust.org/>]. Accessed 1 Sept 2014.
- Semlitsch T, Blank WA, Kopp IB, Siering U, Siebenhofer A. Evaluating Guidelines: A Review of Key Quality Criteria. *Dtsch Arztebl Int.* 2015; 112(27–28):471–8.
- Ernst E. Assessments of complementary and alternative medicine: the clinical guidelines from NICE. *Int J Clin Pract.* 2010;64(10):1350–8.
- Ernst E, Terry R. NICE guidelines on complementary/alternative medicine: more consistency and rigour are needed. *Br J Gen Pract.* 2009;59(566):695.
- Lorenc A, Leach J, Robinson N. Clinical guidelines in the UK: Do they mention Complementary and alternative medicine (CAM) – Are CAM professional bodies aware? *Eur J Intern Med.* 2014;6(2):164–75.
- Robinson N, Lui J, Lee MS. Clinical guidelines: The way for best practice. *Eur J Intern Med.* 2014;6(2):133–4.
- Andrews JC, Schunemann HJ, Oxman AD, Pottie K, Meerpohl JJ, Coello PA, Rind D, Montori VM, Brito JP, Norris S, et al. GRADE guidelines: 15. Going from evidence to recommendation: determinants of a recommendation's direction and strength. *J Clin Epidemiol.* 2013;66(7):726–35.
- Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, Nasser M, Meerpohl J, Post PN, Kunz R, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. *J Clin Epidemiol.* 2013;66(7):719–25.
- Oxman AD, Lavis JN, Fretheim A, Lewin S. SUPPORT Tools for evidence-informed health Policymaking (STP) 17: Dealing with insufficient research evidence. *Health Res Policy Syst.* 2009;7 Suppl 1:S17.
- Petitti DB, Teutsch SM, Barton MB, Sawaya GF, Ockene JK, DeWitt T. Update on the methods of the U.S. Preventive Services Task Force: insufficient evidence. *Ann Intern Med.* 2009;150(3):199–205.
- Walach H, Falkenberg T, Fonnebo V, Lewith G, Jonas WB. Circular instead of hierarchical: methodological principles for the evaluation of complex interventions. *BMC Med Res Methodol.* 2006;6:29.
- Walach H, Loeff M. Using a matrix-analytical approach to synthesising evidence solved incompatibility problem in the Hierarchy of Evidence. *J Clin Epidemiol.* 2015.
- Greenhalgh T, Howick J, Maskrey N. Evidence based medicine: a movement in crisis? *BMJ.* 2014;13:348.
- Noblit G, Hare R. *Meta-ethnography: synthesizing qualitative studies.* Newbury Park: Sage; 1988.
- Dixon-Woods M, Cavers D, Agarwal S, Annandale E, Arthur A, Harvey J, Hsu R, Katbamna S, Olsen R, Smith L, et al. Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. *BMC Med Res Methodol.* 2006;6:35.
- Schunemann HJ, Wiercioch W, Etzeandia I, Falavigna M, Santesso N, Mustafa R, Ventresca M, Brignardello-Petersen R, Laisaar KT, Kowalski S, et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. *CMAJ.* 2014;186(3):E123–142.
- Booth A, Harris J, Croot E, Springett J, Campbell F, Wilkins E. Towards a methodology for cluster searching to provide conceptual and contextual

- "richness" for systematic reviews of complex interventions: case study (CLUSTER). *BMC Med Res Methodol.* 2013;13:118.
29. Bodeker GBG. *Traditional, Complementary and Alternative Medicine: Policy and Public Health Perspectives.* UK: Oxford University; 2007.
 30. Paterson C, Baarts C, Launso L, Verhoef MJ. Evaluating complex health interventions: a critical analysis of the 'outcomes' concept. *BMC Complement Altern Med.* 2009;9:18.
 31. Coulter ID, Willis EM. The rise and rise of complementary and alternative medicine: a sociological perspective. *Med J Aust.* 2004;180(11):587–9.
 32. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, DeBeer H, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol.* 2011; 64(4):383–94.
 33. Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, Vist GE, Falck-Ytter Y, Meerpohl J, Norris S, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol.* 2011;64(4):401–6.
 34. Brunetti M, Shemilt I, Pregno S, Vale L, Oxman AD, Lord J, Sisk J, Ruiz F, Hill S, Guyatt GH, et al. GRADE guidelines: 10. Considering resource use and rating the quality of economic evidence. *J Clin Epidemiol.* 2013;66(2):140–50.
 35. GRADE Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. In: Schünemann H, Brozek J, Guyatt G, Oxman A, editors. 2013.
 36. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, Alderson P, Glasziou P, Falck-Ytter Y, Schunemann HJ. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *J Clin Epidemiol.* 2011;64(4): 395–400.
 37. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Falck-Ytter Y, Jaeschke R, Vist G, et al. GRADE guidelines: 8. Rating the quality of evidence—indirectness. *J Clin Epidemiol.* 2011;64(12):1303–10.
 38. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Glasziou P, Jaeschke R, Akl EA, et al. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. *J Clin Epidemiol.* 2011; 64(12):1294–302.
 39. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, Alonso-Coello P, Djulbegovic B, Atkins D, Falck-Ytter Y, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol.* 2011;64(12):1277–82.
 40. Guyatt GH, Oxman AD, Santesso N, Helfand M, Vist G, Kunz R, Brozek J, Norris S, Meerpohl J, Djulbegovic B, et al. GRADE guidelines: 12. Preparing summary of findings tables-binary outcomes. *J Clin Epidemiol.* 2013;66(2):158–72.
 41. Guyatt GH, Oxman AD, Schunemann HJ. GRADE guidelines—an introduction to the 10th–13th articles in the series. *J Clin Epidemiol.* 2013;66(2):121–3.
 42. Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, Atkins D, Kunz R, Brozek J, Montori V, et al. GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol.* 2011;64(12):1311–6.
 43. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, Montori V, Akl EA, Djulbegovic B, Falck-Ytter Y, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). *J Clin Epidemiol.* 2011; 64(4):407–15.
 44. Guyatt GH, Thorlund K, Oxman AD, Walter SD, Patrick D, Furukawa TA, Johnston BC, Karanicolas P, Akl EA, Vist G, et al. GRADE guidelines: 13. Preparing summary of findings tables and evidence profiles-continuous outcomes. *J Clin Epidemiol.* 2013;66(2):173–83.
 45. GRADEpro GDT [www.gradepr.org]. Accessed 4 April 2016
 46. Lavis JN, Oxman AD, Lewin S, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP). *Health Res Policy Syst.* 2009;7 Suppl 1:11.
 47. Fretheim A, Munabi-Babigumira S, Oxman AD, Lavis JN, Lewin S. SUPPORT Tools for evidence-informed health Policymaking (STP) 6: Using research evidence to address how an option will be implemented. *Health Res Policy Syst.* 2009;7 Suppl 1:56.
 48. Fretheim A, Oxman AD, Lavis JN, Lewin S. SUPPORT Tools for evidence-informed health Policymaking (STP) 18: Planning monitoring and evaluation of policies. *Health Res Policy Syst/BioMed Centra.* 2009;7 Suppl 1:518.
 49. Lavis JN, Boyko JA, Oxman AD, Lewin S, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP) 14: Organising and using policy dialogues to support evidence-informed policymaking. *Health Res Policy Syst.* 2009;7 Suppl 1:514.
 50. Lavis JN, Oxman AD, Grimshaw J, Johansen M, Boyko JA, Lewin S, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP) 7: Finding systematic reviews. *Health Res Policy Syst.* 2009;7 Suppl 1:57.
 51. Lavis JN, Oxman AD, Lewin S, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP) 3: Setting priorities for supporting evidence-informed policymaking. *Health Res Policy Syst.* 2009;7 Suppl 1:53.
 52. Lavis JN, Oxman AD, Souza NM, Lewin S, Gruen RL, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP) 9: Assessing the applicability of the findings of a systematic review. *Health Res Policy Syst.* 2009;7 Suppl 1:59.
 53. Lavis JN, Permanand G, Oxman AD, Lewin S, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP) 13: Preparing and using policy briefs to support evidence-informed policymaking. *Health Res Policy Syst.* 2009;7 Suppl 1:513.
 54. Lavis JN, Wilson MG, Oxman AD, Grimshaw J, Lewin S, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP) 5: Using research evidence to frame options to address a problem. *Health Res Policy Syst.* 2009;7 Suppl 1:55.
 55. Lavis JN, Wilson MG, Oxman AD, Lewin S, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP) 4: Using research evidence to clarify a problem. *Health Res Policy Syst.* 2009;7 Suppl 1:54.
 56. Lewin S, Oxman AD, Lavis JN, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP) 8: Deciding how much confidence to place in a systematic review. *Health Res Policy Syst.* 2009;7 Suppl 1:58.
 57. Lewin S, Oxman AD, Lavis JN, Fretheim A, Garcia Marti S, Munabi-Babigumira S. SUPPORT Tools for evidence-informed health Policymaking (STP) 11: Finding and using evidence about local conditions. *Health Res Policy Syst.* 2009;7 Suppl 1:511.
 58. Oxman AD, Fretheim A, Lavis JN, Lewin S. SUPPORT Tools for evidence-informed health Policymaking (STP) 12: Finding and using research evidence about resource use and costs. *Health Res Policy Syst.* 2009;7 Suppl 1:512.
 59. Oxman AD, Lavis JN, Fretheim A, Lewin S. SUPPORT Tools for evidence-informed health Policymaking (STP) 16: Using research evidence in balancing the pros and cons of policies. *Health Res Policy Syst.* 2009;7 Suppl 1:516.
 60. Oxman AD, Lavis JN, Lewin S, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP) 10: Taking equity into consideration when assessing the findings of a systematic review. *Health Res Policy Syst.* 2009;7 Suppl 1:510.
 61. Oxman AD, Lavis JN, Lewin S, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP) 1: What is evidence-informed policymaking? *Health Res Policy Syst.* 2009;7 Suppl 1:51.
 62. Oxman AD, Lewin S, Lavis JN, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP) 15: Engaging the public in evidence-informed policymaking. *Health Res Policy Syst.* 2009;7 Suppl 1:515.
 63. Oxman AD, Vandvik PO, Lavis JN, Fretheim A, Lewin S. SUPPORT Tools for evidence-informed health Policymaking (STP) 2: Improving how your organisation supports the use of research evidence to inform policymaking. *Health Res Policy Syst.* 2009;7 Suppl 1:52.
 64. NHMRC. A guide to the development, evaluation and implementation of clinical practice guidelines. National Health & Medical Research Council. Canberra: Commonwealth of Australia; 1999.
 65. Hillier S, Grimmer-Somers K, Merlin T, Middleton P, Salisbury J, Tooher R, Weston A. FORM: an Australian method for formulating and grading recommendations in evidence-based clinical guidelines. *BMC Med Res Methodol.* 2011;11:23.
 66. Merlin T, Weston A, Tooher R, Middleton P, Salisbury J, Coleman K, Norris S, Grimmer-Somers K, Hillier S. NHMRC additional levels of evidence and grades for recommendations for developers of guidelines. Australia: National Health and Medical Research Council; 2009.
 67. German Association of the Scientific Medical Societies (AWMF) - Standing Guidelines Commission. AWMF Guidance Manual and Rules for Guideline Development, 1st Edition 2012. English version. Available at: <http://www.awmf.org/leitlinien/awmf-regelwerk.html>. (Accessed 9 Sept 2014).
 68. SIGN. SIGN 50 - A guideline developer's handbook. Edinburgh: Scottish Intercollegiate Guidelines Network (SIGN); 2014.
 69. USPSTF: U.S. Preventive Services Task Force Procedure Manual. AHRQ Publication No. 08-05118-EF edn. 2008
 70. Grade Definitions After July 2012. [<http://www.uspreventiveservicestaskforce.org/Page/Name/grade-definitions>]. Accessed 10 April 2014
 71. Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, Atkins D. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med.* 2001;20(3 Suppl):21–35.
 72. NICE. Developing NICE guidelines: the manual. Process and methods guide. London: National Institute for Health and Care Excellence; 2014.

73. NICE. The guidelines manual. Process and methods guide. London: National Institute for Health and Care Excellence; 2012.
74. Thornton J, Alderson P, Tan T, Turner C, Latchem S, Shaw E, Ruiz F, Reken S, Mugglestone MA, Hill J, et al. Introducing GRADE across the NICE clinical guideline program. *J Clin Epidemiol*. 2013;66(2):124–31.
75. Boon H, Macpherson H, Fleishman S, Grimsgaard S, Koithan M, Norheim AJ, Walach H. Evaluating Complex Healthcare Systems: A Critique of Four Approaches. *Evid Based Complement Alternat Med*. 2007;4(3):279–85.
76. Deng G, Weber W, Sood A, Kemper KJ. Research on integrative healthcare: context and priorities. *Explore (NY)*. 2010;6(3):143–58.
77. Hunter J, Corcoran K, Leeder S, Phelps K. Integrative medicine outcomes: What should we measure? *Complement Ther Clin Pract*. 2013;19(1):20–6.
78. MacPherson H. Acupuncture for depression: state of the evidence. *Acupunct Med*. 2014;32(4):304–5.
79. Smith Caroline A, Hay Phillipa PJ, MacPherson H. Acupuncture for depression. In: *Cochrane Database of Systematic Reviews*. John Wiley & Sons, Ltd; 2010.
80. Costanzo S, di Niro V, Di Castelnuovo A, Gianfagna F, Donati MB, de Gaetano G, Iacoviello L. Prevention of postoperative atrial fibrillation in open heart surgery patients by preoperative supplementation of n-3 polyunsaturated fatty acids: an updated meta-analysis. *J Thorac Cardiovasc Surg*. 2013;146(4):906–11.
81. Xin W, Wei W, Lin Z, Zhang X, Yang H, Zhang T, Li B, Mi S. Fish oil and atrial fibrillation after cardiac surgery: a meta-analysis of randomized controlled trials. *PLoS One*. 2013;8(9):e72913.
82. Liu H. The Effect of Fish Oil on Hemostasis and Coagulations During Cardiac Surgery. *SCA Bulletin: Drug & Innovation Update*; 2013. 12(2). <http://www.scahq.org/sca3/newsletters/2013apr/diu.html>.
83. Makarski J, Brouwers MC. The AGREE Enterprise: a decade of advancing clinical practice guidelines. *Implement Sci*. 2014;9:103.
84. MacLennan AH, Broadbent JL, Lester S, Moore V. Oral oestrogen and combined oestrogen/progestogen therapy versus placebo for hot flushes. *Cochrane Database Syst Rev*. 2004;4:CD002978.
85. Rossouw JE, Manson JE, Kaunitz AM, Anderson GL. Lessons learned from the Women's Health Initiative trials of menopausal hormone therapy. *Obstet Gynecol*. 2013;121(1):172–6.
86. Beer AM, Osmers R, Schnitker J, Bai W, Mueck AO, Meden H. Efficacy of black cohosh (*Cimicifuga racemosa*) medicines for treatment of menopausal symptoms - comments on major statements of the Cochrane Collaboration report 2012 "black cohosh (*Cimicifuga* spp.) for menopausal symptoms (review)". *Gynecol Endocrinol*. 2013;29(12):1022–5.
87. Leach MJ, Moore V. Black cohosh (*Cimicifuga* spp.) for menopausal symptoms. *Cochrane Database Syst Rev*. 2012;9:CD007244.
88. Peng W, Adams J, Sibbritt DW, Frawley JE. Critical review of complementary and alternative medicine use in menopause: focus on prevalence, motivation, decision-making, and communication. *Menopause*. 2014;21(5):536–48.
89. Posadzki P, Lee MS, Moon TW, Choi TY, Park TY, Ernst E. Prevalence of complementary and alternative medicine (CAM) use by menopausal women: a systematic review of surveys. *Maturitas*. 2013;75(1):34–43.
90. Gentry-Maharaj A, Karpinskyj C, Glazer C, Burnell M, Ryan A, Fraser L, Lanceley A, Jacobs I, Hunter MS, Menon U. Use and perceived efficacy of complementary and alternative medicines after discontinuation of hormone therapy: a nested United Kingdom Collaborative Trial of Ovarian Cancer Screening cohort study. *Menopause*. 2015;22(4):384–90.
91. Gores KM, Hamieh TS, Schmidt GA. Survival following investigational treatment of amanita mushroom poisoning: thistle or shamrock? *Chest*. 2014;146(4):e126–129.
92. Mengs U, Pohl RT, Mitchell T. Legalon(R) SIL: the antidote of choice in patients with acute hepatotoxicity from amatoxin poisoning. *Curr Pharm Biotechnol*. 2012;13(10):1964–70.
93. Wilson JX. Mechanism of action of vitamin C in sepsis: ascorbate modulates redox signaling in endothelium. *Biofactors*. 2009;35(1):5–13.
94. Fowler 3rd AA, Syed AA, Knowlson S, Sculthorpe R, Farthing D, DeWilde C, Farthing CA, Larus TL, Martin E, Brophy DF, et al. Phase I safety trial of intravenous ascorbic acid in patients with severe sepsis. *J Transl Med*. 2014;12:32.
95. Jacobs C, Hutton B, Ng T, Shorr R, Clemons M. Is there a role for oral or intravenous ascorbate (vitamin C) in treating patients with cancer? A systematic review. *Oncologist*. 2015;20(2):210–23.
96. Kongtharvonkul J, Anothaisintawee T, McEvoy M, Attia J, Woratanarat P, Thakkinstian A. Efficacy and safety of glucosamine, diacerein, and NSAIDs in osteoarthritis knee: a systematic review and network meta-analysis. *Eur J Med Res*. 2015;20:24.
97. Bruyere O, Cooper C, Pelletier JP, Branco J, Luisa Brandi M, Guillemin F, Hochberg MC, Kanis JA, Kvien TK, Martel-Pelletier J, et al. An algorithm recommendation for the management of knee osteoarthritis in Europe and internationally: a report from a task force of the European Society for Clinical and Economic Aspects of Osteoporosis and Osteoarthritis (ESCEO). *Semin Arthritis Rheum*. 2014;44(3):253–63.
98. Linde K, Berner MM, Kriston L. St John's wort for major depression. *Cochrane Database Syst Rev*. 2008;4:CD000448.
99. Raja M, Azzoni A. Hypericum-induced mood disorder: switch from depression to mixed episodes in two patients. *Int J Psychiatry Clin Pract*. 2006;10(2):146–8.
100. Gunnell D, Saperia J, Ashby D. Selective serotonin reuptake inhibitors (SSRIs) and suicide in adults: meta-analysis of drug company data from placebo controlled, randomised controlled trials submitted to the MHRA's safety review. *BMJ*. 2005;330(7488):385.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

