

Supplementary file S2: Overview of studies examining WHO's guideline development process

Study	Aim	Method	Main findings
<p>Dedios MC, Esperato A, De-Regil LM, Peña-Rosas JP, Norris SL Improving the adaptability of WHO evidence-informed guidelines for nutrition actions: results of a mixed methods evaluation. Implement Sci. 2017 Mar 21;12(1):39. doi: 10.1186/s13012-017-0571-2.</p>	<ul style="list-style-type: none"> • To independently evaluate the guideline development process implemented by the Department of Nutrition for Health and Development • It specifically focused on assessing the adaptability of the nutrition guidelines • Adaptability was defined as the guidelines having methodological quality and implementability • The study investigated three main questions: <ol style="list-style-type: none"> (1) How adaptable are the guidelines in their current form? (2) What are the key characteristics of the nutrition guideline development process? (3) How can nutrition guideline development be modified in order to attain more adaptable guidelines? 	<ul style="list-style-type: none"> • The evaluation followed a rapid assessment procedure structure (RAP), starting with a pre-established question and conceptual model, and incorporated RAP specific features including <ol style="list-style-type: none"> (a) a focus on a narrow question; (b) small samples of key informants; (c) a short period of field research; (d) interview guides focusing on specific topics; (e) multiple data collection methods, including desk review and semi-structured interviews (n=12) <ul style="list-style-type: none"> • Guideline adaptability was assessed quantitatively using two standardized instruments: the Appraisal Guideline for Research and Evaluation questionnaire, version II (AGREE II) was used to assess guideline quality, while implementability was assessed with the electronic version of the GuideLine Implementability Appraisal (eGlia) 	<ul style="list-style-type: none"> • The evaluation of methodological quality indicated that the nutrition guidelines had increased in quality over the years • The scoping and defining the purpose of the guidelines, the use of systematic reviews, and the management of conflict of interests were indicated to be the main factors improving methodological quality • The study suggested that the process of formulating recommendations can be improved, including categorising a recommendation as conditional versus strong, and communicating more clearly the level of uncertainty implied in the conditional category • The evaluation of implementability suggested that applicability and executability of guidelines should be improved, including using specific strategies and tools to implement recommendations, that guidelines should more explicitly state the actions to be taken, and that guidelines must assess implementation barriers and consider costs, resources, and potential impact on health systems • The study also identified two additional aspects of the guideline development process that merit improvement <ol style="list-style-type: none"> (1) Improving the diversity of the guideline development group, particularly striking a sound balance between academics and implementers (2) Preparing the guideline development panel better for the evidence that is to be presented,

			<p>so that decision-making processes during guideline development meetings more effectively can be managed</p> <ul style="list-style-type: none"> Finally, the study identified that the interviewees indicated need for strengthening guidance related to the implementation of guidelines, including developing separate implementation guidelines or providing implementation guidance within the current guidelines
<p>Garritty CM, Norris SL, Moher D. Developing WHO rapid advice guidelines in the setting of a public health emergency. <i>J Clin Epidemiol.</i> 2017 Feb;82:47-60. doi: 10.1016/j.jclinepi.2016.08.010. Epub 2016 Aug 31.</p>	<ul style="list-style-type: none"> To describe the criteria that WHO staff use to assess the need for developing a rapid advice guideline and to outline the steps and methods for developing such a guideline in the context of a public health emergency 	<ul style="list-style-type: none"> The guidance on how to develop rapid advice guidelines was informed by an existing rapid review approach, and discussions with WHO staff involved in emergency response about staff roles, experiences, and needs with regard to development of rapid advice guidelines 	<ul style="list-style-type: none"> The paper discusses the criteria for considering if a rapid advice guideline is appropriate and feasible, outlines the roles of various contributors across the phases of development, and describe the methods and steps involved in performing rapid reviews The paper describes some important differences between rapid reviews and systematic reviews conducted for standard guideline development processes—including a shorter timeline, more limited scope and fewer outcomes of interest, more restricted search criteria, a more targeted and iterative procedure for screening the literature and for data analysis and synthesis, places less emphasis on meta-analyses, and involves a concise and abbreviated report In addition, in a rapid review, the search process is more iterative and hierarchical, depending on the findings at each step, and the types of publication and study designs included and the bibliographic databases searched may change as the evidence is explored The paper underlined that the core principles for WHO guidelines apply to rapid advice guidelines including minimizing bias, applying transparent processes and the use of explicit methods

<p>Glenton C, Lewin S, Gülmezoglu AM. Expanding the evidence base for global recommendations on health systems: strengths and challenges of the OptimizeMNH guidance process. <i>Implement Sci.</i> 2016 Jul 18;11:98.</p>	<ul style="list-style-type: none"> • To describe innovative strategies used to expand the evidence used as a basis for WHO recommendations, and the experience with these strategies 	<ul style="list-style-type: none"> • Experience with innovative strategies for broadening the evidence base beyond data on effectiveness alone— implemented during the development of the WHO recommendations on optimizing health worker roles for maternal and newborn health through task shifting— was described 	<ul style="list-style-type: none"> • During the stage of guidance development, systematic reviews of qualitative studies, multi-country case study syntheses, and qualitative evidence syntheses of the views of programme managers, policy makers and other stakeholders were strategies used to broaden the range of evidence included when developing the guidance • The inclusion of evidence from qualitative evidence syntheses and country case study syntheses reduced the use of anecdotal evidence by guidance panel members when assessing acceptability and feasibility issues • The GRADE approach is not designed to be applied to qualitative research, and GRADE-CERQual was developed as an approach to assess the confidence in findings from qualitative evidence syntheses • The use of GRADE-CERQual addressed the need to systematically incorporate a wider range of evidence into the guidance process. • An early version of the evidence-to-recommendations table named DECIDE Evidence to Decision framework was implemented to support guidance panel members move from evidence to health system recommendations • The use of the DECIDE framework helped ensure that panel members took account of key considerations when making recommendations, and improved the transparency of the decision-making process, including enabling access to end users of the guidance of the record of recommendation decisions and how these were made
--	---	---	---

<p>Alexander PE, Li S-A, Gionfriddo MR, Stoltzfus RJ, Neumann I, Brito JP, et al. Senior GRADE methodologists encounter challenges as part of WHO guideline development panels: an inductive content analysis. <i>J Clin Epidemiol.</i> 2016 Feb;70:123–8.</p>	<ul style="list-style-type: none"> To understand the experience of GRADE methodologists who have worked on WHO guidelines that made discordant recommendations (strong recommendations despite low or very low confidence in estimates of effect) 	<ul style="list-style-type: none"> Interviews with five senior GRADE methodologists Inductive, content analysis of the qualitative data 	<ul style="list-style-type: none"> Methodologists experienced tensions between themselves and WHO panels Methodologists felt that major drivers for discordant recommendations were limited understanding of GRADE, conflicts of interest among panel members, and political pressure Conflicts of interest included financial and nonfinancial conflicts of interest among panel members Methodologists felt the need for better understanding and increased support of the role as co-chairs of WHO panels
<p>Alexander PE, Brito JP, Neumann I, Gionfriddo MR, Bero L, Djulbegovic B, et al. World Health Organization strong recommendations based on low-quality evidence (study quality) are frequent and often inconsistent with GRADE guidance. <i>J Clin Epidemiol.</i> 2016 Apr;72:98–106.</p>	<ul style="list-style-type: none"> A previous study identified that 55.5% of 289 strong recommendations were discordant recommendations This study classified WHO recommendations according to possible explanations for issuing discordant recommendations 	<ul style="list-style-type: none"> All WHO guidelines approved by the WHO GRC between 2007 and 2012 that applied GRADE methods were included Thirty-three guidelines containing 160 discordant recommendations were identified, and the investigators evaluated the extent to which each guideline documented the reasons for the discordant recommendations The rationale was considered to be transparent when the guideline development group provided a rationale (in some instances a few lines; in others, a more extensive description) Investigators classified each of the 160 recommendations as either consistent with one of the five previously 	<ul style="list-style-type: none"> 82.5% of the included guidelines provided some form of rationale for making discordant recommendations 25 (15.6%) of the 160 discordant recommendation were judged to be consistent with one of the five paradigmatic situations in which it is appropriate to offer discordant recommendations 33 (20.6%) of the discordant recommendations were judged to represent a misclassification of confidence, 29 recommendations (18.1%) were judged to be good practice statements; and 73 (45.6%) as warranting conditional recommendations Overall, the study suggested that there are major limitations in the extent to which WHO sponsored guideline development groups adhere to GRADE guidance when issuing discordant recommendations

		<p>identified optimal categories for discordant recommendations, or three categories where the discordant recommendations were judged to be inconsistent with GRADE guidance: 1) misclassification of evidence; 2) good practice statements; 3) more likely conditional recommendations.</p>	
<p>Guyatt GH, Schünemann HJ, Djulbegovic B, Akl EA. Guideline panels should not GRADE good practice statements. J Clin Epidemiol 2014;68:597e600.</p>	<ul style="list-style-type: none"> To increase awareness about good practice statements, and clarify how guideline panels should identify such statements 		<ul style="list-style-type: none"> A common misunderstanding is that evidence should be classified as low or very low quality in the absence of randomized trial or other formal studies addressing the question of interest Guideline panels using GRADE have commonly issued strong recommendations with the evidence classified as warranting low or very low certainty Good practice statement typically represent situations in which a large body of indirect evidence, made up of linked evidence including several indirect comparisons, strongly support the net benefit of the recommended action It might be reasonable to consider making good practice statements when it would be an unproductive and wasteful exercise to collect the indirect linked evidence supporting the recommendation There will be instances where guideline panels have high level of certainty in estimates, and that high level of certainty is based on a large body of linked evidence. It may be appropriate to forego the formal GRADE process and make a good practice statement in cases where formally collecting and summarizing the evidence will be a poor use of time and energy

			<ul style="list-style-type: none"> • Good practice statements may be abused if guideline panels issue strong recommendations that are unwarranted, and guideline panels should use them sparingly • Authors propose key questions to ask when considering a good practice statement
<p>Public Health. 2014 May;128(5):444-74. doi: 10.1016/j.puhe.2014.01.002. Epub 2014 May 22. Appraisal of guidelines developed by the World Health Organization. Burda BU(1), Chambers AR(2), Johnson JC(3).</p>	<ul style="list-style-type: none"> • To appraise the quality of all WHO guidelines approved by the GRC 	<ul style="list-style-type: none"> • All guidelines approved by the Guidelines Review Committee between 2007 till May 2013 were assessed for their quality with the AGREE II instrument • In total, 124 guidelines were reviewed 	<ul style="list-style-type: none"> • After scoring with AGREE II, 58 of the 124 guidelines were recommended for use, while 58 were recommended with modifications and eight were not recommended • Scope and purpose, and clarity of presentation were AGREE domains with the highest scores • It was unclear whether members of the target population and affected communities participated in any of the guideline development processes • The score for rigor of development was variable (range 5-92) and frequently low, and only 77% of the guidelines conducted or commissioned a review of the evidence, and it was often not possible from the reporting to determine whether the reviews were systematic • In guideline documents where decision tables and detailed evidence profiles were absent, the link between the evidence and recommendations was not explicit, and the assessment of the overall balance of health benefits and risks were unclear • Only 49% used GRADE to assess the quality of the evidence • Stakeholder engagement, editorial independence, and applicability were identified to be domains in need for improvement
<p>Sinclair D, Isba R, Kredon T, Zani B, Smith H, Garner P. World Health Organization guideline development: an</p>	<ul style="list-style-type: none"> • To evaluate WHO guideline development against international standards pre and post formation of the GRC 	<p><u>Guideline appraisal</u></p> <ul style="list-style-type: none"> • WHO guidelines prior to and after the guideline development reforms were sampled 	<ul style="list-style-type: none"> • Mean scores for all six domains of the AGREE II appraisal were higher for the ten guidelines approved by the GRC compared with ten matched guidelines published between 2003-2008

<p>evaluation. PLoS ONE. 2013;8(5):e63715.</p>	<ul style="list-style-type: none"> • Explore senior staff's perceptions of the GRC and the new procedures 	<ul style="list-style-type: none"> • AGREE II appraisal tool was used to assess the methods and presentation of each guideline using 23 criteria across six domains: scope and purpose, stakeholder involvement, rigour of development, clarity of presentation, applicability, and editorial independence <p><u>Interviews</u></p> <ul style="list-style-type: none"> • Interviews with 20 senior staff across 16 WHO departments • Qualitative data from the interviews were analyzed using framework analysis where the main themes were identified by coding the data, and assembling the main themes 	<ul style="list-style-type: none"> • Interviewees expressed that the high standards set by the guideline development reforms were essential for the credibility of WHO's recommendations • Mixed views were identified about whether WHO needed a single quality assurance mechanism, and uncertainties and lack of capacity among staff in applying the GRADE approach were described • Overall, the study concluded that the quality assurance standards set by the GRC were yet to be fully embedded within the organization
<p>BMC Public Health. 2013 Jan 8;13:9. doi: 10.1186/1471-2458-13-9. Current experience with applying the GRADE approach to public health interventions: an empirical study. Rehfuss EA(1), Akl EA.</p>	<ul style="list-style-type: none"> • The objectives of the study were to review current use of and experience with the GRADE approach in rating the quality of evidence in the field of public health, and to identify challenges encountered 	<ul style="list-style-type: none"> • Semi-structured interviews with two groups were conducted: individuals that have applied the GRADE approach in the context of systematic reviews or guidelines in the field of public health (n=15) and selected individuals representing public health organizations that actively decided against the use of the GRADE approach (n=3) • The experience with the use of GRADE was categorized according to five different categories: <ul style="list-style-type: none"> (i) applied without significant challenges; 	<ul style="list-style-type: none"> • Overall, interviewees with experience with the GRADE approach emphasised that core GRADE principles—the systematic and transparent process of assessing the quality of evidence—apply to public health interventions, and that the GRADE approach has improved the quality of guideline development • Specific challenges were related to: <ol style="list-style-type: none"> (1) decisions about which studies to include or exclude, how to interpret heterogeneity, and how to make careful judgement about the degree of indirectness when assessing the evidence for public health interventions; (2) choice of outcomes and outcome measures; (3) ability of the GRADE approach to distinguish between public health interventions that are reasonably well-supported by evidence and those that are less supported by evidence, and a particular concern over the current approach not being able to appropriately discriminate

		<ul style="list-style-type: none"> (ii) applied with minor challenges; (iii) applied with major challenges; (iv) currently being considered/tested; and (v) not currently applied 	<p>between stronger and weaker observational study designs, nor offering any possibility to upgrading the quality of the evidence from observational studies;</p> <ul style="list-style-type: none"> (4) that the GRADE approach does not allow for the integration of non-epidemiological evidence, such as laboratory, mechanistic or animal studies, principles from other disciplines, and evidence on implementation and context; (5) concerns about policymakers misinterpreting terms such as “low quality evidence” and/or “weak recommendation” to justify inaction, and that that GRADE does not appropriately distinguish between different kinds of observational studies; (6) applying GRADE criteria were challenging for systematic reviews where the results are very heterogeneous, and that the process for going from evidence to recommendations was felt ad hoc in terms of taking into consideration values and preferences, cost effectiveness and feasibility issues <ul style="list-style-type: none"> • Interviewees made suggestions for fine-tuning how GRADE rate the quality of the evidence, which fell into two categories: (1) pragmatic guidance on the application of the GRADE approach for complex interventions and; (2) modifications to the existing GRADE criteria
<p>Akl EA, Kennedy C, Konda K, Caceres CF, Horvath T, Ayala G, Doupe A, Gerbase A, Wiysonge CS, Segura ER, Schünemann HJ, Lo YR. Using GRADE methodology for the development of public health guidelines for the</p>	<ul style="list-style-type: none"> • The experience with developing WHO Department of HIV/AIDS’s guidelines for delivery of essential package of interventions for the prevention and treatment of HIV and other sexually transmitted infections among men who have sex with men and transgender people was used as a case study for 	<ul style="list-style-type: none"> • Through discussions of methodological challenges, the core working group of the guideline development process identified nine challenges 	<ul style="list-style-type: none"> • Nine methodological challenges were identified during the guideline development. Four of the challenges were related to the grading of the quality of the evidence: <ul style="list-style-type: none"> (1) Heterogenous and complex interventions, including whether interventions are similar enough to consider together in a meta-analysis; (2) Paucity of trial data, including the need to consider the broader evidence base for public health interventions;

<p>prevention and treatment of HIV and other STIs among men who have sex with men and transgender people. BMC Public Health. 2012 May 28;12:386. doi: 10.1186/1471-2458-12-386.</p>	<p>providing lessons learned and insights into the use of GRADE when developing public health guidelines</p>		<ul style="list-style-type: none"> (3) Selecting outcomes of interest, including the development of an outcome framework to help consider whether the indirectness of outcomes warrants downgrading the quality of evidence; (4) Using indirect evidence, including considering whether the indirectness of population and the setting warrants downgrading the quality of evidence <ul style="list-style-type: none"> • The other five challenges were related to grading the strength of the recommendation (5) Integrating values and preferences of target populations affected by the implementation of the recommendation; (6) Considering resource use when determining the direction and strength of recommendations in the GRADE framework; (7) Addressing social and legal barriers, including considerations to be made when framing recommendations as good practice recommendations; (8) Wording of recommendations, including the use of the word “conditional” to characterize the strength of the recommendation; (9) Developing global guidelines, including being sensitive to the problem addressed by the guideline, and the implications of an interventions varying substantially across settings
<p>Glob Public Health. 2010;5(4):395-412. doi: 10.1080/17441690903473253.</p> <p>Developing WHO guidelines with pragmatic, structured, evidence-based processes: A case study.</p>	<ul style="list-style-type: none"> • Concrete examples that describe challenges and successes in developing guidelines may provide important insights into how guideline development processes in WHO and other public health organizations can be improved • To highlight several issues related to the current debate about guideline development, 		<ul style="list-style-type: none"> • The authors describe that three principles guided their approach to guideline development: (1) transparency – each part of the guideline process should be documented, open and accessible to all participants; (2) fairness – all participants were able to produce and comment on guideline statements and critique draft recommendations; (3) leadership – transparency and fairness was balanced against leadership needed to move the process forward

<p>Chang LW(1), Kennedy CE, Kennedy GE, Lindegren ML, Marston BJ, Kaplan JE, Sweat MD, Bunnell RE, O'Reilly K, Rutherford GW, Mermin JH.</p>	<p>the authors conducted a narrative case study of an effort by WHO and its partners to use pragmatic, structured, evidence-based processes to develop guidelines on prevention and care interventions for adults and adolescents living with HIV in RLS</p>		<ul style="list-style-type: none"> • The guideline process was divided into three phases—pre-conference, conference, and post-conference—and the authors describe the steps underlying each of these phases, including successes and challenges associated with each of these • It was described that the GRADE approach was not used since it was felt that the approach would be difficult to implement due to the large number of diverse interventions the guidelines aimed to cover, the lack of expertise and familiarity with the GRADE approach, and the lack of resources needed to produce GRADE-level evidence summaries, and the significant emphasis on evidence from non-randomized study designs
<p>Paul E. Alexandera,* Lisa Berob, Victor M. Montoric,d,e,f, Juan Pablo Britog, Rebecca Stoltzfush,i, Benjamin Djulbegovicj, Ignacio Neumanna,k, Supriya Ravel, Gordon Guyattm,* World Health Organization recommendations are often strong based on low confidence in effect estimates. Journal of Clinical Epidemiology 67 (2014) 629e634.</p>	<ul style="list-style-type: none"> • To examine all WHO guidelines developed with the GRADE approach and describe the classifications of strong and weak recommendations and their associated confidence in effect estimates (high, moderate, low, and very low) 	<ul style="list-style-type: none"> • All available WHO public health guidelines issued between January 2007-December 2012 were retrieved, and 43 guidelines which had used the GRADE approach were included • The recommendations were classified according to quality of evidence (high, moderate, low, and very low), strength of recommendation (strong or weak/conditional), and whether the recommendation was for or against an action 	<ul style="list-style-type: none"> • Of 456 recommendations identified in the 43 included guidelines, 289 (63.4%) were strong recommendations and 167 (36.6%) were weak • 160 (55.5%) of the 289 strong recommendations were identified to be based on evidence warranting low or very low confidence in estimates • The study raised questions about whether GRADE is being applied appropriately and the extent to which WHO guideline panelists neglect uncertainties in the evidence when considering the strength of recommendations
<p>Alexander PE, Gionfriddo MR, Li SA, Stoltzfus RJ, Neumann I. A number of factors explain why WHO</p>	<ul style="list-style-type: none"> • To elucidate explanations for discordant recommendations, in order to inform ongoing improvement of the GRADE process and efforts to improve 	<ul style="list-style-type: none"> • Interviews conducted with 13 individuals involved with development of guidelines (11 content area expert panel chairs and 2 WHO technical officers) 	<ul style="list-style-type: none"> • The use of the GRADE approach was viewed as a positive step to improve evidence-informed guideline development, but unclear or insufficient GRADE guidance was viewed as a challenge

<p>guideline developers make strong recommendations inconsistent with GRADE guidance. J Clin Epidemiol 2015. http://dx.doi.org/10.1016/j.jclinepi.2015.09.006.</p>	<p>and facilitate the uptake and implementation of GRADE</p>		<ul style="list-style-type: none"> • Developing appropriate PICO questions, and considering contextual factors was experienced as challenging • Explaining what conditional recommendation means would require time and training, and labelling recommendations as strong was viewed to be more straightforward • Power dynamics within the guideline development group where those more experienced dominated could affect considerations • Clinical and practical experience sometimes took precedence over evidence during discussions about recommendations • Explanations for discordant recommendations included skepticism about the value of making conditional recommendations; that political considerations sometimes could push guideline panelists to make strong recommendations; high certainty in benefits despite assessing evidence as low certainty, for example because evidence of higher quality is unlikely to be available; that strong recommendation is warranted if no or minimal harm exists; reluctance in delivering conditional recommendations for established practices; and concerns that conditional recommendations will be ignored
<p>Hoffman SJ, Lavis JN, Bennett S. The Use of Research Evidence in Two International Organizations' Recommendations about Health Systems. Healthc Policy. 2009 Aug;5(1):66-86.</p>	<ul style="list-style-type: none"> • To systematically compare health systems recommendations by WHO and the World Bank with the research evidence available at the time of their formulation 	<ul style="list-style-type: none"> • A series of inventories were developed that facilitated the purposive sampling of two international organizations (WHO and the World Bank), five health topics, ten relevant publications (two per topic) and 30 recommendations (three per publication) based on explicit selection criteria 	<ul style="list-style-type: none"> • A total of 14 WHO and seven World Bank recommendations from the eight publications were compared to the research evidence from systematic reviews that were available at the time of their formulation • Two of the eight publications examined were found to cite systematic reviews • Five of the 14 WHO and two of the seven World Bank recommendations were consistent

		<ul style="list-style-type: none"> • To compare the chosen recommendations with the nature and direction of effect claims made in systematic reviews, the research evidence with which to compare was compiled for each topic using existing overviews of systematic reviews related to the topics and recommendations issued by the organizations, as well as updates of each of these searches and new searches for tuberculosis and tobacco control on MEDLINE, CINAHL and EMBASE • Where research evidence from systematic reviews existed at the time that recommendations were written and it was not utilized, an explanation for this discrepancy was sought within the publication 	<p>with both the direction and nature of effect claims from systematic reviews</p> <ul style="list-style-type: none"> • A total of 10 WHO and five World Bank recommendations were consistent with the direction of effect claims • No explanation was found within any of the WHO or World Bank publications for the discrepancies between the recommendations and the existing research evidence from systematic reviews • Overall, systematic reviews were found not to be consistently used (or at least reported as having been used and then weighed explicitly against competing social, political, economic or ethical considerations) in the development of recommendations from WHO and the World Bank • Overall, there appeared to be no clear rationale for the consistency between recommendations and research evidence that occurs with some health topics but not others
--	--	---	--