WHO

Handbook for Guideline Development

2nd edition

18. Incorporating a complexity perspective into WHO guidelines

World Health Organization

Contents

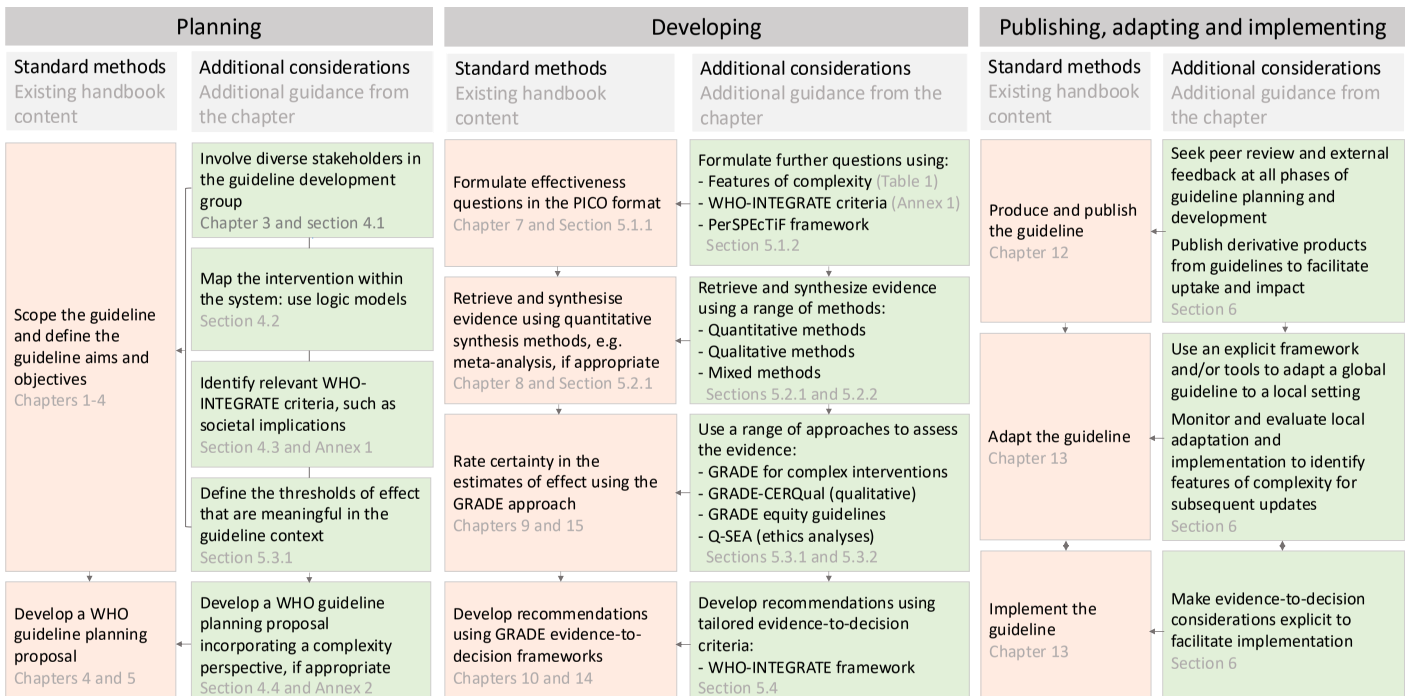
18. Incorporating a complexity perspective into WHO guidelines

18.1 Introduction

As outlined in Chapter 1 of the *WHO handbook for guideline development* (1), WHO guidelines aim to provide recommendations for decision-makers on what to do or how to choose among a range of interventions and policies to tackle specific health problems and achieve the best outcomes possible. To develop guidelines, WHO follows a systematic and transparent process derived from that used for clinical guidelines. However, this process does not fully attend to the complex nature of public health and health system decisions and the measures needed to assess these. Health is increasingly being viewed as an outcome of multiple interlinked factors, including biological, socio-economic and environmental determinants. Interventions to address health often need to include multiple components to tackle these different factors (commonly referred to as "complex multi-component interventions") and are implemented in "complex systems" with specific contextual features (e.g. epidemiological, socio-cultural, socio-economic, geographical, ethical, political and legal). To inform effective decision-making, it is therefore important to know *when, why, how* and *in what circumstances* interventions work; otherwise, decision-makers will have only limited confidence in whether the effects would be the same in their own context.

This chapter aims to demonstrate the value of considering a complexity perspective in WHO guidelines. It is based on the series published in *BMJ Global Health* entitled *Complex health interventions in complex systems: improving the process and methods for evidence-informed health decisions* (2-10). Figure 1 provides an overview of the entire guideline development process from scoping the guideline to implementing recommendations at regional, national or sub-national levels and describes how complexity can be factored into specific steps. This chapter describes when and how to address complexity when developing WHO guidelines.

Figure 1. Incorporating a complexity perspective into the WHO guideline development process



CERQUAL, Confidence in the Evidence from Reviews of Qualitative research; GRADE, Grading of Recommendations Assessment, Development, and Evaluation; PerSPEcTiF, Perspective, Setting, Phenomenon of interest, Environment, Comparison, Time, and Findings; PICO, Population, Intervention, Comparison, Outcomes; Q-SEA, Quality Standards for Ethics Analyses.

18.2 What is a complexity perspective?

The phrase *complex intervention* is often used to describe health service and public health interventions, including behavioral, educational, psychological, occupational and organizational interventions (11). These interventions may: 1) have many interacting components in their design; 2) include complex behaviors in the delivery and receipt of the intervention; 3) target different groups and levels; 4) involve many health and non-health outcomes; and/or 5) have flexible (i.e. non-standardized) implementation across different contexts (11). Examples of such complex interventions include health promotion interventions (e.g. sexual health education) and health system and organizational interventions (e.g. chronic disease management) (3).

Complex systems, on the other hand, refer to dynamic networks of interactions (e.g. among people, groups, communities, schools or occupational settings) in which interventions take place (12, 13). While the intervention itself may be simple or complex in design (i.e. mono-component or multi-component), when delivered through a system, it may bring about wider changes than just those directly related to the health problem. An example of this is smoke-free legislation (a simple intervention in design) which resulted not only in changes in smoking-related health outcomes, but also in the patterns of socializing and drinking in the community (i.e. wider system changes) (3, 14). The definition of complex interventions is often contrasted with that of *complex systems thinking* because of because of differences in emphasis: the first highlights the complexity of the intervention design while the second highlights the complexity of the functioning of systems, including changes in system dynamics and networks (15).

For the purposes of this chapter, we will use the term *complexity* to highlight a perspective that would allow a more nuanced consideration of the aspects of interventions and/or the properties of the wider systems in which interventions take place (see Table 1).

Table 1. Defining the complexity perspective through features of the intervention and/or the system (adapted from Lewin et al. 2017 (16), Petticrew et al. 2019 (3), Rehfuess et al. 2019 (4))

Features of the intervention	Description
Many interacting components in the intervention	Interventions may include multiple components which may have synergistic or dysynergistic interactions. Multi-component interventions can target individuals as well as entire populations or sub-groups.
Many organizational levels targeted by the intervention	Interventions may target multiple levels. This is more common in population and system-level interventions which may target individuals in households located in communities which are further influenced by national-level interventions.
Focus on behavior change	Interventions may require behavior change among recipients. These interventions can target individuals as well as entire populations or sub-groups.
High level of skill required by persons delivering the intervention	The skills required by persons delivering the intervention may be high. These interventions can target individuals, entire populations or sub-groups.
High level of skill required by persons receiving the intervention	The skills required by persons receiving the intervention may involve specific abilities or broader resources and capacities. These interventions can target individuals as well as entire populations or sub-groups.
Interaction of interventions with context	Interventions may be context-dependent, i.e. their effectiveness relies on tailoring of their design and delivery strategies to specific contexts. These interventions more frequently target entire populations and communities.
Multiple (health and non-health) outcomes and complex causal pathways	Interventions such as those involving multiple components often impact a large number of health and non-health outcomes and involve complex causal pathways. These interventions can target individuals as well as entire populations or sub-groups.
Features of the system	**Description**
Adaptivity (how the system responds)	Interventions may influence the context of implementation (directly or indirectly). This is more common for population- and system-level interventions such as public health policies. The entire system adapts and responds in expected or unexpected ways.
Emergent properties	Intervention effects may emerge from self-organization among the interacting agents. Emergence is more commonly observed in population-level interventions (or when individual-level interventions are implemented or assessed at the population-level).
Non-linearity and phase changes	Interventions may demonstrate effects once they have reached a certain scale. Non-linearity or phase changes are more specific for population-level interventions.
Feedback loops	Interventions comprised of different components can produce feedback loops reducing the overall effect (negative), or conversely, enhancing the effect beyond what might be expected (positive). These interventions can target individuals as well as entire populations or sub-groups.

18.2.1 Why is it important to consider complexity in WHO guidelines?

The WHO guideline development process draws on the methods and procedures predominantly developed for and used in clinical guideline development such as the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach and the GRADE evidence-to-decision (EtD) frameworks. While these frameworks are largely consistent with WHO norms and values, they do not fully address all factors and considerations relevant for public health and health system decision-making. In the standard guideline development process, questions regarding the comparative effects of interventions and the systematic reviews of effectiveness are the main considerations when formulating recommendations (17). Recommendations on public health and health system interventions, however, are shaped by a range of considerations which go beyond questions on intervention effectiveness and safety, including for example, the role of social and economic health determinants and the implications of interventions for society as a whole (2, 4).

Public health and health system interventions are often described as context-dependent, that is to say, they often interact and adapt to the context within which they are implemented (18). Context reflects a set of characteristics and circumstances that consists of active and unique factors within which the implementation of the intervention is embedded (19). Decision-makers at local, national or international levels therefore need to have a comprehensive understanding of the intervention and its broad impact, including why, how, when and under what circumstances the intervention works in order to decide whether it is worth implementing at all and whether they might need to make modifications to better fit the local context (see section 6) (7).

To accommodate the needs of decision-makers in public health, it is important that the standard guideline development methods focusing on comparative effectiveness and safety of interventions are enhanced and expanded to include questions on the broader impact of interventions within the system in which they are implemented. Taking a complexity perspective allows for rebalancing the emphasis in the guidelines by highlighting the important factors within interventions and the broader context and allowing for their explicit consideration and assessment.

18.3 When should a complexity perspective be considered in WHO guidelines?

Not all guidelines may need to incorporate a complexity perspective. The decision to adopt a complexity perspective and the degree to which this perspective needs to be incorporated in a guideline, should be taken in line with the specific topic area, needs and aims as well as in consultation with the stakeholders including end-users of the guideline and representatives of those affected by the guideline recommendations (see section 4.1).

To help determine whether a complexity perspective is useful in a guideline, it is first necessary to consider the priority question for the guideline. If users only want to know about the effects of an intervention on specific individual- or population-level outcomes, then a complexity perspective may not be necessary. In this case, the standard question for the guideline is: *does intervention x effectively reduce outcome y?* On the other hand, if interventions have wider societal implications and interact with a specific context, then a complexity perspective would be important to avoid simplistic and misleading conclusions. In this case, the priority question for the guideline is: *what happens in the system when intervention x is introduced?* For example, childhood obesity is a complex public health problem with multiple determinants, such as biological (e.g. intergenerational passage of obesity risk), socio-cultural (e.g. socio-cultural context not encouraging physical activities) and environmental (e.g. lack of access to healthy diets, food and drink options). A standard approach tackling soft drink consumption as a well-established determinant of childhood obesity would consider only a simple linear model of cause (e.g. interventions to reduce soft drink consumption) and effect (e.g. excess weight gain). On the other hand, a complexity perspective would encourage asking a broader question: *what else might be happening in the system that needs consideration in the guideline?* A range of other important factors and their potential interactions would thus be identified, including safety, access to water and the role of industry. The conclusions of a guideline taking a complexity perspective on childhood obesity might thus be different from those reached by taking a standard perspective which examines only one link between soft drink consumption and weight gain.

Further questions to help determine whether a complexity perspective is useful for a guideline include:

- Does the intervention affect the context into which it is introduced?
- Does the intervention involve system-level changes (i.e. changes to wider structures or processes which affect health, such as through regulation, healthcare reorganization and introduction of new policies)?

- Does the intervention bring about changes through system-level mechanisms (e.g. in order to influence substance use outcomes among students the entire school ethos might need to be changed)?

Positive answers to these questions signify the added value of a complexity perspective. Use of logic models and a mapping exercise during guideline planning will also help to decide whether a complexity perspective is appropriate (see section 4.2).

18.4 How should a complexity perspective be considered when planning a guideline?

The basic steps for planning WHO guidelines also apply to guidelines that take a complexity perspective, including the development of a guideline planning proposal. There are however additional aspects and steps that need to be considered when planning the guideline using a complexity perspective. These are outlined in Figure 1 and further described below.

18.4.1 Involve stakeholders in all steps of guideline planning and development

WHO guidelines involve multidisciplinary guideline development groups to finalize the guideline scope and develop recommendations and this is particularly important when taking a complexity perspective. For public health and health system interventions, stakeholders include end-users of the guideline such as providers or organizations that deliver or finance the recommended interventions as well as persons directly affected by the recommendations. An additional, important stakeholder group to consider in many policy-level interventions is industry, as it may also be directly affected by the recommendations (e.g. introduction of an industry levy to reduce the consumption of soft drinks). The views of various relevant stakeholders can be further integrated by conducting surveys or needs assessments, for example. Stakeholder input needs to occur early and throughout the guideline planning and development process. This could include, for example, targeted peer review of intermediate products such as the planning proposal, draft key questions (in PICO or other appropriate format), systematic review search strategies and the results of primary data collection on acceptability.

18.4.2 Map the intervention within the system and define the guideline scope

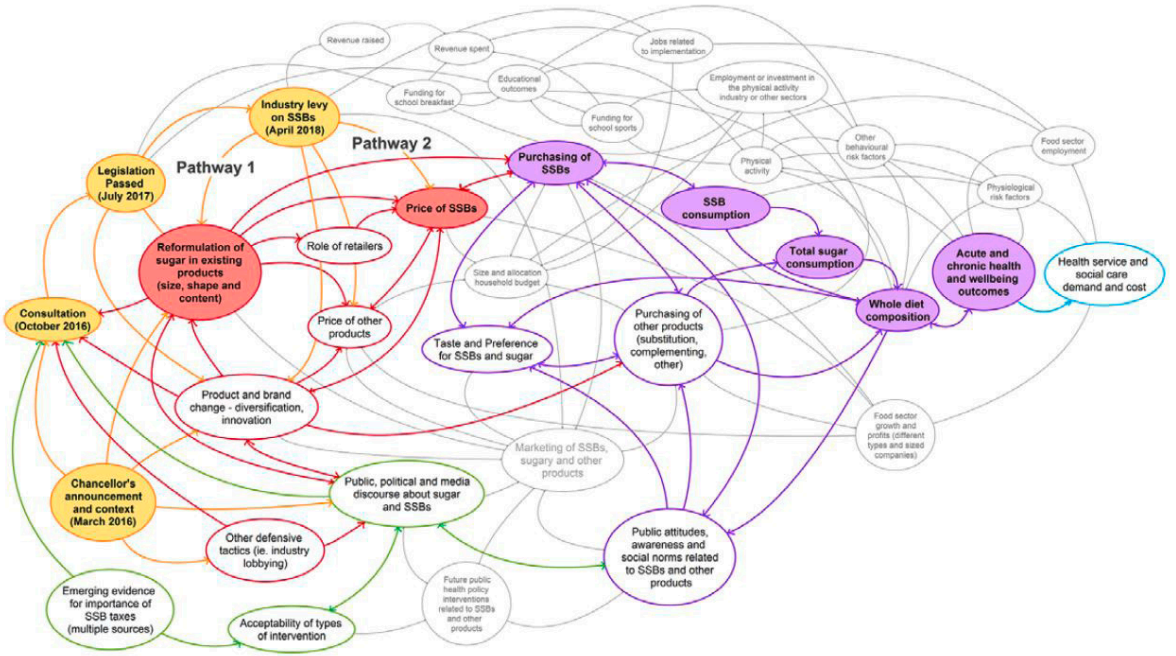
As described above, public health problems occur in broad systems, and interventions to address them often impact the system well beyond specific health outcomes. The important first step in scoping a guideline is therefore to comprehensively describe the intervention and the wider system around it. This initial mapping exercise helps guideline developers to decide whether producing the guideline will involve synthesising evidence on the effects of a specific intervention in relation to specific health outcomes or whether evidence synthesis should extend beyond these questions to other important factors (3). Consideration of the guiding questions outlined above (see section 3) along with the mapping exercise will help to inform the decision as to whether taking a complexity perspective is appropriate for the guideline, and thereby to define the guideline scope and questions.

Guideline development groups can map an intervention and the system around it by using a graphical display of interventions, different elements of the system and the relationships among them. These displays are referred to as logic models (also termed conceptual frameworks, causal loop diagrams and system-based logic models) (3, 4, 20). For example, when considering the effects of a levy on sugar-sweetened beverages (SSBs) on childhood obesity, a logic model outlines different elements of the system around the levy and helps to illuminate how the industry might adapt to the levy by reformulating the products (see Figure 2) (i.e. system adaptivity (see Table 1)). This highlights a potential question to consider in the guideline: *how might the system change when a levy is imposed on the soft drink industry?*

Logic models can be constructed through a combination of literature reviews (both empirical and theoretical literature on the impact of interventions), consultations with key stakeholders such as those affected by the intervention, and expert team discussions.

Figure. 2. Logic model on the effects of industry levy on sugar sweetened beverages

(Reproduced from (21) with permission.)



This logic model (also termed a causal loop diagram) maps the system around the industry levy on sugar-sweetened beverages (SSBs). A range of factors and their potential interactions are outlined extending beyond simple linear model of cause (industry levy on SSBs) and effect (sugar consumption). These include the potential reformulation of products, brand changes and industry lobbying. This mapping leads to the question: what else might be happing in the system that needs consideration in the guideline? The guideline scope is thereby defined and the key questions are outlined.

18.4.3 Use the WHO-INTEGRATE framework to identify important guideline questions

Recommendations on public health and health system interventions are influenced by a range of factors in addition to the balance of their benefits and harms, thus feasibility, societal implications, equity and other issues become important when making public health and health system decisions (4, 9). Most importantly, these considerations need to be identified early in the guideline planning phase, so that key questions can be defined and the relevant evidence gathered and summarized in a timely manner. This evidence will then be used later in the process to inform specific recommendations (see section 5.4).

The WHO-INTEGRATE (INTEGRATe Evidence) framework offers a tool to help identify important considerations for public health and health system guidelines in addition to intervention effectiveness. It includes six criteria to examine in a guideline which can be applied in the context of both individual and population-level interventions: balance of health benefits and harms, human rights and sociocultural acceptability, health equity, equality and non-discrimination, societal implications, financial and economic considerations, and feasibility and health system considerations (4). The seventh criterion, quality of evidence, is a meta-criterion that should be applied to the evidence gathered for each of the six substantive criteria (4). Table 2 describes more specific sub-criteria and outlines a range of methods that could be used to collect, synthesize and assess evidence for each specific criterion (see sections 5.2. and 5.3 below). Annex 1 provides further details on the criteria as well as examples of how these can be translated into specific guideline questions.

When scoping a guideline, guideline development groups should prioritize the most relevant criteria and/or sub-criteria to derive key questions and collect and synthesize evidence, rather than aiming to thoroughly address all of them. While all the criteria are important to reflect upon when scoping a guideline and making a recommendation, it would be a very time-consuming and probably unnecessary task to gather evidence towards every single criterion. Moreover, even when a specific criterion is prioritized and unpacked, not all of the sub-criteria may apply. The prioritization should be driven by the guideline topic (e.g. complexity of the health problem considered), the type of intervention (e.g. individual-level versus population-

level intervention; health sector versus inter-sectoral intervention), time and resources and the amount and type of evidence available for each criterion. Importantly, this prioritization should consider the views of key stakeholders (see section 4.1).

18.4.4 Consider using a complexity perspective to develop the guideline planning proposal

The next step after the mapping exercise and defining guideline scope is to prepare the guideline planning proposal (see Chapter 4). The essential components and the procedures for developing the planning proposal apply to all types of guidelines no matter the specific perspective taken. However, when a guideline development group decides to take a complexity perspective, there are a number of additional considerations that need to be addressed in the proposal, for example, mapping of the system around the intervention in the background and scope section. Similarly, the scope and key questions section may include additional questions on wider impacts of the intervention beyond its effects on specific outcomes (i.e. those formulated through the PICO elements). Annex 2 provides extended guidance on what should be reported in a planning proposal when a guideline takes a complexity perspective.

18. 4.5 Further information

Further information on how to scope a guideline using a complexity perspective can be found in Petticrew et al. 2019 (3), including use of logic models to map systems around interventions. Details on different types and uses of logic models can be found in Allender et al. 2015 (20), Bonell et al. 2015 (22), Mills et al. 2019 (23), Rehfuess et al. 2018 (24) and Rohwer et al. 2017 (25). Finally, for a detailed description of the WHO-INTEGRATE framework, its criteria and sub-criteria, see Rehfuess et al. 2019 (4).

Table 2. WHO-INTEGRATE framework and suggested methods for evidence synthesis and assessment of quality of evidence (adapted from Rehfuess et al. 2019 (4))

Criteria and definitions	Sub-criteria	Evidence synthesis methods	Approaches to assessing quality of evidence[1]
Balance of health benefits and harms The balance of health benefits and harms reflects the magnitude and types of health impact of an intervention on individuals or populations, taking into account how those affected value different health outcomes.	• Efficacy or effectiveness on health of individuals. • Effectiveness or impact on health of population. • Patients'/beneficiaries' values in relation to health outcomes • Safety-risk-profile of intervention. • Broader positive or negative health-related impacts.	• Systematic reviews of efficacy/effectiveness for anticipated effects (26). • Qualitative evidence syntheses (6, 27) and mixed-method reviews (5) or cross-sectional studies (28) for patients'/beneficiaries' values in relation to health outcomes. • Scoping reviews for unanticipated effects (29).	• GRADE (30). • GRADE CERQual (where applicable) (31).
Human rights and sociocultural acceptability This criterion encompasses two distinct constructs: The first refers to an intervention's compliance with universal human rights standards and other considerations laid out in international human rights law beyond the right to health (as the right to health provides the basis of other criteria and sub-criteria in this framework). The second, sociocultural acceptability, is highly time-specific and context-specific and reflects the extent to which those implementing or benefiting from an intervention as well as other relevant stakeholder groups consider it to be appropriate, based on anticipated or experienced cognitive and emotional responses to the intervention.	• Accordance with universal human rights standards. • Socio-cultural acceptability of intervention by patients/ beneficiaries and those implementing the intervention. • Socio-cultural acceptability of intervention by the public and other relevant stakeholder groups. • Impact on autonomy of concerned stakeholders. • Intrusiveness of intervention.	• Ethics syntheses (32, 33) for accordance with universal human rights standards • Qualitative evidence syntheses (6, 9, 27) and mixed-method reviews (5) for socio-cultural acceptability and impact on autonomy of concerned stakeholders and intrusiveness of interventions	• GRADE CERQual (where applicable) (31). • Q-SEA for ethics analyses (34).

continues ...

... continued

Criteria and definitions	Sub-criteria	Evidence synthesis methods	Approaches to assessing quality of evidence[1]
Health equity, equality and non-discrimination Health equity and equality reflect a concerted and sustained effort to improve health for individuals across all populations, and to reduce avoidable systematic differences in how health and its determinants are distributed. Equality is linked to the legal principle of non-discrimination, which is designed to ensure that individuals or population groups do not experience discrimination on the basis of their sex, age, ethnicity, culture or language, sexual orientation or gender identity, disability status, education, socioeconomic status, place of residence or any other characteristics.	• Impact on health equality and/or health equity. • Distribution of benefits and harms of the intervention. • Affordability of the intervention. • Accessibility of the intervention. • Severity and/or rarity of the condition. • Lack of a suitable alternative.	• Quantitative systematic reviews (35) using PROGRESS or PROGRESS PLUS (36, 37), where possible using pre-specified sub-group analyses. • Quantitative systematic reviews targeting disadvantaged groups. • Equity weights and social welfare functions in economic analyses (see Financial and economic considerations). • Qualitative evidence syntheses (6, 9, 27) and mixed-method reviews (5). • Ethics syntheses (32, 33).	• No standardized approach. • GRADE for subgroup analyses (where applicable) (30). • Relevant considerations, such as including health equity as an outcome, in Welch et al. (38).
Societal implications Societal implications recognize that health interventions do not take place in isolation and may enhance or inhibit broader social, environmental or economic goals in the short or long term. It also reflects the fact that many regulatory, environmental or other population-level health interventions are directly aimed at system-level rather than individual-level changes.	• Social impact. • Environmental impact.	• Systematic reviews of effectiveness (35). • Qualitative evidence syntheses (6, 9). • Mixed-method reviews (5). • Health technology assessments (39).	• No standardized approach. • GRADE (where applicable) (30).

continues ...

Criteria and definitions	Sub-criteria	Evidence synthesis methods	Approaches to assessing quality of evidence[1]
Financial and economic considerations Financial and economic considerations acknowledge that available financial (budgetary) resources are constrained and take into account the economic impact of an intervention on the health system, government or society as a whole.	• Financial impact. • Impact on economy. • Comparison of costs to benefits.	• Comprehensive or representative cost or budget impact data at the appropriate level (global, regional, national, sub-national). • Economic burden of disease studies undertaken at the appropriate level (global, regional, national, sub-national). • Economic analyses undertaken at the appropriate level (40) or economic analysis reviews (41).	• No standardized approach. • Relevant considerations in Drummond et al. (40) and Brunetti et al. (42).
Feasibility and health system Feasibility and health system considerations recognize that the most appropriate and feasible interventions may vary significantly across different contexts, both across countries and across jurisdictions within countries. Legislation and governance, the structure of the health system and existing programmes as well as human resources and infrastructure should be taken into account.	• Legislation. • Leadership and governance. • Interaction with and impact on health system. • Need for, usage of and impact on health workforce and human resources. • Need for, usage of and impact on infrastructure.	• Qualitative evidence syntheses (6, 9). • Mixed-method reviews (5).	• No standardized approach. • GRADE CERQual (where applicable) (31).

[1] Quality of evidence, also referred to as certainty of evidence, reflects the confidence that the available evidence is adequate to support a recommendation. In principle, quality of evidence can be applied across all six criteria in the WHO-INTEGRATE framework version 1.0.

CERQUAL, Confidence in the Evidence from Reviews of Qualitative research; GRADE, Grading of Recommendations Assessment, Development, and Evaluation; PROGRESS PLUS, Place of residence, Race, Occupation, Gender/sex, Religion, Education, Socioeconomic status, Social capital; Q-SEA, Quality Standards for Ethics Analyses.

18.5 How to consider complexity when developing WHO guidelines

As in the case of planning WHO guidelines, the basic steps for developing WHO guidelines also apply to guidelines that take a complexity perspective. However, there are additional considerations and procedures within each of the steps as shown in Figure 1; these are further described below.

18. 5.1 Formulate guideline questions

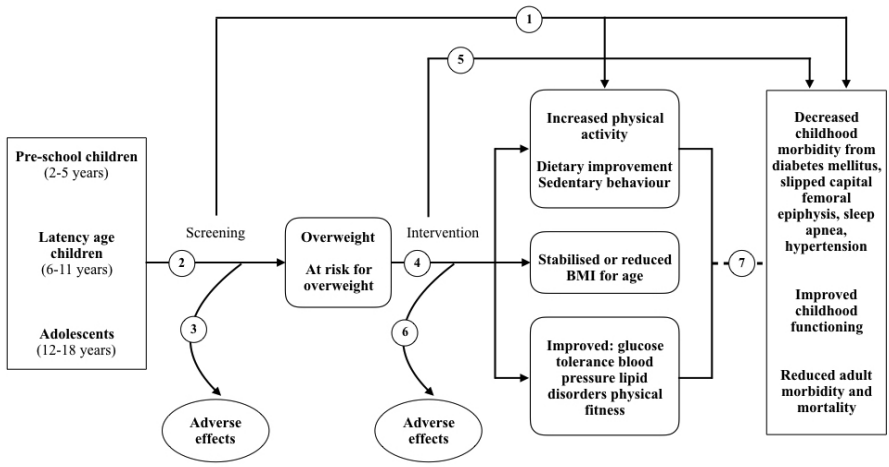
18. 5.1.1 Formulate questions on intervention effectiveness

The starting point in any guideline development process, including one which incorporates a complexity perspective is the formulation of specific questions on intervention effectiveness. However, public health interventions are often context-dependent, that is to say, their effects depend on a specific combination of contextual factors (7). Instead of or in addition to asking broad questions as to whether a public health intervention works compared with an alternative intervention, it is often important to examine the specific conditions in which the intervention has a larger or smaller effect. In this light, the broad effectiveness questions can be broken down into more specific questions exploring possible variation of the effects based on the features of the interventions or systems in which they are implemented (3). These questions can be formulated by dissecting one or more of the PICO (Population, Intervention, Comparison, Outcome) elements (8). Examples include:

- What are the effects of the intervention across different population groups or different contextual factors (dissecting the "P" element)?
- What is the independent effect of the individual components (or a combination of components) of the intervention (dissecting the "I" element)?
- What are the effects of the intervention across different implementation/delivery modes (dissecting the "I" element)?
- What are the effects of the intervention on outcomes measured at different time-points (dissecting the "O" element)?
- What are the effects of the intervention as assessed by different outcomes measures (dissecting the "O" element)?

Guideline development groups can also use analytic frameworks (also termed process-based logic models) to formulate guideline questions (24). Analytic frameworks graphically display the hypothesised processes that lead from the intervention to its outcomes, including how specific intervention components may interact with each other and broader contextual factors to produce intermediate and distal outcomes. They can help to depict intervention components, contextual factors and the relationships among them and to explicate the underlying assumptions about causal pathways (25). In this way, analytic frameworks can be useful in identifying important outcomes to examine in the guideline as well as in formulating specific questions based on relevant mediating and moderating factors (e.g. Figure 3) (43).

Figure 3. Analytic framework for screening and interventions for overweight in children (from Whitlock et al. 2005 (44))



The analytic framework elucidates the causal chain from screening and intervention for overweight to child health outcomes. In addition to assessing the direct effect of the intervention (arrow 5), it allows formulating further questions, including: are there differences in effects among population sub-groups (arrow 4)? What are the adverse effects of the intervention (arrow 6)? Are improvements in intermediate outcomes associated with improvements in health outcomes (arrow 7)?

18. 5.1.2 Formulate questions beyond intervention effectiveness

In addition to questions on intervention effectiveness, further questions will need to be formulated in line with the specific criteria and sub-criteria of the WHO-INTEGRATE framework (see Table 2 and Annex 1). Examples include:

- To what extent do stakeholders value different outcomes (benefits and harms)?

- What are stakeholders' views about acceptability, preferences, or appropriateness of the intervention (sociocultural acceptability of the intervention)?
- How will the intervention impact household health expenditures (health equity, equality and non-discrimination)?
- What is the social impact of the intervention: are there features of the intervention that increase or reduce stigma and that lead to social consequences (societal implications)?
- What is the cost of the intervention (financial impact)?
- What aspects of the health system influence implementation of the intervention (feasibility and health system considerations)?

These questions may benefit from frameworks other than PICO. One such framework is PerSPEcTiF (9) which can be used for formulating guideline questions such as those related to stakeholders' experiences with the intervention in a specific context. Table 3 outlines the elements of the PerSPEcTiF framework and provides a worked example.

Asking questions that extend beyond intervention effectiveness or harms can be resource-intensive. Not every guideline, even one which adopts a complexity perspective, needs to address all the considerations in Table 2. Guideline development groups therefore will need to prioritize these (3). For example, WHO recommendations on antenatal care for a positive pregnancy experience conducted multiple qualitative evidence syntheses to identify outcomes important to pregnant women; these outcomes then informed prioritization of the critical outcomes for the review of effectiveness (6, 45).

Table 3. Worked example of a question using the PerSPEcTiF framework
(adapted from Booth et al. 2019 (9))

Per	S	P	E	(C)	Ti	F
Perspective	Setting	Phenomenon of interest or problem	Environment	Comparison (optional)	Time / timing	Findings
From the perspective of a pregnant woman	In the setting of rural communities	How does the phenomenon of facility-based care	Within an environment of poor transport, infrastructure and geographically remote facilities	Compare with traditional birth attendants at home	In the time period up to and including childbirth	In relation to the woman's perceptions and experiences

18. 5.1.3 Further information

For further information on how to break down broad effectiveness questions into more specific questions, see Higgins et al. 2019 (8) and to formulate questions beyond intervention effectiveness, see Booth et al. 2019 (9), Flemming et al. 2019 (6), Noyes et al. 2019 (5) and Petticrew et al. 2019 (3). Further details on and examples of analytic frameworks can be found in Butler et al. 2017 (46), Kneale et al. 2015 (43), Lin et al. 2011 (47) and Whitlock et al. 2005 (44).

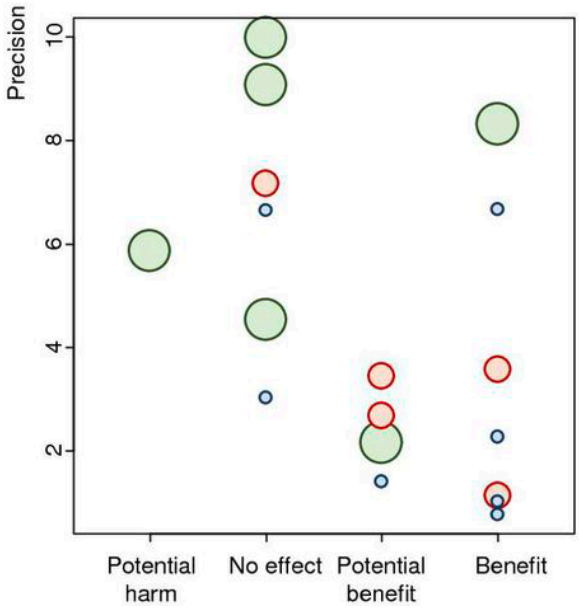
18. 5.2 Retrieve and synthesize evidence

18. 5.2.1 Synthesize evidence on intervention effectiveness

Systematic review teams can use standard evidence synthesis approaches, such as subgroup analyses and meta-regressions to explore variation of intervention effects across different conditions and population groups (8). A component-level approach and network meta-analysis can also be used to separate out the effects of individual components of the intervention or their combinations (48). It should, however, be noted that these methods can be fraught with dangers associated with having few primary studies in the review and many sources of variation. In this light, it is important that a small number of plausible sources of diversity are pre-specified in the guideline (8). When there are few primary studies, qualitative comparative analysis (QCA) is an alternative approach which involves cross-tabulation of evidence to identify configurations of interventions and various contextual factors that may explain the observed effects (49).

When studies are too diverse to combine or effect size estimates cannot be obtained from the original studies, evidence will often need to be synthesized in a narrative manner by describing findings across studies. This can be challenging, especially when the number of primary studies is large. A helpful approach is to use graphical displays, including forest, albatross, harvest or bubble plots to illustrate patterns in the retrieved studies (e.g. Figure 4). When effect size estimates are not reported, minimally reported information from each study can be used, such as the direction of effect in each study to make statistical inferences (8).

Figure 4. A bubble plot (Reproduced from Higgins et al. (8) under Creative Commons Attribution - Noncommercial IGO License (CC BY-NC 3.0 IGO))



This plot conveys three types of information: direction of each finding (horizontal scale), volume of evidence (vertical scale) and measure of credibility based on study design (bubble size): randomized trials (large, green), quasi-experimental studies (medium, red) and observational studies (small, blue). Precision is defined as inverse of the standard error of each effect estimate (derived from the confidence intervals).

18. 5.2.2 Synthesize evidence on broader questions beyond intervention effectiveness

Systematic review teams can use a range of approaches to synthesize evidence on questions beyond intervention effectiveness. This may include quantitative, qualitative and mixed methods synthesis.

From quantitative synthesis methods, a model-driven meta-analysis can be used to explore how an intervention works and which aspects of the intervention are driving the overall effect (8). It is an explanatory analysis based on causal path models (8). Model-based approaches can also be used to examine the wider system changes as a result of implementing the intervention. These provide mathematical representations of analytic frameworks and may incorporate empirical data (e.g. from existing systematic reviews), computer simulation, direct computation or a mixture of these.

Qualitative evidence synthesis (QES) methods play a critical role in answering guideline questions beyond intervention effectiveness (see Chap-

ter 15). QES is an overall term referring to all methods that involve bringing together diverse types of qualitative evidence from primary studies (6). The choice of a specific QES method should be driven by a guideline question and scope (see Table 2). For example, thematic synthesis would work well for questions relating to socio-cultural acceptability of an intervention, as it aims to develop descriptive or analytic themes (see Box 1) (50). Framework synthesis would be more suitable for questions relating to feasibility or health system considerations (51). Finally, meta-ethnography would be suitable for questions aiming to examine why and how intervention components work together as it aims to create new explanations about a phenomenon (6, 52).

Several guideline questions might require synthesis of both quantitative and qualitative evidence, that is a mixed-method synthesis. An example of such a question is: *how does intervention x impact on socioeconomic inequalities in outcome y* (5). For this question, quantitative evidence can inform whether effects are likely to be different for people from certain backgrounds. Qualitative evidence can further help to understand the reasons behind these differences. There are different ways that quantitative and qualitative evidence may be integrated in a guideline. They may be collated and analysed in a parallel or complementary way (i.e. convergent synthesis) or conducted in a sequence with one synthesis informing the other (i.e. sequential synthesis) (53). This integration can occur in a single synthesis, or two or more stand-alone reviews may be conducted first and then the findings combined in a cross-study synthesis (5). Integration of quantitative and qualitative evidence can happen at different points of the guideline process (see Box 1).

Box 1. Using qualitative and quantitative synthesis in a guideline to determine the sociocultural acceptability of an intervention (5, 54)

In a guideline on long-term rehabilitation after stroke, developers wanted to determine whether using goal-setting with patients during the planning of their rehabilitation activities leads to an improvement in psychological well-being, functioning and activity. The guideline development team conducted quantitative and qualitative evidence syntheses to answer this question. The findings from the seven studies included in the quantitative synthesis showed that goal-setting used by health professionals did not incorporate a patient-centred approach. In the meantime, the findings from the qualitative synthesis (using thematic synthesis) revealed that patients considered active participation in goal-setting as vital to their rehabilitation. This led to formulation of guideline recommendations which were driven by a patient-centred approach to stroke rehabilitation.

18.5.2.3 Further information

For further information on quantitative evidence synthesis approaches to explore heterogeneity in the effects as well as graphical displays and model-based synthesis approaches see Higgins et al. 2019 (8) and Melendez-Torres et al. 2015 (48). Examples of use of model-based synthesis can be found in Briggs et al. 2017 (55) and Brown et al. 2015 (56). Qualitative comparative analysis is comprehensively described by Thomas et al. 2014 (49). For different methods and uses of qualitative evidence synthesis see Chapter 15, Booth et al. 2018 (27), Carroll et al. 2014 (51), Flemming et al. 2019 (6), Noblit and Hare 1988 (52) and Thomas et al. 2008 (50). Finally, for further guidance on mixed-methods evidence synthesis, see Noyes et al. 2019 (5).

18. 5.3 Assessing the body of evidence

As described in section 4.3 above, the quality of evidence should be assessed for each question and each type of evidence gathered and synthesized in the guideline. Table 2 shows different approaches that can be used to assess quality of evidence based on the question and type of synthesis. For many questions, GRADE is the most appropriate approach.

18.5.3.1 Rate the quality of evidence in intervention effects (the GRADE approach)

GRADE was originally designed for rating the quality of evidence of intervention effects (see Chapter 9). The GRADE Working Group refers to this as certainty of evidence assessment. When taking a complexity perspective, there are a few additional issues that should be considered in the GRADE assessment. This will help to avoid inappropriate downgrading of the certainty of evidence.

When the GRADE approach is used in the context of decision-making (i.e. fully-contextualized use of GRADE), certainty of evidence reflects the confidence in where the true effect lies relative to specified thresholds which are meaningful for the specific decision context (57). The certainty of evidence assessment thus reflects the confidence that the effect lies above a threshold that makes implementation of the intervention worthwhile. In a guideline development context, these thresholds should be specified by considering all critical outcomes (e.g. potential harms).

In systematic reviews which are not performed with a specific decision context in mind, thresholds are set based on a pre-specified magnitude of effect (i.e. partly contextualized use of GRADE). Another approach is to use the non-null effect as a meaningful threshold in the review (i.e. non-

contextualized use of GRADE). In this latter case, the certainty of evidence rating reflects the confidence that there is an effect in the desired direction independent of the size of the effect (10, 57). It should be noted that the non-null effect can also be used in public health and health system guidelines, if the guideline development group identifies it as a meaningful threshold informing intervention implementation. This could be particularly relevant for global guidelines, for which specific implementation contexts will vary. The non-null effect can serve as a general threshold which will be further revised when adapting and implementing the guideline in a specific context (see section 6 below). Guideline development groups need to be explicit about the threshold and its justification as these also inform judgments on specific domains of GRADE, such as inconsistency and imprecision (see Table 4).

18.5.3.2 Assess the quality of the findings from other types of evidence synthesis

As shown in Table 2, guideline development groups will need different approaches to assess the quality of the evidence for different criteria (and sub-criteria) in the WHO-INTEGRATE framework (4). For this, using extensions to the GRADE approach will be appropriate. For example, for guideline questions requiring a QES approach, GRADE-CERQual (Confidence in the Evidence from Reviews of Qualitative Research) is the most appropriate method (31). Similarly, the GRADE equity guidelines can be used to assess evidence in relation to the WHO-INTEGRATE criterion of health equity, equality and non-discrimination (38). One approach is to include health equity as an outcome in the guideline (38).

Approaches may be required that extend beyond the GRADE framework. For example, Quality Standards for Ethics Analyses (Q-SEA) can be used as an instrument to assess quality of ethics analyses conducted for a guideline (see Table 2). Q-SEA consists of two main domains, the process domain and the output domain, each of which includes further elements for assessment (34).

18.5.3.3 Further information

For further guidance on how to set meaningful thresholds (including differentiation between non-contextualized, partly-contextualized and fully-contextualized ratings) in GRADE assessment see Hultcrantz et al. 2017 (57). Montgomery et al. 2019 (10) provide further details for using GRADE from a complexity perspective. Further information on GRADE-CERQual can be found in Chapter 15 and the series by Lewin et al. (31). For GRADE equity guidelines, see the series by Welch et al. 2017 (38).

Table 4. Key considerations for GRADE certainty of evidence ratings in guidelines adopting a complexity perspective (adapted from Montgomery et al. 2019 (10))

Recommendation	Rationale
Guideline scoping phase	
Define certainty of evidence in a manner that matches the needs of the intended users of the guideline	• Specify the meaningful threshold used to rate certainty of evidence considering all critical outcomes. • In the context of global guideline development, consider whether the non-null effect is a meaningful threshold.
Using Table 1, identify important features that may help explain heterogeneity and which will need separate GRADE certainty of evidence ratings	• Dissect the PICO question to examine in both the intervention and the system in which it is being used all the relevant features that guideline users will want to know about: • For intervention features, consider all aspects of its implementation, process, components, implementers, moderators, causal pathways (linear and non-linear) and important process outcomes. • For system features, consider important contextual interactions.
Rate certainty of evidence using GRADE	
Initially rate any body of evidence as high if a rigorous tool is used to assess risk of bias in NRSs (i.e. ROBINS-I). Otherwise, use the standard GRADE guidance (see Chapter 9)	• Consider using Cochrane Risk of Bias (RoB 2.0) tool for randomized controlled trials (58). • Consider using ROBINS-I for cohort-type studies (59).
Give extra scrutiny to the impact of lack of blinding providers/participants on overall risk of bias for outcomes	• If lack of blinding of either participants or providers is unlikely to affect assessment of outcome (such as when using objective outcome measures, e.g. mortality), then do not downgrade evidence for lack of blinding for that outcome.
Consider multiple criteria for assessing inconsistency of the evidence	• Assessment of heterogeneity should always start off with an appraisal of study heterogeneity, including heterogeneity in PICO elements as well as methodological aspects. • Assessment of heterogeneity should take account of multiple rather than single criteria for inconsistency (such as, I2 and its p value, overlap of CIs and degree of variation within chosen thresholds). • Assess inconsistency according to the pre-selected threshold for assessing the certainty of evidence (e.g. when effect sizes across all studies are consistently in the same direction outside of the null effect or another specified threshold, then downgrading for inconsistency is not warranted despite other measures). • Consider different analytic methods to explain heterogeneity (e.g. subgroup analysis, meta-regression, qualitative comparative analysis).

continues ...

... continued

Recommendation	Rationale
Rate certainty of evidence using GRADE	
Rate imprecision of evidence according to the threshold selected.	• If the threshold is the null effect, downgrading evidence for imprecision will not be warranted if the confidence interval does not include the null effect. If the confidence interval includes the null effect, consider whether the evidence is: 1) imprecise (due to small number of events or participants); or 2) precise and the intervention is not effective. • When rating certainty in the context of a specific threshold, consider where the effect and its confidence interval lie in relation to that threshold to determine precision.
Examine indirectness of evidence by way of assessing important differences in the evidence base beyond what is expected	• Consider grouping studies, synthesizing evidence and rating certainty in the estimates of effect for separate outcomes according to the relevant features of complexity identified at the start of the guideline. • Consider splitting the questions to answer subset conditions, downgrading only for those with less certain evidence. Do not downgrade for indirectness if observed differences are unlikely to affect the outcome.
Consider publication bias	• Conduct extensive grey literature searches and contact experts to identify unpublished data or other information. • Consider sponsorship of studies by vested industries as well as potential allegiance bias.
Upgrading evidence	• Consider upgrading certainty of evidence for a dose-response relationship related to the level of implementation (better implementation, larger effects). • Consider upgrading evidence for a body of evidence from studies with low implementation fidelity yielding positive results (this counteracts plausible residual bias or confounding).

CI, confidence Intervals; GRADE, Grading of Recommendations Assessment, Development and Evaluation; NRS, non-randomized studies; PICO, Population/Problem, Intervention, Comparison, Outcome.

18. 5.4 Develop recommendations integrating evidence across multiple criteria in the WHO-INTEGRATE framework

WHO recommendations are usually formulated using the GRADE Evidence-to-Decision (EtD) framework which suggests important factors that may affect the direction and strength of a recommendation (see Chapter 10) (60). Recommendations on public health and health system interventions, however, often require consideration of a broader range of factors or an emphasis that is not primarily on benefits and harms. The WHO-INTEGRATE framework explicitly considers the broad implications of public health and health system interventions and therefore is well-suited for guidelines that take a complexity perspective (4).

When formulating recommendations, the guideline development group should consider each criterion of the WHO-INGRATE framework in turn (see Table 2 and section 4.3), along with the collected, synthesized and assessed evidence supporting each criterion and judge how it affects the recommendation. The source of evidence used and the panel judgements should be made explicit. Evidence contributing to each criterion may be of different types (a range of quantitative and qualitative, normative statements and analyses can be used). As described above, it is not feasible to collect evidence to inform every single criterion (and sub-criterion); nevertheless, guideline development groups should still reflect on all of them when making a recommendation. Box 2 illustrates how taking a complexity perspective may involve consideration of all WHO-INTEGRATE criteria when making a guideline recommendation. Table 5 further describes how the criteria influence the guideline recommendations.

Box 2. **Illustration of how all criteria of the WHO-INTEGRATE framework may be taken into account when making a guideline recommendation** (4)

If one takes the usual perspective when examining the effects of raising taxes on sugar-sweetened beverages (SSBs) as a policy to tackle childhood obesity, the criterion of balance of health benefits and harms would warrant the most attention. However, when taking a complexity perspective, this single criterion would not be regarded as the most influential, but all criteria would be carefully examined. For example, acceptability among different groups of stakeholders will be examined as well as potential negative impacts of the intervention on health equity, equality, and non-discrimination (e.g. changes in consumption patterns across different socioeconomic groups). Positive and negative social and environmental impacts will also be examined (e.g. changes in social norms in relation to SSBs and reductions in aluminum and plastic waste), along with estimation of the financial and economic impacts of the intervention (e.g. distribution of costs and benefits of the raised taxes among different stakeholders) and its feasibility and health system considerations (e.g. implications of human resources involved with other ongoing efforts to reduce consumption of SSBs and childhood obesity).

Table 5. **WHO-INTEGRATE criteria and their implications for a recommendation** (from Rehfuess el al. 2019 (4))

Criterion	Implications for a recommendation
Balance of health benefits and harms	The greater the net health benefit associated with an intervention, the greater the likelihood of a general recommendation in favor of this intervention.
Human rights and sociocultural acceptability	All recommendations should be in accordance with universal human rights standards and principles. The greater the sociocultural acceptability of an intervention to all or most relevant stakeholders, the greater the likelihood of a general recommendation in favor of this intervention.
Health equity, equality, and non-discrimination	The greater the likelihood that the intervention increases health equity and/or equality and that it reduces discrimination against any particular group, the greater the likelihood of a general recommendation in favor of this intervention.
Societal implications	The greater the net societal benefit associated with an intervention, the greater the likelihood of a general recommendation in favor of this intervention.
Financial and economic considerations	The more advantageous the financial and economic implications of an intervention, the greater the likelihood of a general recommendation in favor of this intervention.
Feasibility and health system considerations	The greater the feasibility of an option from the perspective of all or most stakeholders, the greater the likelihood of a general recommendation in favor of the intervention. The more advantageous the implications for the health system as a whole, the greater the likelihood of a general recommendation in favor of the intervention.
Quality of evidence	The greater the quality of the evidence across different criteria in the WHO-INTEGRATE framework, the greater the likelihood of a general recommendation.

18.6 What are the implications of complexity on production, adaptation and implementation of WHO guidelines?

All WHO draft guidelines must undergo targeted, external peer review process before submission to the Guidelines Review Committee (see Chapter 12). This applies equally to guidelines which adopt a complexity perspective. However, since these guidelines are generally larger in scope and include a broader range of considerations, it is important that peer reviewers encompass additional key perspectives such as systems thinking and expertise across a range of practice fields. For example, WHO guidelines on sexual and reproductive health and rights of women living with HIV involved an external review group with a broad range of expertise in issues related to sexual and reproductive health as well as equity, gender and human rights (61). It is important that feedback from this group is sought not only at the end of the guideline development process, but most importantly also at earlier phases of guideline planning and development to help to identify any important questions that might have been overlooked by technical experts.

Implementation of a guideline in a specific national or local context must be taken into account from the beginning of guideline planning. As described above, the complexity perspective aims to provide a more nuanced understanding of *when, why, how and in what circumstances* interventions work. Taking this perspective is thus expected to produce guidelines that are better suited to the context within which they are to be implemented. This process is facilitated by considering the voices of key stakeholders affected by the intervention at various important points and decisions in guideline planning and development (see section 4.1). To facilitate guideline implementation in a specific context, it is important that guideline development groups make explicit remarks on all the contextual factors and conditions considered in guideline development, as these factors will have important implications for guideline recommendations (see Table 5).

The guideline development group and the steering group should additionally aim to document aspects of the guideline that would need further (re)consideration when implementing and evaluating it at a regional, national or subnational level (see Chapter 13). One such aspect relates to the meaningful thresholds of effect. While guideline development groups may choose one threshold as meaningful (e.g. the non-null effect), other thresholds may be deemed more relevant by national- or local-level programme managers for their specific contexts. These issues and the reasonable options should be explicitly described in the guideline to facilitate local adaptation

and implementation. This can be achieved by using systematic approaches such as the EtD and the WHO-INTEGRATE frameworks (see sections 4.3 and 5.4). Importantly, guideline adaptation and implementation should be adequately monitored, documented and evaluated. This will inform the additional features of interventions and local systems to consider in future guideline updates. While there is some ongoing work in this area (62), further research and evaluation are needed on optimal approaches for adaption of global public health and health system guidelines to local contexts.

Acknowledgements

This chapter was written by Dr Ani Movsisyan, Ludwig Maximilian University of Munich, Germany, with input from Professor Eva Rehfuess, Ludwig Maximilian University of Munich, Germany, and Dr Susan L Norris and Dr Annie Portela of the World Health Organization, Geneva Switzerland.

References

1. WHO Handbook for Guideline Development - 2nd edition. Geneva: World Health Organization; 2014. (https://apps.who.int/iris/bitstream/handle/10665/75146/9789241548441_eng.pdf;jsessionid=0B1B3B6FDC8E02D80D1486701EB1AD4F?sequence=1, accessed 20 June 2019).
2. Norris SL, Rehfuess EA, Smith H, Tuncalp O, Grimshaw J, Ford N, et al. Complex health interventions in complex systems: improving the process and methods for evidence-informed health decisions. BMJ Glob Health. 2019;4:e000963.
3. Petticrew M, Knai C, Thomas J, Rehfuess EA, Noyes J, Gerhardus A, et al. Implications of a complexity perspective for systematic reviews and guideline development in health decision-making. BMJ Glob Health. 2019;4:e000899.
4. Rehfuess EA, Stratil JM, Scheel IB, Portela A, Norris S, Baltussen R. The WHO-INTEGRATE evidence to decision framework version 1.0: Intergrating WHO norms and values and a complexity perspective. BMJ Glob Health. 2019;4:e000844.
5. Noyes J, Booth A, Moore G, Flemming K, Tuncalp O, Shakibazadeh E. Synthesising quantitative and qualitative evidence to inform guidelines on complex interventions: clarifying the purposes, designs and outlining some methods. BMJ Glob Health. 2019;4:e000893.
6. Flemming K, Booth A, Garside R, Tuncalp O, Noyes J. Qualitative evidence synthesis for complex interventions and guideline development: clarification of the purpose, designs and relevant methods. BMJ Glob Health. 2019;4:e000882.
7. Booth A, Moore G, Flemming K, Garside R, Rollings N, Tuncalp O, et al. Taking account of context in systematic reviews and guidelines considering a complexity perspective. BMJ Glob Health. 2019;4:e000840.
8. Higgins JPT, Lopez-Lopez JA, Becker BJ, Davies SR, Dawson S, Grimshaw J, et al. Synthesising quantitative evidence in systematic reviews of complex health interventions. BMJ Glob Health. 2019;4:e000858.

9. Booth A, Noyes J, Flemming K, Moore G, Tuncalp O, Shakibazadeh E. Formulating questions to explore complex interventions within qualitative eviedence synthesis. BMJ Glob Health. 2019;4:e001107.
10. Montgomery P, Movsisyan A, Grant S, Macdonald G, Rehfuess E. Considerations of complexity in rating certainty of evidence in systematic reviews: a primer on using the GRADE approach in global health. BMJ Glob Health. 2019;4:e000848.
11. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions. Medical Research Council; 2008 (https://mrc.ukri.org/documents/pdf/complex-interventions-guidance/, accessed 27 May 2019).
12. Galea S, Riddle M, Kaplan GA. Causal thinking and complex system approaches in epidemiology. Int J Epidemiol. 2010;39(1):97-106.
13. Shiell A, Hawe P, Gold L. Complex interventions or complex systems? Implications for health economic evaluation. BMJ. 2008;336(7656):1281-3.
14. Gruer L, Tursan d'Espaignet E, Haw S, Fernandez E, Mackay J. Smoke-free legislation: global reach, impact and remaining challenges. Public Health. 2012;126(3):227-9.
15. Moore G, Evans R, Hawkins J, Littlecott H, Melendez-Torres GJ, Bonell C, et al. From complex social interventions to interventions in complex social systems: Future directions and unresolved questions for intervention development and evaluation. Evaluation. 2018;25(1):23-45.
16. Lewin S, Hendry M, Chandler J, Oxman AD, Michie S, Shepperd S, et al. Assessing the complexity of interventions within systematic reviews: development, content and use of a new tool (iCAT_SR). BMC Med Res Methodol. 2017;17(1):76.
17. Djulbegovic B, Kumar A, Kaufman RM, Tobian A, Guyatt GH. Quality of evidence is a key determinant for making a strong guidelines recommendation. J Clin Epidemiol. 2015;68(7):727-32.
18. Craig P, Di Ruggiero E, Frohlich KL, Mykhalovskiy E, White M. on behalf of the Canadian Institutes of Health Research (CIHR)–National Institute for Health Research (NIHR) Context Guidance Authors Group. Taking account of context in population health intervention research: guidance for producers, users and funders of research: Southampton: NIHR Evaluation, Trials and Studies Coordinating Centre; 2018.
19. Pfadenhauer LM, Gerhardus A, Mozygemba K, Lysdahl KB, Booth A, Hofmann B, et al. Making sense of complexity in context and implementation: the Context and Implementation of Complex Interventions (CICI) framework. Implement Sci. 2017;12(1):21.
20. Allender S, Owen B, Kuhlberg J, Lowe J, Nagorcka-Smith P, Whelan J, et al. A Community Based Systems Diagram of Obesity Causes. PLoS One. 2015;10(7):e0129683.
21. White M. Evaluation of the health impacts of the UK Treasury Soft Drinks Industry Levy (SDIL). Protocol. ISRCTN: 18042742 Funded by NIHR Public Health Research Programme Study number: 16/130/01; 2017. (https://njl-admin.nihr.ac.uk/document/download/2010886, accessed 18 June 2019).
22. Bonell C, Jamal F, Melendez-Torres GJ, Cummins S. 'Dark logic': theorising the harmful consequences of public health interventions. J Epidemiol Community Health. 2015;69(1):95-8.
23. Mills T, Lawton R, Sheard L. Advancing complexity science in healthcare research: the logic of logic models. BMC Med Res Methodol. 2019;19(1):55.
24. Rehfuess EA, Booth A, Brereton L, Burns J, Gerhardus A, Mozygemba K, et al. Towards a taxonomy of logic models in systematic reviews and health technology assessments: A priori, staged, and iterative approaches. Res Synth Methods. 2018;9(1):13-24.

25. Rohwer A, Pfadenhauer L, Burns J, Brereton L, Gerhardus A, Booth A, et al. Series: Clinical Epidemiology in South Africa. Paper 3: Logic models help make sense of complexity in systematic reviews and health technology assessments. J Clin Epidemiol. 2017;83:37-47.
26. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions [Version 5.1.0]; 2011. (www.cochrane-handbook.org, accessed 18 April 2019).
27. Booth A, Noyes J, Flemming K, Gerhardus A, Wahlster P, van der Wilt GJ, et al. Structured methodology review identified seven (RETREAT) criteria for selecting qualitative evidence synthesis approaches. J Clin Epidemiol. 2018;99:41-52.
28. Siegfried N, Narasimhan M, Kennedy CE, Welbourn A, Yuvraj A. Using GRADE as a framework to guide research on the sexual and reproductive health and rights (SRHR) of women living with HIV - methodological opportunities and challenges. AIDS Care. 2017;29(9):1088-93.
29. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol. 2005;8(1):19-32.
30. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strenght of recommendations. BMJ. 2008;336:924-6.
31. Lewin S, Booth A, Glenton C, Munthe-Kaas H, Rashidian A, Wainwright M, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings: introduction to the series. Implement Sci. 2018;13(Suppl 1):2.
32. Droste S, Dintsios CM, Gerber A. Information on ethical issues in health technology assessment: how and where to find them. Int J Technol Assess Health Care. 2010;26(4):441-9.
33. Mertz M, Kahrass H, Strech D. Current state of ethics literature synthesis: a systematic review of reviews. BMC Med. 2016;14(1):152.
34. Scott AM, Hofmann B, Gutierrez-Ibarluzea I, Bakke Lysdahl K, Sandman L, Bombard Y. Q-SEA - a tool for quality assessment of ethics analyses conducted as part of health technology assessments. GMS Health Technol Assess. 2017;13:Doc02.
35. Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.
36. O'Neill J, Tabish H, Welch V, Petticrew M, Pottie K, Clarke M, et al. Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. J Clin Epidemiol. 2014;67(1):56-64.
37. Campbell and Cochrane Equity Methods Group. PROGRESS- PLUS; 2017. (https://methods.cochrane.org/equity/projects/evidence-equity/progress-plus, accessed 10 April 2019).
38. Welch VA, Akl EA, Pottie K, Ansari MT, Briel M, Christensen R, et al. GRADE equity guidelines 3: considering health equity in GRADE guideline development: rating the certainty of synthesized evidence. J Clin Epidemiol. 2017;90:76-83.
39. EUnetHTA Joint Action 2 WP. HTA Core Model ® version 3.0, 2016.
40. Drummond M, Sculpher M, Claxton Kea. Methods for the economic evaluation of health care programmes; 4th ed. Oxford: Oxford University Press; 2015.
41. Shemilt I, McDaid D, Marsh K, Henderson C, Bertranou E, Mallander J, et al. Issues in the incorporation of economic perspectives and evidence into Cochrane reviews. Syst Rev. 2013;2:83.
42. Brunetti M, Shemilt I, Pregno S, Vale L, Oxman AD, Lord J, et al. GRADE guidelines: 10. Considering resource use and rating the quality of economic evidence. J Clin Epidemiol. 2013;66(2):140-50.

43. Kneale D, Thomas J, Harris K. Developing and Optimising the Use of Logic Models in Systematic Reviews: Exploring Practice and Good Practice in the Use of Programme Theory in Reviews. PLoS One. 2015;10(11):e0142187.
44. Whitlock E, P., Williams S, B., Gold R. Screening and interventions for childhood overweight. Rockville (MD): U.S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews; 2005.
45. Downe S, Finlayson K, Tuncalp, Metin Gulmezoglu A. What matters to women: a systematic scoping review to identify the processes and outcomes of antenatal care provision that are important to healthy pregnant women. BJOG. 2016;123(4):529-39.
45. Butler M, Epstein RA, Totten A, Whitlock EP, Ansari MT, Damschroder LJ, et al. AHRQ series on complex intervention systematic reviews-paper 3: adapting frameworks to develop protocols. J Clin Epidemiol. 2017;90:19-27.
47. Lin JS, Eder M, Weinmann S, Zuber SP, Beil TL, Plaut D, et al. Behavioral Counseling to Prevent Skin Cancer: Systematic Evidence Review to Update the 2003 US Preventive Services Task Force Recommendation. U.S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews. Rockville (MD); 2011.
48. Melendez-Torres GJ, Bonell C, Thomas J. Emergent approaches to the meta-analysis of multiple heterogeneous complex interventions. BMC Med Res Methodol. 2015;15:47.
49. Thomas J, O'Mara-Eves A, Brunton G. Using qualitative comparative analysis (QCA) in systematic reviews of complex interventions: a worked example. Syst Rev. 2014;3:67.
50. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol. 2008;8:45.
51. Carroll C, Booth A, Leaviss J, Rick J. "Best fit" framework synthesis: refining the method. BMC Med Res Methodol. 2013;13:37.
52. Noblit G, W., Hare R, D. Meta-ethnography: synthesizing qualitative studies. Newbury Park: SAGE; 1988.
53. Hong QN, Pluye P, Bujold M, Wassef M. Convergent and sequential synthesis designs: implications for conducting and reporting systematic reviews of qualitative and quantitative evidence. Syst Rev. 2017;6(1):61.
54. Rosewilliam S, Roskell CA, Pandyan AD. A systematic review and synthesis of the quantitative and qualitative evidence behind patient-centred goal setting in stroke rehabilitation. Clin Rehabil. 2011;25(6):501-14.
55. Briggs ADM, Mytton OT, Kehlbacher A, Tiffin R, Elhussein A, Rayner M, et al. Health impact assessment of the UK soft drinks industry levy: a comparative risk assessment modelling study. Lancet Public Health. 2017;2(1):e15-e22.
56. Brown SA, Becker BJ, Garcia AA, Brown A, Ramirez G. Model-driven meta-analyses for informing health care: a diabetes meta-analysis as an exemplar. West J Nurs Res. 2015;37(4):517-35.
57. Hultcrantz M, Rind D, Akl EA, Treweek S, Mustafa RA, Iorio A, et al. The GRADE Working Group clarifies the construct of certainty of evidence. J Clin Epidemiol. 2017;87:4-13.
58. Vrabec JT, Backous DD, Djalilian HR, Gidley PW, Leonetti JP, Marzo SJ, et al. Facial Nerve Grading System 2.0. Otolaryngology-Head and Neck Surgery. 2009;140(4):445-50.
59. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919.

60. Alonso-Coello P, Schunemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. BMJ. 2016;353:i2016.
61. Consolidated guideline on sexual and reproductive health and rights of women living with HIV. Geneva: World Health Organization: Licence: CC BY-NC-SA 3.0 IGO; 2017.
62. The ADAPT Study. DECIPHer: Development and Evaluation of Complex Interventions for Public Health Improvement. A UKCRC Public Health Research Centre of Excellence. (http://decipher.uk.net/research-page/adaptation-of-population-health-interventions-for-implementation-and-or-re-evaluation-in-new-contexts-development-of-guidance/, accessed 22 June 2019).

Annex 1

WHO-INTEGRATE framework version 1.0: criteria, definitions and example questions

Criteria Definition	**Sub-criteria and example questions** All criteria are relevant for all interventions in health decision or guideline development processes. For sub-criteria there should be a discussion as to which are most relevant and if or how evidence should be collected to inform these.	
Balance of health benefits and harms The balance of health benefits and harms reflects the magnitude and types of health impact of an intervention on individuals or populations, taking into account how those affected value different health outcomes. Both positive and negative impacts on health must be considered, aiming to measure outcomes rather than surrogate or process markers and including those health outcomes most valued by patients/beneficiaries. Where possible, evidence of real-world effectiveness rather than efficacy under controlled circumstances should be used and a long-term perspective adopted. The greater the net health benefit associated with an intervention, the greater the likelihood of a general recommendation in favor of this intervention.	• **Efficacy or effectiveness on health of individuals**	• What is the efficacy (under controlled, often ideal circumstances) or effectiveness (in a real-life setting) of the intervention on the health of individuals, including patient-reported outcomes? Does efficacy or effectiveness vary in the short- versus longer-term?
	• **Effectiveness or impact on health of population**	• What is the effectiveness or impact of the intervention on the health of the population, including on beneficiary-reported outcomes? Can individual-level effects be aggregated at the population level, or do important system dynamics (e.g. positive or negative feedback loops) play a role? Does effectiveness or impact vary in the short- versus longer-term?
	• **Patients'/beneficiaries' values in relation to health outcomes**	• To what extent do patients/beneficiaries value different health outcomes?
	• **Safety-risk-profile of intervention**	• Which adverse events are associated with the intervention, including the risk of the intervention being misused?
	• **Broader positive or negative health-related impacts**	• Which broader positive or negative health-related impacts, such as impact on other diseases, over-diagnosis and spillover effects beyond patients/beneficiaries, are associated with the intervention? Are there features of the intervention that increase or reduce stigma associated with the disease and that lead to health consequences (see Societal implications)?

continues ...

... continued

Criteria Definition	**Sub-criteria and example questions** All criteria are relevant for all interventions in health decision or guideline development processes. For sub-criteria there should be a discussion as to which are most relevant and if or how evidence should be collected to inform these.	
Human rights and socio-cultural acceptability This criterion encompasses two distinct constructs: The first refers to an intervention's compliance with universal human rights standards and other considerations laid out in international human rights law beyond the right to health (as the right to health provides the basis of other criteria and sub-criteria in this framework). The second, socio-cultural acceptability, is highly time- and context-specific and reflects the extent to which those implementing or benefiting from an intervention as well as other relevant stakeholder groups consider it to be appropriate, based on anticipated or experienced cognitive and emotional responses to the intervention. Socio-cultural acceptability is affected by socio-cultural norms and preferences as well as power dynamics in relation to sex, age, ethnicity, culture or language, sexual orientation or gender identity, disability status, education, socio-economic status, place of residence or other characteristics. It may also be affected by the different values assigned to considerations of autonomy and intrusiveness (including privacy and dignity), freedom of information and movement as well as the distribution of benefits, harms and costs. Socio-cultural acceptability can vary greatly between stakeholder groups, and health interests should always take precedence over commercial interests. Where applicable, aspects of socio-cultural acceptability are to be assessed in comparison with usual care/the status quo or alternative interventions and should keep likely changes over time and across different population and stakeholder groups in mind. All recommendations should be in accordance with universal human rights standards and principles. The greater the socio-cultural acceptability of an intervention to all or most relevant stakeholders, the greater the likelihood of a general recommendation in favor of this intervention.	• **Accordance with universal human rights standards**	• Is the intervention in accordance with universal human rights standards and principles?
	• **Socio-cultural acceptability of intervention to patients/ beneficiaries and those implementing the intervention**	• Is the intervention socio-culturally acceptable to patients/ beneficiaries as well as to those implementing it? To which extent do patients/beneficiaries value different non-health outcomes?
	• **Socio-cultural acceptability of intervention to the public and other relevant stakeholder groups**	• Is the intervention socio-culturally acceptable to the public and other relevant stakeholder groups? Is the intervention sensitive to sex, age, ethnicity, culture or language, sexual orientation or gender identity, disability status, education, socio-economic status, place of residence or any other relevant characteristics?
	• **Impact on autonomy of concerned stakeholders**	• How does the intervention affect an individual's, population group's or organization's autonomy, i.e. their ability to make a competent, informed and voluntary decision?
	• **Intrusiveness of intervention**	• How intrusive is the intervention, ranging from low intrusiveness (e.g. providing information) to intermediate intrusiveness (e.g. guiding choices) to high intrusiveness (e.g. restricting or eliminating choices)? Where applicable, are high intrusiveness and/or impacts on the privacy and dignity of concerned stakeholders justified?
Health equity, equality, and non-discrimination Health equity and equality reflect a concerted and sustained effort to improve health for individuals across all populations, and to reduce avoidable systematic differences in how health and its determinants are distributed. Equality is linked to the legal principle of non-discrimination which is designed to ensure that individuals or population groups do not experience discrimination on the basis of their sex, age, ethnicity, culture or language, sexual orientation or gender identity, disability status, education, socio-economic status, place of residence or any other characteristic. Firmly rooted in the right to health and international human rights law and aiming to promote social justice, this criterion is concerned with the distribution of a condition, of its determinants and of the effects of interventions across different population groups. Sometimes interventions improve health for the population as a whole (see Balance of health benefits and harms) but – either in the short term, or over prolonged periods of time – negatively affect the distribution of health. Health inequalities constitute systematic differences between population groups; health inequities refer to those inequalities that are avoidable and deemed unfair. Interventions may either decrease or increase such differences through their affordability (including protection against unwanted financial and social consequences of taking up services) and accessibility (which is generally described as the physical and information-based distribution of health-relevant goods and services but may also encompass non-financial investments from recipients, e.g. time, energy) or acceptability (see Human rights and socio-cultural acceptability). In addition, this criterion recognizes that society may (or may not) attach greater value to interventions that target severe conditions, rare diseases, or conditions for which there is no suitable alternative. The greater the likelihood that the intervention increases health equity and/or equality and that it reduces discrimination against any particular group, the greater the likelihood of a general recommendation in favor of this intervention.	• **Impact on health equality and/or health equity**	• How are the condition and its determinants distributed across different population groups? Is the intervention likely to reduce or increase existing health inequalities and/or health inequities? Does the intervention prioritize and/or aid those furthest behind? How do such impacts on health inequalities and /or health inequities vary over time, e.g. are initial increases likely to balance out over time, as interventions are scaled up?
	• **Distribution of benefits and harms of intervention**	• How are the benefits and harms of the intervention distributed across the population? Who carries the burden (e.g. all), who benefits (e.g. a very small sub-group)?
	• **Affordability of intervention**	• How affordable is the intervention for individuals, households or communities? How will it impact household health expenditures, including risk of catastrophic health expenditures and health-related financial risks?
	• **Accessibility of intervention**	• How accessible - in terms of physical as well as informational access - is the intervention across different population groups?
	• **Severity and/ or rarity of the condition**	• Does the intervention address a particularly severe condition (e.g. life-threatening, end-of-life, affecting individuals with a low pre-existing health status)? Does it address a rare condition?
	• **Lack of a suitable alternative**	• Is there any suitable alternative to addressing the condition, does the intervention represent the only available option? Is this option proportionate to the need, and will it be subject to periodic review?

continues ...

... continued

Criteria Definition	**Sub-criteria and example questions** All criteria are relevant for all interventions in health decision or guideline development processes. For sub-criteria there should be a discussion as to which are most relevant and if or how evidence should be collected to inform these.	
Societal implications Societal implications recognize that health interventions do not take place in isolation and may enhance or inhibit broader social, environmental or economic goals in the short- or long-term. It also reflects the fact that many regulatory, environmental or other population-level health interventions are directly aimed at system-level rather than individual-level changes. This criterion acknowledges that the right to health embraces a wide range of socio-economic and other underlying determinants of health that may or may not lead to conditions in which people can lead a healthy life; these determinants operate across different sectors and organizational levels. The criterion also acknowledges that promoting health must go hand-in-hand with strategies to end poverty and address a broad range of social needs and to build economic growth (see Impact on economy under Financial and economic considerations), while tackling climate change and environmental protection. The greater the net societal benefit associated with an intervention, the greater the likelihood of a general recommendation in favor of this intervention.	• **Social impact**	• What is the social impact of the intervention: Are there features of the intervention that increase or reduce stigma and that lead to social consequences (see Balance of health benefits and harms)? Does the intervention enhance or limit social goals, such as education, social cohesion and the attainment of various human rights beyond health? Does it change social norms at individual or population level? Does it impact research and innovation?
	• **Environmental impact**	• What is the environmental impact of the intervention? Does it contribute to or limit the achievement of goals to protect the environment and efforts to mitigate or adapt to climate change?
Financial and economic considerations Financial and economic considerations acknowledge that available financial (budgetary) resources are constrained and take into account the economic impact of an intervention on the health system, government or society as a whole. Embedded in this criterion is the idea of progressive realization which implies being resource-conscious in moving as expeditiously and effectively as possible towards the full realization of the right to health. The criterion also captures the notion of opportunity costs, operationalized through the cost-effectiveness or cost-benefit of an intervention – these reflect the health gains that would be foregone, if resources were spent on alternative interventions. Positive and negative financial and economic implications must be considered – from a health system or broader societal perspective, depending on the intervention concerned. Where possible and appropriate, a long-term perspective should be adopted and a formal economic evaluation conducted. The more advantageous the financial and economic implications of an intervention, the greater the likelihood of a general recommendation in favor of this intervention.	• **Financial impact**	• What is the cost of the intervention? What is the overall budget impact of implementing the intervention? Do cost and budget impacts vary in the short- versus longer-term, and are they sustainable?
	• **Impact on economy**	• What is the overall economic impact of the intervention? How are different types of economic impact distributed, how does the intervention influence different sectors at different organizational levels? Does it contribute to or limit the achievement of broader development and poverty reduction goals? How does it impact the working population, for example in terms of who participates in the workforce and their level of engagement?
	• **Ratio of costs and benefits**	• What is the value-for-money of the intervention, based on an appropriate choice of method, e.g. cost-effectiveness, cost-benefit or cost-utility?

continues ...

... continued

Criteria Definition	**Sub-criteria and example questions** All criteria are relevant for all interventions in health decision or guideline development processes. For sub-criteria there should be a discussion as to which are most relevant and if or how evidence should be collected to inform these.	
Feasibility and health system considerations Feasibility and health system considerations recognize that the most appropriate and feasible interventions may vary significantly across different contexts, both across countries and across jurisdictions within countries. Legislation and governance, the structure of the health system and existing programmes as well as human resources and infrastructure should be taken into account. Barriers (e.g. lack of human resources, opposing legislation) as well as facilitators (e.g. an intervention fitting with previous spending patterns and/or existing programmes) should be recognized. When considering the fit of the intervention with the health system and its likely impacts on human resources and infrastructure at various levels, a broad societal and longer-term perspective should be adopted, where appropriate. The greater the feasibility of an option from the perspective of all or most stakeholders, the greater the likelihood of a general recommendation in favor of the intervention. The more advantageous the implications for the health system as a whole, the greater the likelihood of a general recommendation in favor of the intervention.	• **Legislation**	• Are there any legal barriers or facilitators to the implementation of the intervention?
	• **Leadership and governance**	• Might governance aspects, such as past decisions and strategic considerations, positively or negatively impact the implementation of the intervention? Are formal or informal institutions available to provide effective leadership, oversight and accountability in implementing the intervention?
	• **Interaction with and impact on health system**	• How does the intervention interact with the existing health system? Is it likely to fit well or not, is it likely to impact on it in positive or negative ways?
	• **Need for, usage of and impact on health workforce and human resources**	• How does the intervention interact with the need for and usage of the existing health workforce and broader human resources (in the health sector or other sectors), at national and sub-national levels? Is it likely to impact on these in positive or negative ways, for example by affecting the number or distribution of staff, their skills, responsiveness or productivity?
	• **Need for, usage of and impact on infrastructure**	• How does the intervention interact with the need for and usage of the existing health system infrastructure (e.g. types of health facilities, health information system, medical products and technologies) as well as other relevant infrastructure (e.g. transportation, energy), at national and sub-national levels? Is it likely to impact on these and their performance in positive or negative ways?

Meta-criterion: Quality of evidence

Quality of evidence, also referred to as certainty of evidence or strength of evidence, reflects the confidence that the available evidence is adequate to support a recommendation. In principle, quality of evidence can be applied across all criteria in the WHO-INTEGRATE framework – balance of health benefits and harms, human rights and socio-cultural acceptability, health equity, equality and non-discrimination, societal implications, financial and economic considerations and feasibility and health system considerations. As a large number of criteria are integrated in the decision-making process, evidence is interpreted in the broadest sense, and allows for relevant contributions from a variety of disciplinary approaches. Moreover, decision-making under uncertainty often involves stakeholder experience and judgement, when stronger evidence is unavailable.

In relation to effectiveness and impact, quality of evidence or certainty of evidence has variably been interpreted as confidence in (i) point estimates, (ii) the true effect lying above (or below) a certain threshold, (iii) the true effect lying within a 95% confidence interval, and (iv) the intervention being effective or not (i.e. a non-null effect). In relation to human rights and socio-cultural acceptability, quality of evidence has been described as the extent to which a finding is a reasonable representation of the phenomenon of interest. Quantity, quality (often described as internal validity) and consistency of evidence are among the most widely described underlying concepts; in addition, relevance of evidence (often referred to as external validity or generalizability) plays an important role. How quality of evidence is assessed depends on the criterion in question and the nature of a given body of evidence, e.g. GRADE is widely used for questions of effectiveness, whereas GRADE CERQual is suitable for rating qualitative evidence (Table 3).

The greater the quality of the evidence across different criteria in the WHO-INTEGRATE framework, the greater the likelihood of a general recommendation.

Annex 2

Reporting the essential components of a guideline planning proposal when taking a complexity perspective

This table provides an extension to the current planning proposal template and guidance, highlighting additional items to consider when taking a complexity perspective. Depending on the nature of the planned guideline, some items may not be relevant: in such cases please indicate "not applicable".

This planning and reporting tool is applicable to all types of guidelines: standard, consolidated, interim and rapid advice guidelines (see Chapter 1 of the *WHO handbook for guideline development*, 2nd edition). The components and content of the planning proposal will vary somewhat depending on the type of guideline, as indicated in the column entitled "Instructions for reporting". In particular, for guidelines that are consolidated, interim or rapid advice, the proposal should explain under "Rationale" why that type of guideline is being undertaken. Any deviations from standard guideline development approaches should be explained in detail in the appropriate part of the planning proposal.

It is important to emphasize that this is a *planning and reporting tool* and is not intended to indicate how to produce a high-quality guideline: such guidance is provided in the other chapters of the *Handbook*.

Essential components of a planning proposal for a guideline which incorporates a complexity perspective

Topic	Instructions for reporting	Lead* Input*	Handbook chapter Comments
A. Background and scope			
Public health problem	Provide a brief overview of the clinical or public health problem addressed by this guideline. The level of detail should be sufficient for a public health expert who is not familiar with the specific focus of the guideline, to understand the problem that the guideline addresses.	RTO None	None
Intervention(s) within the system	Describe the intervention(s) and the wider system around it using logic models. Highlight all important elements of the system and the relationships around them.	RTO SG, GDG	Chapter 18
WHO-INTEGRATE criteria and sub-criteria	Reflect upon all criteria of the WHO-INTEGRATE framework and highlight the most relevant criteria and/or sub-criteria that should be thoroughly examined in the guideline.	RTO SG, GDG	Chapter 18 Additional consultation (internal and/or external to WHO) may be needed.
General scope	Provide a brief description of the population, intervention(s), outcomes, context and setting for the proposed guideline. Indicate any major exclusions from scope. Also briefly describe why a complexity perspective was chosen for the guideline and what value it adds.	RTO SG	None

continues ...

... continued

Topic	Instructions for reporting	Lead* Input*	Handbook chapter Comments
History and related guidelines	Indicate if this is an updated guideline or one developed *de novo.* Describe the relationship of the proposed guideline to other existing or planned WHO guidelines. If the proposed guideline is a consolidated or updated guideline, indicate which recommendations or guidelines will be incorporated without review and which will be re-examined.	RTO None	None Provide the URL and/or citation to any referenced documents. There is no need to list all related WHO guidelines: list only those that are, or may be perceived to be, directly relevant.
Type of guideline	Indicate which type of guideline is proposed: standard, consolidated or rapid advice. Indicate also if this is an interim guideline. For consolidated guidelines: Why is a consolidated guideline the optimal information product for intended end-user? For rapid advice guidelines: Why is a rapid advice guideline needed rather than a standard guideline? Please justify by addressing the following questions: What is the type of emergency and the risk to public health? Is the event novel? Why does the uncertainty need to be urgently addressed? What is the anticipated time frame for the event? How will the guideline be disseminated, adopted/adapted, and implemented? For interim guidelines: Why is an interim guideline needed rather than waiting until more additional, perhaps more definitive data are available?	RTO SG	Chapter 1 The definitions of the types of guidelines are found in the *WHO Handbook for guideline development,* 2nd edition. Interim guidelines are a sub-type of standard or rapid advice guidelines where it is clear at the start of the development process that all data and research evidence are not available. Thus, an interim guideline has a short shelf-life which is specified in the information product. For example, a pivotal clinical trial may be in progress, or data are being collected in the field during an infectious disease outbreak.
B. Rationale, objectives and target audience			
Rationale	Provide a brief, cogent rationale for this guideline by addressing the following questions: • Who requested that this guideline be developed? Why? • What gaps in guidance exist and how will this guideline fill those gaps? • What evidence supports the need for this guideline? • Why does the guideline need to be developed now?	RTO SG	Chapter 2, 11 Gaps in guidance may arise from: • uncertainty reflected in suboptimal or varied practice • new interventions or approaches • new evidence on existing interventions • new regulations or policies • changes in resource availability or access to services • other sources. • The rationale should reference country priorities or global public goods.
Objectives	List two to five specific objectives for this guideline which address the following questions: • What WHO priority area(s) is (are) addressed? • What population health outcomes or indicators are expected to improve with implementation of this guideline? These objectives should reflect and link the more general goal(s) to specific outcomes. Do not include objectives that are not meaningful such as to update an existing guideline, to develop an evidence-based guideline using GRADE, or to meet Members States' needs.	RTO SG, GDG	Chapter 2 Link the objectives to the Sustainable Development Goals, Universal Health Coverage, General Programme of Work, or other WHO priorities or indicators as appropriate.
Target audience	Describe the planned end-user or target audience for this guideline. Be specific; do not list every possible end-user but rather indicate the primary and any key secondary end-users of this guideline.	RTO SG	Chapter 2 End-users implement the recommendations in a guideline, to be distinguished from the recipients of the recommended interventions. The latter do not need to be described here; they are encompassed by the description of the population(s) when formulating key questions.

continues ...

... continued

Topic	Instructions for reporting	Lead* Input*	Handbook chapter Comments
C. Contributors and funders			
WHO steering group (SG)	Provide a list of confirmed members, including their cluster/department/unit. Describe the role of this group with respect to the following as relevant: • defining the scope and key questions (e.g. PICO, PerSPEcTiF); • assessing the quality of the body of evidence; • assessing DOI of contributors external to WHO; • drafting the guideline document; • assessing and managing peer review; • reviewing the final guideline document; • managing the publication process; • disseminating the guideline; and • evaluating its impact.	RTO TU	Chapter 3 SG members can be from WHO headquarters or Regional Offices, or from other United Nations agencies; they are not individuals from outside organizations, except in the case of a guideline developed in collaboration with a third party. SG members' roles may vary across members. For example, some may be more advisory while the lead (responsible technical officer) may play a role at every stage). Be specific about who is doing what. Individuals should not be listed if they have no defined role.
Guideline development group (GDG)	Provide a list of confirmed and potential members, including their institutional affiliation, gender, relevant expertise, and WHO region of primary work or residence. If all members have not yet been confirmed, indicate when and how you will seek additional members, the expertise that is needed, and how you will achieve representation from all WHO regions. Indicate the number of individuals planned for this group. Describe the role of the members of this group specifically with respect to: • defining the scope and key) questions (e.g. PICO, PerSPEcTiF); • assessing the quality of the evidence (e.g. GRADE, CERQUal); • formulating recommendations; • reviewing and approving the final guideline document; • disseminating the guideline; and • other tasks or roles.	RTO SG	Chapter 3 GDG members are most clearly presented in a table. If global representation is not essential for this guideline because of the topic, please explain.
Guideline methodologist(s)	Indicate who will assist you in adhering to the principles and methods of evidence-based decision-making in the development of your guideline. Describe the role of the guideline methodologist(s) in the guideline development process, indicating if they will be a Technical Advisor or a member of the GDG.	RTO SG	Chapter 3 The methodologist(s) should be identified early in the guideline development process and should assist in formulating the key questions and the planning proposal as well as subsequent steps in development.
Systematic review team	Indicate who will be performing the systematic review(s) and why they were selected.	RTO (SG)	Chapter 3 Note that the systematic review team should assess the quality of the body of evidence using GRADE or other approaches as appropriate. A separate person or team should not do this assessment. (Quality assessments should also be independently reviewed by the guideline methodologist.)
External review group (ERG)	Provide a list of confirmed or potential members, including their institutional affiliation, academic degrees, gender, and WHO region of work or residence. If you have not confirmed all members, indicate who is confirmed, who is a potential member, and when and how you will seek additional members. Indicate the number of individuals planned for this group. Describe the role of the members of this group, specifically with respect to: • defining the scope and key questions (e.g. PICO, PerSPEcTiF); • reviewing the final guideline document; and • other tasks or roles.	RTO SG, GDG	Chapter 3 See comments for the GDG.

continues ...

... continued

Topic	Instructions for reporting	Lead* Input*	Handbook chapter Comments
External partners	List any external individuals or organizations that you are collaborating with and provide a rationale for their involvement. Describe the role of any external partners.	SG GDG	Chapter 3
Funder(s)	List all funders for development of this guideline. Indicate if the funder is a public or private entity For private entities, indicate if the organization is for-profit or not-for-profit. Indicate if any unfunded or in-kind services will be used or if volunteer or intern support will be used and for which specific task(s).	RTO SG	Chapter 6 Note that industry funding of WHO guidelines is prohibited and that external funders should have no role in formulating recommendations. Although representatives of entities that have contributed funding to a guideline may observe GDG meetings, they cannot participate in, or influence in any way, the deliberations and the recommendations
D. Management of contributors			
Declaration of interests	Declarations of interests for external experts: Describe who will complete the WHO DOI form and who will collect and manage them. GDG meetings: Indicate that the DOI of each GDG member will be presented and updated at the beginning of each meeting. Indicate how new interests will be assessed and managed when declared at meetings. Updating: Indicate that GDG members will be instructed to update their DOI with any potentially relevant change by notifying the RTO. Public notice and comment: Describe how you will implement WHO's requirement for public notice of potential GDG members. Indicate where you will post the notice and for how long, how you will receive, track, assess and respond to comments. Due diligence on potential GDG members: Describe how you will implement WHO's requirement for gathering information on potential GDG members, including how you will assess any information retrieved that might interfere with the potential member's ability to contribute in an objective way to guideline development.	RTO	Chapter 6 The WHO policy on COI for external experts is available at (http://intranet.who.int/homes/cre/ethics/doiexperts/. Note the following WHO policies which apply to GDG members (and not to other external contributors): ". . . technical units are required to publish the names and brief biographies of individuals considered for participation on WHO's advisory bodies together with a description of the objectives of relevant meetings. They will be made public ahead of the first meeting planned to allow time for "public notice and comment. . .. The technical unit should ensure that the public is afforded a period of not less than 2 weeks to provide information on any interests or biases relating to the individuals being considered for appointment." "As the WHO technical unit identifies or invites individuals to serve on advisory bodies or perform advisory services, it should gather information (e.g. from the internet or public media) in order to identify any obvious public controversies or interests that may lead to compromising situations for WHO and the expert concerned."
Conflicts of interest	Assessing DOI for conflicts of interests: Indicate who will determine if a disclosure is a COI and by what criteria. Define what is a (significant) COI and why, in the context of this specific guideline, including both financial and nonfinancial (intellectual and professional) interests.	RTO SG, Director TU, CRE	Chapter 6 WHO's COI policy is available at http://intranet.who.int/homes/cre/ethics/doiexperts/).
Leadership	Name the chair and co-chair(s) of the GDG, if those appointments have been made. Indicate how and why they were selected. If the chair (and co-chair) has/have not yet been selected, describe the process that will be used to select them.	SG (GDG)	Chapter 3

continues ...

... continued

Topic	Instructions for reporting	Lead* Input*	Handbook chapter Comments
Group processes/ decision-making	Describe how the GDG will make decisions, particularly at GDG meetings when the scope is defined or recommendations are formulated. If recommendations are formulated using "consensus": • define consensus; • describe how consensus be achieved; and • indicate if informal voting be used to assess agreement. What approach(es) will be used if the group cannot reach consensus? Will voting be used? If so, who will vote? How will voting be executed? Raising of hands? By anonymous ballot? What are the decision rules for voting if it occurs? If consensus cannot be reached, what approach, if any, will be used to present the minority view in the guideline?	SG (GDG)	Chapter 3 GDG need to try and reach consensus. If that is not possible then voting is permitted. Pre-defined decision rules are essential and are usually for a super majority such as 70-80% in agreement.
Observers	Indicate if observers will be permitted at GDG meetings and how they will be managed if they try to assume a role or provide input. Indicate who is invited and their institutional affiliation.	SG	Observers have no role in GDG meetings: they cannot participate in or influence in any way, the deliberations and the recommendations.
Confidentiality	Indicate that you will ask (or have asked) each member of the GDG, ERG and SR teams to sign the standard WHO confidentiality agreement.	RTO None	None The WHO confidentiality agreement (entitled "Confidentiality undertaking) is available at http://intranet.who.int/homes/cre/ethics/doiexperts/.
E. Scope and key questions			
Background questions	Describe any questions that will inform the guideline but for which a systematic review of the evidence will not be performed. Explain why these questions are important for the recommendations but are not key questions.	SG GDG	Chapter 7, 18 Background questions inform the topic and the context and do not generally require a systematic review (e.g. the disease burden or mechanisms, or intervention costs).
Analytic framework	Process-based logic models or analytic frameworks, are a very helpful tool for articulating hypotheses, examining the multiple factors affecting the ultimate health outcomes, defining scope of the guideline and the specific key questions, depicting theories of change, and identifying relevant system properties.	SG GDG (ERG)	Chapter 18 Useful references: Anderson L, Petticrew M, Rehfuess E, et al. Using logic models to capture complexity in systematic reviews. Research Synthesis Methods 2011;2(1):33–42. Rohwer A, Pfadenhauer L, Burns J, et al. Logic models help make sense of complexity in systematic reviews and health technology assessments. J Clin Epidemiol 2017;83:37-47. Rehfuess E, Booth A, Brereton L, et al. Towards a taxonomy of logic models in systematic reviews and health technology assessments: A priori, staged, and iterative approaches. Res Syn Meth 2017:1-12.
Key questions	Indicate all of the key questions that will be used to underpin the recommendations. For effectiveness questions, use PICO format, and/or break it down into more specific questions. Use another approach, such as PerSPEcTiF format to formulate key questions beyond intervention effectiveness. In any case, use a format and framework that clearly presents the question at hand. Include both benefits and harms as appropriate for the key (PICO) questions related to intervention effectiveness. Using the WHO-INTEGRATE criteria and sub-criteria, indicate questions which help to inform other considerations for decision-making by the GDG, such as human rights and sociocultural acceptability, feasibility, societal implications, and equity, equality and non-discrimination.	SG GDG	Chapter 7, 15, 18 A tabular format, with columns for each of PICO/ PerSPEcTiF can be helpful, but is not mandatory. Do not present the questions in multiple formats: present a single format that conveys to the GRC that you have clear, answerable questions that will directly inform the recommendations.

continues ...

... continued

Topic	Instructions for reporting	Lead* Input*	Handbook chapter Comments
Important and critical outcomes	Describe how outcomes will be prioritized and a subset selected as the focus for the evidence review and for formulating recommendations, and indicate who will perform this exercise.	SG GDG (ERG)	Chapter 7
Humanitarian and other emergencies	Describe how the recommendations in the planned guideline are or may be relevant in the emergency context, including all hazards (infectious disease outbreaks, environmental exposures, natural disasters, forced migration, etc.).	SG GDG	None Most all WHO guidelines are potentially relevant in humanitarian settings and in public health emergencies, particularly in the context of protracted emergencies. It may not be feasible to develop the proposed guideline with consideration of such contexts in addition to the non-emergency context, as this involves additional expertise, evidence reviews, and resources. Nonetheless, the justification for NOT including this setting in the guideline needs to be carefully considered and presented. Options such as tools for adaptation in the emergency setting, prioritization of interventions when resources are severely constrained, or expert opinion in the form of remarks or considerations, may facilitate implementation of the guideline in these contexts.
F. Systematic review methods			
Information specialist	Indicate who will be primarily responsible for developing the search strategies, describe their experience and expertise, and indicate who will check and verify the validity of draft search strategies.	RTO Library scientist, SR, MX,	Chapter 8 It is mandatory that draft search strategies developed by an expert are reviewed by one or more information specialists and by content experts. At a minimum, WHO staff should review the search strategy with a WHO information specialist in the library.
Sources of evidence	Indicate which bibliographic databases you will search and why. Describe any other data sources such as citations from experts or hand-searching. If this information is not yet available, indicate who will perform these tasks and who will ensure the comprehensiveness and validity of the methods.	RTO Library scientist, SR, MX, GDG	Chapter 8
Types of evidence	Indicate if you plan to restrict your searching to existing systematic reviews or if you will likely need to examine primary studies. If you will be using existing systematic reviews obtained via a systematic search, explain how you will assess the quality of the review and determine if the review is up to date. If the selected existing systematic review is not deemed current, describe how you will update it. Do you plan to use qualitative or mixed-methods research to address some of the questions used to inform the recommendations? If so, describe how will you identify such evidence and evaluate confidence in the findings. Indicate if you plan to include only RCTs, or also controlled clinical trials (experimental studies with investigator assignment to the intervention by a means other than randomization), or observational studies in the effectiveness review. Provide a brief rationale for the selected designs.		Chapter 18 Because study design terminology is problematic, use descriptor labels such as randomized (or not), comparative, investigator-assigned groups, etc. If you use labels, define them in the proposal. The study design algorithm in *American Journal of Preventive Medicine* 2000;18(1S): 335-43 may be helpful. Distinguish study design criteria by outcomes (benefit, harms, other). You may want to consider observational data for harms outcomes, but not for benefits, for example. Choice of study design may depend on what is found in the first round of searches. For example, for outcomes of benefits you might start with RCTs, then modify your criteria if no data are available. These possibilities and the rationale should be outlined. GRADE CERQual: Confidence in the evidence from reviews of qualitative research: information and publications at: https://www.cerqual.org/publications/.

continues ...

... continued

Topic	Instructions for reporting	Lead* Input*	Handbook chapter Comments
Quality assessment of primary studies	Describe which tools will be used to assess the risk of bias (quality, internal validity, limitations) of individual studies and why that tool was selected.	SR team	Chapter 9
Quality assessment of the body of evidence	Indicate if you will use GRADE to assess the confidence in effect estimates for each outcome for intervention and diagnostic studies. If you plan to modify GRADE or not use it, detail the planned approach and provide a rationale for deviating from GRADE. Also describe whether you plan to use extensions to the GRADE approach, such as GRADE equity guidelines and GRADE-CERQual for different types of evidence synthesis.	SR RTO MX, SG (GDG)	Chapter 9, 18
G. Formulation of recommendations and peer review			
General approach	Describe the framework and methods that you will use to translate evidence to recommendations, along with the considerations or constructs that you will include.	RTO SR, MX, GDG	Chapter 10, 18 GRADE/DECIDE WHO-INTEGRATE
Specific considerations	Gender and other social determinants of health, and their impact on health equity, equality and non-discrimination: Describe how you will incorporate relevant issues into the recommendations. Human rights: Describe how these will be incorporated into the recommendations. Relative values of important and critical outcomes: Whose values will be considered? From where will you obtain these data? Preferences regarding the intervention (sociocultural acceptability): Whose values will be considered? From where will you obtain these data? If your only option is to use the preferences of the guideline development group members, justify this approach. Financial and economic considerations: Indicate what resources are the most relevant and how you will estimate the cost and/or cost-effectiveness. If you plan to use modelling for the assessment of cost effectiveness, indicate what approach you plan to use and who will perform the modelling. Societal implications: Describe how the net societal benefit associated with the intervention will be incorporated into the recommendations. Feasibility and health system considerations: Describe whose perspectives will be considered when evaluating intervention's feasibility. Indicate how health system considerations will be considered for making recommendations?	RTO SR, MX, GDG	Chapter 5, 10, 18
Peer review	Indicate: • Who will perform peer review. • How you will document the peer review comments and their disposition. • Who will assess the comments and incorporate them as appropriate into the guideline.	RTO SG	Chapter 12 Usually peer review is performed by the External Review Group. WHO staff from relevant departments and Regional Offices should also be included as indicated.
H. Project management			
Meetings	Indicate what meetings will be held that involve external experts (whether virtual or in-person), their purpose and which contributors will be invited.		None
Timeline	Provide a detailed timeline, including the dates for GDG meetings, performance and completion of the systematic reviews, peer review, submission to the GRC, and other key events.	RTO SG, SR, MX	None
Budget	Indicate the total funds available for development of this guideline and provide a detailed budget using the template.	RTO TU	None
Writer	Indicate who will draft the guideline document and who will revise and finalize it prior to submission to the GRC.	RTO SG	Chapter 3

continues ...

Topic	Instructions for reporting	Lead* Input*	Handbook chapter Comments
Publication	Consultation with WHO Press: Confirm that you have met with staff in WHP regarding formats for publication, inclusion in IRIS, and the management of annexes. Media and formats: Describe where the guideline will be published and in what formats. Translations: Indicate what translations are planned and how you will fund them.		None Consultation with WHO Press is essential at an early stage in guideline development to ensure that the final product is compatible with and takes advantage of WHO's publication processes and procedures. This will save you time and effort at the end of the development process.
I. Uptake and evaluation			
Dissemination	Briefly describe how you will make the final guideline accessible to the target audience.	SG GDG	Chapter 12
Derivative products	Describe derivative products that are planned, provide a rationale and brief description, and indicate the target audience for each product.	SG GDG	None Derivative products include brief summaries, tool kits, algorithms, "how-to manuals, posters, applications for mobile devices, among many others.
Model List of Essential Medicines	If the proposed guideline will include recommendations on medicines, indicate the relationship of those medicines to the WHO Model List of Essential Medicines. If the medicine is not on the current Model List, describe any plans for having it reviewed by the WHO Expert Committee on the Selection and Use of Essential Medicines.	RTO	None Information can be found at http://www.who.int/selection_medicines/list/en/.
Adaptation	Indicate the key issues that end-users will likely have to consider when adopting or adapting this guideline and describe how the final guideline will address these issues.	SG GDG	Chapter 13, 18
Implementation	Briefly describe what implementation strategies programme managers and other individuals could use and discuss potential barriers and facilitating factors.	SG GDG	Chapter 13, 18
Evaluation	Indicate how you will assess dissemination of, and access to, your guideline. Indicate how you plan to evaluate the impact of your guideline on the outcomes that you hope to improve. Link this plan to existing data collection structures and programmes in Member States and at WHO headquarters.	RTO SG	Chapter 13 Although an assessment of guideline quality with AGREE-II may be helpful (https://www.agreetrust.org/agree-ii/), it is not required and does not constitute an adequate evaluation plan.
Updating	Indicate when you plan to update this guideline (the review-by date) and provide a rationale. Indicate plans for ongoing monitoring for new studies or other relevant data and how that information will be used to inform the timing of an update. For consolidated guidelines, indicate how you will address recommendations contained within the guideline that need updating at different times. For rapid advice guidelines, indicate if you anticipate that a standard guideline will need to be produced in the foreseeable future. For interim guidelines, indicate when you plan to publish an updated interim guideline or a standard guideline (depending on the rate of production of new research).	SG GDG	Chapter 12

(*) This column indicates who is primarily responsible (the lead) for the task and who else provides input into the item of the planning proposal. Groups indicated within parenthesis are discretionary and may not need to be involved.

Abbreviations: COI, conflict of interest; CRE, WHO office of Compliance, Risk Management and Ethics; DOI, disclosure of interest; GDG, Guideline Development Group; GRADE, Guideline Recommendations Assessment, Development and Evaluation; IRIS, Institutional Repository for Information Sharing; MX, guideline methodologist; SG, WHO Steering Group; RCT, randomized controlled trial; RTO, responsible technical officer; SR, systematic review team; TU, Technical Unit originating the guideline.