World Health Organization

GUIDANCE FOR THE DEVELOPMENT OF EVIDENCE-BASED VACCINATION-RELATED RECOMMENDATIONS

Version 8
31 January 2017

This guidance applies to the development of recommendations by the Strategic Advisory Group of Experts (SAGE) on Immunization and the development of WHO vaccine position papers. Its aim is to facilitate the work of SAGE, its working groups and the WHO Secretariat. Additionally, its description of the recommendation development process will inform the wider readership. The document will continue to be updated as necessary as the methodology for evidence based-decision making evolves. Comments and suggestions for improvement are welcome, and should be sent to sageexecsec@who.int.

Content

1. Introduction

Vaccines are one of the most successful public health interventions of all time. Millions of lives have been saved and substantial disability averted due to the advent of critical vaccines. Much work is devoted to the development and testing of vaccines, leading ultimately to their licensure and use in populations. However, availability of the products does not ensure their appropriate use. The World Health Organization (WHO) is tasked to provide leadership in global health, to shape research agendas, to provide guidance and standards for public health practice, and to provide support to country programmes with global recommendations for vaccine use.

The Strategic Advisory Group of Experts (SAGE) on Immunization is an independent advisory committee with a mandate to advise WHO on the development of policy related to vaccines and immunization. SAGE, as outlined in its terms of reference provides recommendations to WHO on vaccination-relevant topics identified as priorities of public health importance. SAGE functions with the help of Working Groups (WGs), which are established to review the evidence and propose draft recommendations for SAGE consideration. A list of current WGs as well as detailed information on purpose, structure and functioning of the SAGE WGs can be found on the SAGE website.
After discussion and deliberation, SAGE issues recommendations captured in the SAGE meeting reports and published following each meeting in the WHO Weekly Epidemiological Record (WER). All reports, meeting presentations and background documents are available online. Recommendations on specific vaccines are adopted as WHO policy and published as WHO vaccine position paper. Since 1998, to fulfil its mission for vaccines, WHO has published WHO vaccine position papers. These vaccine-specific position papers are comprised of four sections: an introduction; a section providing information on the respective disease (disease epidemiology, the pathogen, the disease); a section providing information on the available vaccines (composition, safety, immune response, efficacy and effectiveness, cost-effectiveness and any other relevant issues), and the WHO position on optimal vaccine use.

The entire process leading to WHO recommendations on the use of vaccines is compliant with principles set out in the WHO Handbook for Guidelines Development. Careful review and consideration of the scientific evidence is an essential step in recommendations and guidelines development. The results from the review of evidence on a given topic should be carefully considered to identify magnitude of the effect, geographic variability, and other factors that are important for assessing impact and generalizability. It should also be noted that although evidence is produced by relatively objective scientific endeavours, the evaluation of the evidence quality and the making of recommendations are activities that require expert interpretation and judgement in addition to rigorous scientific review. In developing the most appropriate recommendations, committees should weigh the desirable and undesirable

consequences of potential recommendations based on the best available evidence, while taking into account numerous additional factors.

Factors that are taken into consideration when making recommendations include:, disease epidemiology and clinical profile; the benefits and harms of the options; values pertaining to the importance of the desirable and undesirable effects; equity considerations; feasibility and resource implications including economic considerations; social values and preferences, and acceptability; health-system opportunities, and interaction with other existing intervention and control strategies. In addition to study results themselves, consideration is given to methodology and study design. While it is generally accepted that randomized controlled trials (RCTs) are the gold standard study design because of their ability to minimize various forms of bias, there are many characteristics of RCTs or observational studies that determine their quality and relevance to the formulation of policy recommendations as outlined in section 3 below. For example, faulty randomization or blinding may reduce the quality of an RCT below that of a well-designed observational study. The quality of evidence reflects the extent to which confidence in the estimation of effect is adequate to support a particular decision or recommendation. Hence, a review of the potential risks for bias and other aspects of study design quality are crucial when drawing conclusions from a study of any type.

The Grading of Recommendations Assessment, Development and Evaluation (GRADE)[1] approach is the most prominent of the many frameworks developed over the years to assess the quality of evidence, and has been adopted by WHO and many other national and international organizations. The use of the GRADE methodology to rate the quality of evidence in support of key recommendations included in the WHO vaccine position papers began in April 2007. Evidence underlying the critical recommendations are rated using the GRADE framework with formal scoring to assess the quality of related evidence. Although all SAGE recommendations and vaccine position papers are evidence-based and follow an evidence-based, systematic process to retrieve and assess available evidence, SAGE also makes strategic recommendations regarding public health programmes and research priorities, which are not subject to a formal GRADE scoring. In some instances, SAGE issues good-practice statements. Good practice statements typically represent situations in which a large body of indirect evidence, often composed of several bodies of evidence linked together in a causal pathway, including indirect comparison, unequivocally demonstrates the net benefit of the recommended action. These types of recommendations are then labelled as such.

The formal GRADE process has been described elsewhere. In short, questions of importance related to a recommendation are identified, a systematic literature review is

[1]Guyatt GH et al. for the GRADE Working Group. Rating quality of evidence and strength of recommendations GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal*, 2008, 336:924–926.

conducted to identify the evidence available to answer the question(s), the quality of relevant evidence is assessed and rated using the GRADE evidence grading system, and the results of the process are summarized for effective communication. Five criteria (limitations in study design commensurate with the type of study, inconsistency, indirectness, imprecision and publication bias) are used to downgrade the quality of evidence when studies do not meet the published standards, and three criteria (large magnitude of effect, dose-response gradient, and ability of the study to limit biases and control for confounding) are used to upgrade the quality of evidence when study results increase confidence in their validity. The quality of the evidence is assessed to be of high, moderate, low or very low quality. It is to make a recommendation based on low or very low quality evidence.

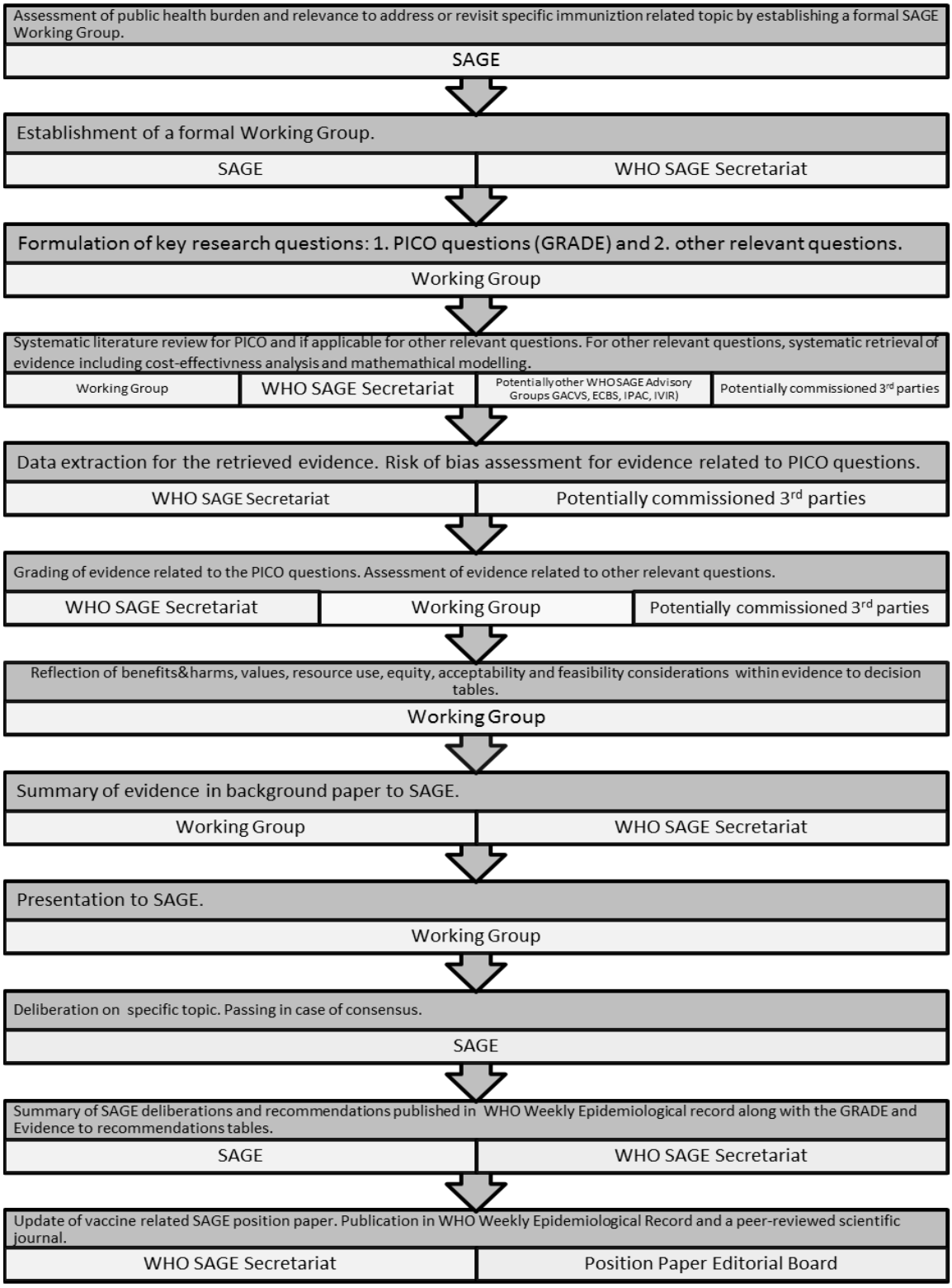
The GRADE tables are factored into the overall decision-making process which is reflected in the Evidence to Recommendation tables. These are based on the tables provided by the "Developing and Evaluating Communication Strategies to Support Informed Decisions and Practice Based on Evidence" (DECIDE) collaboration which is aimed to improve the dissemination of evidence-based recommendations by building on the work of the GRADE Working Group to develop and evaluate methods that address the targeted dissemination of guidelines. A hallmark of these tables is its aim to improve transparency in decision-making as interested parties are able to follow the logic and processes that led to a given conclusion, recommendation and/or guideline. Such a process also promotes useful dialogue and opportunities to reassess the evidence as required.
Since 2007, GRADE tables and since 2014, Evidence to Recommendation tables, have accompanied WHO vaccine position papers and are made available online. GRADE and Evidence to Recommendation tables attempt to apply the GRADE and DECIDE framework as strictly as possible, although GRADE evidence profiles, summary of findings[2] and Evidence to Recommendation tables have been adjusted to the specific needs of vaccination-related recommendations, and provide additional information in footnotes and narrative text where considered necessary.

SAGE continues to follow the evolution of methodological processes and refines and adjusts the approach to ensure its relevance to immunization public health policy.

Figure 1 depicts the entire evidence to recommendation process applied by SAGE.

[2] Guyatt GH et al. GRADE Guidelines: 1. Introduction GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*, 2011, 64:383–394.

Figure 1: SAGE process to obtain immunization-related recommendations.



2. SAGE mode of functioning

SAGE meets biannually, usually in April and October. SAGE meeting dates are set 3 years in advance and are published on the SAGE website. During or in-between the meetings, SAGE identifies specific immunization-related public health priorities for which SAGE may decide to establish a WG related to the identified topic. SAGE then, jointly with WHO lead technical staff, develops the terms of reference for this WG. A public call for nominations soliciting 8-12 international experts in the specific field is launched, which can be accessed on the WHO SAGE website. If the issue is less complex, in exceptional cases, SAGE may decide to rely on a consultation of selected subject-matter experts rather than a formal SAGE WG, or have the evidence prepared by the WHO Secretariat. This does not imply that the evidence review is less thorough.

As the evidence and/or other factors change, the need to update existing policy recommendations is reviewed periodically by the WHO Secretariat at a minimum frequency of two years and potentially sooner, depending on the availability of new scientific evidence and public health priorities. A review of the vaccine position papers may also be requested directly by SAGE, WHO Regional Technical Advisory Groups on Immunization, country authorities, or key partners. The need for updating is brought to SAGE which then decides whether a comprehensive review of the evidence is required.

Although it could easily be argued that all important public health decisions that may lead to the savings of many lives are a matter of urgency in themselves, there are exceptional situations (such as the influenza pandemic or Ebola) that require more rapid decisions. In such situations, extraordinary SAGE meetings or teleconferences may have to be held and recommendations issued rapidly and revised as the context changes and/or additional data becomes available.

3. SAGE process from evidence to recommendation

In general, following its establishment, WG is tasked to conduct the initial review of evidence pertaining to a given topic. The principals of evidence-based medicine[3] are used when systematically assessing the available evidence. WGs present proposals for recommendations to SAGE, which in turn discusses, deliberates and ultimately provides recommendations to WHO. In some instances, WGs build on specific reviews of the data or data collection tools done by other technical advisory groups (e.g. The Global Advisory Committee on Vaccine Safety (GACVS) for vaccine safety assessment).

The key steps involved in creating evidence-based SAGE recommendations are as follows:

1. Definition of the questions to inform recommendations including identification of the critical questions and outcomes for which an in-depth review of evidence is needed.
2. Execution of a systematic review of the literature with or without meta-analysis and, where necessary, commissioning research to address gaps in evidence.
3. Review of the quality of the evidence, in particular through assessment of the risk of bias and confounding.
4. Rating of the quality of the evidence (using the GRADE approach for data on safety and effectiveness).
5. Reflection of benefits & harms, values, resource use, equity, acceptability and feasibility considerations of the intervention within Evidence to Recommendation tables.
6. Discussion and deliberation leading to the development of proposed recommendations.
7. Presentation of proposed recommendations, along with their supporting evidence to the entire SAGE membership at SAGE meetings.
8. SAGE discussion, deliberation and decision regarding the proposed recommendations to WHO.

Each of these steps is discussed in the sections that follow. The process of public health decision-making is often stepwise, multifaceted and complex. Decision-making under uncertainty is part of public health. To explicitly deal with uncertainty, it is necessary to transparently and honestly inform policy-makers and the public. The guiding principles of the review process are that careful review and consideration of the evidence should precede development of recommendations, and that the entire process should be transparent, robust and reproducible.

[3] Cook DJ, Jaeschke R, Guyatt GH (1992). "Critical appraisal of therapeutic interventions in the intensive care unit: human monoclonal antibody treatment in sepsis. Journal Club of the Hamilton Regional Critical Care Group". J Intensive Care Med 7 (6): 275–82.

For all recommendations, the steps listed above are always conducted, with the exception of the formal GRADE rating (step 4). Examples of recommendations for which a formal rating is not done include strategic recommendations, recommendations on the removal of well-defined barriers with respect to vaccine implementation, and other programmatic recommendations. As mentioned previously, SAGE in some instances issues good-practice statements.

3.1 Definition of the key questions to inform recommendations.

After its establishment, the initial aim of the WG is to reach consensus on the key questions that may be relevant to making immunization recommendations. This may include questions on the burden of the disease, the effectiveness and safety of a vaccine and the optimal schedule for protection, programmatic consideration such as acceptability, resource use and feasibility. All of these may need to be considered in light of values and preferences and equity within the general population, in different geographic regions, and in various subpopulations. Thus, a broad range of contextual issues should be taken into account when making recommendations, including the epidemiologic and clinical features of the disease, vaccine and immunization characteristics, economic considerations, potential interactions with other existing interventions/control strategies, and social, legal and ethical considerations (see Appendix 1 for a detailed list).

<u>PICO questions</u>

Key questions include questions that can be addressed using a systematic review of the evidence. The standard is to use the "PICO" format, which is a well-accepted methodology for framing of questions for systematic reviews (see Appendix 2 for standard PICO question format).
This approach is used to ensure questions are formulated and framed effectively. PICO refers to the following elements:

P	• Population
I	• Intervention (or exposure)
C	• Comparator
O	• Outcome of interest

On specific vaccine-related topics, three issues that will generally require systematic review of literature and GRADE scoring are (1) vaccine efficacy/effectiveness, (2) vaccine safety, and (3) duration of protection.

The framing of questions relating to vaccine safety is of particular importance. Formal GRADE scoring should focus on evidence related to the potential occurrence of serious and specific adverse events. However, other factors such as variations in vaccine reactogenicity and more minor local or systemic reactions (e.g. fever) can lead to decreased vaccine acceptance, and must also be factored into recommendation-making. Evidence of causality between vaccination and adverse events must be sought. For considerations of adverse events following immunization, SAGE usually refers to vaccine safety reviews and statements from the GACVS, when available.

Because GRADE decisions about the overall quality of evidence supporting a recommendation are dependent on which outcomes are selected for review, it is important to choose relevant outcomes for assessment that are important to the target population and the broader community. All identified potentially important outcomes are classified into the following categories: critical; important but not critical; limited importance. Evidence regarding critical and important outcome types may have bearing on SAGE recommendations, but evidence relating to outcomes of limited importance will not. If important outcomes are represented by a surrogate, they will frequently require down-rating of the quality of evidence for indirectness.

As a general rule, an initial rating of the importance of outcomes should be done prior to the evidence review. It is also advisable that the questions identified by the WG be validated by SAGE early in the process to ensure that resources are not invested in exercises that would not satisfy the needs of SAGE.

Because there are many factors to be considered when making recommendations, WGs will often identify many questions for which answers will be sought. However, it is impossible to address all issues to a minute level, and hence questions must be prioritized by the WG. A systematic evidence review, should no high quality systematic literature review already exist, will be conducted for all relevant questions, but only questions or outcomes determined to be critical for intervention implementation decisions are prioritized for formal GRADE scoring of the quality of evidence. Unless there are unusual circumstances, GRADE should be applied to no more than five questions.

Other relevant key policy questions

Further, WGs may identify other relevant key questions where the PICO format and/or formal GRADE scoring might not be applicable, such as key questions on disease burden, economic considerations or strategic recommendations (e.g. research gaps, decision to

pursue an eradication goal, etc.). WGs are asked to identify appropriate tools to address these questions, such as systematic literature review, mathematical modelling and cost-effectiveness evaluations.

It is important that the quality of the data be assessed and reflected upon. For assessing the quality of economic and cost-effectiveness evaluations, other guidelines (e.g. WHO Guide for Standardization of Economic Evaluations of Immunization Programmes) can be followed.

3.2 Systematic review of the literature

A detailed methodological overview of systematic literature reviews can be found e.g. in Methods for the development of NICE public health guidance and in the Cochrane handbook for systematic reviews of interventions. Briefly, there are 5 phases involved in carrying out a literature search that is carefully documented, transparent and reproducible. First, a review protocol is developed, in which the objective of the review, PICO questions, study inclusion and exclusion criteria, search strategy, data collection, quality assessment, and data synthesis are specified. Next, the systematic literature search is done, which involves identification of information sources, development of a search strategy, management of references, and documentation of the search procedure. Following this, the study selection step identifies those search results that meet the specified inclusion criteria; data on study characteristics and outcomes from included studies are then extracted using a standardized data extraction tool. Finally, if appropriate, data may be synthesized by meta-analysis (for quantitative studies) or other approaches (for qualitative or mixed-method studies).

Unless a relevant recent review is available, comprehensive systematic literature reviews are carried out by the WG or the WHO SAGE Secretariat. Systematic literature reviews may be commissioned to 3rd parties. No specific historical time limit is set for the retrieval of information and no language restrictions are applied. Critical publications are translated into English as necessary. In certain instances when high-quality systematic reviews are available, one proceeds to update these reviews to reflect new publications.

Completed systematic literature reviews, including any pre-existing systematic reviews, are assessed by the WG to ensure completeness. Data should be extracted and consolidated using a data extraction tool and a list of relevant papers (including access to full content of the manuscripts) should be provided to the WG for review. When differences in interpretation arise, adjudication is made on the basis of WG consensus.

In some instances, WGs build on specific reviews of the data or data collection tools done by other technical advisory groups such as the GACVS, the WHO Expert Committee on Biological Standardization (ECBS), the WHO Immunization and vaccines related

implementation research advisory committee (IVIR-AC) and the WHO Immunization Practices Advisory Committee (IPAC).

The WGs may also turn to reviews of data by National Immunization Technical Advisory Groups (NITAGs). In some cases WGs may decide to commission an update of an existing systematic literature review or a full systematic review to independent third parties, who are required to submit a declaration of their potential interests based on WHO principles.

Data considered in SAGE and WHO evidence reviews may be published or unpublished, and concerted efforts should be made to identify any unpublished but relevant data that would inform WG and SAGE deliberations. For unpublished data to be taken into consideration, it is essential that SAGE and WHO are provided with enough information on the methodology to meaningfully assess (study) quality, and that such unpublished data be properly referenced (e.g. as 'in press' publications or by reference to the host web address).

While RCTs are considered the gold standard for intervention assessment, observational studies, including outbreak investigations, disease surveillance and post-market surveillance data represent important sources of data for vaccine effectiveness and safety, and constitute a significant component of the body of evidence used for SAGE recommendations. Other types of data, such as programme evaluations, cost-effectiveness analyses, forecasting and landscape analyses, may also be relevant.

In most cases, final decisions regarding recommendations to use particular vaccines are not made until the critical missing data are made available. In rare instances, recommendations may be needed for interventions about which there is a very limited evidence base. In these circumstances, what little evidence is available may come from related but indirect studies (e.g. studies evaluating other live vaccines given to pregnant women), and after careful considered by the key experts, may form the foundation for a recommendation. When recommendations based on limited evidence must be formulated and clear explanations should be provided.
Strong recommendations based on low or very low quality evidence are classified as discordant recommendations and should if possible be avoided, though may be warranted in 5 different scenarios:

- Life-threatening situations
- When uncertain benefit, but certain harm
- When potentially equivalent options, one clearly less risky or costly than the other
- High confidence in benefits being similar, but one option potentially more risky or costly
- Potential catastrophic harm.

Conditional recommendations may also be considered for specific populations in such situations.

For example, in the 2007 Vaccine position paper on rotavirus vaccine, WHO stated *"…until the full potential of the current rotavirus vaccines has been confirmed in all regions of the world, in particular in Asia and Africa, WHO is not prepared to recommend global inclusion of rotavirus vaccines into national immunization programmes."* WHO later amended the recommendation once data supporting widespread use were available. This recommendation was classified as conditional pending further evidence on the effectiveness of the intervention.

Literature searches are also important for identifying knowledge gaps and helping prioritize future research agendas. Those areas where data are lacking should be highlighted by both the WG and SAGE to encourage additional research.

3.3 Identifying study limitations

Studies identified in the systematic literature review should be documented in a summary table and associated with an evaluation of methodological quality[4] (e.g. Appendix 4). This process allows for easier comparison and evaluation of studies when scoring the quality of scientific evidence.

There are a number of factors that may put studies at a higher risk of bias (i.e. systematic error) and they need to be considered when determining the quality of the evidence. Both the Cochrane Collaboration and the Critical Appraisals Skills Programme have developed useful tools for evaluating study quality. Standardized approaches to evaluating the quality of non-randomized trials are also available but generic tools to evaluate the quality of observational studies are difficult to develop because observational studies encompass a wide variety of study designs.

The tools listed in Appendix 6 are adapted from the Cochrane Handbook and Critical Appraisal Skills Programme (CASP). As noted in the Cochrane Handbook, there are other important aspects of study quality (e.g. reporting quality and ethical approval) which are not addressed in this section. Rather, the primary focus is on the risk of bias that could affect the interpretation of study results.

Appendix 5 provides a list of data items to consider for extraction from included studies and Appendix 6 provides checklists developed by the GRADE, CASP, the Cochrane Effective Practice and Organization of Care Group, and other groups. These may be used and adapted to assess study methods and potential limitations of vaccine studies.

[4] Mlika-Cabanne N et al. Guidelines International Network (GIN) Working Group on Evidence Tables. *BMJ Quality & Safety*, 2011, 20(2):141–145.

3.3.1 Risk of bias in RCTs

When properly conducted and of adequate size, RCT study designs have the lowest risk for bias. The Cochrane Collaboration highlights six characteristics to consider concerning the risk of bias in RCTs.

- Sequence generation refers to the method of randomly allocating an intervention to study participants.
- Allocation sequence concealment refers to the prevention of knowledge (or prediction) of intervention assignment by study participants and investigators.
- Blinding refers to the masking of the intervention to assigned study participants and investigators.
- Incomplete outcome data may be the result of participant drop-out (missing data) or exclusion of data from the study results.
- Selective reporting (i.e. reporting bias) is the incomplete publication of results based on their results.
- Other sources of reporting bias may include design-specific risks of bias, early stopping, baseline imbalance, blocking of experimental units in unblinded studies and differential diagnostic activity.

For more detail on each of these, see the Cochrane Handbook (Chapter 8). Each feature should be evaluated to determine the risk of bias in each study (using the data extraction tool and checklist) and then documented in the summary table for evidence review.

3.3.2 Risk of bias in observational studies

Observational studies are particularly susceptible to selection bias and confounding. As different types of observational studies carry different risks of bias, it is more challenging to standardize the evaluation of bias across study types. Checklists have been included in Appendix 6 for reviewing the quality and risk of bias in case-control and cohort studies. These can be modified for other types of observational studies. According to the Cochrane Collaboration (Chapter 13), consideration should be given to differences in the comparison groups or within participants over time, potential temporal and geographic differences in group allocation and interventions that could bias the results, and the prospective and retrospective aspects of the studies. A clear description of potential confounders, along with what the authors did to address confounding (e.g. matching, stratification, etc.) should also be clearly outlined.

All of these features are included in the data extraction tool and checklists to aid reviewers' assessments of potential risks of bias. The collective results should then be reflected in the GRADE scoring under the "limitations" criterion (see section 3.4).

3.3.3 Impact of bias

After carefully reviewing each study for potential bias, an overall assessment of the evidence for risk of bias as well as the likely direction(s) and magnitude of the bias(es) should be reached. If many of the studies that constitute the evidence base have a high risk of bias, any conclusions from that body of evidence must be considered with caution. Studies at high risk for bias may be excluded if the results are deemed unreliable. At times, sensitivity analysis can be performed with and without biased studies to test the robustness of the decisions made from the systematic review.

3.3.4 Quality of systematic reviews and meta-analyses

Systematic reviews and meta-analyses can be useful tools for evaluating effects across studies. Their validity will depend on the completeness of the literature search, the thorough assessment of study quality, the appropriateness of combining data across studies and the relevance of the outcomes considered. In assessing the quality of an existing systematic review, careful attention should be paid to the following: search methodology; heterogeneity, and inclusion/exclusion criteria (particularly for observational studies), in addition to the quality of the design and methodology of individual studies. If any of these are in question, the results of the systematic review should be viewed cautiously. Some reviews do not consider all of the data that may be relevant to an assessment of vaccine efficacy and safety (e.g. observational studies, outbreak investigations, surveillance reports, etc.). Appendix 6 provides links to key quality appraisal tools, among other the AMSTAR tool for the assessment of the quality of systematic reviews. The Cochrane Adverse Effects Methods Group (AEMG)was also established to develop methods for producing high-quality systematic reviews of adverse effects, and provides useful information on this topic.

In some cases, a systematic literature review may already have been done by WHO or another group (e.g. Cochrane Collaboration), independent of, or on behalf of WHO. Previous reviews may serve as the basis for analysing the evidence base, but an updated search should be conducted to ensure studies published since the previous review are not missed.

3.4 Using GRADE to rate the quality of evidence

General considerations

Throughout the evidence review process, expert opinion is critical in the assessment of these factors and their importance to the question under consideration. The application of the GRADE criteria and the inferences that may be drawn from the studies relating to the question under consideration are inherently subjective, and rely on the judgement of skilled and experienced public health professionals. Active participation of the WGs is essential to ensure that the most appropriate studies are utilized and that the results

are carefully considered; in addition to formulating the questions for GRADE assessment, the WGs will review the provided evidence and the resulting GRADE tables.

GRADE quality assessment

Only primary data sources should be entered into GRADE tables. Both published and unpublished studies may be included as long as they are in press or accessible through a link on the website. Mathematical models do not represent primary data, but build on other sources of information, and should therefore not be included in the GRADE tables. Nevertheless, mathematical models are used as part of the decision making process by SAGE. IVIR-AC assists SAGE with implementation research questions including mathematical models in addition to reviewing and advising on quantitative methods in vaccine research.

Each study should be reviewed using the following criteria, with recognition that application of the criteria is a subjective process and open to individual interpretation. For example, decisions on the degree of similarity in study estimates of effect or on the appropriate thresholds for downgrading due to inconsistency can be guided by a review of point-estimates, confidence intervals and values of i^2 statistic of heterogeneity, but are inherently subjective issues.[5] Furthermore, it may be very difficult to conclusively assess whether selective outcome reporting bias is present in a body of evidence as this may require detailed information held by a specific study team, and close attention to missing data that should have been collected during the study. However, because of their content expertise, WGs are particularly well-positioned to comment on this parameter. Documenting the process of quality assessment in an open and transparent manner allows others to review the process and propose alternative interpretations for consideration.

Studies enter into the GRADE system at an initial level based on their study design. To begin however, all RCTs enter at level 4 (⨁⨁⨁⨁) i.e. the highest level of quality of evidence and observational studies and surveillance data enter at level 2 (⨁⨁) i.e. low level of quality of evidence. As not all studies of a particular design are equal, the GRADE approach provides a framework to then upgrade or downgrade the rating of the evidence, based on methodological and quantitative assessment.

Boxes 1 and 2 (below) outline the criteria for downgrading and upgrading the strength of evidence after its initial entry into the framework at level 4 (RCTs) or level 2 (observational studies including surveillance data) (see also Appendix 6). Each downgrading or upgrading of evidence needs to be succinctly footnoted and justified in the GRADE summary table. The brief associated descriptions provide specific

[5] Huedo-Medina T et al. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *CHIP Documents*, Paper 19, 2006 (http://digitalcommons.uconn.edu/chip_docs/19, accessed Dec 2016).

instructions on how to apply GRADE to the area of vaccines and vaccination. More detailed information may also be found in GRADE-related publications.

Box 1: **Criteria used to downgrade the quality level of evidence**[6]

Limitations: Quality rating may be downgraded by one or two levels for serious or very serious methodological limitations in the studies. Examples of these limitations include: inappropriate randomization; lack of concealment; violation of the intention to treat principle; inadequate blinding; substantial loss to follow-up; and early stopping for benefit. (See section 3.4 for how to evaluate risks of bias due to methodological limitations.)

Inconsistency: Quality rating may be downgraded by one or two levels if the effect is not similar and heterogeneous across studies, and if inconsistencies are serious or very serious.

Indirectness: Quality rating may be downgraded by one or two levels if there are serious or very serious issues with indirectness. Examples of indirectness may include: using surrogate end-points; use of immunogenicity versus clinical end-point; indirect comparisons between two treatments; potential problems with generalizability to the population of interest; and test inaccuracies. It is suggested that when assessing clinical protection, there is no downgrading for immunogenicity studies when there are well-established standard correlates of protection.

Imprecision: Quality ratings may be downgraded by one or two levels if there is serious or very serious imprecision (i.e. confidence intervals are wide or very wide). When possible, imprecision should be assessed using 95% confidence intervals of pooled relative risks or odds ratios (using meta-analysis techniques), as opposed to looking at 95% confidence intervals of individual studies.

Reporting Bias: Quality ratings may be downgraded by one or two levels if publication bias (i.e. failure to report studies), and selective outcome reporting bias (i.e. failure to report outcomes) are likely or very likely.

[6] Guyatt et al. GRADE Guidelines: 4. Rating the quality of evidence — study limitations (risk of bias). *Journal of Clinical Epidemiology*, 2011, 64(4):407–415.

Box 2: **Criteria used to upgrade the quality level of evidence**[7]

Large effect/strength of association:

Quality rating may be upgraded by one level if there is evidence from RCTs or observational (including surveillance) studies of vaccine effectiveness of 50% or higher (OR/RR >=2 or =< .5[8]) with no major[9] residual confounders.

Quality rating may be upgraded by two levels if there is strong evidence from RCTs or observational studies of a vaccine effectiveness of 80% or higher (or depending on the outcome an OR/RR >=5 or =< .2) with no major residual confounders. If RCTs suffer very serious methodological limitations, then upgrading for large effect should not be applied.

Population effect (dose-response gradient at population level):

Quality rating may be upgraded if there is evidence of a dose response gradient at the population level, i.e.

- Increase by one level if there is evidence of risk reduction in disease incidence with increasing population vaccine coverage. Evidence of decreased risk with increased vaccine coverage includes evidence of reversal at population level (where there is unfortunately a programme failure leading to a decrease in vaccine coverage, and subsequent disease return), and evidence of risk reduction in older or younger age groups not targeted for the intervention, but who benefit from herd immunity.

- Increase by two levels if there is very strong evidence of population risk reduction with increasing population vaccine coverage in many different settings and many years of evidence, and/or evidence of reversal at population level where programme failure results in decrease in vaccine coverage followed by return of disease.[10]

Mitigated bias and confounding:

Major confounders:[11] Quality rating may be upgraded by one level if all major confounders would have reduced the demonstrated effect (or increased the effect if no effect was observed).

or

Good quality study design: Quality rating may be upgraded by one level if there was a good quality of study(ies) design to control for confounding and selection biases among cases and controls e.g. with population-based record linkage, self-controlled case series or other appropriate designs.

The quality rating may be further upgraded by one point if there is consistency between studies across different settings, different investigators and different designs.[12]

[7] Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. The GRADE Working Group. GRADE guidelines: 9. Rating up the quality of evidence. J Clin Epidemiol 2011;64:1311-6.

[8] These thresholds refer to risk ratios. When baseline risk is low (i.e. below 20%), odds and risk ratios are very similar and one can comfortably apply this criteria. When the baseline risk is high (a rare occurrence for vaccine preventable diseases) and the effect size is large, ORs can be far larger in magnitude than risk ratios. Under such circumstances, a higher threshold for ORs may be appropriate (Davies H et al. When can odds ratios mislead? BMJ 1998;316:989 http://www.bmj.com/content/316/7136/989)

[9] Changed from "plausible" confounders in the formal GRADE framework.

[10] This increase by 2 levels is not directly reflected in the current GRADE rating scheme and collaboration with the GRADE working group will continue to further optimize the process. The GRADE working group, however recognizes that in some circumstances other considerations may lead to upgrading as appropriate. This is an example of other criteria that have been determined by SAGE to be applicable for upgrading.

[11] This criterion has been slightly modified from the GRADE criteria, which specify that all "plausible" confounders would have reduced the effect.

[12] Criterion not included in the formal GRADE framework and only applicable to observational studies.

Ratings of quality should be clearly displayed in the GRADE table. For reductions in quality levels, possible ratings include "none serious" (no downgrade), "serious" (downgrade by one level) or "very serious" (downgrade by two levels). For upgrading the quality level, possible ratings include "not applicable" (no upgrade), "strong evidence" (upgrade by one level) or "very strong evidence" (upgrade by two levels). Final quality levels cannot exceed four or drop below one. If there are major limitations in the study design commensurate with the design, then upgrading criteria should not be applied.

Whenever a downgrade or upgrade is applied, a footnote is needed to explain the rationale for the change in rating.

Example

Studies aiming to evaluate vaccine efficacy may be downgraded under the criterion of "indirectness" due to the use of surrogate end-points such as immunogenicity data used to measure vaccine efficacy.

In some cases, studies may not be downgraded, but footnotes should still be used to highlight potential issues. This promotes transparency and shows readers that the full range of issues has been considered.

When the GRADE criteria are applied, studies should not be repeatedly penalized for limitations already factored into their starting rating. As an example, a controlled observational study that enters into the rating system at a level 2 (⊕⊕) should not be further downgraded because it was not randomized. On the other hand, it would be appropriate for passive surveillance data of uncertain quality to be downgraded through application of relevant limiting factors.

The decision to downgrade or upgrade a body of evidence depends on individual judgement. While two individuals may agree on the study limitations during a review of the evidence, it may not be clear whether or not such limitations warrant a change in rating. Similarly, the amount of variation in results from multiple studies allowed before they are deemed inconsistent, may be contentious. These examples illustrate the subjective nature of the exercise, the importance of expert opinion in interpretation and assessment of the criteria, and the need to explain the thought process used throughout the evaluation, so that areas of agreement and disagreement are evident.

Quality of evidence rating

Using the criteria described above, both individual studies and the collective body of evidence should be evaluated. The overall collection of studies will receive a rating

based upon analysis of the component studies. The quality of scientific evidence is rated using the GRADE scale, as follows[13]:

- Evidence supports a high level of confidence that the true effect lies close to that of the estimate of the effect on the health outcome (level 4, or ⊕⊕⊕⊕).
- Evidence supports a moderate level of confidence that the true effect lies close to that of the estimate of the effect on the health outcome (level 3, or ⊕⊕⊕).
- Evidence supports a limited level of confidence that the true effect lies close to that of the estimate of the effect on the health outcome (level 2, or ⊕⊕).
- Evidence supports a very low level of confidence that the true effect lies close to that of the estimate of the effect on the health outcome (level 1, or ⊕).

The GRADE tables explicitly provide the rating of quality of evidence for the outcomes critical to recommendations. These factors help inform whether or not a recommendation should be made.

Application of GRADE by SAGE

SAGE has adjusted the GRADE methodology to strengthen its relevance and facilitate its use by SAGE and SAGE WGs. These adjustments to the traditional presentation of the GRADE tables are an attempt to clarify its application to vaccines/vaccination recommendations without changing the intent. The adjustments ensure that the many types of data that comprise immunization-related research are adequately taken into consideration in the decision-making process. Vaccine development and testing has occurred over many decades, and many older vaccines are still used today. Therefore, the evidence base that is used to formulate recommendations often includes studies spanning a long time-horizon, and as RCTs are unethical once the impact of protection is evident, observational studies heavily contribute to the evidence base. When robust RCTs exist, the scoring of the evidence concerning efficacy is performed using only those RCTs, without the inclusion of observational studies. However, when observational studies are an important part of the body of evidence used to formulate recommendations, both RCTs and observational studies are taken into consideration and reviewed in totality. Unlike in the conservative use of GRADE, SAGE, if considered as suitable, combines RCTs and observational studies within the same GRADE table. See Appendices 9a and 9b for a template and an example of a SAGE-modified GRADE table.

Application of GRADE to recommendations

In the formal GRADE approach developed by the GRADE Working Group, a strength of recommendation score is also given (i.e. strong versus weak or conditional recommendations). For immunization, SAGE has made the decision not to apply the

[13] See Appendix 7 for details of modification of GRADE scale terminology for SAGE use.

differentiation between strong and weak vaccine recommendations according to the GRADE methodology, as weak recommendations are of little value to country immunization programmes.

It is the goal of WHO and SAGE to provide only strong recommendations, which may either be for or against an activity, or may be condition-dependent.[14]

3.5 SAGE Evidence to Recommendation tables

To increase transparency and systematically consider predefined criteria leading to recommendations, SAGE uses "Evidence to Recommendation tables" based on DECIDE framework.

SAGE requires that WGs provide an Evidence to Recommendation table using the standard SAGE format (see Appendix 10). The table should contain the following: background information and the research question, the specific criteria to consider and the related judgements that are made for each criterion, research evidence to support each judgement, and additional information to justify the judgements and decisions made for each criterion. The table concludes with the balance of consequences of benefits and harms, the recommendation made and justification for the aforementioned recommendation, implementation considerations and research priorities.

With the help of the table to structure and facilitate their discussions, WGs will develop draft recommendations to propose to SAGE. SAGE will then deliberate on the proposed recommendation which may lead to an adjustment to the evidence to recommendation table after the final recommendation by SAGE is issued.

Access to GRADE and Evidence to Recommendation tables

GRADE and Evidence to Recommendation tables (Appendix 9a. Template of a GRADE table used to rate the quality of evidence; Appendix 9b. Sample completed GRADE table; Appendix 10. Sample on an Evidence to Recommendation table) are available on the WHO website together with the vaccine position paper and also as part of the background information presented to SAGE. Within the position papers, GRADE and Evidence to Recommendation tables are cited as footnotes.

[14] In the formal GRADE framework, a conditional recommendation is synonymous with a weak recommendation. For WHO and SAGE, a conditional recommendation is a strong recommendation constrained to a particular subpopulation or country after having met given criteria. For example, a second dose of measles vaccine in national schedules is not recommended until a country has achieved 80% coverage of the first dose of measles vaccine for the last three years.

Updating of GRADE and Evidence to Recommendation tables

If additional evidence provides further scientific support for the recommendations, the GRADE and Evidence to Recommendation tables may be updated by WHO without updating the position paper. If new evidence arises that necessitates a re-evaluation of the vaccine position paper recommendations, a more formal update process will be initiated.

3.6 Discussion and deliberation by the WGs leading to the development of proposed recommendations

WGs meet on a regular basis until they have completed all the objectives in their terms of reference, which typically takes 12-18 months. WG often meet up to three times in person, participate in frequent (often monthly) conference calls via telephone, video or web conferencing. Both meetings and conference calls are only open to WG members and invited contributors. WG members review evidence provided in the form of presentations from WHO, experts and/or WG members, highlight issues and make proposals for recommendations. Draft documents (such as background papers, summaries of the evidence, evidence to recommendation tables, etc.) and presentations to SAGE are discussed and vetted by the WG, and proposed recommendations are developed through WG consensus. For additional information on WGs, please see the SAGE Terms of Reference.

Much work goes into information-gathering and synthesis that forms the basis of vaccine recommendations and guidance. Even recommendations that do not utilize a formal GRADE evaluation are the product of data review, evaluation of data quality, discussion and deliberation.

When recommendations are proposed, a number of factors need to be put in consideration:

1. The quality of the evidence i.e. the degree of confidence in the estimates of effect needs to be assessed as a key factor in determining the strength of a recommendation.
2. Values and preferences of individuals and populations affected by the recommendations pertaining to their relative importance assigned to the outcomes associated with the intervention or exposure.
3. Balance of benefits and harms taking into account the magnitude of the effects and the relative importance of the outcomes.
4. Resource implications which can be informed by a formal economic evaluation based on estimates collected during evidence retrieval and by modelling of cost–benefit and cost–effectiveness.
5. Priority of the problem pertaining to the burden of disease, prevalence, incidence or baseline risk.
6. Equity and human rights as the options given in a guideline can reduce or increase health inequities.
7. Acceptability of an intervention to the stakeholders. Acceptability is affected by several factors, such as who benefits from an intervention and who is harmed by it; who pays for it or saves money on account of it; and when the benefits, harms and costs occur.
8. Feasibility which is influenced by the resources available, programmatic considerations, the existing and the necessary infrastructure and training, and many other factors.

Of all the factors, questions of costs and resources at the global level are particularly challenging, again highlighting the need for transparent review of the data and key issues, so that countries may make their own decisions and prioritize health interventions. WGs should also consider the social perspective when evaluating cost and resource implications. Social values are critical factors that have a strong impact on the vaccine policy decisions, such as timing of vaccination, whether it is mandated, number of doses for optimal protection and the goals of a programme. The decision to implement a programme will always entail trade-offs, which must be carefully reviewed at the national level prior to adoption of recommendations.

Proposed recommendations are only made after careful review and judgement of the evidence, risk-benefit ratio, values and feasibility.

3.7 Presentation of proposed recommendations to SAGE along with the supporting evidence

SAGE receives regular updates throughout the WG progress. After finalization of the review of evidence and formulation of the proposed recommendations, WG Chairs, who are required to be SAGE members, or their delegates formally present to SAGE at one of their bi-annual SAGE meetings. For each recommendation proposed by WGs, a written rationale should be provided with supporting evidence (for an example, see the SAGE WG Background Paper on Yellow Fever Vaccine) along with the important underlying considerations. WG summary reports, detailing recommendations and their underlying rationale are provided in advance of SAGE meetings.

The format for how data and their synthesis are provided and presented to the WG will depend upon the terms of reference for the WG. In addition to the point-by-point recommendations and justifications and evidence to recommendations tables for critical recommendations/questions, additional background materials may often be appropriate. In general, when providing evidence in support of recommendations for new vaccines, an in-depth background paper should be provided to SAGE. For many recommendations, an appropriate format for displaying the evidence may include following major categories: disease epidemiology; clinical characteristics; vaccine and immunization characteristics; economic considerations; health-system opportunities; and existence of, and interaction with, other existing intervention and control strategies (see Appendix 1). Consideration is given to cost-effectiveness by benchmarking the opportunity costs against other vaccines and/or other relevant health interventions at local level. Particular attention is given to affordability and opportunity costs. The amount of information and level of detail presented will depend on the topic at hand.

The Evidence to Recommendation tables depicts how the factors that determine the direction and strength of a recommendation inform the process of developing the recommendation. These tables enhance the transparency of the process, focus the discussions and permit recording of the judgments made about each factor and how each one contributed to the recommendation.

3.8 SAGE discussion, deliberation and ultimate decision regarding the proposed recommendations to WHO

Prior to the SAGE meetings, SAGE members will have received updates from WGs, meeting minutes from all teleconferences and in person meetings, and background materials important to the deliberations. During the SAGE session on the topic at hand, SAGE members will discuss and deliberate upon WG proposed recommendations in the open forum of a SAGE meeting. SAGE members may adopt WG proposed recommendations or make necessary adjustments.

The language used to convey recommendations should be unambiguous and only strong recommendations are provided. Recommendations encompass following formulations: is recommended, is not recommended, recommend against, should or should not. Recommendations can be conditional, i.e. recommendations targeted at restricted groups or circumstances (after having met certain pre-specified criteria).

At times SAGE may issue a Good Practice Statement. These represent recommendations that SAGE feels are important, but that are not appropriate for formal ratings of quality of evidence. Good Practice Statements characteristically represent situations in which a large and compelling body of evidence, made up of triangulated evidence including several indirect comparisons, strongly supports the net benefit of the recommended action.
Good Practice Statements are practical when SAGE has confidence that the benefits of the recommended vaccine/intervention clearly outweigh the harms, but where evidence may be triangulated, difficult or resource-intensive to find or synthesize, and when the potential to cause harms is low. For further information on Good Practice Statements, see Appendix 11.

SAGE adopts recommendations by consensus and transmits these to WHO. A summary of the deliberations are published after the meeting in the WHO Weekly Epidemiological Record. The full recommendations are then incorporated into WHO vaccine position papers published on the WHO SAGE website.

More information on SAGE and its role in policy development is available on the SAGE website or in Duclos et al.[15]

[15] Duclos P, Okwo-Bele JM, Salisbury D. Establishing global policy recommendations and achieving global goals: the Strategic Advisory Group of Experts on immunization. *Expert Review of Vaccines*, 2011, 10(2):163–173.

4. Conclusions

Providing policy guidance on the use of vaccines and on immunization-related topics in different geographic and cultural contexts is a challenging but important public health endeavour that must have its foundation in the highest quality scientific evidence available.

The approach described above represents a consensus of a range of immunization experts on how best to apply a rigorous approach to evaluating the quality of scientific evidence. As judgement will always be necessary in policy development, transparency is required throughout the process.

This guidance document is intended to increase transparency and standardization of the development of WHO vaccine and immunization recommendations, and will continue to be regularly fine-tuned as improvements are identified and as the methodology evolves.

SAGE fosters continuing dialogue with experts in the field of public health methodology and the GRADE working group. Based on these interactions as well as on experience with the outlined processes, SAGE aims to a further optimization of its evidence review process.

5. Appendices

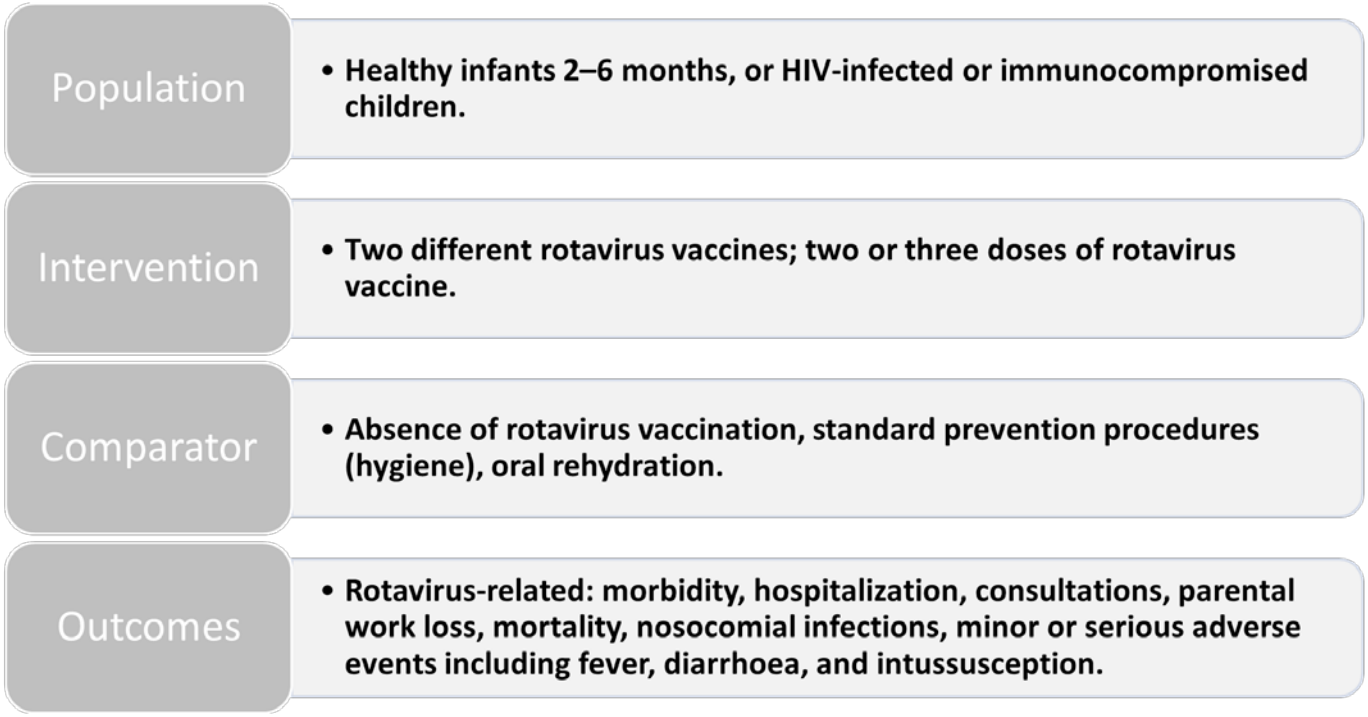
Appendix 1. Specific factors which underpin the development of SAGE recommendations [16]

Main factors	Specific elements
Epidemiologic features of the disease	-disease burden, including age specific mortality, morbidity, and social impact. -specific risk groups. -epidemic potential. -disease occurrence over time (i.e. secular trends). -serogroup or serotype distribution (for serogroup or serotype specific vaccines). -changes in epidemiological features over time.
Clinical characteristics of the targeted disease	-clinical management. -disease severity and fatality. -primary/secondary/tertiary care implications. -long-term complications and medical care requirements.
Other options for disease control and prevention	-existence of other prevention and control options.
Vaccine and immunization characteristics	-efficacy. -effectiveness and population impact of the vaccine (including herd immunity). -safety (serious adverse events and reactogenicity profile). -indirect effects (potential impact on strain selection, herd immunity, potential safety concerns of live attenuated vaccines in contacts of vaccines, serotype replacement). -cold chain and logistical concerns. -vaccine availability. -vaccine schedule(s). -social and programmatic acceptability of the schedule(s). -ability to reach the target populations. -ability to monitor programme impact.
Economic considerations	-cost of illness. -vaccine and vaccine delivery costs. -potential for vaccine price reductions. -cost-effectiveness of immunization programmes. -affordability of immunization.
Health system considerations	-possible interactions with other interventions and control strategies. -possible impact of vaccine introduction on the wider health system.
Social impacts	-possible impact on social equality and inequality.
Legal considerations	-possible legal requirements for implementation.
Ethical considerations	-possible ethical considerations.

[16] Further factors to be considered are reflected in the Evidence to Recommendation table.

Appendix 2. PICO elements for assessment of rotavirus vaccine

To illustrate the application of the framing in the SAGE process, the PICO elements for an evidence review for general rotavirus vaccine recommendations are provided here:



Appendix 3: Core PICO questions and example of PICO question

Efficacy/Effectiveness

Population: Specific population of interest (+/- age group, gender, immune status, geography, previous vaccination, etc.)
Intervention: ____ vaccination (+/- number of doses, etc.)
Comparison: No ____ vaccination (+/- alternative schedule, existing vaccine, etc.)
Outcome: Occurrence of ____ disease (+/- specific outcomes, mortality, etc.)

Safety

Population: Specific population of interest (+/- age group, gender, immune status, geography, previous vaccination etc.)
Intervention: ____ vaccination (+/- number of doses, etc.)
Comparison: No ____ vaccination (+/- alternative schedule, existing vaccine, etc.)
Outcome: Serious adverse events (+/- specific adverse events)

Duration of Protection

Population: Specific population of interest (+/- age group, gender, immune status, geography, previous vaccination etc.)
Intervention: ____ vaccination (+/- number of doses, etc.)
Comparison: No ____ vaccination (+/- alternative schedule, existing vaccine, etc.)
Outcome: Duration of immunity (e.g. +/- specific immunological or clinical endpoints)

Appendix 4. Draft summary table for evidence review

Study authors	Year	Location	Study population	Vaccination/ intervention	Methods	Limitations/ potential sources of bias	Relevant outcomes, measure of association and impact/effect	Comments

Appendix 5. Data items to consider for extraction from included studies

Data extraction forms should be tailored for each systematic review. The data items below represent key fields to consider including, when appropriate, in the data extraction form.

1. Study author, year **2. Name of reviewer** **3. Date of review** **4. Methods** 4.1. Study design 4.2. Source of sample(s) 4.3. Sampling method 4.4. Sample size 4.5. Entry criteria/exclusions 4.6. Non-respondents/loss to follow up 4.7. Which parts of the study were prospective **5. Participants** 5.1. Setting 5.2. Country 5.3. Age (range and mean/median) 5.4. Gender (% male/female) 5.5. Ethnicity 5.6. Control group 5.7. Definition of controls 5.8. Source of controls 5.9. Comparability 5.9.1. Potential confounders identified 5.9.2. Baseline assessment of outcome variables	**6. Group allocation** 6.1. Randomization 6.1.1. Sequence generation 6.1.2. Allocation sequence concealment 6.1.3. Blinding 6.2. Allocation by 6.2.1. Quasi-randomization 6.2.2. Time differences 6.2.3. Location differences 6.2.4. Treatment decisions 6.2.5. Participants' preferences 6.2.6. On the basis of outcome 6.2.7. Other important processes **7. Intervention** 7.1. Vaccine (formulation, dose, etc.) 7.2. Length of follow-up **8. Outcomes** 8.1. How defined 8.2. Intervals at which outcomes were assessed 8.3. Validity 8.4. Reproducibility 8.5. Quality control 8.6. Missing/incomplete data 8.7. Selective reporting 9. **Summary of results** 10. **Summary of possible risks of bias** 10.1. Selection bias 10.2. Information bias 10.3. Confounding

Appendix 6. Checklists for reviewing study quality

There are many quality appraisal tools available to assess for risk of bias in various study designs. SAGE does not take a prescriptive approach and has used a variety of tools/checklists in its past work.

GRADE Working Group

Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise

Critical Appraisal Skills Programme (CASP)

Checklist for RCTs

Checklist for systematic reviews

Checklist for case-control studies

Checklist for cohort studies

EPOC (Cochrane)

Checklist for RCTs/NRCTs/controlled before-after studies/interrupted time series

ROBINS-I tool

Tool to evaluate the risk of bias in the results of non-randomized studies of interventions (NRSI) that compare the health effects of two or more interventions.

NICE

Checklist for randomized controlled trials

Checklist for case-control studies

Checklist for cohort studies

Checklist for systematic reviews and meta-analysis

SIGN

Critical appraisal checklists (systematic reviews and meta-analysis, RCTs, cohort and case control studies, diagnostic and economic studies)

AMSTAR

Measurement tool to assess the methodological quality of systematic reviews

Meta-Analyses of observational studies in epidemiology

http://jama.ama-assn.org/content/283/15/2008.full

Consolidated standards of reporting trials

http://www.consort-statement.org/.

Strengthening the reporting of observational studies in epidemiology

http://www.strobe-statement.org/index.php?id=available-checklists.

Preferred reporting items for systematic reviews and meta-analyses

http://www.prisma-statement.org/.

Appendix 7. Rating the quality of the evidence

Quality of evidence	**Quality starting factor is first assigned base on study design**	**Quality rating is lowered[17] if**	**Quality rating is raised[36] if**
Evidence supports a high level of confidence that the true effect lies close to that of the estimate of the effect on the health outcome. (4)	Randomized trials	**1)Limitation of design:[18]** *-1 Serious* *-2 Very serious* ***2)*Inconsistency:** *-1 Serious* *-2 Very serious* **3)Indirectness:[37]** *-1 Serious* *-2 Very serious* **4)Imprecision:** *-1 Serious* *-2 Very serious* **5)Publication bias:** *-1 Likely* *-2 Very likely*	**1)Strength of association[19]:** *+1 RR or OR[20]>=2 (or =<0.5)* *+2 RR or OR >=5 (or =<0.2)* **2)Dose response (population based):** *+1 Evidence of decreased risk with increased vaccine coverage, including evidence of reversal at population level* *+2 Very strong evidence of decreased risk with increased coverage, including evidence of reversal at population level* ***3)Mitigated bias and confounding:*** *+1 All major confounders would have reduced the effect or increased the effect if no effect was observed* *or +1 ability of design to control for confounding and avoid biases* *+2 If in addition to design, consistency across different settings, different investigators, and possibly different designs*
Evidence supports a moderate level of confidence that the true effect lies close to that of the estimate of the effect on the health outcome.(3)			
Evidence supports a limited level of confidence that the true effect lies close to that of the estimate of the effect on the health outcome. (2)	Observational studies including disease surveillance and post-market safety surveillance data		
Evidence supports a very low level of confidence that the true effect lies close to that of the estimate of the effect on the health outcome. (1)			

[17] 1= move up or down one level (for example from high (4) to intermediate (3). 2= move up or down two levels (for example from low (2) to high (4)).

[18] Should be commensurate with study design.

[19] GRADE refers to it as large effect.

[20] These thresholds refer to risk ratios. When baseline risk is low i.e. below 20%, odds and risk ratios are very similar and one can comfortably apply this criteria. When the baseline risk is high (a rare occurrence for vaccine preventable diseases) and the effect size is large, OR can be far larger in magnitude than risk ratios. Under such circumstances a higher threshold for ORs may be appropriate

Appendix 8. GRADE levels of evidence scale terminology

Original GRADE[21] levels of evidence scale wording:

High = We are **very confident** that the true effect lies close to that of the estimate of the effect Moderate = We are **moderately confident** in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different. Low = Our **confidence** in the effect estimate **is limited**: The true effect may be substantially different from the estimate of the effect. Very low = We have **very little confidence** in the effect estimate: The true effect is likely to be substantially different from the estimate of effect.

Prior SAGE guidance document wording:

High = We are **very confident** that the true effect lies close to that of the estimate of effect on health outcome (level 4, or ⊕⊕⊕⊕). Moderate = We are **moderately confident** in the estimate of effect on health outcome. The true effect is likely to be close to the estimate of the effect (level 3, or ⊕⊕⊕). Low = Our **confidence** in the estimate of the effect on the health outcome **is limited** (level 2, or ⊕⊕). Very low = We have **very little confidence** in the estimate of the effect on the health outcome (level 1, or ⊕).

Current SAGE guidance document wording:

High = Evidence supports a **high level of confidence** that the true effect lies close to that of the estimate of the effect on the health outcome (level 4, or ⊕⊕⊕⊕). Moderate = Evidence supports a **moderate level of confidence** that the true effect lies close to that of the estimate of the effect on the health outcome (level 3, or ⊕⊕⊕). Low = Evidence supports a **limited level of confidence** that the true effect lies close to that of the estimate of the effect on the health outcome (level 2, or ⊕⊕). Very low = Evidence supports a **very low level of confidence** that the true effect lies close to that of the estimate of the effect on the health outcome (level 1, or ⊕).

[21] GRADE guidelines: 3. Rating the quality of evidence. Howard Balshem, Mark Helfand, Holger J. Schünemann, Andrew D. Oxman, Regina Kunz, Jan Brozek, Gunn E. Vist, Yngve Falck-Ytter, Joerg Meerpohl, Susan Norris, Gordon H. Guyatt. Journal of clinical epidemiology 1 April 2011 (volume 64 issue 4 Pages 401-406 DOI: 10.1016/j.jclinepi.2010.07.015)

Appendix 9a. Template of a GRADE table used to rate the quality of evidence

Different study designs may be graded separately in different tables (e.g. RCT and observational studies), or only the highest quality design used while including consideration of other sources of evidence through footnotes and adjusting the rating as appropriate.

Question necessary for recommendation development:				
			Rating	Adjustment to rating
Quality Assessment	No of studies/starting rating			
	Factors decreasing confidence	Limitation in study design		
		Inconsistency		
		Indirectness		
		Imprecision		
		Publication bias		
	Factors increasing confidence	Strength of association		
		Dose-response		
		Mitigated bias and confounding		
	Final numerical rating of quality of evidence			
Summary of Findings	**Statement on quality of evidence**			
	Conclusion			

Appendix 9b. Sample completed GRADE table

Level of clinical protection conferred by a complete primary series of *Haemophilus influenzae* type b (Hib) vaccination

Population : Immunocompetent individuals
Intervention: Complete primary series of Hib vaccination (≥2 doses)
Comparison: No vaccination
Outcome : Hib meningitis

PICO Question: What is the level of clinical protection conferred by a complete primary series of Hib vaccination (≥2 doses) in preventing Hib meningitis in immunocompetent individuals?				
			Rating	Adjustment to rating
Quality Assessment	No of studies/starting rating		2 RCT/ 6 observational[22]	4
	Factors decreasing confidence	Limitation in study design	Serious[23]	-1
		Inconsistency	None serious	0
		Indirectness	None serious	0
		Imprecision	None serious	0
		Publication bias	None detected	0
	Factors increasing confidence	Strength of association/ large effect	High[24]	+1
		Dose-response	Not applicable	0
		Antagonistic /mitigated bias and confounding	Not applicable	0
	Final numerical rating of quality of evidence			**4**
Summary of Findings	**Statement on quality of evidence**			We are very confident that the true effect lies close to that of the estimate of effect on health outcome
	Conclusion			We are very confident that a primary series of vaccination against Hib (≥2 doses) confers high levels of clinical protection against Hib meningitis. Vaccine efficacy ranged from 67%-95% and effectiveness from 65%-99%.

[22] Evidence retrieved from two systematic reviews (Jackson et al.2013; Low et al.2013). For 2p+0 vs 0 doses, intention to treat (ITT) vaccine efficacy was 96% (95%CI 37-100%) against Hib meningitis (Santosham 1991). ITT efficacy of a 3p vs 0 dose schedule against meningitis was calculated to be 67% (95%CI 22-86%) (Mulholland 1997). Observational studies confirm vaccine effectiveness against Hib meningitis ranging from 65% (95%CI -190-100%) (Baqui 2007) to 99% (95%CI 92-100%) (Lee 2008) after two or more doses.

[23] Unclear allocation concealment and blinding of participants in the larger of the two RCTs (21490 participants) (Mulholland 1997)

[24] Evidence from RCTs and observational studies suggest vaccine efficacy and effectiveness over 50%

Reference List

1. Jackson C, et al. Effectiveness of *Haemophilus influenzae* type b Vaccines Administered According to Various Schedules: Systematic Review and Meta-Analysis of Observational Data. *Pediatr Infect Dis J,* 2013 Jul 9.

2. Low N, et al. Comparing *Haemophilus influenzae* type b Conjugate Vaccine Schedules: A Systematic Review and Meta-Analysis of Vaccine Trials. *Pediatr Infect Dis J* 2013 Jul 9

3. Adegbola RA et al. Elimination of *Haemophilus influenzae* type b (Hib) disease from The Gambia after the introduction of routine immunisation with a Hib conjugate vaccine: a prospective study. *The Lancet*, 2005. 366(9480): 144-150.

4. Baqui AH et al. Effectiveness of *Haemophilus influenzae* type b conjugate vaccine on prevention of pneumonia and meningitis in Bangladeshi children: a case-control study. *Pediatric Infectious Disease Journal*, 2007. 26(7) 565-571.

5. Daza P et al. The impact of routine infant immunization with *Haemophilus influenzae* type b conjugate vaccine in Malawi, a country with high human immunodeficiency virus prevalence. *Vaccine*, 2006. 24:6232-6239.

6. Lee EH et al. Impact of *Haemophilus influenzae* type b conjugate vaccine on bacterial meningitis in the Dominican Republic. *Rev Panam Salud Publica*, 2008. 24(3):161-168

7. Lee EH et al. *Haemophilus influenzae* type b conjugate vaccine is highly effective in the Ugandan routine immunization program: a case-control study. *Tropical Medicine & International Health*, 2008. 13(4): 495-502.

8. Lewis RF et al. Action for child survival: elimination of *Haemophilus influenzae* type b meningitis in Uganda. *Bulletin of the World Health Organization*, 2008. 86(4): 292-301

9. Mulholland K et al. Randomised trial of *Haemophilus influenzae* type b tetanus protein conjugate vaccine [corrected] for prevention of pneumonia and meningitis in Gambian infants. *The Lancet*, 1997. 349(9060): p. 1191-7

10. Santosham M et al. The efficacy in Navajo infants of a conjugate vaccine consisting of *Haemophilus influenzae* type b polysaccharide and Neisseria meningitidis outer-membrane protein complex. *The New England Journal of Medicine*, 1991. 324(25): 1767-1772.

Appendix 10. Sample SAGE Evidence to Recommendation table

SAGE evidence to recommendations table[1]

Question: Overarching policy question to be answered by the guideline panel (SAGE) using the evidence to recommendations framework. The question should be very precise and should identify among other, the specific intervention, comparison and outcome as well as the target population as well as the setting (global; low/medium/high income countries; specific subpopulations).
Example: Should one dose of varicella-containing vaccine be recommended, over no vaccination or available treatment- and prevention options, to be administered to immunocompetent children (<6years of age) globally to mitigate burden of severe varicella disease.
Population: Target population for vaccine (age-range, sex, immune-status) *Example: Immunocompetent children (<6years of age)*
Intervention: Vaccination (if applicable dosage and schedule) *Example: One dose of Varicella-containing vaccine*
Comparison(s): No Vaccination/ Placebo/Control/ Standard care/Other prevention options *Example: No vaccination, control vaccine, acyclovir, Varicella-zoster immunoglobulin (VZIG)*
Outcome: Outcome(s) to be prevented by vaccine *Example: Severe varicella with 500 or more lesions or any complications such as bacterial superinfection, varicella pneumonitis, encephalitis, hospitalization, or death.*

When available, please refer to background papers on the underlying evidence. *Example: More evidence that was made available to SAGE to support their recommendations on the use of varicella vaccine can be found in the background paper on the WHO website.*

Background:

The addressed PICO question should be described in details and the important public health background information for the understanding of the question and why a recommendation or decision is need, should be briefly provided.

Example: Varicella (chickenpox) is an acute, highly contagious viral disease with worldwide distribution. While mostly a mild disorder in childhood, varicella tends to be more severe in adults. It may be fatal, especially in neonates and in immunocompromised persons. Following infection, the virus remains latent in neural ganglia, and upon subsequent reactivation Varicella Zoster Virus (VZV) may cause herpes zoster disease (shingles), a painful vesicular rash with dermatomal distribution mainly affecting the elderly and immunocompromised persons. In about 10%-20% of the cases, varicella is followed later in life by herpes zoster after the age of 50 or in immunocompromised persons.

Although individual cases may be prevented or modified by varicella-zoster immune globulin or treated with antiviral drugs, control of varicella can be achieved only by widespread vaccination. Decision-makers considering the use of varicella vaccine in routine immunization programs must take into account the epidemiology and the public health and socioeconomic impact of varicella relative to other health concerns competing for scarce resources.

	CRITERIA	JUDGEMENTS	RESEARCH EVIDENCE	ADDITIONAL INFORMATION
PROBLEM	Is the problem a public health priority?	*No* ☐ *Uncertain* ☐ *Yes* ☐ *Varies by setting* ☐	Provide available scientific evidence on burden of disease, preferably within the target population the recommendations are aimed at. If no evidence available, provide expert judgment on the public health priority considerations. *Example: In temperate climates of the northern hemisphere the majority of adults are seropositive when tested. The epidemiology is less well understood in tropical areas, where a relatively large proportion of adults in some countries are seronegative. Before*	Provide additional information on public health priority considerations. *Example: So far, several industrialized countries have introduced varicella vaccines into their routine immunization programs. The priority varies between countries. Especially in countries with few resources, varicella vaccination competes with other public health priorities.*

			the varicella vaccination program in US, approximately 4 million varicella cases occurred each year, resulting in 10600 hospitalizations and 100 deaths. The majority of cases occurred in children (Seward 2004). Before introduction of vaccination into routine immunization schedule in 2004, Germany reported 760 000 new cases in a population of 82 million in 1999 (Wagenpfeil 2004). *In Australia, more than 240,000 cases, 1,500 hospitalizations and 4 fatalities have been estimated to occur annually due to infection with varicella zoster virus (Vally 2007). In Spain, Varicella causes approximately 400,000 cases, 1500 hospitalizations and 15 deaths every year (Lenne 2006). In Canada, from year 1994 to 2000 showed that over 1,550 varicella hospitalizations occur annually for all age groups. For the most recent period, 1999 to 2009, a total of 2,297 pediatric varicella related hospitalizations were reported, averaging 208 hospitalizations annually for children (PHAC,2012).*	
BENEFITS & HARMS	Benefits of the intervention Are the desirable	*No* ☐ *Uncertain* ☐ *Yes* ☐ *Varies* ☐	How large are the beneficial effects of the intervention on individual (vaccine effectiveness) and population level (herd immunity).	To take into consideration: Is the baseline benefit similar across subgroups (age, gender, pregnancy, lactation, healthcare

	anticipated effects large?		*Example: Pooled one-dose VE for prevention of combined moderate and severe varicella: 98% (95% CI: 97%-99%., based on "Varicella Vaccine Effectiveness Worldwide: a Systematic Review and Meta-Analysis" (Marin et al. under review)* *On the population level, in countries, where varicella vaccination is already recommended, a decline in incidence in all age groups over time, not only age-group targeted by vaccination program, suggests induction of community protection (Marin et al 2008, Marin et al 2011, Lopez et al 2011, Guris et al 2008).*	workers, immunostatus, disability, race, and SES, and other groups such as refugees and asylum seekers)? *Example: Significant risk exists for nosocomial varicella when susceptible health care personnel are exposed to VZV with transmission to vulnerable populations.* Should there be separate recommendations for subgroups based on benefit or disease severity levels? *Example: Yes, recommendations on varicella immunization targeted at health care workers*
	<u>Harms of the intervention</u> Are the undesirable anticipated effects small?	*No* ☐ *Uncertain* ☐ *Yes* ☐ *<u>Varies</u>* ☐	Are there deleterious effects of the intervention, either on the individual ((serious) adverse events following immunization) or on the population level (age-shift of disease, serotype replacement, etc.). *Example:* *On the individual basis, monovalent varicella vaccine was well tolerated during pre-licensure clinical trials. This was confirmed by postlicensure data.*	To take into consideration: Is the baseline risk for harm similar across subgroups (see above)? Should there be separate recommendations for subgroups based on harms? *Example: A review of data reported after 16 years of pregnancy registry for VZV-containing vaccines that follows up pregnancy outcomes in women who inadvertently received Merck*

			On a population basis, an undesired effect of a routine varicella vaccine program in children may be increased morbidity and mortality due to a shift of varicella infection to a higher age group if high vaccination coverage ≥80-90% cannot be achieved. Further, re-exposure to circulating varicella disease can inhibit varicella reactivation and consequently also decrease the incidence of herpes zoster in immune individuals. Decrease of varicella circulation due to childhood varicella immunization programs may temporarily increase the incidence of herpes zoster. (Report of the Varicella Working Group to SAGE)	*varicella vaccine during pregnancy showed no congenital varicella syndrome among 157 live born infants of seronegative women (Rate=0 per 100, 95% CI 0.0, 2.4) or in the overall registry (735 live births) (Merck/CDC Pregnancy Registry for Varicella-Zoster Virus-Containing Vaccines. The 16th Annual Report, 2011).* *Vaccination in populations of patients who are immunosuppressed or are receiving or have received medications that may be immunosuppressive should be considered.*
	Balance between benefits and harms	*Favours intervention* ☐ *Favours comparison* ☐ *Favours both* ☐ *Favours neither* ☐ *Unclear* ☐	Please balance benefits of the intervention with possible harms (this applies to individual and population level). *Example: Modeling indicated that for countries, at vaccine coverage levels of <30% and ≥80% very little risk exists of increased morbidity due to shifts in the age of infection.*	

			However, at moderate coverage levels (30%-70%), there may be a risk of increased morbidity due to shifts in the age at infection.	
	What is the overall quality of this evidence for the critical outcomes?	Effectiveness of the intervention *No included studies* ☐ *Very low* ☐ *Low* ☐ *Moderate* ☐ *High* ☐ Safety of the intervention *No included studies* ☐ *Very low* ☐ *Low* ☐ *Moderate* ☐ *High* ☐	Please provide GRADE (safety and effectiveness) tables with respective rating of the intervention. For more information (e.g. upgrading of effectiveness), please see the SAGE Guidance document. *Example: GRADE high quality evidence for the critical outcomes of severe varicella and serious adverse events.*	

VALUES & PREFERENCES	How certain is the relative importance of the desirable and undesirable outcomes?	Important uncertainty or variability ☐ Possibly important uncertainty or variability ☐ Probably no important uncertainty or variability ☐ No important uncertainty or variability ☐ No known undesirable outcomes ☐	Please provide information on the relative importance the target population attributes to the desirable and the undesirable outcomes related to the intervention as well as the comparison. *Example:* *The relative importance or values of the main outcomes of interest:* *Outcome* / *Relative importance* / *Certainty of the evidence* *Severe varicella disease* / *No evidence* / *-*	*Example: For some parents/caregivers severe varicella disease may be acceptable whereas other parents/ caregivers might consider important to avoid (due to incremental costs, avoidance to not put the child at risk of severe disease,..)*

	Values and preferences of the target population: Are the desirable effects large relative to undesirable effects? *No* *Probably No* *Uncertain* *Probably Yes* *Yes* | *Varies* ☐ ☐ ☐ ☐ ☐ ☐	Provide evidence on target population value& preferences related to intervention as well as the comparative health outcomes. Describe the source (consultations with populations, direct and indirect research). Is there uncertainty or variability in the preference target groups attribute to the harms and benefit outcomes? When no evidence available, provide transparent reflection by guideline panel on this matter. *Example: No evidence was retrieved on the values and preferences of the target population, but it would be assumed that on the individual's level, avoidance of varicella related disease would outweigh the undesirable effected related to immunization (pain during immunization, AEFIs). Herd protection would further reduce the risk of those who cannot receive the intervention.*	Are the benefits, harms and costs of the intervention valued differently by disadvantaged populations compared to the privileged populations? All critical outcomes relevant measured? Source of variability, if any: Methods for determining values satisfactory for this recommendation? If no, conducting systematic assessment of values and preferences of target group needs to be considered and could be done using qualitative research. If target group doesn't value the intervention or attributes little to the harms and benefits outcomes- are advocacy measures needed?

RESOURCE USE	Are the resources required small?	*No* ☐ *Uncertain* ☐ *Yes* ☐ *Varies* ☐	Provide data on interventions costs as well as programmatic costs (e.g. employing/ training of health care workers; supply chain expenses). *Example: Price range of varicella vaccine is dependent on country, from 10 US$ to 150 US$. The cost of varicella vaccine compared to other vaccines in the routine immunization program is high. High income countries might be able to allocated resources for this vaccine, whereas most middle and low income countries without external resources most likely will not be able to afford to include the vaccine in their routine immunization program.* *In terms of additional resources, vaccine could be delivered within EPI schedule at time of MMR vaccination, no additional visits needed. Stress on the supply chain needs to be assessed.*	Opportunity cost: Is this intervention and its effects worth withdrawing or not allocating resources from other interventions?
	Cost-effectiveness	*No* ☐ *Uncertain* ☐ *Yes* ☐ *Varies* ☐	Provide cost-effectiveness data of the intervention in the target population. *Example: Cost-effectiveness of varicella childhood immunization was predominantly assessed in high income countries with consistent results:*	

			– Cost saving (or cost-effective) under the societal perspective *– Cost-effective under the health payer perspective when excluding potential impact on zoster* *– Cost-ineffective (or increased morbidity) when including potential impact on zoster* *Cost-effectiveness in low- and middle income countries still needs to be assessed. (Thiry Pharmacoeconomics 2003; Rozembaum Expert Reviews 2008; Valentim Vaccine 2008; Bonnani Vaccine 2008; van Hoek Vaccine 2012)*	

EQUITY	What would be the impact on health inequities?	*Increased* ☐ *Uncertain* ☐ *Reduced* ☐ Varies ☐	Questions to Determine Equity Focus: What would be the impact of this recommendation on health equity? • Is the condition more common in certain disadvantaged group? or • Is its severity greater, in people from specific group or with a particular disability? Is there a risk that discrimination could impact outcomes? Are there significant differences resulting in varying levels of access to intervention or coverage levels? *Example: Compared to temperate climate, in tropical climate the majority of studies have described later acquisition of varicella in childhood with a higher proportion of cases and higher susceptibility among adults. These features lead to a higher overall mean age at varicella infection compared with temperate climates and associated higher morbidity.* *Further, treatment options may be reduced or non-existent in low-income settings.*	

ACCEPTABILITY	Which option is acceptable to key stakeholders (Ministries of Health, Immunization Managers)?	*Intervention* ☐ *Comparison* ☐ *Both* ☐ *Neither* ☐ *Unclear* ☐	Assessment, whether intervention would be acceptable to stakeholders (ethically, programmatically, financially, etc.) *Example: Countries should assess whether adequate resources can be allocated to implement and sustain varicella vaccination in the routine immunization schedule. This especially applies to low and middle income countries with limited resources, where varicella vaccination might be competing with other important public health interventions.*	
	Which option is acceptable to target group?	*Intervention* ☐ *Comparison* ☐ *Both* ☐ *Neither* ☐ *Unclear* ☐	Assessment, whether intervention would be acceptable to target group (ethically, religious, related to opportunity costs, financially, etc.) *Example: It is presumed that the option would be acceptable to the target group if no additional visit at the health clinic is needed and the costs are covered by the health care provider.*	

FEASIBILITY	Is the intervention feasible to implement?	*No* ☐ *Probably No* ☐ *Uncertain* ☐ *Probably Yes* ☐ *Yes* ☐ *Varies* ☐	Feasibility: Is this intervention accessible, acceptable to target groups and providers and affordable to disadvantaged as well as advantaged populations? Providers: Are programmatic issues considered (e.g. costs related to health care worker's training and employment, logistic/cold-chain considerations), etc. Target population: Opportunity costs (e.g. additional visits to health care clinic), community demand, etc. *Example: The option is feasible to high income countries rather than to middle and lower income countries due to financial constraints and other public health priorities.*	To take into consideration: Is there a risk of discrimination? Analysis of opportunity cost of equity. Equity weighing of health outcomes. Is there variability in resource requirements and feasibility across settings and populations? Is there a need for additional recommendations?

Balance of consequences	Undesirable consequences *clearly outweigh* desirable consequences in most settings ☐	Undesirable consequences *probably outweigh* desirable consequences in most settings ☐	The balance between desirable and undesirable consequences *is closely balanced or uncertain* ☐	Desirable consequences *probably outweigh* undesirable consequences in most settings ☐	Desirable consequences *clearly outweigh* undesirable consequences in most settings ☐
Type of recommendation	We recommend the intervention ☐	We suggest considering recommendation of the intervention ☐ Only in the context of rigorous research ☐ Only with targeted monitoring and evaluation ☐ Only in specific contexts or specific (sub)populations		We recommend the comparison ☐	We recommend against the intervention and the comparison ☐
Recommendation (text)	Please provide the draft recommendations proposed to SAGE. *Example: There is strong scientific evidence that varicella vaccine is safe and effective in preventing varicella related morbidity and mortality in immunocompetent individuals. WHO recommended that routine childhood immunization against varicella could be considered in countries where the disease has an important public health impact. Resources should be sufficient to support sustained vaccine coverage ⩾80%. Settings where varicella vaccine coverage levels are less than 80% are at risk of an increase of severe disease and mortality in adults.* *Those countries deciding to introduce routine childhood varicella immunization, should administer vaccination at 12-18 months of age. The number of doses administered is dependent on the goal of the vaccination program. One dose is sufficient to reduce mortality and severe morbidity from varicella. Two doses induce higher effectiveness and should therefore be recommended in countries where the programmatic goal is, in addition to decreasing mortality and severe morbidity, to further reduce the number of cases and outbreaks.* *Due to the increase in severity of varicella in immunocompromised, certain groups of immunocompromised should be considered for VZV vaccination. Limited data on the immunization of health care workers (HCW) are available, yet countries should consider vaccination of non-immunized health care workers without a history of varicella with two doses of varicella vaccine, even in absence of varicella vaccination in the routine immunization schedule, if the risk of severe varicella in the population in direct contact with the HCW is high (e.g. immunocompromised).*				

Implementation considerations	Please consider aspects related to implementation (communication, advocacy, etc.) *Example: Recommendations will be made available in the standard WHO format (WHO position paper).*
Monitoring and evaluation	Please outline monitoring and evaluation considerations *Example: Monitoring of immunization coverage and disease surveillance*
Research prioritie s	Please outline research priorities *Example:* • *long-term duration of protection of varicella vaccine* • *effect of vaccination in immunocompromised populations* • *the consequences of reactivation of vaccine virus later in life need to be assessed, in particular reactivation after acquired natural immunity and in immunosuppressed individuals and impact of vaccination programs on herpes zoster incidence*

[1] This Evidence to Recommendation table is based on the DECIDE Work Package 5: Strategies for communicating evidence to inform decisions about health system and public health interventions. Evidence to a recommendation (for use by a guideline panel). http://www.decide-collaboration.eu/, accessed January 2017.

Appendix 11. Strategic Advisory Group of Experts (SAGE) on Immunization principles on the use of Good Practice Statements.

Formulation of recommendations based on GRADE

The GRADE approach creates an evidence profile for systematic review of literature with an underlying PICO (population, intervention, comparison and outcomes) question. Outcomes for interventions should be prioritized *a priori*, and then the recommendations can be developed based on the certainty of the balance of benefits and harms, and evidence on patient values and preferences, feasibility, acceptability, and costs.

When SAGE develops recommendations on vaccine-specific topics, the issues that will generally require systematic review of literature, appraisal of evidence and formal GRADE scoring are (1) vaccine efficacy/effectiveness, (2) vaccine safety, if applicable including special risk groups (e.g. immunocompromised or pregnant), (3) duration of protection and (4) potential further questions critical for development of recommendations.

In general, direct evidence related will be graded. Nevertheless, in some instances e.g. when direct evidence is poor or lacking, indirect evidence can support the development of a recommendation and will be taken into consideration for formal GRADE scoring (e.g. immunogenicity data supporting the efficacy of a vaccine).

Further, there may be SAGE recommendations for which no formal grading is conducted although using GRADE may be feasible. These recommendations do not automatically qualify to be a Good Practice Statement, which are identified as such (see below).

Development of Good Practice Statements

At times it will make more sense to formulate a Good Practice Statement rather than the application of GRADE. Good Practice Statements represent recommendations that SAGE feels are important, but that are not appropriate for formal ratings of quality of evidence. SAGE aims to use Good Practice Statements sparingly and whenever warranted will aim for formal grading of the evidence.

Good Practice Statements characteristically represent situations in which a large and compelling body of indirect evidence, made up of linked evidence including several indirect comparisons, strongly supports the net benefit of the recommended action. Good Practice Statements are practical when SAGE has confidence that the benefits of the recommended vaccine/intervention clearly outweigh its harms, but where the evidence is linked i.e. representing several separate bodies of evidence that may be difficult or resource-intensive to find or synthesize. However, lack of resources or time alone is not sufficient to support the development of a Good Practice Statement instead of using GRADE.

A set of questions will guide SAGE in determining the need for a Good Practice Statement:

Guiding questions particular to good practice statements, modified from Guyatt et al 2016[25]:

(i) Is the message really necessary? [26]
(ii) After consideration of all relevant outcomes and potential downstream consequences, will implementing the good practice statement result in large net positive consequences?
(iii) Is collecting and summarizing the evidence a poor use of a guideline panel's limited time and energy (opportunity cost is large)?
(iv) Is there a well-documented clear and explicit rationale connecting the linked, indirect evidence? [27]

The answers to all questions (i) to (iv) should be yes, and the statement should be clear and actionable as for all recommendations, to proceed with a Good Practice Statement.
A Good Practice Statement will be a WHO recommendation; however, it would differ from a WHO recommendation using the formal application of GRADE. Good Practice Statements will be identifiable as such through thorough documentation of the responses to the guiding questions. These will be published as supporting document along with the specific WHO vaccine position paper.
SAGE will continuously monitor the applicability of the guiding questions to its Good Practice Statements and adjust these in the future should there be the need.
SAGE will give due consideration to neither overuse Good Practice Statements nor trivialize these for common sense policy statements. Where ample previously graded evidence on specific recommendations is available, there may not be the need to re-grade such statements nor label these as Good Practice Statements (Example: Benefit of education of health-care workers to ensure informed interactions with their patients).

[25] Guyatt GH, Alonso-Coello P, Schunemann HJ, et al. Guideline panels should seldom make good practice statements: guidance from the GRADE Working Group. J Clin Epidemiol 2016 Jul 22.
[26] Originial question: Is the message really necessary in regard to actual health care practice?
[27] Originial question: Is there a well-documented clear and explicit rationale connecting the indirect evidence?