Conference report

Using existing systematic reviews for developing vaccination recommendations: Results of an international expert workshop

Catherine L. Jo,[a] Helen Burchett,[b] Magdalena Bastías,[c] Pauline Campbell,[d] Deepa Gamage,[e] Louise Henaff,[f] Benjamin Kagina,[g] Carole Lunny,[h] Melanie Marti,[f] Rudzani Muloiwa,[g] Dawid Pieper,[i] James Thomas,[j] Matthew C. Tunis,[k] Ole Wichmann,[a] Zane Younger,[a] Thomas Harder[a]

[a] Robert Koch Institute, Seestrasse 10, 13353 Berlin, Germany; joc@rki.de, wichmanno@rki.de, hardert@rki.de

[b] London School of Hygiene and Tropical Medicine, 15-17 Tavistock Place, London WC1H 9SH, United Kingdom; helen.burchett@lshtm.ac.uk

[c] Comité Asesor en Vacunas y Estrategias de Inmunización (CAVEI), Ministerio de Salud, Monjitas 565, p7, Santiago, Chile; bastiasmalu@gmail.com

[d] Nursing, Midwifery and Allied Health Professions Research Unit, Glasgow Caledonian University, Govan Mbeki Building, Glasgow G4 0BA, United Kingdom; pauline.campbell@gcu.ac.uk

[e] Epidemiology Unit and Advisory Committee on Communicable Diseases, Ministry of Health, #231, De Saram Place, Colombo 10, Sri Lanka; deepagamage@gmail.com

[f] World Health Organization, Avenue Appia 20, 1211 Geneva, Switzerland; henaffl@who.int, martim@who.int

[g] University of Cape Town, Faculty of Health Sciences, Observatory, 7925, Cape Town, South Africa; benjamin.kagina@uct.ac.za, rudzani.muloiwa@uct.ac.za

[h] Cochrane Hypertension Review Group, University of British Columbia, 2176 Health Sciences Mall, Vancouver, BC Canada V6T1Z2; carole.lunny@ti.ubc.ca

[i] Witten/Herdecke University, Ostmerheimer Str. 200, Haus 38, 51109 Cologne, Germany; dawid.pieper@uni-wh.de

[j] Evidence for Policy and Practice Information and Co-ordinating (EPPI-) Centre, UCL Social Research Institute, University College London, 10 Woburn Square, London WC1H 0NR, UK; james.thomas@ucl.ac.uk

[k] Public Health Agency of Canada, Centre for Immunization and Respiratory Infectious Diseases, 130 Colonnade Road, A.L. 6501H, Ottawa, Ontario K1A 0K9, Canada; matthew.tunis@canada.ca

*Corresponding author: Dr. med. Thomas Harder, Robert Koch Institute, Seestrasse 10, 13353 Berlin, Germany; hardert@rki.de

Abstract

National immunization technical advisory groups (NITAGs) develop immunization-related recommendations. Systematic reviews are recommended to be used in this process, but conducting them requires significant resources, which many NITAGs lack. Using existing systematic reviews could help address this problem.

The Robert Koch Institute and collaborators set up the SYSVAC2 project to facilitate the retrieval of existing systematic reviews and offer guidance on using them. This will include an online registry of systematic reviews relevant to immunization policy and an online course on how to use existing reviews. This report describes an international expert workshop held in December 2019 to develop consensus on methods for using existing reviews and other relevant factors for the registry and course.

Members from NITAGs representing different regions of the world presented their experiences of using systematic reviews and reflected on challenges inhibiting use. Three methodologists considered different aspects of using systematic reviews. Interactive sessions followed, where implications for SYSVAC2 were discussed. Participants supported having critical appraisal ratings, plain language summaries, keyword search, and data visualization functions in the registry. They suggested tailoring course content to different audiences and including overviews of reviews as a topic and examples of how NITAGs have used or could use existing reviews. Participants agreed that whether a review is out-of-date should be decided by those using the review rather than registry staff. The registry could help by highlighting the date of literature search or included primary studies. Participants recommended a visualization function to highlight overlap across reviews and guidance on handling challenges to using reviews, ideally, involving a practical element. No consensus was reached on which critical appraisal tool to use for reviews in the registry, but a majority of participants wanted registry staff to perform appraisals. Formative research is planned before the registry and online course are launched in 2020.

Keywords: Evidence-based medicine, immunisation recommendation, methodology, systematic reviews, vaccination

1. Background and objectives

The role of national immunization technical advisory groups (NITAGs) is to develop recommendations to support national immunization program decision-making [1]. NITAGs are independent expert committees comprising members from disciplines relevant for immunization such as pediatrics, immunology, epidemiology, internal medicine and virology. They are nominated by the ministry of health of their respective country to provide independent and evidence-based advice to national decision-makers. On a global scale, NITAGs have varying resources, ranging from very limited staff to large secretariats. As recommendations made by NITAGs should reflect the best available evidence, it is suggested that systematic reviews be used in this process, since they synthesize findings from numerous primary studies and can provide more precise estimates of intervention effects than individual studies [2, 3]. However, conducting systematic reviews requires significant time, expertise, and human resources, which many NITAGs do not have.

Using existing systematic reviews could help address these problems but is not without its challenges. It can be resource-intensive and difficult to synthesize multiple reviews on the same topic and reconcile discrepancies across them [4]. The trustworthiness of existing reviews' findings may not be clear. Retrieving reviews can also be challenging without access to academic databases and journals, or in-depth knowledge of literature searching techniques [5].

In 2019, the Robert Koch Institute (RKI), in collaboration with the World Health Organization (WHO) and London School of Hygiene and Tropical Medicine (LSHTM), launched the SYSVAC2 project to help address the challenges of retrieving and using existing systematic reviews. The SYSVAC2 project builds on LSHTM's original SYSVAC project, which is described elsewhere [5]. This first version of SYSVAC was limited in technical functionality

(e.g., filtering options) and not accompanied by any teaching material to support users. SYSVAC2 was initiated to make it easier for NITAGs to (a) identify relevant systematic reviews and (b) access guidance on how to use existing reviews when developing recommendations. The goal of the project is to create a free, regularly updated, user-friendly online registry, or database, of systematic reviews on vaccine-related topics and an online training course on how to use existing reviews in developing recommendations for vaccine policy. By developing SYSVAC2, the project group aims at balancing the trade-offs between the lack of resources available to conduct new systematic reviews versus the investment of new resources needed to establish and maintain the registry and the course.

The RKI has planned multi-method formative research to inform the development of the registry and course, the first of which was an international expert workshop, which took place in Berlin, Germany on 12-13 December 2019. The purpose of the workshop was to develop expert consensus on methods for using existing systematic reviews and to discuss implications for the design of the online registry and course. Workshop objectives were to:

1. Share NITAGs' experiences in using existing systematic reviews in vaccine decision-making
2. Present guidance on methods for using existing systematic reviews
3. Agree on how the registry and course could help NITAGs navigate the evidence and deal with common challenges in using existing systematic reviews
4. Determine how best to assess the methodological quality and indicate the quality rating of systematic reviews in the registry

This report describes the methods involved in the workshop and summarizes the results.

2. Methods

Twenty-three experts participated in the workshop, representing the following entities and countries:

- NITAGs and their secretariats: Australia, Canada, Chile, China, Germany, South Africa, Sri Lanka, USA
- Multilateral Organizations: WHO, European Centre for Disease Prevention and Control
- Academia: Glasgow Caledonian University, LSHTM, University College London, University of British Columbia, University of Cape Town, Witten/Herdecke University

(see Supplementary file S1 for a list of participants).

Speakers included representatives from NITAGs and NITAG secretariats, who described their experiences using systematic reviews, and methodologists, who discussed methodological aspects of using existing systematic reviews. RKI staff corresponded and held planning meetings with the speakers prior to the workshop to communicate workshop objectives and ensure complementarity across talks. Each methodological talk was followed by an interactive brainstorming session, in which facilitators used modified Nominal Group Technique to translate insights from the talks into concrete ideas for the design of the registry and course [6, 7]. Facilitators posed brainstorming questions and allowed five to ten minutes for participants to gather their thoughts. In the first two sessions, participants presented their ideas in a round-robin session. The facilitators led a discussion and then participants “voted” on the three to five ideas they liked best. In the third session, facilitators led a discussion of each brainstorming question and requested voting only in the event that a decision had to be made. Neither RKI project staff nor facilitators of the session participated in voting.

3. Results

3.1 Use of existing systematic reviews in immunization-related decision-making

Five NITAG representatives presented their experiences of using existing systematic reviews when developing vaccination recommendations. Magdalena Bastías described how the Comité Asesor en Vacunas y Estrategias de Inmunización (CAVEI) (Chile) uses existing systematic reviews to orient themselves to the research in a particular area but relies mainly

on the primary studies included in the reviews. One challenge they face when using existing reviews is the heterogeneity across primary studies. Systematic reviews often use different measures for the same outcome, which makes interpreting and synthesizing results difficult. Another challenge is that most existing reviews are published in English; reviews in local or regional languages would be used more often. CAVEI supplements data from primary studies with evidence from other sources (e.g., surveillance and epidemiological data, Global NITAG Network resources, WHO Strategic Advisory Group of Experts (SAGE) on Immunization reviews, and vaccine recommendations from other countries).

Like CAVEI, Deepa Gamage described how the National Advisory Committee on Communicable Diseases (NACCD) (Sri Lanka) consults a wide range of sources beyond systematic reviews. The type of evidence consulted depends on their research question but may include local data on vaccine coverage and disease burden; vaccine effectiveness studies; risk profile assessments; published and unpublished literature about other countries' experiences, particularly in Southeast Asia; WHO position papers and recommendations; and cost-effectiveness studies. They consult existing systematic reviews, mainly from Cochrane and SAGE, to compare the results of their country-specific research with results from global reviews and guide decision-making.

Rudzani Muloiwa explained that the National Advisory Group on Immunization (NAGI) (South Africa) typically base recommendations on data on disease burden, effectiveness, cost-effectiveness, feasibility and affordability of the introduction of the vaccine, and the impact of including a new vaccine on the expanded program on immunization schedule. Local data are critical for their work, but systematic reviews are not available; existing reviews often only include studies from high-income countries. NAGI has found data from a local or similar context to be more useful than a systematic review from elsewhere, so they tend to rely on expert opinion, surveillance data, and primary studies rather than systematic reviews. The exception is systematic reviews on vaccine effectiveness, which, despite taking place in other contexts, remain useful in estimating impact.

With limited resources (e.g., smaller secretariats or no standalone secretariats), the NITAGs in Chile, Sri Lanka and South Africa reported having no means to conduct reviews themselves (*de novo* systematic reviews). In contrast, the US and Canadian NITAGs do conduct *de novo* systematic reviews. Jessica MacNeil reported that the Advisory Committee on Immunization Practices (ACIP) (USA) has 12 to 15 work groups, each led by experts and assisted by a librarian. These groups summarize published and unpublished data and prepare GRADE (Grading of Recommendations Assessment Development and Evaluation) evidence profiles and Evidence to Recommendations frameworks [8]. Non-systematic and systematic reviews are performed as part of this process, so existing systematic reviews are typically not used. Matthew Tunis mentioned that the National Advisory Committee on Immunization (NACI) (Canada) relies predominantly on reviews conducted by the secretariat at the Public Health Agency of Canada or through affiliated academic groups. Their reviews are systematic but would not always meet Cochrane review gold standards (e.g., they may have one data extractor and another spot-checking a sample, rather than double data extraction).

Despite having the capacity to conduct *de novo* reviews, NACI increasingly uses existing systematic reviews when developing vaccination recommendations. Using existing reviews authored from within Canada, produced by other high-income countries, or retrieved from SAGE, has increased NACI's efficiency. Since 2017 NACI has used a formal process, based on previously published approaches [9-11], to decide when and to what extent existing reviews should be used. If no relevant, high-quality reviews exist, then NACI initiates a *de novo* review. If relevant reviews of sufficient quality do exist, then NACI determines which elements of these reviews to use (i.e., search strategy, quality assessment, synthesis). If the search strategy from an existing review is older than six months, NACI will update it. Tunis noted that updating existing reviews can be complex, as many diverse risk of bias tools are used for observational studies [12], which are common in the vaccine literature. NACI has faced challenging decisions whether to update using the original study risk of bias tools or to apply tools that are preferred by NACI.

The question of whether data or results from existing systematic reviews can be “trusted” arose in multiple presentations. CAVEI has found discrepancies between information about a primary study reported in a review and information in the primary study itself, which creates mistrust of review findings. Systematic review authors are sometimes authors of included studies as well, a conflict of interest that may lead to bias. NITAGs reported trusting reviews conducted by certain groups, such as SAGE or other known NITAGs (e.g., ACIP, STIKO) but acknowledged that, even in these cases, NITAGs must carefully consider each component of existing reviews (e.g., search strategy, risk of bias assessments) before determining which elements to adopt. Tunis described NACI’s experience using a high-quality SAGE review on the HPV vaccine dose schedule [13]. NACI adopted all elements of this review, however upon later re-analysis, concluded that SAGE’s interpretation of the data differed from their own [14]. NAGI also expressed questioning estimates from SAGE reviews when based on WHO epidemiological estimates that differ from NAGI’s own estimates.

3.2 Navigating the evidence

James Thomas (EPPI-Centre at University College London) presented the first methodological talk. Thomas described the context within which systematic reviews are produced and how this has evolved, challenges in navigating systematic review evidence, and implications for the design of the SYSVAC2 registry. Research takes place in an evidence ecosystem in which those producing the research and those using research results (e.g., decision-makers) engage with each other and affect and are affected by broader socio-political factors. Against this backdrop, two models of reviews have emerged: the knowledge-driven model, which is driven by *research producers* and their use of the existing literature, and the problem-solving model, which is driven by *research users* and the problems they are facing.

Interactions between research producers and users in both models influence review aims and methods. Decision-makers are commissioning reviews at an increasing rate and demanding immediate and easy access to the evidence base [15], which has led to the

emergence of rapid reviews, living systematic reviews, and reviews of reviews ('overviews') [16]. Review questions have grown in range and complexity, which has led to the synthesis of a wide variety of study designs (e.g., randomized and non-randomized trials, qualitative research, economic data) using different methods (e.g., network meta-analysis, translational reviews, automation). There is also increased awareness that many factors can influence intervention outcomes (e.g., frequency or duration of delivery, level of participant engagement) [17]. Reviews now not only investigate whether an intervention worked but how and under what conditions [16, 18-20]. Reflecting these trends, the SYSVAC2 registry will include different types of systematic reviews, including rapid reviews, meta-analyses, and overviews of reviews, addressing a wide variety of research questions.

Decision-makers face several challenges when attempting to use existing reviews. They may have questions that are not directly addressed by any single review. For example, although a decision-maker might find an up-to-date, high-quality review that answers their question, particularly if they were involved in defining the scope of the review, the review may not directly address the decision-maker's context, constraints, or assumptions. As a result, rather than using the review in its entirety, it might be more appropriate to use a subset of studies from it. Alternatively, one might supplement the review with additional studies or take subsets of results from different reviews that, together, address a decision-maker's question and parameters.

Another challenge is when multiple relevant reviews exist. Decision-makers could, for example, synthesize them in an overview, use the most recent or highest quality review, or the most comprehensive. Weighing the tradeoffs associated with each course of action is a difficult task.

A third challenge is how to proceed if no relevant reviews on the decision-maker's topic are found. Decision-makers may consult guidance documents, NITAG documents [21], WHO position papers [22], the European Medicines Agency website (https://www.ema.europa.eu/en), or the Vaccine Adverse Event Reporting System database

for information relevant to vaccine recommendation development [23]. They could conduct a *de novo* systematic review. If existing systematic review evidence lacks local data, they could consider using population impact analysis, which incorporates local data (e.g., population size and demographics) with the results of meta-analyses to estimate an intervention's risks and benefits [24]. Alternatively, review results could be recalibrated to weight studies differentially based on their similarity to the inference population. Decision-makers could also map interventions in a review against what is locally available.

The registry's interface could help address some of these challenges by curating existing review evidence to help users find the evidence most relevant to their needs. One potentially useful function would be to map evidence and gaps visually. The Campbell Collaboration's evidence and gap maps (https://campbellcollaboration.org/evidence-gap-maps.html), Epistemonikos' matrix of evidence (https://www.epistemonikos.org/) [25], and the COVID-19 living systematic map [26] are examples of such a function.

3.2.1 Interactive session: Navigating the evidence

This session aimed to develop a ranked list of ideas on how the registry and course could most effectively help NITAGs find relevant evidence. Tables 1 and 2 list the ideas mentioned for the registry and course respectively, along with the votes that each idea received. Ideas receiving one or more votes are listed.

The most popular idea for the registry was to quality-appraise included reviews. Participants debated the merits of including poor-quality reviews in the registry and ultimately decided to retain them because they could be useful, for example, for pointing one to other studies. There is also value in knowing that reviews exist, despite receiving poor ratings. Participants supported having plain language summaries of reviews and the ability to search by a variety of keywords. Participants wanted a data visualization function built into the registry.

For the online course, the most popular idea was to tailor content to different audiences, e.g., by professional role (i.e., NITAG member vs. NITAG secretariat) or by level of experience

(i.e., new to using existing systematic reviews vs. experienced user). Participants were keen to learn about overviews and to read examples – either real or fictional – of how NITAGs have used or might use existing reviews. Examples of both successes and failures were regarded as useful.

3.3 Addressing common challenges in the use and synthesis of systematic reviews

Overviews of reviews summarize the results of multiple systematic reviews. Carole Lunny (University of British Columbia) spoke about common challenges encountered when synthesizing systematic reviews for an overview of reviews and ways to address them. Her talk, which was based on the Methods for Overviews of Reviews (MoOR) Framework [27, 28], focused on methods for addressing three out of seven challenges that authors face when synthesizing existing systematic reviews: overlapping primary studies data from multiple systematic reviews, out-of-date reviews, and discordant results and conclusions across systematic reviews.

Overlap in data can arise when systematic reviews on the same topic include one or more identical primary studies. Overlapping data may include overlapping risk of bias assessments, pooled effect estimates across similar outcomes, meta-analysis results (e.g., I^2 heterogeneity statistics), or certainty of the evidence assessments (e.g., GRADE). Overlap is problematic because effect estimates from pooled meta-analyses give undue statistical weight to and produce overly precise effect estimates for duplicated studies. These errors could result in incorrect results and conclusions about the effects of an intervention. Methods for dealing with overlap can be employed at various stages of conducting an overview. For example, at the eligibility criteria stage, one could either select one or a subset of reviews based on pre-specified inclusion criteria or include all systematic reviews and deal with the overlapping study data at the synthesis stage. At the synthesis stage, one can quantify the amount of overlap, visually present the overlap using tables and figures, select only one review to analyze (e.g., highest quality and most comprehensive), or use statistical approaches to deal with overlap, such as sensitivity analyses. Other solutions can be used at

the data extraction, risk of bias assessment, or certainty of the evidence stages, as noted in the MoOR Framework [27, 28].

The main challenge when reviews are out-of-date is that they provide incomplete and outdated evidence. Evidence may be out-of-date due to continually evolving research or when significant time has elapsed between completion of searches and production of the final report. This can be addressed at the search strategy stage and through pre-specification of eligibility criteria. For example, one can select the most recent review that fits one's Population Intervention Comparison Outcome (PICO) question and update the search strategy with primary studies that have been recently published.

The last challenge is discordance, which can arise for a number of reasons, for example, because reviews have different PICO questions, eligibility criteria, or search strategies; search different databases and sources; use different risk of bias tools, statistical models, or meta-analysis software; or interpret their results differently. Errors in data extraction could result in discordance as well, as could different approaches to retrieving missing data from the primary studies (e.g., search clinical trial registries or contact study authors).

There are solutions to discordance at multiple stages and with various methods. At the data extraction stage, decision-makers could extract data from all reviews or from only one review, selected according to pre-specified criteria. Alternatively, at the synthesis stage, one could examine and record the discordance, use decision rules or tools (e.g., Jadad algorithm [29]) to select one review, and/or use graphs and tables to depict discordance.

Notably, there is neither expert consensus about the optimal methods in terms of efficiency, usability, and resource use for dealing with these challenges nor empirical data on the validity and reliability of particular methods. Tradeoffs should be considered when choosing one method over another. Choosing one review from among many would result in a loss of information (e.g., the highest quality review may have fewer studies than a lower quality review, one review might have the most studies but miss more recent trials), which may lead

to uncertainty about the true effects of the intervention. However, including all reviews may introduce overlap, discordance, and possibly other challenges, and would require more resources to synthesize. Updating reviews is also resource-intensive, as it requires assessing the risk of bias of the new primary studies and, possibly, a new meta-analysis and incorporation of new studies into certainty of evidence assessments (e.g., GRADE). Doing nothing to resolve overlap, out-of-dateness, or discordance may affect the validity and reliability of the findings of an evidence review.

3.3.1 Interactive session: Addressing common challenges

This session aimed to develop a ranked list of ideas on how the registry and course could most effectively help NITAGs deal with common challenges.

The challenge of out-of-date reviews dominated the discussion around the registry. Participants agreed that whether a systematic review is out-of-date should be decided by those using the review. Popular ideas included highlighting the date of the last literature search or the range of dates of included primary studies (see Table 3). To address the challenge of overlapping data, participants supported including a function that would allow users to visualize the overlap in primary studies across reviews and, ideally, import this analysis into Excel. Participants felt that discordance across reviews could not be addressed by the registry but rather covered in the online course.

Another popular topic of discussion was how to keep the registry itself up-to-date. Participants supported engaging the community, pointing to Epistemonikos as a model. They also supported linking the registry to the course, such that exercises performed when completing the course could serve to maintain the registry (e.g., course participants could tag a review for keywords when reading it).

The most popular ideas for the course were the use of consistent terminology and the inclusion of specific training on overlapping data, out-of-date reviews, and discordance (see Table 4). Participants wanted guidance on how to handle these challenges, ideally, involving

a practical element where they could try out different solutions and learn about the tradeoffs involved.

3.4 Appraising systematic reviews

In the final session, Dawid Pieper (Witten/Herdecke University) presented on the appraisal of systematic reviews, a key aspect of using existing reviews. Pieper outlined available critical appraisal tools, reviewed their strengths and weaknesses and highlighted considerations when performing and reporting quality appraisals.

Three critical appraisal tools could be applied to the reviews housed in the registry: A MeaSurement Tool to Assess systematic Reviews (AMSTAR), Risk of Bias in Systematic Reviews (ROBIS), and AMSTAR 2. Since AMSTAR 2 is the revised version of AMSTAR and allows the appraisal of reviews containing both randomized and non-randomized studies, it is more up-to-date and comprehensive than AMSTAR. AMSTAR 2 and ROBIS measure slightly different, but related, concepts. AMSTAR 2 assesses methodological quality (i.e., how well a review was designed and conducted) [30]. ROBIS assesses risk of bias, which refers to the extent to which systematic flaws or limitations in the design, conduct, or analysis of a review might influence the results or conclusions [31]. Despite this distinction, the tools have considerable overlap, and empirical evidence suggests high correlation in ratings for the two tools [32-34].

AMSTAR 2 is a 16-item tool that provides a summary of confidence in the overall findings of the review [35]. Strengths include its relative ease and efficiency of use. Interrater-reliability is slightly better for AMSTAR 2 than for ROBIS [32, 36]. Furthermore, one can use the tool without in-depth content knowledge, methodological expertise, or training. Its primary weakness is that several items are vague or broad, so users have considerable latitude in interpreting their meaning. For example, item eight in AMSTAR 2 asks if the review authors described included studies in “adequate detail.” Moreover, guidance is lacking regarding how to interpret flaws identified by the tool. The AMSTAR 2 developers highlight seven domains as being “critical” and suggest tallying the flaws in these domains and in the remaining (“non-

critical") domains to gauge overall confidence in review results [35]. However, they leave it up to users of the tool to determine whether the domains highlighted as "critical" are indeed the most important for users.

ROBIS is a domain-based tool, which is completed in three phases: (1) assess relevance (i.e., directness) of one's question to the review being assessed (optional), (2) identify concerns with the review process, and (3) judge risk of bias in the review. There are four domains (i.e., study eligibility criteria, identification and screening, data collection and study appraisal, synthesis and findings), each of which includes signaling questions [31]. A key strength of ROBIS is its versatility. In contrast to AMSTAR 2, which was designed for reviews of healthcare interventions, ROBIS can be applied to reviews spanning a broader set of topics, such as diagnostic test accuracy or prediction models. However, the time required to complete a ROBIS assessment is longer than for AMSTAR 2, and more in-depth content knowledge and methodological expertise are required [32, 36]. For instance, item 1.2 on whether the eligibility criteria used in the review were appropriate requires an understanding of the kinds of studies – for example, in terms of population, setting, and intervention dose – suitable for answering the research question. Similarly, item 3.3 on whether relevant study results were collected for use in the synthesis requires knowing what constitutes "relevant" study results, which will vary based on the subject matter of the review and included study designs [37].

Both tools have limitations. For example, they are more expert- than evidence-based, and their overall ratings depend on reporting quality. Moreover, they fail to capture some issues, such as when reviews have incorrect data or do not include relevant studies. Critical appraisal tools cannot capture flaws in data extraction and use in meta-analyses, nor bias from conflicts of interest. Research suggests that authors tend to assess the quality of their own studies higher than those of others [38]. One option for the SYSVAC2 registry is to include a commentary alongside the results of the critical appraisal tool, highlighting problematic issues not captured by the tool.

Since many systematic reviews have already been assessed by others in overviews of reviews, clinical guidelines, and databases (e.g., https://www.healthevidence.org/), one question is whether to use existing critical appraisals for reviews in the registry. Pieper noted that risk of bias assessments of randomized controlled trials included in multiple reviews have been found to be inconsistent [39, 40], and the situation is likely to be similar for AMSTAR 2 and ROBIS assessments of reviews conducted by different groups. To ensure consistency in quality/risk of bias judgments across reviews in the registry, the same team should conduct the assessment of all included reviews, with independent appraisal by two people, who then compare their assessments and resolve differences in judgments. Alternatively, one person can perform the assessment with a second person checking a sample to ensure consistency.

3.4.1 Interactive session: Appraising systematic reviews

The last session aimed to determine (1) which critical appraisal tool should be applied to reviews in the registry, (2) how the results from critical appraisal should be communicated in the registry, and (3) which critical appraisal topics the course should cover.

The first question sparked a broad-ranging discussion that compared the tools but, ultimately, did not result in consensus around a particular tool. Participants regarded the setup of domains in ROBIS, its applicability to grey literature, and the fact that it does not confuse reporting quality with risk of bias, as advantages. Participants also appreciated that the optional relevance question could be used to compare vaccine-related reviews to registry users' research questions. Disadvantages included the more in-depth content knowledge and methodological expertise required to use ROBIS. Participants liked AMSTAR 2 for how easy and intuitive it is to use and for its item on conflict of interest, which ROBIS does not have. Participants noted that AMSTAR 2 could be supplemented with the ROBIS question on relevance, or NITAGs could simply assess relevance by comparing their PICO question against the PICO question of existing reviews. Although designed to have broader applicability than the healthcare-focused AMSTAR 2, ROBIS has not been validated with

non-healthcare reviews (e.g., economics). Thus, in practice, both tools seem best suited for reviews of healthcare interventions.

A few participants questioned whether critically appraising reviews in the registry was worthwhile. Relevance to a registry user's research question might be a bigger deciding factor in whether to use a review than quality. Others proposed performing critical appraisal on some, but not all, reviews in the registry.

Facilitators asked participants to vote for one of three options: perform critical appraisal for all reviews, offer a critical appraisal "on demand" service, or do not offer critical appraisal. Results, shown in Table 5, revealed participants overwhelmingly wanted registry staff to undertake critical appraisal, with more than half participants supporting an "on demand" service.

The remaining questions on how quality should be depicted in the registry and what critical appraisal topics should be included in the course were briefly discussed. Participants recommended avoiding a color coding system when communicating judgments on quality ratings (e.g., red indicating a high risk of bias rating, green indicating a low risk of bias rating) and enabling users to access the ratings for all quality appraisal items easily. Regarding the course, participants suggested training on both AMSTAR 2 and ROBIS and explaining their differences, similarities, strengths, and weaknesses.

4. Next steps

RKI will conduct a survey with NITAGs globally to learn about their experiences in retrieving scientific literature online and, specifically, using existing systematic reviews to formulate vaccine recommendations. Insights from this workshop, as well as from the published literature and survey, will inform the development of the registry and online course, which RKI plans to launch in 2021. Future plans include refining the online course content and further adapting the search platform of the registry based on users' experiences.

5. Summary and conclusions

This workshop brought together experts in immunization policy and methodologists to share their experiences and expertise and brainstorm ideas regarding the design of an online registry of systematic reviews on vaccine-related topics and a complementary course. NITAGs use a suite of evidence (e.g., primary studies, WHO vaccine position papers, surveillance data) when developing immunization-related recommendations. While existing systematic reviews can be retrieved and included as part of this process, they are not always freely and publically accessible, perceived as being relevant to a user's question, or considered trustworthy. Identifying relevant reviews is challenging because often there is not a direct match between a decision-maker's research question and the existing evidence. Sometimes systematic reviews only include global data or data from high-income countries, which may have limited applicability to one's local context. A lack of guidance on how to proceed when there are multiple, relevant reviews can also inhibit their use. Conversely, sometimes relevant reviews do not exist. Synthesizing existing reviews can be difficult, with challenges such as overlapping, out-of-date, and discordant data. Although multiple methods have been used to address these challenges, there is neither consensus nor empirical evidence to support the use of one method over another.

The SYSVAC2 registry and online course could help users resolve some of the challenges associated with retrieving, synthesizing, and using reviews. For example, the user interface could help identify out-of-date reviews and visualize overlapping primary study data across reviews on the same topic. The course could help users understand the tradeoffs between methods used to deal with these challenges. Registry staff could critically appraise reviews in the registry to help users choose among reviews and understand each review's strengths and limitations. Both AMSTAR 2 and ROBIS were considered acceptable critical appraisal tools.

Insights from this workshop, results from a survey with NITAGs, and published literature will inform the development of the registry and online course, which will be launched in 2021.

Acknowledgements

We would like to thank Charbel El Bcheraoui, Kari Johansen, Judith Koch, Chao Ma, Jessica MacNeil, Sarah Sheridan, and Sabine Vygen-Bonnet for participating in the workshop. We would also like to thank the following RKI staff for assisting with workshop preparations: Yvonne Bichel, Denise Mehlitz, and Sarah Eva Wetzel.

Funding: This workshop was supported by the German Federal Ministry of Health through the Global Health Protection Program.

The authors alone are responsible for the views expressed in this article and they do not necessarily represent the views, decisions, or policies of the institutions with which they are affiliated.

Conflict of interest

The authors declare no conflicts of interest.

All authors attest they meet the ICMJE criteria for authorship.

Table 1. Navigating the evidence: Design ideas for the online registry (n=18)[1]

Idea	n (%)
Appraise included reviews with AMSTAR 2 or ROBIS	12 (67)
Include a plain language summary of the review	7 (39)
Allow searching by keywords (e.g., disease, population characteristics)	6 (33)
Include visualization to help users interact with the evidence	5 (28)
Keep registry up-to-date with automation	4 (22)
Include date of search for review as keyword or filtering option	4 (22)
Make full text of reviews open access	4 (22)
Include papers beyond published reviews (e.g., NITAG reports or reviews)	3 (17)
List aims and objectives of reviews in each entry	3 (17)
Link to PROSPERO	2 (11)
Include a version for mobile phones/smart devices	2 (11)
Allow users to filter results by whether or not an author has a conflict of interest	1 (6)
Indicate whether the results of a systematic review are conclusive or stable	1 (6)
Allow email notifications (e.g., if a new review is uploaded that fits particular criteria)	1 (6)
Highlight gaps in the evidence that reviews identify	1 (6)
Allow users to comment on reviews (e.g., “This review was useful to me or not”)	1 (6)
Exclude low-quality reviews	1 (6)

[1]n represents total number of people who participated in voting.

513 **Table 2. Navigating the evidence: Design ideas for the online course (n=18)[1]**

Idea	n (%)
Tailor course content to different audiences (e.g., NITAG member vs. NITAG secretariat)	13 (72)
Include information about conducting overviews of reviews	8 (44)
Include examples from NITAGs' own experiences. Include best and worst case examples.	8 (44)
Include tools for assessing risk of bias of systematic reviews and tutorials for performing these assessments	4 (22)
Link to other courses, when possible	3 (17)
Include information about software available to assist with systematic reviews, like Covidence, Distiller, and RevMan	3 (17)
Include reviews in languages other than English	3 (17)
Include templates, when possible. For example, a blank ROBIS form used for assessment of the risk of bias of a systematic review and blank Excel sheets used for data extraction.	3 (17)
Do not make the course too long	2 (11)
Enable people to access materials offline	2 (11)
Have the course accredited so that it could count as continuing medical education	2 (11)
Follow up with users six months afterwards, perhaps with a mentoring session, to find out about their experiences with using systematic reviews and how they have applied what they learned	2 (11)
Include information about how to update reviews	2 (11)
Include tests throughout the course – not just at the end	1 (6)
Include a module on reporting quality and transparency of methods	1 (6)
Allow users to interact with each other	1 (6)
Make the online course a podcast so that people can listen to it in the car	1 (6)

514 [1]n represents total number of people who participated in voting.

515

Table 3. Addressing common challenges: Design ideas for the online registry (n=17)[1]

Idea	n (%)
Highlight the date of last search performed in a review	8 (47)
Enlist the community to keep the registry up-to-date	8 (47)
Link the registry to the training. Consider how tasks in the online course could feed into maintenance of the registry.	6 (35)
Provide a visual of overlap of primary studies across reviews and make it available for export	6 (35)
Include an “online communication with an expert” function	6 (35)
Highlight the range of dates for when primary studies included in a review were conducted	5 (29)
Do not try to deal with discordance in findings across reviews in the registry	3 (18)
Allow users to access/click on primary studies included in reviews	3 (18)
Allow users to show all studies that would fit the inclusion criteria of a systematic review	2 (12)
Allow sorting/filtering of search results by last search performed in review	2 (12)
Distinguish overlap of primary study data across reviews at the PICO level and at the level of results	1 (6)
Do not try to set criteria for whether a review is out-of-date. It should be decided on a case-by-case basis.	1 (6)
Consider a collaboration with Epistemonikos	1 (6)
Include GRADE assessments when systematic review authors have performed them	1 (6)

Table 4. Addressing common challenges: Design ideas for the online course (n=17)[1]

Idea	n (%)
Use consistent terminology when describing methods for course users	10 (59)
Synthesize three systematic reviews and make sure there is discordance in findings and overlapping primary studies. Show the tradeoffs associated with choosing different methods to address these challenges.	10 (59)
Explain what it means for a review to be “out-of-date” and how to deal with it. Link the registry with the course when discussing this.	9 (53)
Explain what to do when there is overlap in primary studies across reviews	4 (24)
Explain what to do in the case of discordance in findings and conclusions across similar reviews	4 (24)
Highlight challenges in using overviews of reviews	4 (24)
Explain how to update a review	3 (18)
Include a chat box or service where users can get advice on out-of-dateness, discordance, etc.	2 (12)
Include an introduction to different types of reviews	1 (6)
Be clear about the time required for the training	1 (6)
Have students do a short pre-test before starting the course to help them determine what sections would be most relevant to them	1 (6)
Consider the Cochrane Crowd training interface for inspiration	1 (6)

[1]n represents total number of people who participated in voting.

Table 5. Options for addressing critical appraisal in the registry (n=16)[1]

Options	n (%)
Offer a critical appraisal “on demand” service	9 (56)
Perform critical appraisal for all reviews in the registry	5 (31)
Do not offer critical appraisal for reviews in the registry	0 (0)

[1]n represents total number of people who participated in voting.

Supplementary File S1. List of participants

Magdalena Bastías

Comité Asesor en Vacunas y Estrategias de Inmunización (Chile)

Helen Burchett

London School of Hygiene and Tropical Medicine

Pauline Campbell

Nursing, Midwifery, and Allied Health Professions Research Unit, Glasgow Caledonian University

Charbel El Bcheraoui

Robert Koch-Institut, Centre for International Health Protection

Deepa Gamage

Ministry of Health and Advisory Committee on Communicable Diseases (Sri Lanka)

Thomas Harder

Robert Koch-Institut, Immunization Unit

Louise Henaff

World Health Organization (WHO) Headquarters

Catherine Jo

Robert Koch-Institut, Immunization Unit

Kari Johansen

European Centre for Disease Prevention and Control and WHO Strategic Advisory Group of Experts (SAGE)

Benjamin Kagina

University of Cape Town

Judith Koch

Robert Koch-Institut, Immunization Unit and Standing Committee on Vaccination (STIKO) Secretariat

Carole Lunny

University of British Columbia

Chao Ma

Chinese Centers for Disease Control and Prevention and the National Immunization Advisory Committee (China)

Jessica MacNeil

Centers for Disease Control and Prevention and the Advisory Committee on Immunization Practices (USA)

Melanie Marti

WHO SAGE Secretariat

Rudzani Muloiwa

University of Cape Town and the National Advisory Group on Immunization (South Africa)

Dawid Pieper

Witten/Herdecke University

Sarah Sheridan

National Centre for Immunisation Research and Surveillance (Australia)

James Thomas

Evidence for Policy and Practice Information and Co-ordinating Centre, University College London

Matthew Tunis

Public Health Agency of Canada and National Advisory Committee on Immunization (Canada)

Sabine Vygen-Bonnet

Robert Koch-Institut, Immunization Unit and STIKO Secretariat

Ole Wichmann

Robert Koch-Institut, Immunization Unit

Zane Younger

Robert Koch-Institut, Immunization Unit and Centre for International Health Protection

613 References

614

615 [1] Duclos P. National Immunization Technical Advisory Groups (NITAGs): guidance for their
616 establishment and strengthening. Vaccine. 2010;28 Suppl 1:A18-25.
617 [2] Lavis JN, Posada FB, Haines A, Osei E. Use of research to inform public policymaking.
618 Lancet. 2004;364:1615-21.
619 [3] Moat KA, Lavis JN, Wilson MG, Rottingen JA, Barnighausen T. Twelve myths about
620 systematic reviews for health system policymaking rebutted. J Health Serv Res Policy.
621 2013;18:44-50.
622 [4] Ioannidis JP. The Mass Production of Redundant, Misleading, and Conflicted Systematic
623 Reviews and Meta-analyses. The Milbank quarterly. 2016;94:485-514.
624 [5] Fernandes S, Jit M, Bozzani F, Griffiths UK, Scott JAG, Burchett HED. A bibliometric
625 analysis of systematic reviews on vaccines and immunisation. Vaccine. 2018;36:2254-61.
626 [6] Gallagher M, Hares T, Spencer J, Bradshaw C, Webb I. The nominal group technique: a
627 research tool for general practice? Family practice. 1993;10:76-81.
628 [7] Jones J, Hunter D. Consensus methods for medical and health services research. BMJ.
629 1995;311:376-80.
630 [8] Lee G, Carr W, Group AE-BRW. Updated framework for development of evidence-based
631 recommendations by the Advisory Committee on Immunization Practices. MMWR Morb
632 Mortal Wkly Rep. 2018;67:1271-2.
633 [9] Harder T, Remschmidt C, Haller S, Eckmanns T, Wichmann O. Use of existing systematic
634 reviews for evidence assessments in infectious disease prevention: a comparative case
635 study. Systematic reviews. 2016;5:171.
636 [10] Robinson KA, Chou R, Berkman ND, Newberry SJ, Fu R, Hartling L, et al. Twelve
637 recommendations for integrating existing systematic reviews into new reviews: EPC
638 guidance. Journal of clinical epidemiology. 2016;70:38-44.

[11] Robinson KA, Whitlock EP, Oneil ME, Anderson JK, Hartling L, Dryden DM, et al. Integration of existing systematic reviews into new reviews: identification of guidance needs. Systematic reviews. 2014;3:60.

[12] Farrah K, Young K, Tunis MC, Zhao L. Risk of bias tools in systematic reviews of health interventions: an analysis of PROSPERO-registered protocols. Systematic reviews. 2019;8:280.

[13] Human papillomavirus vaccines: WHO position paper, October 2014-Recommendations. Vaccine. 2015;33:4383-4.

[14] National Advisory Committee on Immunization (NACI). Amendment to the 2015 "Update on the recommended Human Papillomavirus (HPV) vaccine immunization schedule". https://www.canada.ca/en/public-health/services/immunization/national-advisory-committee-on-immunization-naci/amendment-2015-update-on-recommended-human-papillomavirus-hpv-vaccine-immunization-schedule.html. 2015 [accessed 3 September 2020].

[15] Oliver s, Bangpan M, Dickson A. Producing policy relevant systematic reviews: Navigating the policy-research interface. Evidence and Policy. 2017.

[16] Gough D, Thomas J, Oliver S. Clarifying differences between reviews within evidence ecosystems. Systematic reviews. 2019;8:170.

[17] Carroll C, Patterson M, Wood S, Booth A, Rick J, Balain S. A conceptual framework for implementation fidelity. Implement Sci. 2007;2:40.

[18] Pawson R, Manzano-Santaella A. A realist diagnostic workshop. Evaluation. 2012;18:176-91.

[19] Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review--a new method of systematic review designed for complex policy interventions. J Health Serv Res Policy. 2005;10 Suppl 1:21-34.

[20] Pawson R. Evidence-Based Policy: A realist perspective. London: Sage Publications; 2006.

[21] NITAG Resource Center. Media Center. https://www.nitag-resource.org/media-center. 2019 [accessed 3 September 2020].

667 [22] World Health Organization. Immunization, Vaccines and Biologicals - WHO vaccine
668 position papers. https://www.who.int/immunization/policy/position_papers/en/. 2020
669 [accessed 3 September 2020].
670 [23] Halsey NA, Proveaux T. Value of an in-depth analysis of unpublished data on the safety
671 of influenza vaccines in pregnant women. Vaccine. 2017;35:6154-9.
672 [24] Verma A, Torun P, Harris E, Edwards R, Gemmell I, Harrison RA, et al. Population
673 Impact Analysis: a framework for assessing the population impact of a risk or intervention.
674 Journal of public health (Oxford, England). 2012;34:83-9.
675 [25] El-Khayat YM. Epistemonikos. Journal of the Medical Library Association : JMLA.
676 2017;105:431-2.
677 [26] EPPI-Centre. COVID-19: a living systematic map of the evidence.
678 http://eppi.ioe.ac.uk/cms/Projects/DepartmentofHealthandSocialCare/Publishedreviews/COV
679 ID-19Livingsystematicmapoftheevidence/tabid/3765/Default.aspx. 2020 [accessed 3
680 September 2020].
681 [27] Lunny C, Brennan SE, McDonald S, McKenzie JE. Toward a comprehensive evidence
682 map of overview of systematic review methods: paper 1-purpose, eligibility, search and data
683 extraction. Systematic reviews. 2017;6:231.
684 [28] Lunny C, Brennan SE, McDonald S, McKenzie JE. Toward a comprehensive evidence
685 map of overview of systematic review methods: paper 2-risk of bias assessment; synthesis,
686 presentation and summary of the findings; and assessment of the certainty of the evidence.
687 Systematic reviews. 2018;7:159.
688 [29] Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic
689 reviews. CMAJ. 1997;156:1411-6.
690 [30] Pussegoda K, Turner L, Garritty C, Mayhew A, Skidmore B, Stevens A, et al. Systematic
691 review adherence to methodological or reporting quality. Systematic reviews. 2017;6:131.
692 [31] Whiting P, Savovic J, Higgins JP, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: A
693 new tool to assess risk of bias in systematic reviews was developed. Journal of clinical
694 epidemiology. 2016;69:225-34.

695 [32] Pieper D, Puljak L, Gonzalez-Lorenzo M, Minozzi S. Minor differences were found
696 between AMSTAR 2 and ROBIS in the assessment of systematic reviews including both
697 randomized and nonrandomized studies. Journal of clinical epidemiology. 2019;108:26-33.
698 [33] Banzi R, Cinquini M, Gonzalez-Lorenzo M, Pecoraro V, Capobussi M, Minozzi S. Quality
699 assessment versus risk of bias in systematic reviews: AMSTAR and ROBIS had similar
700 reliability but differed in their construct and applicability. Journal of clinical epidemiology.
701 2018;99:24-32.
702 [34] Lorenz RC, Matthias K, Pieper D, Wegewitz U, Morche J, Nocon M, et al. A
703 psychometric study found AMSTAR 2 to be a valid and moderately reliable appraisal tool.
704 Journal of clinical epidemiology. 2019;114:133-40.
705 [35] Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical
706 appraisal tool for systematic reviews that include randomised or non-randomised studies of
707 healthcare interventions, or both. BMJ. 2017;358:j4008.
708 [36] Gates M, Gates A, Duarte G, Cary M, Becker M, Prediger B, et al. Quality and risk of
709 bias appraisals of systematic reviews are inconsistent across reviewers and centers. Journal
710 of clinical epidemiology. 2020;125:9-15.
711 [37] Whiting P, Savović J, Higgins J, Caldwell D, Reeves B, Shea B, et al. ROBIS: Tool to
712 assess risk of bias in systematic reviews - Guidance on how to use ROBIS.
713 https://www.bristol.ac.uk/media-library/sites/social-community-
714 medicine/robis/robisguidancedocument.pdf. n.d. [accessed 3 September 2020].
715 [38] Pieper D, Waltering A, Holstiege J, Buchter RB. Quality ratings of reviews in overviews:
716 a comparison of reviews with and without dual (co-)authorship. Systematic reviews.
717 2018;7:63.
718 [39] Bertizzolo L, Bossuyt P, Atal I, Ravaud P, Dechartres A. Disagreements in risk of bias
719 assessment for randomised controlled trials included in more than one Cochrane systematic
720 reviews: a research on research study using cross-sectional design. BMJ Open.
721 2019;9:e028382.

722 [40] Konsgen N, Barcot O, Hess S, Puljak L, Goossen K, Rombey T, et al. Inter-review
723 agreement of risk-of-bias judgments varied in Cochrane reviews. Journal of clinical
724 epidemiology. 2020;120:25-32.

725