



Population Council Knowledge Commons

Reproductive Health

Social and Behavioral Science Research (SBSR)

2014

Second consultation on developing standards for identifying evidence-based practices in reproductive health

STEP UP Research Programme Consortium

Follow this and additional works at: https://knowledgecommons.popcouncil.org/departments_sbsr-rh

 Part of the [Demography, Population, and Ecology Commons](#), [Family, Life Course, and Society Commons](#), [International Public Health Commons](#), [Maternal and Child Health Commons](#), and the [Women's Health Commons](#)

Recommended Citation

STEP UP Research Programme Consortium. 2014. "Second consultation on developing standards for identifying evidence-based practices in reproductive health," STEP UP Report. New York: Population Council.

This Report is brought to you for free and open access by the Population Council.

SECOND CONSULTATION ON DEVELOPING STANDARDS FOR IDENTIFYING EVIDENCE-BASED PRACTICES IN REPRODUCTIVE HEALTH

18TH–19TH SEPTEMBER 2013

SELSDON PARK HOTEL, CROYDON, UK

SECOND CONSULTATION ON DEVELOPING STANDARDS FOR IDENTIFYING EVIDENCE-BASED PRACTICES IN REPRODUCTIVE HEALTH

18TH–19TH SEPTEMBER 2013

SELSDON PARK HOTEL, CROYDON, UK

The **STEP UP (Strengthening Evidence for Programming on Unintended Pregnancy) Research Programme Consortium** generates policy-relevant research to promote an evidence-based approach for improving access to family planning and safe abortion. STEP UP focuses its activities in five countries: Bangladesh, Ghana, India, Kenya, and Senegal. STEP UP is coordinated by the Population Council in partnership with the African Population and Health Research Center; icddr,b; the London School of Hygiene and Tropical Medicine; Marie Stopes International; and Partners in Population and Development. STEP UP is funded by UK aid from the UK Government. www.stepup.popcouncil.org



The London School of Hygiene & Tropical Medicine is a world-leading centre for research and postgraduate education in public and global health. Our mission is to improve health and health equity in the UK and worldwide; working in partnership to achieve excellence in public and global health research, education and translation of knowledge into policy and practice. www.lshtm.ac.uk



The Population Council confronts critical health and development issues—from stopping the spread of HIV to improving reproductive health and ensuring that young people lead full and productive lives. Through biomedical, social science, and public health research in 50 countries, we work with our partners to deliver solutions that lead to more effective policies, programs, and technologies that improve lives around the world. Established in 1952 and headquartered in New York, the Council is a nongovernmental, nonprofit organization governed by an international board of trustees. www.popcouncil.org



This document was made possible by the generous support of the American people through the United States Agency for International Development (USAID). The contents are the responsibility of the authors and do not necessarily reflect the views of USAID or the United States Government.

Suggested citation: STEP UP Research Programme Consortium. 2013. “Second consultation on developing standards for identifying evidence-based practices in reproductive health,” *STEP UP Report*. New York: Population Council.

Please address any inquiries about STEP UP to the RPC co-directors:

Dr. Harriet Birungi, hbirungi@popcouncil.org

Dr. Ian Askew, iaskew@popcouncil.org

© 2014 The Population Council, Inc.

Funded by:



CONTENTS

INTRODUCTION	1
Overview of the consultation	1
HIGHLIGHTS.....	2
Types of design and quality of evidence generated	2
Synthesising and summarising bodies of evidence	4
SUMMARY AND RECOMMENDATIONS	7
Designing implementation research to maximise the quality of evidence generated and its utility for decision-making	7
Recommendations for synthesising and grading bodies of evidence on reproductive health	9
Funding and research implementation structures and procedures to generate quality evidence and strong recommendations	11
LIST OF PARTICIPANTS.....	13
LIST OF PRESENTATIONS.....	16
REFERENCES	17

INTRODUCTION

The STEP UP consortium and the High Impact Practices (HIP) initiative, together with other partners, are supporting three consultation meetings to inform use of appropriate standards of evidence for recommending best practices in reproductive health (RH). The first consultation (held on 18–19 June 2013 in New York, USA) focused on the generation and synthesis of evidence that can inform the production of HIP briefing documents. These documents describe service delivery and enabling environment practices that, if fully implemented, will maximise investments in a comprehensive family planning (FP) strategy. The report is available [here](#).

The second consultation (held on 18-19 September 2013 in Croydon, UK) built on issues discussed during the first consultation by focussing on the generation and synthesis of evidence for practice recommendations. This meeting brought together researchers and research funders to review the research designs and methodologies that can be used to generate evidence on the impact of FP/RH interventions and on their implementation, the mechanisms and structures through which such evidence is reviewed and translated into recommendations, and the implications for organizing and funding evidence generation to maximise its quality and utility.

The third consultation, which will be held during the second quarter of 2014 in Nairobi, Kenya, will focus on communicating and using practice recommendations.

Overview of the consultation

Thirty-seven researchers and research funders participated in the two-day meeting, which combined panel presentations with group discussions that proposed various ways to improve evidence to inform RH/FP programming and policy. The objectives for this consultation were:

1. Suggestions for **designing** implementation research to maximise the quality of evidence generated and its utility for decision-making to strengthen reproductive health practices.
2. Recommendations for **synthesising and grading** bodies of evidence on a) reproductive health service delivery, and b) other health system building blocks to improve reproductive health services.
3. Proposals to guide appropriate **funding** and **research implementation structures and procedures** to generate strong evidence for practice recommendations in reproductive health.

The meeting provided a forum to present and discuss a wide range of perspectives. Agreement was reached that a consensus was unlikely – and probably unnecessary given the many differing evidence needs of decision-makers, the critical role of context in framing the application of evidence, the complexity of most FP/RH interventions, and the challenges in ensuring that an intervention is implemented at scale in the same way as it was designed and evaluated during pilot-testing. Meeting participants discussed the complexity of evidence generation and synthesis for identifying evidence-based practices in FP/RH and valued the opportunity to learn from a highly experienced group of participants.

HIGHLIGHTS

The meeting began with an overview of the HIP initiative, which identifies and generates a strong evidence base for service delivery and enabling environment practices that are most likely to support a FP/RH programme in achieving high impact (defined as improved health behaviours and/or outcomes). Depending on the strength of evidence available, HIPs are categorised as proven, promising, or emerging. Five challenges to determining the standards of evidence for guiding FP/RH programming were identified during the first consultation [Shawn [Malarcher](#)]: 1) research on FP lacks a common language; 2) there are no absolute agreed outcome measures; 3) strength of evidence varies greatly; 4) contexts are varied and complex; and 5) insufficient lessons are learnt from failures.

Angela [Baschieri](#) described DFID's multifaceted approach to generating and using evidence throughout its policy and programming cycle. Strategies include an 'Evidence into Action' team, commissioning evidence products (e.g. systematic reviews, rapid reviews, literature reviews, evidence papers), a policy research fund, several research programme consortia (e.g. STEP UP), a maternal health platform, and core funding for WHO's Human Reproduction Programme. DFID's 'Framework for Results' for improving reproductive, maternal, and newborn health in the developing world was informed by numerous evidence products to determine which implementation strategies represented the best value for money (VfM).

John [Cleland](#) highlighted several differences between research on FP/RH interventions and on disease/death prevention interventions: the efficacy, effectiveness, and safety of contraceptive products are well established; FP/RH research is primarily focused on optimizing supply and demand interventions; intensity of demand varies widely, whereas disease/death prevention is universal; evidence is largely derived from observational studies with relatively few experimental studies [see [Mwaikambo et al.](#) for a review]; and programming is informed more by accumulated experience and evidence of implementation than by systematic reviews or randomised controlled trials (RCTs). Cleland concluded by identifying common weaknesses in FP/RH evaluation designs and recommended that future FP/RH research should seek to use quasi-experimental or experimental designs when possible, take advantage of strengthened monitoring and evaluation (M&E), include economic evaluations, and be implemented through closer collaboration between implementers and researchers.

Types of design and quality of evidence generated

Karen [Hardee](#) provided an overview of systems used to rate the strength of evidence that informs decision-making. Hierarchies for rating evidence have predominated in healthcare decision-making, drawing primarily from Gray's rating which ranks evidence from systematic reviews of RCTs as strong and evidence from quasi-experimental and non-experimental studies as weak. These hierarchies, designed originally for clinical practice, have been developed to rate evidence of intervention effectiveness. The challenges of interpreting evidence that is rated as 'weak' were discussed extensively, especially as related to recommending interventions for FP/RH programming. The limitations of this hierarchy in rating evidence to inform research questions other than intervention effectiveness (e.g. intervention implementation, salience, acceptability,

etc.) were discussed, and an alternative typology was presented [[Petticrew & Roberts](#)] that rates evidence according to the type of question it is being used to inform.

James [Hargreaves](#) started his presentation from the premise that randomised trials are the most efficient and simplest to understand design option for addressing questions about impact and effectiveness. However, they are not always feasible and may be inappropriate for other types of questions. Alternative designs were discussed (e.g. pragmatic randomisation, process evaluation, stepped wedge/phased implementation, randomised 'encouragement' to perform an intervention, interrupted time series, and systematic non-random allocation with regression discontinuity), with examples provided for illustration.

Many FP/RH interventions are multi-faceted because they need to intervene at multiple levels in a causal chain, operating through distal causes and affecting distal outcomes [[Campbell](#)]. However, evaluation of complex interventions using RCTs is difficult for many reasons and thus is usually not an appropriate design. An additional alternative to those described by Hargreaves is to strengthen data collection to measure outputs and, if possible, outcomes through an M&E system designed for an FP/RH intervention implemented at scale in different settings and populations and with variations in the intervention design. Such data can provide an understanding of how the intervention functions in 'real life'; however, the M&E system must be able to measure outcomes as well as inputs, process, and outputs. Doing so can be challenging as most M&E systems do not collect population-level data. This evaluation approach is dependent on a well-constructed and clearly understandable *theory of change* [[Vogel, I.](#)] that explains how the intervention is meant to work, its purpose and goal, and the assumptions underlying the causal pathways that translate implementation into outputs and outcomes.

Theories of change are often not specified in sufficient detail prior to implementing an intervention, and/or there may be different understandings among implementers, researchers, and funders of how an intervention is meant to work and its purpose. A case study of a theory of change for evaluating a safe abortion intervention in India was presented by Carine Ronsmans. Intervention implementers are trying to maximise the programme's outputs; evaluators are seeking to maximise attribution of the observed outcomes to the programme; and the funder wants both. Tensions can emerge as a result of unclear expectations; for example, implementers may want to introduce the intervention into locations most likely to serve the largest population so as to maximise outputs, while evaluators would prefer to randomise placement of the intervention to maximise validity of the findings. A well-constructed theory of change should clarify to implementers, funders, and evaluators which questions a research study or evaluation can and cannot answer and enable the perspectives of each group to be considered jointly to permit consensus on intervention and research designs.

Ulla Kou [Griffiths](#) presented an overview of economic evaluation, defined as the comparative analysis of alternative interventions in terms of their costs and consequences, which can inform evidence-based decisions on the most effective way of using scarce resources. Key components of an economic evaluation are a *cost-effectiveness ratio* of the incremental cost per outcome unit that can be used to compare alternative interventions (including the status quo) and a clearly designed '*decision analytic model*' by which this ratio can be estimated. Economic evaluations

can usually be built into impact evaluations, but have rarely been used when assessing FP/RH interventions. While the cost-effectiveness of providing FP services to improve maternal health has been rigorously documented (see presentation by Darroch), decision-makers frequently lack information on how best to use funds within a FP/RH programme, how interventions are most efficiently delivered in various settings, and how related services can be cost-effectively integrated.

Jacqueline [Darroch](#) presented the economic evaluation (known as 'Adding it Up') that is used by DFID, USAID, UNFPA, and others to explain the comparative advantage of investing in FP/RH services [[Singh et al.](#)] and to justify such investments to governments and other donors. The evidence from this analysis has also been used effectively in supporting the rationale for the FP2020 initiative. The 'decision analytic model' for this economic evaluation compares various scenarios in terms of the impact of meeting women's needs for contraception on averting undesired events (e.g. unintended pregnancies, maternal deaths/DALYs, neonatal and infant deaths) and then estimates the cost per 'event averted.'

Synthesising and summarising bodies of evidence

Systematic reviews of bodies of evidence are usually considered the highest-ranked methodology for synthesising evidence of intervention **effectiveness** (see [Hardee](#)) for narrowly defined measures of impact and for relatively simple interventions that can be feasibly evaluated by an RCT. As indicated above, RCTs have numerous limitations for evaluating FP/RH interventions, given their complexity and need to adapt to specific contexts. Rachel [Isba](#) described the benefits and shortcomings of using the standard systematic review process (based on the experiences of the Cochrane Infectious Diseases Group), and compared this with other review processes commonly used by DFID, USAID, and others, such as 'expert reviews' and 'structured literature reviews'. While recognising that systematic reviews synthesise evidence drawn primarily from RCTs, Isba outlined conditions under which evidence from quasi-experimental and non-randomised designs could be incorporated into standard systematic reviews, by ensuring that the risks of bias and confounders are carefully assessed when summarising the findings. The utility of formal systematic reviews for synthesising evidence of the impact and implementation of FP/RH interventions stimulated extensive discussion and revealed differences of opinion. The recommendations presented later represent a meeting consensus, but one that was not unanimous.

Communicating evidence-informed recommendations to decision-makers needs to be undertaken carefully, especially when 'low'-quality evidence supports a 'strong' recommendation or 'high'-quality evidence supports a 'weak' recommendation. Indeed, the terminology used to describe the relative 'strength' of evidence and of the recommendations that can be made stimulated intense discussion. Will [Evans](#) discussed modalities for summarizing and presenting a body of evidence, including detailed narratives proposed by DFID for describing the strength of the evidence, and diagrams that can present evidence strength visually. For DFID, the strength of evidence should be described and judged in terms of five domains: the size of the body of evidence; its quality; the context to which the evidence applies; the consistency of the evidence; and the diversity of research methods on which the evidence is based.

There are many ways in which a body of evidence can be translated into programmatic recommendations. Joshua [Vogel](#) described the ‘Grading of Recommendations Assessment, Development and Evaluation’ (GRADE) approach used by WHO, which has two sequential phases: i) assessing the quality of the evidence (high, moderate, low, very low) in terms of the extent to which a decision-maker can be confident that an estimate of effect or association is correct as determined by the methodological quality of evidence and the likelihood of bias (all evidence from non-RCTs is considered low quality); and ii) the strength of the recommendation that can be made for or against using the intervention based on this evidence, usually categorised as strong or weak/ conditional/ qualified. WHO is also using the ‘Developing and Evaluating Communication Strategies to Support Informed Decisions and Practice Based on Evidence’ (DECIDE) framework [\[Vogel, J. et al.\]](#) to improve the dissemination of recommendations developed through GRADE. The framework includes 10 criteria (seriousness of the problem, number of people affected, quality of the evidence, size of the benefits, size of the adverse effects, resource use (costs), value for money, impacts on equity, feasibility, and acceptability) that help determine which health system decisions can be made, by using evidence to inform judgements about each criterion.

Mary Lyn [Gaffield](#) provided further insights into the procedures followed by WHO in developing practice guidelines, which are defined as any document that contains WHO recommendations about health interventions. A recommendation “provides information about what policy-makers, health-care providers or patients should do. It implies a choice between different interventions that have an impact on health and that have ramifications for the use of resources.” Various types of guidelines have been developed, all of which have to follow a standard process of development [\[WHO 2012\]](#) that is managed by a Guideline Development Group, reviewed by an External Review Group, formulates PICO (population, intervention, comparator, outcome) questions to be addressed, uses GRADE, and can take 3–37 months.

David [Ross](#) described another method for synthesising evidence and developing recommendations that adapted the principles of systematic review to summarise evidence to inform specific questions posed by policymakers. This method synthesises evidence from evaluations of specific interventions designed for particular populations and delivered in defined contexts. It assumes that different intervention types will need different strengths of evidence in order to recommend scaling up, and so assesses the evidence in terms of whether it reaches a particular threshold from which a recommendation can be made. The evidence is reviewed in relation to a hierarchy of contexts, firstly determining whether there is sufficient evidence of the intervention’s effect within a specific population group and context, and then assessing evidence from the intervention in similar populations in other contexts. The strength of evidence available for each type of intervention / context is then assessed using a hierarchy similar to that outlined by Hardee, and a decision is made as to whether the threshold of evidence needed for a recommendation on implementation and scale-up has been reached based on six domains (feasibility, cost, potential for adverse outcomes, acceptability, potential size of the effect, and other health or social benefits). Depending on the threshold of evidence reached, the recommendation is framed as ‘Steady’ (more evidence needed), ‘Ready’ (proceed with careful M&E), ‘Go’ (proceed with widespread implementation), or ‘Do not go.’

The fields of implementation science (IS) and health policy and systems research (HPSR) offer several methodologies that can generate evidence to enable decision-makers to understand *why* and *how* interventions work. Nhan [Tran](#) provided an overview of implementation research which draws on both fields. Health systems are unique and continually changing. Consequently, intervention implementation is necessarily adaptive, iterative, and dynamic; and decisions need to be *evidence-informed* rather than *evidence-based*, drawing from a range of options and accounting for factors other than solely research-based evidence. Implementation research is the scientific study of the processes used in the implementation of interventions as well as the contextual factors that affect these processes. The evidence generated and summarised is used to support and promote the effective application and scale-up of interventions that have been demonstrated to be effective in a range of contexts. The research design and methods used for implementation research, and thus the quality of evidence generated, depend entirely on the nature of the implementation question being addressed; [Peters et al.](#) provide a review of approaches used for implementation research. Standards for reporting and publishing implementation research results have been developed, e.g. SQUIRE [[Davidoff et al.](#)], which can also be used to assess the quality of evidence generated; these focus on clarity of description and appropriateness of the research for the implementation question being addressed.

Because systematic reviews of RCTs are not necessarily appropriate for informing decisions on *how* an FP/RH intervention should best be designed and implemented, alternative approaches were discussed for synthesising the evidence from implementation research. Marjolein [Dieleman](#) presented the realist synthesis approach, which seeks to understand and summarise what characteristics of an intervention enable it to work (or not), for whom, and in what circumstances. A realist synthesis starts by specifying the theory of change that explains how an intervention is expected to work and assumes that FP/RH interventions may be implemented differently depending on the context and the implementers. A systematic review process is then followed to identify and extract evidence from published studies on the intervention. This body of evidence is analysed by searching for patterns in the way the mechanism functions and the outcomes it produces in different contexts. The search is guided by the theory of change and results in a description that incorporates both the outcomes associated with the intervention and how it functions in various contexts.

A common feature of the various approaches to summarising a body of evidence is the expectation that decisions should be based on an *evidence framework*, which uses multiple domains to arrive at a summary judgement of the evidence. Ian [Askew](#) summarised the key characteristics of a typical evidence framework as being systematic and rigorous, having transparent procedures to reach a summary judgement, and rating evidence across multiple domains. A study by [Luoto et al.](#), which summarised the same body of evidence using six different evidence frameworks (including those of the U.S. Community Preventive Services Task Force, UK National Health Service Health Development Agency, WHO (GRADE), and the Australian National Health and Medical Research Council), found that the choice of framework influences the type of recommendations that can be made. While variability in how the frameworks are applied may influence the evidence summary and recommendations, the choice of domains included in the framework and how they are rated and weighted are critically

important, especially for domains relating to implementation, context, and cost. Choice of domains to include in an evidence framework should be influenced by whether the decision requires evidence of efficacy (e.g. choice of service delivery intervention), of feasibility and implementation (e.g. population level impact of a service delivery model), or of sustainability at scale (e.g. context and cost).

SUMMARY AND RECOMMENDATIONS

Drawing from these presentations, three plenary discussions were held during which participants sought clarification and critiqued the issues presented. Discussions reflected a diversity of views on what constitutes high-quality evidence, which type of evidence is appropriate for which type of recommendations, and which approaches should be followed to translate evidence into practice recommendations. During each discussion, recommendations were made for consolidating standards of evidence and formulating guideline recommendations for high-impact practices in FP/RH programming and policy making as follows:

- Suggestions for designing implementation research to maximise the quality of evidence generated and its utility for decision-making to strengthen reproductive health practices;
- Recommendations for rigorously synthesising and grading bodies of evidence;
- Proposals to guide appropriate funding and research implementation structures and procedures to generate strong evidence for practice recommendations in FP/RH.

The recommendations from these discussions are presented below.

Designing implementation research to maximise the quality of evidence generated and its utility for decision-making

1. Much of the confusion concerning quality of evidence stems from applying an inappropriate standard of evidence to a particular research question. Most of the evidence generated and communicated for guiding FP/RH programming recommendations has sought to answer questions such as “what works?”, “does this work better than that?”, or “is it safe?”, i.e. questions of impact, effectiveness, and safety. While questions about intervention effectiveness are still being asked and require a standard of evidence with high internal validity, policymakers, programme managers, and funders are increasingly seeking to better understand how and why an intervention works; to strengthen implementation procedures and service quality; to determine whether an intervention is acceptable, appropriate, and affordable for different populations; and to forecast the resources required to sustain and scale-up effective interventions [[Hardee](#); [Cleland](#)].

- **Different decisions require different types of evidence generated through different study designs; consequently the utility of research-based evidence will depend on its capacity to inform a particular decision.**

2. The highest quality of evidence for evaluating the **impact and effectiveness of an intervention** is generated by designs that produce strong internal validity, i.e. which can demonstrate that an intervention is the only or main cause of the outcomes observed [[Hardee](#); [Hargreaves](#)]. Internal validity is strongest when the research compares the impact of an intervention on individuals or on 'clusters' (groups of individuals or units such as health facilities) that have been randomly assigned to receive or not receive the intervention, or to receive different interventions.
 - **When evaluating impact and effectiveness, and when randomisation does not compromise the utility of the findings and is affordable, randomisation should be used to reduce the possibility that factors other than the intervention influence the outcomes.**
3. However, randomisation is generally not feasible if an intervention is complex, if the implementation site has been pre-determined, or if it would be unethical in certain contexts. Moreover, implementation may require a specific context that cannot be replicated or scaled-up [[Campbell](#)]. Furthermore, intervention sites are often intentionally selected by funders, implementers, and other decision-makers, and so it may not be feasible to randomise study sites. A range of **quasi-experimental / non-randomised designs** use different means of comparison to reduce the threats to internal validity and so can produce high-quality evidence of intervention effectiveness and impact. Such designs include systematically matching intervention and comparison populations through use of intervention and control groups, and/or using statistical adjustments to create an equivalent comparison [[Hargreaves](#)].
 - **Evaluations of intervention effectiveness that cannot randomise the unit of analysis can instead use systematic comparisons of intervention and control groups through matching or statistical adjustment to generate high-quality evidence of impact. Evaluations that use neither randomisation nor systematic comparison generate lower-quality evidence that should be interpreted with caution.**
4. Designs to evaluate the impact of interventions, especially complex interventions with multiple outcomes, or to understand implementation procedures, should be guided by an explicit **theory of change (ToC)** [[Cleland](#); [Campbell](#)]. A ToC should clearly describe an intervention, how it is expected to influence a change, and the type and level of change that it can realistically achieve. Impact should also be measured in terms of equity in access to and use of services among the beneficiaries. A strong ToC also describes the implementation processes needed for the intervention to achieve the desired outcomes. Without a clear ToC and a study design to evaluate it, attribution of impact to a particular intervention cannot be proven.

- **Agreement on the theory of change and alignment of the evaluation design with the ToC should be reached between the implementing partner, evaluator, and funder before research is initiated.**
5. An economic evaluation measures the **costs** and **cost-effectiveness ratio** of implementation, that is, the incremental cost of achieving the impact by implementing the intervention compared with the existing service delivery approach and/or with other comparable interventions; this is usually measured as a cost per outcome unit (e.g. per unintended pregnancy averted) [Griffiths]. These costs and ratios can then be modelled to inform decisions on improving implementation efficiencies, forecasting resource needs for scale-up, and allocating resources between alternative interventions to avert the adverse events of unintended pregnancy [Darroch]. Incorporating evidence of costs into decision-making must be carefully balanced with ensuring that a cost-effective intervention also reduces inequities.
- **Whenever affordable, feasible, and appropriate, an impact evaluation should include an economic evaluation component so that the feasibility of scale-up and sustainability of a proven high-impact practice can be determined.**
6. When decision-makers need guidance on how a high-impact intervention can be implemented at scale, evidence can be generated through documenting both the **process of implementation** and the **contextual factors** influencing procedures for intervention implementation [Campbell; Tran]. Implementation research can provide evidence to inform guidance for institutionalizing and strengthening an intervention within a health system, for guiding and assessing quality improvements, and for supporting an intervention's scale-up and replication in other systems and contexts.
- **Implementation research using mixed-methods design should be used to generate high-quality evidence describing intervention procedures and context to inform decisions about scale-up and sustainability; the specific study design should be determined by the information needed by decision-makers.**

Recommendations for synthesising and grading bodies of evidence on reproductive health

1. Recommendations for guiding decisions on **practice effectiveness** are based on bodies of evidence that have been summarised and synthesised and whose findings have been rated by an expert panel. These processes of summarising and synthesising evidence on *practice effectiveness* draw from evidence with high internal validity, including from both randomised and rigorously implemented comparative non-randomised designs. Non-randomised designs should specify explicit inclusion and exclusion criteria, specify potential biases and confounders, and describe the probability of an intervention causing a desired impact [Isba; Ross].

- **Bodies of evidence that inform decision-makers on the effectiveness of interventions are best summarised using a transparent, structured review process that includes evidence from both randomised and rigorous non-randomised designs with systematic comparisons.**
2. Deciding whether to introduce or scale-up an intervention needs to be informed by evidence other than solely its effectiveness or cost-effectiveness, however. Practice recommendations for guiding the **introduction and scaling up** of interventions demonstrated to have high impact should also be informed by bodies of evidence derived from **implementation research** [Tran] that include feasibility, generalizability, and scalability [Ross]. While such bodies may include evidence from randomised and/or comparative designs, research methods that describe implementation procedures, and explain them within the context in which they are operating, usually provide higher-quality evidence for informing such 'how to' decisions. A rigorous review process should also be followed to summarise evidence from implementation research. A **realist synthesis** [Dieleman] that specifies a ToC and assumes that FP/RH interventions may be implemented differently depending on the context and the implementers can provide a structured framework for such reviews.
- **Bodies of evidence to inform implementation and scaling-up decisions can be derived from implementation research and economic evaluations. Highest-quality data are generated when the decision question is clearly stated and the research design tailored to generate evidence that will address that question.**
 - **Such bodies of evidence should be guided by a theory of change, reviewed rigorously, synthesised systematically, and summarised to inform implementation decisions identified by decision-makers.**
3. Summarising the key findings from a body of evidence into recommendations that can be translated into practice guidelines requires an **evidence framework** that rates evidence across multiple domains to arrive at a summary judgement [Askew]. Examples of such frameworks include GRADE [Vogel], "Steady, Ready, Go!" [Ross], and DFID's assessment framework [Evans]. The domains used within an evidence framework will influence the conclusions that can be drawn and the recommendations that can be made; thus careful attention must be paid to the configuration and clear explanation of the evidence framework used.
- **A systematic, transparent, and replicable process, guided by an explicit evidence framework, should be followed when developing practice recommendations from a body of evidence. The evidence framework should incorporate those domains that are of specific interest to particular decision-makers; different evidence frameworks may be appropriate for summarising evidence to inform different types of decisions.**

4. Great care is needed when interpreting an evidence summary and communicating the findings through narrative statements or diagrammatically to ensure that the various domains are clearly described and the rating accurately reflects the available evidence [Evans]. **Formulating recommendations** is usually undertaken by an expert group or panel that is convened for the purpose [Malarcher; Gaffield], and that follows systematic, rigorous, and fully transparent procedures. Communicating recommendations for 'what' to do will necessarily be framed differently than recommendations for 'how' to do it.
 - **Recommendation formulation should be carefully planned and implemented, using a representative and knowledgeable expert group and recommendation statements or diagrams that accurately and unequivocally represent the body of evidence available.**
 - **Given the diversity of contexts in which RH/FP interventions are implemented, recommendations for implementation should offer a choice of options – that is, should be 'evidence-informed' – rather than specify a single 'evidence-based' recommendation for addressing a particular need or problem.**

Funding and research implementation structures and procedures to generate quality evidence and strong recommendations

Panel discussions with representatives from donors (DFID; USAID; Bill and Melinda Gates Foundation; Packard Foundation; Alliance for Reproductive, Maternal and Newborn Health) and from research organisations (London School of Hygiene and Tropical Medicine; Population Council; FHI360; Columbia University; University of North Carolina; University of Ghana; Harvard University), followed by discussions in plenary, led to the following suggestions:

- External evaluations that separate the roles of donor, implementer, and researcher have been the norm for most donor-funded impact evaluations. Experience suggests that **funding and designing research on FP/RH interventions so that donors, researchers, and implementers can work jointly and with complementary roles produces evidence that is relevant and useful for decision-makers, without compromising the quality of evidence.**
- **Challenges encountered when implementing and evaluating interventions, especially complex interventions, must be documented so that lessons learned from 'failure' can be shared and inform the reconfiguration of an intervention's structure and procedures prior to scaling up.**
- **A decision regarding which research design is most appropriate for generating evidence to address decision-makers' specific needs should be made jointly by donors, researchers, and implementers through a consultative process prior to initiating research.** These consultations should be informed by an awareness of the variety of designs that can be used, and the type and quality of evidence they can generate.

- The evidence frameworks for rating evidence used by organisations that set service delivery and programming norms, and those that fund research and guideline development, vary considerably. **Reducing the variability among and consolidating the key domains and rating scales used in different evidence frameworks, together with more transparent evidence review processes, would reduce the likelihood of differing interpretations of a body of evidence, thereby reinforcing the strength of recommendations.**
- As country strategy plans are being reviewed, revised, and costed through initiatives such as *Every Woman, Every Child* and *FP2020*, and investments in large-scale expansion of FP/RH programme coverage are being made, there are multiple needs and opportunities for high-quality evidence on implementation. **Donors and researchers should prioritise investment in and application of implementation research to improve the quality of evidence to guide rapid scale-up of HIPs and other recommended FP/RH practices.**
- Despite decades of investment in building research capacity, a lack of individual and institutional capacity still exists for undertaking impact evaluations and implementation research on FP/RH interventions in most developing countries. **Donors, governments, and research organisations should increase investments in a wide range of skills-building and systems-strengthening activities to enhance and sustain national research capacities.** Priorities should include strengthening HMIS and M&E capacities of service delivery programmes, building researcher skills in design and analysis methodologies for generating internally valid data, and intensifying ethical and technical review processes for national research regulatory bodies.
- Communicating and packaging practice recommendations and guidelines so that they convey appropriate and valid messages and provide decision-makers with sufficient evidence to make a particular decision has been challenging. **Recommendations on FP/RH interventions should include guidance on the contexts and populations for which they may (or may not) be effective and reduce inequities, and on how best to implement them in such contexts. Recommendations should offer a range of options from which to choose, so that programming decisions are evidence-informed.**

LIST OF PARTICIPANTS

Name	Organisation	Position	Email address
Richard Adanu	University of Ghana	Dean, School of Public Health, University of Ghana	rmadanu@yahoo.com
Sabaratnam Arulkumaran	International Federation of Gynaecologists and Obstetricians (FIGO)	President	sarulkum@sgul.ac.uk
Ian Askew	Population Council	Co-Research Director, STEP UP	iaskew@popcouncil.org
Angela Baschieri	Department for International Development (DFID)	Health and Population Advisor	A-Baschieri@dfid.gov.uk
Neal Brandes	United States Agency for International Development (USAID)	Health Research Advisor, Health Specialist	nbrandes@usaid.gov
Win Brown	Bill and Melinda Gates Foundation	Senior Program Officer, Family Health Program	Win.Brown@gatesfoundation.org
Oona Campbell	London School of Hygiene and Tropical Medicine (LSHTM)	Professor of Epidemiology and Reproductive Health	Oona.Campbell@lshtm.ac.uk
John Cleland	LSHTM	Emeritus Professor of Medical Demography	John.Cleland@lshtm.ac.uk
Jacqueline Darroch	Guttmacher Institute	Senior Fellow	JEDarroch@yahoo.com
Marjolein Dieleman	Royal Tropical Institute (KIT)	Senior Advisor, Development Policy & Practice	m.dieleman@kit.nl
Will Evans	DFID	Governance Adviser	w-evans@dfid.gov.uk
Veronique Filippi	LSHTM	Reader in Maternal Health and Epidemiology	Veronique.Filippi@lshtm.ac.uk
Mary Lyn Gaffield	World Health Organisation (WHO)	Epidemiologist	gaffieldm@who.int

Name	Organisation	Position	Email address
Ulla Griffiths	LSHTM	Lecturer in Health Economics	Ulla.Griffiths@lshtm.ac.uk
Metin Gülmezoglu	WHO	Coordinator, Maternal and Perinatal Health and Preventing Unsafe Abortion	gulmezoglum@who.int
Karen Hardee	Futures Group	Senior Fellow and Deputy Director, Health Policy Project	KHardee@popcouncil.org
James Hargreaves	LSHTM	Senior Lecturer, Social Epidemiology and Public Health Evaluation	James.Hargreaves@lshtm.ac.uk
Maggie Hobstetter	David and Lucile Packard Foundation	Research Analyst, Population and Reproductive Health Program	MHobstetter@packard.org
Rachel Isba	Liverpool School of Tropical Medicine (LSTM)	Clinical Lecturer	R.Isba@liverpool.ac.uk
Jill Keesbury	Alliance for Reproductive, Maternal and Newborn Health	Director	jkeesbury@path.org
Sue Kinn	DFID	Team Leader and Research Manager, Human Development Research Team	S-Kinn@dfid.gov.uk
Ana Langer	Harvard University	Director, Women and Health Initiative, Maternal Health Task Force	alanger@hsph.harvard.edu
Kazuyo Machiyama	LSHTM	Research Fellow	kazuyo.machiyama@lshtm.ac.uk
Shawn Malarcher	USAID	Senior Best Practices Utilisation Advisor, Office of Population and Reproductive Health	smalarcher@usaid.gov

Name	Organisation	Position	Email address
Heidi Marriott	International Planned Parenthood Federation (IPPF)	Senior Adviser, Organizational Learning and Evaluation	hgagnebemarriott@ippf.org
Kirsty Newman	DFID	Team Leader, Evidence into Action Team	KC-Newman@DFID.gov.uk
Thoai Ngo	Marie Stopes International (MSI)	Head of Research	Thoai.Ngo@mariestopes.org
James F. Phillips	Columbia University	Professor of Clinical Population and Family Health	jfp2113@cumc.columbia.edu
Timothy Powell-Jackson	LSHTM	Lecturer in Health Economics	timothy.powell-jackson@lshtm.ac.uk
Carine Ronsmans	LSHTM	Professor and Head of Infectious Disease Epidemiology	Carine.ronsmans@lshtm.ac.uk
David Ross	LSHTM	Professor of Epidemiology & International Public Health	david.ross@lshtm.ac.uk
James Shelton	USAID	Science Adviser, Bureau for Global Health	jshelton@usaid.gov
Ilene Speizer	University of North Carolina (UNC)	Research Professor	speizer@email.unc.edu
John Townsend	Population Council	Vice President, Reproductive Health	jtownsend@popcouncil.org
Nhan Tran	Alliance for Health Policy and Systems Research	Manager, Implementation Research Platform	trann@who.int
Johannes van Dam	FHI360	Director, Program Sciences/GHPN	JVanDam@fhi360.org
Joshua Vogel	WHO	Scientist	vogeljo@who.int

LIST OF PRESENTATIONS

[Askew, I. "Using evidence-based recommendations for guidance, guidelines, and scale-up strategies."](#)

[Baschieri, A. "Using evidence to guide investing in and programming of reproductive health services."](#)

[Campbell, O. "Design issues arising when evaluating reproductive health interventions in the real world."](#)

[Cleland, C. "Quality of research on family planning interventions."](#)

[Darroch, JE. "Evidence from scenario modelling for practice recommendations."](#)

[Dieleman, M. "Realist synthesis: Understanding health system interventions."](#)

[Evans, W. "Summarising bodies of evidence."](#)

[Gaffield, ML. "WHO guidelines and guideline procedures."](#)

[Griffiths, U. "Economic evaluation of reproductive health interventions: Overview of design options."](#)

[Hardee, K. "Hierarchical ratings of study designs."](#)

[Hargreaves, J. "Some quantitative evaluation design ideas with examples."](#)

[Isba, R. "Systematic reviews: Benefits and shortcomings."](#)

[Malarcher, S. "Standards of evidence for reproductive health programming: Issues identified by the HIP TAG."](#)

[Ross, D. "Steady...Ready... Go!: A method for synthesizing evidence for policy makers based on systematic reviews."](#)

[Tran, N. "Needs and challenges in summarizing evidence relating to implementation and scale up of reproductive health services."](#)

[Vogel, J. "Using the GRADE and DECIDE frameworks."](#)

REFERENCES

- [Davidoff, F., Batalden, P., Stevens, D., Ogrinc, G., Mooney, S. 2008. "Publication guidelines for quality improvement in health care: Evolution of the SQUIRE project." *Quality & Safety in Health Care* 17: i3-i9.](#)
- [Luoto, J., Maglione, M.A., Johnsen, B., et al. 2013. "A comparison of frameworks evaluating evidence for global health interventions." *PLoS Medicine* 10:7: e1001469.](#)
- [Mwaikambo, L., Speizer, I.S., Schurmann, A., Morgan, G., Fikree, F. 2011. "What works in family planning interventions: A systematic review." *Studies in Family Planning* 42\(2\): 67-82.](#)
- [Peters, D.H., Tran, N.T., Adam, T. 2013. *Implementation research in health: A practical guide*. Alliance for Health Policy and Systems Research and World Health Organisation.](#)
- [Petticrew, M., Roberts, H. 2003. "Evidence, hierarchies, and typologies: Horses for courses." *Journal of Epidemiology and Community Health* 57: 527-529.](#)
- [Singh, S., Darroch, J., Ashford, L. 2013. *Adding It Up: The Need for and Cost of Maternal and Newborn Care—Estimates for 2012*. New York: Guttmacher Institute.](#)
- [Vogel, I. 2012. *Review of the use of 'Theory of Change' in international development*. DFID.](#)
- [Vogel, J.P., Oxman, A.D., Glenton, C., et al. 2013. "Policymakers' and other stakeholders' perceptions of key considerations for health system decisions and the presentation of evidence to inform those considerations: an international survey." *Health Research Policy and Systems* 11: 19.](#)
- [World Health Organisation. 2012. *WHO Handbook for Guideline Development*. Geneva: WHO.](#)

www.stepup.popcouncil.org

The STEP UP (Strengthening Evidence for Programming on Unintended Pregnancy) Research Programme

Consortium generates policy-relevant research to promote an evidence-based approach for improving access to family planning and safe abortion. STEP UP focuses its activities in five countries: Bangladesh, Ghana, India, Kenya, and Senegal.

STEP UP is coordinated by the Population Council in partnership with the African Population and Health Research Center; icddr,b; the London School of Hygiene and Tropical Medicine; Marie Stopes International; and Partners in Population and Development. STEP UP is funded by UK aid from the UK Government.