



Organizational unit: Evaluation Office **Year of report:** 2022

Title of evaluation report: MID-TERM EVALUATION OF THE MATERNAL AND NEWBORN HEALTH THEMATIC FUND PHASE III, 2018-2022

Overall quality of report: **Very good** **Date of assessment:** 15 June 2022

Overall comments: The evaluators used a theory-based and country case-study approach to conduct a thorough analysis of UNFPA's Maternal and Newborn Health Thematic Fund (MHTF) contributions to country health systems. The report is well written and clearly highlights the successes of the MHTF and the contextual factors that support and hinder achievement in enabling countries to meet global standards. In accordance with good practice the theory of change was adapted, clearly explained, and used to structure the evaluation questions and the analysis. Data was gathered through multiple methods (including through global and country-level interviews, group discussions, a survey and site visits/observations) and meticulously recorded in a detailed evaluation matrix. The findings are comprehensive and thoroughly investigate issues of gender and equity. The conclusions and recommendations are well formulated and appear very useful for decision makers. The report could have been more clear about the methodology including the total number of evaluation participants and the sampling strategy used to select them. It could also have provided more information on the rightsholders who participated and on ethical considerations adhered to, as well as being more concisely written as both the overall report and the Executive Summary exceeded the preferred maximum page lengths.

Disability inclusion was a consideration of the evaluation and was reflected in the evaluation questions. It is not clear if people with disabilities (PWD) were included as evaluation participants. The assessment of disability was briefly included as part of the discussion on leave no one behind where two examples of UNFPA's work in reaching PWD are noted. Under conclusions, disability is not explicitly mentioned but is covered under broader issues of reach and equity. There is specific mention as part of the rationale for one of the ten recommendations.

Assessment Levels **Very Good** strong, above average, best practice **Good** satisfactory, respectable **Fair** with some weaknesses, still acceptable **Unsatisfactory** weak, does not meet minimal quality standards

Quality Assessment Criteria *Insert assessment level followed by main comments. (use 'shading' function to give cells corresponding colour)*

I. Structure and Clarity of Reporting	Yes No Partial	Assessment Level:	Good
--	----------------------	-------------------	-------------

To ensure the report is comprehensive and user-friendly

I. Is the report structured in a logical way? Is the report easy to read and understand (i.e. written in an accessible language appropriate for the intended audience) with minimal grammatical, spelling or punctuation errors? Is there a clear distinction made between analysis/findings, conclusions, recommendations and lessons learned (where applicable)?	Yes	The report is very well written. It has a logical structure with clear distinction between sections.
---	-----	--

<p>2. Is the report of a reasonable length? (maximum pages for the main report, excluding annexes: 60 for institutional evaluations; 70 for CPEs; 80 for thematic evaluations) <i>Note: YES - the report is within the indicated maximum page length. PARTIAL - the report exceeds the maximum page limit by 1- 5 pages. NO - the report exceeds the maximum page limit beyond 5 pages.</i></p>	No	The report is longer than the preferred length as it is 96 pages including the Executive Summary.
<p>3. Do the annexes contain – at a minimum – the ToRs; a bibliography; a list of interviewees; the evaluation matrix; methodological and data collection tools used (e.g. interview guides; focus group notes, outline of surveys)? <i>Note: YES - the report contains all the annexes indicated. PARTIAL - if the report is missing the ToRs or the bibliography in the annexes. NO - if the report is missing any of the following annexes: a list of interviewees; the evaluation matrix; methodological and data collection tools used (e.g. interview guides; focus group notes, outline of surveys).</i></p>	Yes	The annexes include a very detailed evaluation matrix spanning over 270 pages, the bibliography, list of interviewees and data collection tools.
<i>Executive summary</i>		
<p>4. Is an executive summary written as a stand-alone section, presenting the i) Purpose; ii) Objectives, scope and brief description of interventions; iii) intended audience; iv) Methodology; v) Main results; Vi) Conclusions and Recommendations?</p>	Yes	The summary is well presented and covers the necessary components for it to be a stand-alone document.
<p>5. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)? <i>Note: YES - the executive summary is within the indicated maximum page limit. PARTIAL - the executive summary exceeds the maximum page limit by 1 to 2 pages. NO - the executive summary exceeds the maximum page limit beyond 2 pages.</i></p>	No	The summary extends over 8 pages. It would be more concise if the findings and conclusions sections were combined as these are both quite lengthy and have some overlap.
<p>2. Design and Methodology</p>	<p>Yes No Partial</p>	<p>Assessment Level: Very good</p>
<i>To ensure that the evaluation is put within its context</i>		
<p>1. Is the development and institutional context of the evaluation clearly described and constraints explained?</p>	Yes	There is a sophisticated and well-referenced (although somewhat lengthy at 15 pages) context section that presents the situation analysis and the MHTF programme, including an indepth explanation of its financial resources and expenditures.
<p>2. Does the evaluation report discuss and assess the intervention logic and/or theory of change?</p>	Yes	The evaluators applied a systematic theory-based approach to review and revise the ToC. It was then used to develop the evaluation questions. The evaluator's process is described and then the overall ToC is presented in Section 3.2. Table 4 provides instructions for reading the refined and very comprehensive ToC; it is accompanied by a separate figure that maps the assumptions to the main outcomes.

<i>To ensure a rigorous design and methodology</i>		
<p>3. Is the evaluation framework clearly described in the text and in the evaluation matrix? Does the evaluation matrix establish the evaluation questions, assumptions, indicators, data sources and methods for data collection?</p>	Yes	The annexed evaluation matrix includes questions, subquestions, related criteria, indicators and assumptions as well as very detailed and meticulously documented findings for each source of data.
<p>4. Are the tools for data collection described and their choice justified? <i>Remember: Please address both aspects of this sub-criteria in the comment: 1) are data collection tools described (i.e. documentary review, interviews, focus group discussions etc.) and 2) is the rationale for their selection detailed</i></p>	Partial	A carefully explained case study approach was used. The data collection processes (document review, in-country interviews/group discussions, global/regional interviews, an online survey and site visits/observations) are all sufficiently described, as is the rationale for their use. However, although the number of survey respondents is identified as being 238, the number of participants involved in interviews and discussions is not provided. It is good practice to state the total number of data sources, preferably by data collection method.
<p>5. Is there a comprehensive stakeholder map? Is the stakeholder consultation process clearly described (in particular, does it include the consultation of key stakeholders on draft recommendations)? <i>Remember: Please address all three aspects of this sub-criteria in the comment: 1) is a comprehensive stakeholder map included (in either the report itself or the annexes) 2) Is the overall stakeholder consultation process described and 3) within the consultation process were key stakeholders consulted on the recommendations specifically? 4) does the evaluation stakeholder mapping and data collection methods involve vulnerable and marginalized groups, including persons with disabilities and their representative organizations?</i></p>	Yes	The context section includes a discussion of the global architecture of the programme and the organizations that are part of this effort at the global level. Although there is no mention of stakeholder consultation specifically on the recommendations, it is noted that preliminary findings were presented to local stakeholders at the end of field missions in each country. This provided an opportunity to comment on and validate the findings, and demonstrates good practice.
<p>6. Are the methods for analysis clearly described for all types of data? <i>Remember: This sub-criteria is asking about the data analysis methods used and whether they are clearly described - was contribution analysis used, or qualitative comparative analysis, for example, or descriptive statistics? Triangulation is not a method of analysis; it is a validation technique.</i></p>	Yes	Contribution analysis was used as the overall approach. It is described, as are the ways that data was organized for each data collection method.
<p>7. Are methodological limitations acknowledged and their effect on the evaluation described? Does the report discuss what was done to minimize such issues?</p>	Yes	The limitations and associated responses are well articulated. The pandemic was noted as the main limitation, particularly for conducting country field studies; issues such as the diverse contexts limiting the generalizable nature of findings and results are also raised.
<p>8. Is the sampling strategy described? <i>This criteria is asking whether the methodological approach to determining the sample of stakeholders consulted and the sample of site visits is described. Reviewers should examine whether the evaluation report includes information on how the universe was determined; the sampling approach used (i.e. purposive); the indicators used to develop the sample to be consulted (or visited); the resulting sample; and importantly limitations to the approach (including any potential resulting bias).</i></p>	Partial	The sampling strategy is partially explained. The sampling considerations are described very clearly for the selection of countries for the case studies (table 5 highlights some of these features). However there is no discussion on how participants were selected for KIIs or group discussions, or how in-country site visit destinations were determined. In particular, the report could be more clear about the number and type of rightsholders and especially vulnerable groups and/or their representatives who were consulted, In respect to the survey, it is noted that the sample frame was developed with programme focal points and the respondents are shown by stakeholder group but no further information on the selection criteria is provided.

<p>9. Does the methodology enable the collection and analysis of disaggregated data? <i>Remember: The default should be to disaggregate by sex. Whenever possible, this sub-criterion is also asking systematic disaggregation of data related to population groups (e.g. persons with disability) where there are implications related to UNFPA's portfolio/interventions for these population groups.</i></p>	Partial	Differences are primarily shown by country. The evaluators had clear intent to use more precise disaggregated monitoring and reporting data but one limitation was that this type of data was not available. Nonetheless, more could have been done in respect to the disaggregation of primary data - the total number of evaluation participants is not disaggregated and the annexed data collection tools do not have a place for indicating gender of respondents.
<p>10. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, disability inclusion, gender equality and human rights)? <i>Remember: This sub-criteria is asking about the evaluation methodology itself – specifically does the evaluation's design / methodology ensure that the evaluation is able to assess the extent to which the country programme integrates cross-cutting issues across its portfolio of work? Therefore, we're looking to see whether, for example, evaluation questions or indicators assess/capture the extent to which a human rights based approach to the development and implementation of the country programme was used (i.e. whether the evaluation queries/assesses whether beneficiaries/partners were consulted and through design process of the country programme); or whether the evaluation's data collection methods capture the voices/perspectives of a range of stakeholders include beneficiaries/vulnerable/marginalized groups.</i></p>	Yes	The evaluators explain that they used an equity-focused approach and that stated that, "Where possible, the evaluation team identified evidence that highlighted different beliefs and practices related to gender roles, ethnicity, age and disability." Marginalized groups, including vulnerable women and girls, were involved as evaluation participants to the extent feasible given the described limitations of access.
<p>3. Reliability of Data</p>	<p>Yes No Partial</p>	<p>Assessment Level: Very good</p>
<p><i>To ensure quality of data and robust data collection processes</i></p>		
<p>1. Did the evaluation triangulate data collected as appropriate?</p>	Yes	The approach to triangulation is well explained in section 3.3.4. Use of multiple data sources is clearly shown in the evaluation matrix.
<p>2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources? <i>Remember: Please address both parts of this sub-criteria, namely do the evaluators identify the sources of the qualitative and quantitative data they used and do they discuss the reliability (or lack thereof) of both?</i></p>	Yes	The data sources appeared appropriate. Evaluators noted the limitations of generalizing the findings based on the sample size.

<p>3. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations? <i>Remember: Ensure that reviewers are assessing the sub-criteria is concerned with whether there is evidence in the report that evaluators' approach to data collection was sensitive to ethical considerations (i.e. consent, confidentiality, etc.) and were not discriminatory against particular groups' participation (i.e. were interviews or focus groups held in a location, at a time, in a setting, using language/translation, that is appropriate and respectful; and facilitates the participation of a full range of stakeholders, including persons with disability). We are also interested if evaluators noted limitations in this regard.</i> <i>Note that mentioning/referencing UNEG standards in the report does not amount to evidence that the data was actually collected with a sensitivity to ethics and discrimination; the reviewer should assess whether there is evidence in the report of the UNEG standards actually being implemented. If the UNEG documents/standards are referenced in the text, but the evaluators do not explain how/show evidence of the data being collected with a sensitivity to issues of discrimination, etc., this should be a "no".</i></p>	Partial	The preamble to the data collection tools discuss anonymity and confidentiality. However, there is no further discussion of ethical practices adhered to or of UNEG ethical guidance being followed.
<p>4. Analysis and Findings</p>	<p>Yes No Partial</p>	<p>Assessment Level: Very good</p>
<p><i>To ensure sound analysis and credible findings</i></p>		
<p>1. Are the findings substantiated by evidence?</p>	Yes	There is general references to data sources in the main report. Under the summarized box of findings for each criterium, the evaluators direct readers to the sections of the evaluation matrix where more detailed evidence for supporting those findings can be found.
<p>2. Is the basis for interpretations carefully described?</p>	Yes	The findings are very well articulated and explained.
<p>3. Is the analysis presented against the evaluation questions?</p>	Partial	Although the findings associated with each evaluation question are shown in the evaluation matrix, the questions are not provided as a way of framing the analysis in the main text of the report.
<p>4. Are the cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted? <i>Remember: Please address both parts of the sub-criteria in the comment, namely: are the cause/effect links (between UNFPA contribution and outputs/contribution to results/outcomes) explained as well as the results and 2) are unintended outcomes discussed. On the latter, please note in the comment whether evaluators considered/looked for unintended outcomes and noted whether there were (or were not) any; or whether the report does not mention unintended outcomes.</i></p>	Yes	Causal relationships are discussed. The findings present many examples of MHTF activities in different countries that have led to improved outcomes (i.e., the discussion on catalytic investments in section 4.7.5). The evaluation also points out the need for monitoring and reporting to go beyond tracking of outputs to focus more on outcomes (p. 70). Unintended consequences are brought out, an example being the conflicts between older midwives who were issued certificates and new midwives who now have degrees.
<p>5. Does the analysis show different outcomes for different target groups, as relevant?</p>	Yes	Differences in achievement of outcomes for different countries is shown. For example, there is a discussion about Zambia being the first to develop respectful maternity care guidelines and awareness of this issue is noted as being absent in Benin and Sudan.

6. Is the analysis presented against contextual factors?	Yes	This is done throughout. An example is the extent of institutionalization of the Emergency obstetric and newborn care (EmONC) network methodology. Covid is highlighted as a major hindering factor in being able to deploy teams to provide expertise, but other factors in each country are also explained.
7. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, disability inclusion, gender equality and human rights?	Yes	Issues of gender equality and GBV are mainstreamed into the analysis (for example, the discussion on midwives, gender attitudinal barriers and respectful care on p 55). There is also a specific section on Strengthening Access and Equity (section 4.6). Evaluators note where disability inclusion is evident in national programmes, for example this was seen in pre- and in-service training of health care providers in Zambia (p. 55), and make further reference to disability, including those with fistula, under the discussion on leaving no one behind (p. 59).
5. Conclusions	Yes No Partial	Assessment Level: Very good
<i>To assess the validity of conclusions</i>		
1. Do the conclusions flow clearly from the findings?	Yes	The conclusions are not explicitly linked to the questions or criteria and, at 6.5 pages, are somewhat longer than usual. However, they do flow logically from the findings.
2. Do the conclusions go beyond the findings and provide a thorough understanding of the underlying issues of the programme/initiative/system being evaluated and reflect as appropriate cross-cutting issues such as equality and vulnerability, disability inclusion, gender equality and human rights?	Yes	The conclusions are well formulated and provide a higher level view of the results discussed in the findings. The strengths and challenges of six main issues in the health system are clearly highlighted in table format. There is a specific conclusion on the success of the programme in addressing cross-cutting issues. Although it does not specify PWD, it does adequately cover leave no one behind which is the section in which disability inclusion was covered in findings.
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	The conclusions are sufficiently evidence based and therefore do not appear to be biased.
6. Recommendations	Yes No Partial	Assessment Level: Very good

To ensure the usefulness and clarity of recommendations			
1. Do recommendations flow logically from conclusions?	Yes	Each recommendation shows the conclusion(s) on which it is based.	
2. Are the recommendations targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?	Yes	The recommendations are clearly structured. For each, the recommendation is succinctly stated, the rationale is provided as well as to whom they are directed, and a section on operational implications is provided. Financial implications are addressed where relevant.	
3. Do recommendations appear balanced and impartial and address, as relevant, key cross cutting issues such as equity and vulnerability, disability-inclusion, gender equality and human rights?	Yes	The recommendations appear balanced as they clearly seek to address the shortcomings found. Cross-cutting issues of gender and equity are covered under multiple recommendations. Disability inclusion is explicitly addressed as part of the rationale of recommendation 6.8 on investing more in the programme's core values.	
4. Are the recommendations prioritized?	Yes	They are prioritized as high, high-medium, and medium.	
7. Gender			
	0 1 2 3 (**)	Assessment Level:	Very good
To assess the integration of Gender Equality and Empowerment of Women (GEEW) (*)			
1. Is GEEW integrated in the evaluation scope of analysis and indicators designed in a way that ensures GEEW-related data to be collected? <i>Remember: Ensure that the reviewer does not confuse the content of the country programme with the approach to the evaluation. This question is asking whether the evaluation criteria and questions are gender responsive and inclusion of GEWE dimensions in its scope of analysis (i.e. in the objectives for example) or the indicators the evaluation selects against which data will be collected so that the evaluation is able to assess whether the country programme is gender responsive.</i> <i>A general note on UNFPA programming: While there may be evidence of gender being referred to as a cornerstone of UNFPA programming - in the sense that most UNFPA programmes target women and girls - this does not necessarily mean that UNFPA's work is gender/human rights responsive. GEEW is about power and shifting resources, social norms, attitudes, laws and policies. One could work on comprehensive sexuality education, for example, in a way that further entrenches gendered norms or power dynamics (i.e. pathologizing LGBTQ communities; or reifying gender binaries by assuming heteronormativity); this would not be GEEW sensitive. Another example: one could deliver sexual and reproductive health care that fails to adequately address the diverse health needs of women (i.e. women who are disabled, older women, LGBTQI women; conceptualizes women as mothers alone (meaning through the spectrum of reproductive rights only, excluding their sexual rights); and/or holding biases against contraceptive options; again this would not be GEEW sensitive.</i>		<p>a. Does the evaluation include an objective specific to assessment of human rights and gender equality considerations or was it mainstreamed in other objectives? (Score: 0-3) The purpose includes the assessment of the MHTF's contribution the promotion of HRGE in the context of newborn and maternal health. = 3</p> <p>b. Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria? (Score: 0-3) The assessment of accessible and accountable access is considered under four criteria. = 3</p> <p>c. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation? (Score: 0-3) Question 6 focuses on equity and access including for PWD. = 3</p> <p>d. Does the evaluation assess whether sufficient information was collected during the implementation period on specific result indicators to measure progress on human rights and gender equality results?(Score: 0-3) The limitations section highlights that monitoring and reporting data did not differentiate among different socially isolated on marginalized groups. = 3</p>	

<p>2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques?</p> <p><i>Remember: Ensure that the reviewer does not confuse the content of the country programme with the approach to the evaluation. This sub-criteria is asking whether the evaluation criteria and evaluation questions (i.e. the evaluation itself) are gender responsive; in other words, are the criteria interpreted/operationalized and evaluations questions developed in a way that is able to capture whether (or not) gender equality/human rights/the empowerment of women has been integrated into UNFPA's country programme/support (in the design/planning, implementation and results)?</i></p>		<p>a. Does the evaluation specify how gender issues are addressed in the methodology, including: how data collection and analysis methods integrate gender considerations and ensure data collected is disaggregated by sex? (Score: 0-3) Section 3.3.5 is focused on how the evaluation addressed HRGE and the limitations faced. = 3</p> <p>b. Does the evaluation methodology employ a mixed-methods approach, appropriate to evaluating GEEW considerations (collecting and analyzing both quantitative and qualitative data, and ensuring the appropriate sample size)? (Score: 0-3) A mixed-methods approach was used but there was minimal presentation of quantitative data beyond what was provided in the background/context section. There could also have been more clarity on sample size. = 2</p> <p>c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility? (Score: 0-3) The evaluation matrix clearly shows triangulation. = 3</p> <p>d. Do the evaluation methods and sampling frame address the diversity of stakeholders affected by the intervention, particularly the most vulnerable, where appropriate? (Score: 0-3) Rights holders were included as evaluation participants within the limitations in reaching these groups due to the pandemic. However, there was no description of the types of rights holders involved. = 1</p> <p>e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality? (Score: 0-3) Although data collection tools show evidence of confidentiality and anonymity being respected, there is no other mention of ethical considerations. = 2</p>
<p>3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?</p>		<p>a. Does the evaluation have a background section that includes an intersectional analysis of the specific social groups affected by the issue or spell out the relevant normative instruments or policies related to human rights and gender equality? (Score: 0-3) The context section includes a concise but solid intersectional analysis of the sexual health issues, GBV including for girls, the social and economic determinants of health, and the combined affect on the most vulnerable populations. = 3</p> <p>b. Do the findings include data analysis that explicitly and transparently triangulates the voices of different social role groups, and/or disaggregates quantitative data, where applicable? (Score: 0-3) Gender issues are reflected throughout the findings. The perspectives of different groups are evident in the evaluation matrix but to a lesser extent in the report. Participant quotes were included in the report but they were not attributed to any stakeholder group. = 2</p> <p>c. Are unanticipated effects of the intervention on human rights and gender equality described? (Score: 0-3) One unanticipated effect related to HRGE is addressed, as noted above. = 3</p> <p>d. Does the evaluation report provide specific recommendations addressing GEEW issues, and priorities for action to improve GEEW or the intervention or future initiatives in this area? (Score: 0-3) GEEW is addressed in multiple recommendations. = 3</p>
<p>(*) This assessment criteria is fully based on the UN-SWAP Scoring Tool. Each sub-criteria shall be equally weighted. (**) Scoring uses a four point scale (0-3).</p>		

Overall Evaluation Quality Assessment

Quality assessment criteria (scoring points*)	Assessment Levels (*)			
	Very good	Good	Fair	Unsatisfactory
1. Structure and clarity of reporting, including executive summary (7)		7		
2. Design and methodology (13)	13			
3. Reliability of data (11)	11			
4. Analysis and findings (40)	40			
5. Conclusions (11)	11			
6. Recommendations (11)	11			
7. Integration of gender (7)	7			
Total scoring points	93	7		

Overall assessment level of evaluation report **Very good**

The evaluation integrates adequately cross cutting issues of gender equality, human rights and disability inclusion, even though is not included as part of the evaluation objective.

Consideration of significant constraints (e.g. COVID-19 or civil unrest)

The quality of this evaluation report has been hampered by exceptionally difficult circumstances: Yes No

If yes, please explain: