



Organizational unit: UNFPA Ethiopia Country Office

Year of report: 2020

Title of evaluation report: Government of Ethiopia/UNFPA 8th Country Programme [2016-2020]

Overall quality of report: **Good**

Date of assessment: 2020 June

Overall comments: This is a well-written evaluation that appears to be very useful for informing the next country programme. The findings are thorough with clearly stated data sources. The conclusions provide a good overview of the accomplishments and challenges of the programme and lead to a clear set of recommendations. The methodology is robust but could benefit by providing more information on qualitative sources including the number and type of stakeholders participating in each of the data collection processes and in the overall evaluation. The analysis of GEEW is well articulated although there could be more clarity on how HRGE was mainstreamed throughout the evaluation process. To further comply with UNFPA evaluation standards, the Executive Summary should be shortened.

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	<i>Insert <u>assessment level</u> followed by main <u>comments</u>. (use 'shading' function to give cells corresponding colour)</i>	
I. Structure and Clarity of Reporting	Yes No Partial	Assessment Level: Fair
<i>To ensure the report is comprehensive and user-friendly</i>		
1. 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?	Yes	The report is clearly written with few errors.
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)	No	The report is over 90 pages excluding annexes, significantly exceeding the maximum limit.
3. Is the report structured in a logical way? Is there a clear distinction made between analysis/findings, conclusions, recommendations and lessons learned (where applicable)?	Yes	The usual structure is followed.
4. Do the annexes contain – at a minimum – the ToRs; a bibliography; a list of interviewees; the evaluation matrix; methodological tools used (e.g. interview guides; focus group notes, outline of surveys) as well as information on the stakeholder consultation process?	Yes	The annexes contain the required elements except for information on the stakeholder consultation process. The latter is adequately discussed under Methodology in respect to the participation of the ERG and staff in guiding the evaluation and in feedback and validation processes.
<i>Executive summary</i>		
5. Is an executive summary included in the report, written as a stand-alone section and presenting the main results of the evaluation?	Yes	The executive summary is comprehensive. It serves as a stand-alone section that covers the evaluation process and main results.
6. Is there a clear structure of the executive summary, (i.e. i) Purpose, including intended audience(s); ii) Objectives and brief description of intervention; iii) Methodology; iv) Main conclusions; v) Recommendations)?	Yes	All elements are included.
7. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	No	At 9 pages, the maximum limit is significantly exceeded. The length is mainly due to the inclusion of findings, conclusions and recommendations for each programme component.

2. Design and Methodology	Yes No Partial	Assessment Level:	Good
<i>To ensure that the evaluation is put within its context</i>			
1. Does the evaluation describe the target audience for the evaluation?	Partial	The audience is specified in the Executive Summary but not in the main report.	
2. Is the development and institutional context of the evaluation clearly described and constraints explained?	Yes	The context and constraints are well laid out in the chapters on Background and UNFPA's programme.	
3. Does the evaluation report describe the reconstruction of the intervention logic and/or theory of change, and assess the adequacy of these?	Yes	The evaluators note that they reviewed the ToCs that existed for each CP component and noted the need for an overall ToC showing the linkages between components. They produced an overall model which is provided on p 26.	
<i>To ensure a rigorous design and methodology</i>			
4. 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?	Yes	The evaluation framework is described and the attached matrix includes all required elements along with a summary of findings for each question.	
5. Are the tools for data collection described and their choice justified?	Yes	The tools identified included desk review, observations, structured and semi-structured interviews, KIIs and FGDs. The rationale for selecting this mix of tools is noted.	
6. 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)?	Partial	A stakeholder map or overview of stakeholders is not included but it is noted that mapping was done during the inception stage. The consultation and validation processes with the EFG and UNFPA are described.	
7. Are the methods for analysis clearly described for all types of data?	Yes	It is stated that content analysis was used to interpret qualitative data. A chart is also provided that links the EQs to the CP phases, evaluation criteria and levels of analysis.	
8. Are methodological limitations acknowledged and their effect on the evaluation described? (Does the report discuss how any bias has been overcome?)	Yes	The limitations identified are the (a) geographic spread affecting representativeness, and (b) selection of programme sites to be visited by IPs potentially introducing selection bias. Mitigation strategies are provided and sufficient. Primarily using triangulation to mitigate effects of bias in data.	

9. Is the sampling strategy described?	Yes	There is a clear description of the criteria/indicators used for selecting sites for data collection and for stakeholders to be interviewed. Purposive sampling was used for both. It is noted that convenience sampling was used to identify beneficiaries to be invited for FGDs. The limitation of the samples not being representative is noted.
10. Does the methodology enable the collection and analysis of disaggregated data?	Partial	The approaches should enable this to be done, however evaluation participants are not disaggregated by stakeholder group or gender and the attached tools do not have a place to indicate gender of respondent. National-level gender disaggregated data is provided in the Context chapter.
11. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?	Yes	There are specific questions on human rights and gender and these are considered as cross-cutting themes.

3. Reliability of Data	Yes No Partial	Assessment Level:	Fair
<i>To ensure quality of data and robust data collection processes</i>			
1. Did the evaluation triangulate data collected as appropriate?	Yes	The evaluators explain how data was triangulated. In several places of the findings section multiple sources are shown for individual findings.	
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Partial	The general sources of data are mostly well described and constraints on reliability are also discussed. However, there is no information provided on the beneficiaries chosen for the FGDs - on the number involved, how they were chosen, from what programme areas, etc. It is only in the Executive Summary that the total number of evaluation participants is specified - 179 - but the annexed List of Persons Interviewed shows that the number of people participating in group discussions was not included in the total. A fuller description of evaluation participants is warranted.	
3. Did the evaluation make explicit any possible limitations (bias, data gaps etc.) in primary and secondary data sources and if relevant, explained what was done to minimize such issues?	Yes	As noted above, issues of reliability and potential selection bias were raised. Mitigation strategies were noted and primarily involved triangulation of sources and methods.	
4. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?	Yes	In a subsection on "Ethics and maintaining the quality of evaluation" it is noted that precautions to protect respondents' rights included informed consent and confidentiality practices, and these are then reflected in the instructions for the evaluation protocols. It is also noted that UNEG and UNFPA guidance was followed in the conduct of the evaluation. Although, there is no further information on practices used to collect information from beneficiary groups (such as location and timing), it is clear they were inclusive and sensitive to discrimination against particular groups.	

4. Analysis and Findings	Yes No Partial	Assessment Level:	Good
<i>To ensure sound analysis and credible findings</i>			
1. Are the findings substantiated by evidence?	Yes	This is well done throughout the section. In addition to document citations, the findings are frequently supported by quotes and examples, i.e. in discussing how mentoring is used to support capacity-building it is noted that 5 senior midwives from Gambella Hospital rotate to 10 catchment health centers, and feedback from health center staff is provided that confirms the usefulness of this activity (p 39).	
2. Is the basis for interpretations carefully described?	Yes	The basis for the findings is consistently well explained.	
3. Is the analysis presented against the evaluation questions?	Yes	The analysis is clearly presented. Each criterion section begins with the evaluation questions and is followed by a concise summary of findings.	
4. Is the analysis transparent about the sources and quality of data?	Yes	Data sources are cited throughout the findings and, in the case of qualitative sources, the specific stakeholder group from which they emerged is regularly provided.	
5. Are cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Partial	Causal effects are shown. This is particularly clear under Effectiveness where tables are used to show the level of achievement for output indicators compared to targets and/or evidence of outcome level achievements. Although the findings do not explicitly discuss unintended outcomes, it appears these were considered as unintended outcomes are listed as an indicator area for each CP component in the results framework and data collection tools included this line of questioning. In addition, the report includes tables which detail the outcomes of activities from the perspective of respondents (some of which could be unintended).	
6. Does the analysis show different outcomes for different target groups, as relevant?	Yes	This is done. Examples include a finding that interventions targeting youth center are unlikely to directly benefit girls due to the lower utilization of these facilities by girls. Gender-disaggregated results are also frequently provided throughout this section.	
7. Is the analysis presented against contextual factors?	Yes	Contextual factors are provided to explain the extent that results were or were not achieved. An example is the challenges in addressing FGM and CM due to social-cultural norms - with informants sharing that government officers publicly denounce these issues but then practice it within their own households (p 50).	
8. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?	Yes	This is done. For example, the analysis considers the extent that gender is mainstreamed throughout the CP (p 50), and issues of equity and human rights are examined in the context of support of refugees communities including the disparity in the conditions and services of health centers serving different refugee groups (p 71).	

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 evaluators clearly specify the relevant questions and criteria for each conclusion.	
2. Do the conclusions go beyond the findings and provide a thorough understanding of the underlying issues of the programme/initiative/system being evaluated?	Yes	The conclusions are clear; they synthesize the main findings, providing a good overview of the accomplishments and challenges of the programme. The structure of this section was unusual as the associated recommendations are placed in the middle of the text for each conclusion, thus interrupting the flow - however, this sub criteria is still met.	
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	There is no evidence of bias.	
6. Recommendations			
<i>To ensure the usefulness and clarity of recommendations</i>			
1. Do recommendations flow logically from conclusions?	Yes	The relevant conclusions and evaluation questions are shown. As with the conclusions, both strategic and programme level recommendations are provided	
2. Are the recommendations clearly written, targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?	Yes	The recommendations are clearly structured and written. The intended users are specified. Each has an Operational Implications section that includes sub-recommendations to support achievement of the main recommendations.	
3. Do recommendations appear balanced and impartial?	Yes	There is no indication of bias and recommendations appear balanced in respect to building on past performance and advising where new emphasis needs to be placed.	
4. Is a timeframe for implementation proposed?	Yes	For most recommendations, the text specifies whether it is directed towards the next CP or whether it should be completed within the timeframe of the current CP.	
5. Are the recommendations prioritized and clearly presented to facilitate appropriate management response and follow up on each specific recommendation?	Yes	Priority levels are specified. The recommendations are framed in a way that supports a management response.	

7. Gender	0 1 2 3 (**)	Assessment Level:	Good
<i>To assess the integration of Gender Equality and Empowerment of Women (GEEW) (*)</i>			
1. Is GEEW integrated in the evaluation scope of analysis and indicators designed in a way that ensures GEEW-related data to be collected?	3	<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) Issues of HRGE are specifically included as part of the scope of the evaluation. = 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) A human-rights based approach and gender mainstreaming are identified as cross-cutting issues. = 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) EQ5 is a dedicated question under Effectiveness that looks at the CP within UNFPA's GEWE framework. Gender issues are also considered as part of other questions under Effectiveness, Relevance and Sustainability. = 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) There is not a question that specifically addresses this issue however it was considered as there is a finding that notes gender-disaggregated that is not always available (p 50). = 2</p>	
2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques?	1	<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) This is not discussed. = 0</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) Mixed methods are noted as being used although quantitative data is only derived from secondary sources. As the number stakeholders involved in each form of data collection is not clear, it is difficult to assess adequacy of the sample size. = 1</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) Diverse sources and processes are used, and the evaluators describe use of triangulation. However, the extent of inclusion is not clear. =2</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) There was representation from a diverse range of stakeholders, including beneficiaries, but the numbers of each are not provided. = 2</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) Attention to ethical issues such as informed consent and confidentiality are discussed, but additional steps taken to ensure the comfort of beneficiaries participating in particular are not discussed. The type of stakeholders / beneficiaries engaged are not made apparent, and therefore it is not clear which type of ethical standards might need to be considered (e.g. if survivors were engaged as focus group participants, special procedures would have been followed.) = 2</p>	

3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?

- 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 chapter includes a subsection on GEWE that clearly articulates issues faced by women and the related policies and instruments. = 3
- 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)** This is done to some extent - there are several direct quotes from evaluation participants and the source of some evidence is identified as being focus group participants and/or beneficiaries. = 2
- c. Are unanticipated effects of the intervention on human rights and gender equality described? (Score: 0-3)**
Interview protocols included this line of questioning, and although unanticipated effects were not explicitly mentioned in the findings, it was noted that girls were not accessing some services to the extent that boys were . = 2
- d. Does the evaluation report provide specific recommendations addressing GEEW issues, and**

(*) This assessment criteria is fully based on the UN-SWAP Scoring Tool. Each sub-criteria shall be equally weighted (in correlation with the calculation in the tool and totaling the scores 11-12 = very good, 8-10 = good, 4-7 = Fair, 0-3=unsatisfactory).

(**) Scoring uses a four point scale (0-3).

0 = Not at all integrated. Applies when none of the elements under a criterion are met.

1 = Partially integrated. Applies when some minimal elements are met but further progress is needed and remedial action to meet the standard is required.

2 = Satisfactorily integrated. Applies when a satisfactory level has been reached and many of the elements are met but still improvement could be done.

3 = Fully integrated. Applies when all of the elements under a criterion are met, used and fully integrated in the evaluation and no remedial action is required.

Overall Evaluation Quality Assessment

	Assessment Levels (*)			
Quality assessment criteria (scoring points*)	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	22	60	18	
Overall assessment level of evaluation report		Good		
	Very good very confident to use	Good confident to use	Fair use with caution	Unsatisfactory not confident to use

- (*) (a) Insert scoring points associated with criteria in corresponding column (e.g. - if 'Analysis and findings' has been assessed as 'Good', enter 40 into 'Good' column).
- (b) Assessment level with highest 'total scoring points' determines 'Overall assessment level of evaluation report'. Write corresponding assessment level in cell (e.g. 'Fair').
- (c) Use 'shading' function to give cells corresponding colour.

If the overall assessment is 'Fair', please explain

- How it can be used?

- What aspects to be cautious about?

Where relevant, please explain the overall assessment Very good, Good or Unsatisfactory

Consideration of significant constraints

The quality of this evaluation report has been hampered by exceptionally difficult circumstances:

Yes No

If yes, please explain: