

Organizational unit: Evaluation Office, UNFPA 2016

Title of evaluation report: Evaluation of the UNFPA support to family planning 2008-2013.

Overall quality of report: Good **Date of assessment:** 1 December 2016

Overall comments: The evaluation looks at the broad spectrum of UNFPA activities concerned with family planning by using a thorough methodology building on intense document review, extensive interviews with key informants, twelve case studies, including five covering all regions that involved field visits, and two surveys. The evaluation constructed a theory of change that sought to consolidate the various strategies that had guided UNFPA family planning work, and used it effectively to show the extent to which expected results had been obtained. The accountability objective of the evaluation was achieved and considerable material was produced on factors explaining how results were obtained, or not, leading to recommendations for the future strategy.

Assessment Levels

<div style="background-color: #ADD8E6; padding: 10px; display: inline-block;">Very good:</div> <p>strong, above average, best practice</p>	<div style="background-color: #90EE90; padding: 10px; display: inline-block;">Good:</div> <p>satisfactory, respectable</p>	<div style="background-color: #FFD700; padding: 10px; display: inline-block;">Fair:</div> <p>with some weaknesses, still acceptable</p>	<div style="background-color: #FF0000; padding: 10px; display: inline-block;">Unsatisfactory:</div> <p>weak, does not meet minimal quality standards</p>
---	---	--	---

Quality Assessment Criteria	Insert <u>assessment level</u> followed by main <u>comments</u> . (use 'shading' function to give cells corresponding colour)	
<p>I. Structure and Clarity of Reporting</p> <p><i>To ensure the report is comprehensive and user-friendly</i></p> <ul style="list-style-type: none"> • Is the report easy to read and understand (i.e. written in an accessible non-technical language appropriate for the intended audience)? • Is the report focused and to the point (e.g. not too lengthy)? • Is the report structured in a logical way? Is there a clear distinction made between analysis/findings, conclusions, recommendations and lessons learned (where applicable)? • Do the annexes contain – at a minimum – the ToRs; a bibliography, a list of interviewees, the evaluation matrix and methodological tools used (e.g. interview guides; focus group notes, outline of surveys)? <p><i>Executive summary</i></p> <ul style="list-style-type: none"> • Is an executive summary included in the report, written as a stand-alone section and presenting the main results of the evaluation? • 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)? • Is the executive summary reasonably concise (e.g. with a maximum length of 5-10 pages)? 	Assessment Level:	Good
	<p>Comment: The report is relatively easy to read and is presented in non-technical language. The report is not too lengthy given its global multi-year scope. It has the minimum content and sequence required for structure: Acronyms (page VII); Executive Summary (pages IX – XIV); Introduction (pages 1-3); Methodology including Approach and Limitations (pages 5-12).</p> <p>There is no separate chapter called “Context,” but findings include description of the context, For instance, context is addressed in the section 3.2.2. “National ownership in different national contexts” (p. 23). Findings and Analysis are presented in the chapter “3. Main findings and analysis.”</p> <p>The report includes Conclusions and Recommendations. There is no chapter “Transferable Lessons Learned,” but the ToR does not require it (Annex 1. Structure for evaluation reports and country case study notes, pp. 112-114).</p> <p>Minimum requirements for Annexes are met: ToRs (annex 2); Bibliography (annex 3); List of interviewees (annex 8); and Methodological instruments used (annexes 9-14).</p> <p>The executive summary is a generally stand-alone section, but with several shortfalls. It does include the evaluation methodology, the conclusions and recommendations. It also has the main findings, which are not necessarily required. The shortfalls include: not identifying the intended audience for the</p>	

	<p>evaluation in the report (even if that means affirming what the Terms of Reference states), nor providing a description of the intervention (except at a high framework level). The Findings in the text are augmented by the findings/analysis at bottom of the Evaluation Matrix (in Annex 1). But absent is any guidance on how to navigate respective pieces of the text and annexes so the report can be read as a coherent whole. Conclusions are explicit (pp. xii - xiii). Recommendations of the evaluation are clearly described too (pp. xiii – xiv).</p>
--	---

<p>2. Design and Methodology</p>	<p>Assessment Level: Good</p>
<p><i>To ensure that the evaluation is put within its context</i></p> <ul style="list-style-type: none"> • Does the evaluation describe whether the evaluation is for accountability and/or learning purposes? • Does the evaluation describe the target audience for the evaluation? • Is the development and institutional context of the evaluation clearly described? • Does the evaluation report describe the reconstruction of the intervention logic and/or theory of change? • Does the evaluation explain any constraints and/or general limitations? <p><i>To ensure a rigorous design and methodology</i></p> <ul style="list-style-type: none"> • Is the evaluation approach and framework clearly described? Does it establish the evaluation questions, assumptions, indicators, data sources and methods for data collection? • Were the methods chosen appropriate for addressing the evaluation questions? Are the tools for data collection described and justified? • Is the methods for analysis clearly described? • Are methodological limitations acknowledged and their impact on the evaluation described? (Does it discuss how any bias has been overcome?) • Is the sampling strategy described? Does the design include validation techniques? • Is there evidence of involvement of stakeholders in the evaluation design? (Is there a comprehensive/credible stakeholder map?) • Does the methodology enable the collection and analysis of disaggregated data? 	<p>Comment: The ToR states that “The particular emphasis of this evaluation will be on learning with a view to informing the implementation of the UNFPA family planning strategy: “Choices not chance 2012-2020” (p. 92).</p> <p>The evaluation describes the target audience for the evaluation, the development and institutional context, and includes a theory of change (Annex 5). It explains constraints and general limitations (Annex 7), including the issue of selection of respondents for the on-line survey of beneficiaries.</p> <p>The context of the evaluation is well-described, including the fact that family planning was the subject of a number of strategies, but had also been somewhat controversial at the country level.</p> <p>The methodology section is well-structured and easy to read: 2.1. Overview of the evaluation process, 2.2. Methods and tools used in evaluation design, 2.3. Methods and tools used in data collection, 2.4. Methods and tools used for data analysis, and 2.5. Assessing assumptions and challenging theories of change. The chapter includes 6 figures to describe the methodology visually. But, the Figure 2 “Evaluation questions, criteria, and data sources” is confusing as there are no actual formulations of evaluation questions. The full list of the evaluation questions is not presented in the methodology. The methods used are carefully chosen, including the purposive sample of sites for the five field visits and seven desk country studies that is explained in depth. The basis for selection of respondents in the beneficiary survey, however, is not explained although in the annex the fact that the respondents</p>

- Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?

were suggested by the country offices was noted as a possible limitation.

The evaluation explains constraints and limitations and their implications. The evaluation approach and framework are described in Annex 1: Evaluation Matrix where data sources and data collection methods are related to each of the evaluation questions. .

Tools used for data collection are presented in order, although in somewhat general way: 2.3.1. Document review, 2.3.2. Country case studies, 2.3.3. Key informant interviews at global and regional level, 2.3.4. Online surveys of 64 countries, 2.3.5. Financial analysis, and 2.3.6. Limitations and mitigation strategies. For instance, section 2.3.1. "Document review" mentions a document database which was a basis for further research activities, including field country studies. However, there is no annex devoted to this database. The references to the annexes would be useful in the section 2.3.2 "Country case studies" (document review, field visits, interviews, and focus group discussions).

Triangulation is explained in the report. The evaluators say that "The information collected through the surveys enabled triangulation of responses...Results were entered into a format corresponding to each of the areas of investigation" (p. 9). But, it is unclear which "format" the evaluators refer to in this text. In the limitation section, it is said that "The mitigation strategy applied throughout the evaluation was to employ a mix-method approach to ensure triangulation of a wide range of information types, range of data collection methods and a variety of sources spanning

across multiple geographical levels, and to focus at the outcome level using a more qualitative approach” (p. 10).

The evaluators do explain how the stakeholders were involved into the evaluation “During an initial evaluation team analysis workshop (June 2015), findings ...were qualitatively analysed...These preliminary findings were further developed by the evaluators and subsequently presented and discussed in a collaborative session with UNFPA staff and partners (November 2015)” (p. 11).

Finally, the evaluators do not comment on how they dealt with involvement of vulnerable groups, youth, and women into the evaluation. It is said in the methodology section that there were evaluation questions on vulnerable and marginalized groups (p. 6). But, the methodology does not specify stakeholders with regards to gender or vulnerable groups (family planning service users, key UNFPA country office staff and external stakeholders (multilateral and bilateral development partners, private foundations, INGOs). In the meantime, ToR requires the evaluation to be “sensitive to fair power relations amongst stakeholders” (p. 100).

3. Reliability of Data	Assessment Level:	Good
<p><i>To ensure quality of data and robust data collection processes</i></p> <ul style="list-style-type: none"> • Did the evaluation triangulate all data collected? • Did the evaluation clearly identify and make use of qualitative and quantitative data sources? • Did the evaluation make explicit any possible issues (bias, data gaps etc.) in primary and secondary data sources and if relevant, explained what was done to minimize such issues? I.e. did the evaluation make explicit possible limitations of the data collected? • Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations? • Is there adequate gender disaggregation of data? And if this has not been possible, is it explained? • Does the evaluation make explicit the level of involvement of different stakeholders in the different phases of the evaluation process? 	<p>Comment: Sources of qualitative and quantitative data have been identified. It is possible to find references in the text such as “²⁸This is in reference to finance department figures in the Statistical and Financial Reviews,” “²⁹See Theory of Change, Diagramme 1, Volume II, Annex 5,” and others.</p> <p>The lack of measurable results data beyond activities and outputs affects the quality of the data and reflects the fact that what are termed outputs in the theory of change are mostly results by governments receiving assistance from UNFPA rather than outputs produced by UNFPA. What are termed outcomes are closer to specific objectives.</p> <p>Annex 3 includes Bibliography for the evaluation matrix. The documents ensure credibility of interviews and focus groups, for instance, Annex 9: In-country key informant interview topic guide; Annex 10: Focus group discussion guide for in-country case studies, Annex 11: Interview guide for UNFPA headquarters, regional offices, country offices and international stakeholders; and Annex 14: Methodological note on financial analysis.</p> <p>The ToR requires data to be “disaggregated by relevant criteria (age, sex, etc. wherever possible)” (p. 100). The evaluators break down data by gender when necessary, for instance, there is Annex 8 “List of people interviewed.” Also, it is possible to find examples in the text such as “However, this strategy had the result of limiting programme focus to married women, and did not meet the needs of unmarried youth or other “non-traditional” groups⁴⁸” (p. 19), “...there are over 200 million women with unmet need for contraception” (p. 21). The evaluation matrix also includes examples of gender disaggregation, for</p>	

	instance, “UNFPA staff, partners’ and users’ (women's and men's) perception of meaning and importance of service integration” (p. 3).
--	---

<p>4. Analysis and Findings</p> <p><i>To ensure sound analysis</i></p> <ul style="list-style-type: none"> • Is information analysed and interpreted systematically and logically? • Are the interpretations based on carefully described assumptions? • Is the analysis presented against the evaluation questions? • Is the analysis transparent about the sources and quality of data? • Are possible cause and effect links between an intervention and its end results explained? • Where possible, is the analysis disaggregated to show different outcomes between different target groups? • Are unintended results identified? • Is the analysis presented against contextual factors? • Does the analysis include reflection of the views of different stakeholders (reflecting diverse interests)? E.g. how were possible divergent opinions treated in the analysis? • Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights? <p><i>To ensure credible findings</i></p> <ul style="list-style-type: none"> • Can evidence be traced through the analysis into findings? E.g. are the findings substantiated by evidence? • Do findings follow logically from the analysis? • Is the analysis of cross-cutting issues integrated in the findings? 	<p>Assessment Level:</p>	<p>Good</p>
<p>Comment: In Section 3, findings follow a common presentation format, starting with each Evaluation Question, and then establishing the underlying programmatic assumptions, the evaluation criteria covered, and a summary of the findings followed by an elaboration of elements of answers to the evaluation question. As a whole, these findings are presented at a global level, with minimal data about country or regional level, variations, of interventions, good practices. Annex 1 includes specific findings at the regional and country level, specific results from other data collection tools, and specific practices worthy of note, leaving specific evidence of UNFPA contributions and practices at the national level largely treated more in the Annex and only briefly alluded to in Section 3. The findings section is in narrative text without any display. The section is more on activities, and outputs and strategies than on outcomes and cause and effect links which are not systematically explained.</p> <p>One of the objectives of the evaluation is to provide a focus on the lessons learned that can help guide future programming and implementation. Although those lessons are implicit in the last two sections of the report, the absence of a set of clearly and explicitly identified lessons learned and good practices reflecting a gender analysis in one part of the report leaves this evaluation objective less fulfilled than the accountability objective.</p> <p>Findings are presented as a narrative text. The text lack of tables or graphs. In the main body of the report the EQ3 in the chapter “3.3. Brokerage and partnerships” sounds the same way</p>		

as EQ2. They should have different formulations according to the evaluation matrix and ToR.

Findings are substantiated by evidence. The analysis includes references to the primary sources such as “Some noted that UNFPA leadership is good at messaging the overarching rationale...Several key informants spoke of the difficulty of “turning around” a large bureaucracy...” (p. 23).

The evaluators note: “...although UNFPA participated in many of the efforts to reposition family planning in 2001-2011, leadership often came from other organisations... .” Then, they clarify that “USAID and the French Development Agency spearheaded the Ouagadougou Partnership, while DFID and the Bill and Melinda Gates Foundation (BMGF) cohosted the London Summit on Family Planning” (p. 22).

ANALYSIS

Interpretations are based on carefully described assumptions which are presented in the beginning of the chapter. Also, there are references in the text to specific assumptions such as “⁷⁰Assumption 2.2.,” “⁷¹Assumption 2.2.,” “⁷²Assumption 2.2, see (Cambodia Country Note 2015: Section 4.2: 21).”

Contextual factors are identified, for instance, there is a section called “3.2.2. National ownership in different national contexts” and “3.2.3. The cultural, institutional and economic sustainability of programmes in different national contexts” (p. 23).

Cause and effect links between an intervention and its end results (including unintended results) are explained. For instance, it is said that “At national level, UNFPA has made important

	<p>contributions to shaping family planning policies in many countries, through advocacy and the provision of technical support to government” (p. 22). However this causality analysis is uneven across the evaluation questions.</p>
--	---

<p>5. Conclusions</p>	<p>Assessment Level:</p>	<p>Good</p>
<p><i>To assess the validity of conclusions</i></p> <ul style="list-style-type: none"> • Are conclusions credible and clearly related to the findings? • Are the conclusions demonstrating an appropriate level of analytical abstraction? • Are conclusions conveying the evaluators' unbiased judgement of the intervention? 	<p>Comment: An innovative approach was to write one-and-a-half-page introduction before listing the conclusions. This introduction serves as useful overview of conclusions derived from credible findings. For instance, the evaluators state that "To advance family planning services requires a more shared vision of how the organisation can contribute strategically, based on a specific comparative advantage" (pp. 63-64). Nevertheless, conclusions are too long (pp. 64-71).</p> <p>The conclusions all flow from the findings. It is said in the report that "from evidence to findings and conclusions. In each evaluation area/question, the evaluators drew on the full set of data sources (document reviews, country desk studies, international key informant interviews, on-line surveys and field country studies) to develop the overall findings associated with the key assumptions. Findings were then reviewed and analysed in order to develop conclusions" (p. 11).</p> <p>Conclusions are not organized in priority order, but rather by subject area: Raising the profile of family planning, Coordination and brokerage, Integration of family planning and sexual and reproductive health, Sustainability, Human rights and vulnerable and marginalized group members, Evidence and learning, Modes of engagement, Contributing to commodity security, and Technical support and oversight (pp. 64-71).</p> <p>Some conclusions do not have references to the findings, so we can assume that they convey evaluators' unbiased judgment of the intervention, because they are based on credible findings. For instance, the evaluators claim that:</p>	

- | | |
|--|--|
| | <ul style="list-style-type: none">• “there is evidence that the 2009 “regionalisation strategy” has been met with some resistance...,” (p. 71);• “UNFPA has carried out effective work with partners to promote reproductive health commodity security with country governments,” “Most importantly, UNFPA has provided effective support to improve the contraceptive method-mix” (p. 70). |
|--|--|

All together, the conclusions are sound and effectively presented.

6. Recommendations	Assessment Level:	Good
<p><i>To ensure the usefulness and clarity of recommendations</i></p> <ul style="list-style-type: none"> • Do recommendations flow logically from conclusions? • Are the recommendations sufficiently clear, targeted at the intended users and operationally-feasible? • Do recommendations reflect stakeholders' consultations whilst remaining balanced and impartial? • Is the number of recommendations manageable? • Are the recommendations prioritised and clearly presented to facilitate appropriate management response and follow up on each specific recommendation? 	<p>Comment:</p> <p>The recommendations relate to the conclusions. They have number of the conclusion they refer to, for instance, the recommendation 1 is "Based on conclusions: 2, 3, 4 and 6" (p. 73). But, it would be useful to have explanation how the recommendation is connected with these conclusions.</p> <p>Recommendations are strategic, targeted and operationally-feasible: they are directed at future actions on specific subject area and include a list of operational implications. They are assigned either priority Very High, High, Medium-High or Medium. All recommendations refer to responsible organizations.</p> <p>The chapter does not state if the recommendations considered the consultations. But the methodology section explains that "These preliminary findings were ... presented and discussed in a collaborative session with UNFPA staff and partners (November 2015). This process allowed the evaluation team to reflect on initial findings and conclusions..." (p. 11).</p>	

7. Gender	Assessment Level:	Fair
<p><i>To assess the integration of Gender Equality and Empowerment of Women (GEEW)¹</i></p> <ul style="list-style-type: none"> • Is GEEW integrated in the evaluation scope of analysis and indicators designed in a way that ensures GEEW-related data to be collected? • Do evaluation criteria and evaluation questions specifically address how GEEW has been integrated into design, planning, implementation of the intervention and the results achieved? • Have gender-responsive evaluation methodology, methods and tools, and data analysis techniques been selected? • Do the evaluation findings, conclusions and recommendations reflect a gender analysis? 	<p>Comment:</p> <p>Although there is no explicit reference to gender equality and empowerment of women in the scope and objectives of the evaluation, gender issues are considered in evaluation questions dealing respectively with vulnerable and marginalized groups and the implementation of a human rights-based approach. Indicators associated to these questions are too broad to allow for specific GEEW-related data to be collected.</p> <p>Two evaluation questions (EQ5 and EQ6) touch upon gender equality issues, however they do not address the integration of GEEW specifically (EQ5 reads: “To what extent has UNFPA focused on the family planning needs of the most vulnerable and marginalised groups, including identification of needs, allocation of resources, and promotion of rights, equity and access?” and EQ6: “To what extent has UNFPA implemented a human rights-based approach to family planning, in particular regarding access to and quality of care, and through support from HQ and RO for a rights-based approach in country?”). Under EQ5, two indicators allow to capture data relating to the integration of gender respectively into the planning and into the implementation of UNFPA support to FP (“Evidence of gender- sensitive needs assessment of target groups for UNFPA supported interventions including identification of rights violations”; “Evidence for gender sensitive participation by VMGs”). However, this does not allow for a comprehensive analysis of the integration of GEEW.</p>	

¹ 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 totalling the scores 11-12 = very good, 8-10 = good, 4-7 = Fair, 0-3=unsatisfactory).

The evaluators claim that “The overall approach to the evaluation was based on evaluating the contribution of UNFPA to family planning and was responsive to both gender and human rights and cultural sensitivity” (p. 5). It is unclear from the Methodology if gender-responsive evaluation methods and tools, and data analysis techniques have been selected. But, the Annex 9 “In-country key informant interview topic guide” includes questions with regards to gender (pp. 175-177).

There are references to gender in the evaluation findings, conclusions and recommendations but they do not reflect a comprehensive gender analysis.

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		93	7	
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 'finding and analysis' 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 explain²:

- How it can be used?
- What aspects to be cautious about?

Where relevant, please explain the overall assessment Very good, Good or Unsatisfactory³:

Consideration of significant constraints⁴

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

yes

no

If yes, please explain:

² The purpose here is to clarify in what way the report can be used. This in order to assist the elaboration of a relevant Management Response and the wider use of the evaluation findings back into programming. When a report has been assessed as Fair, it is obligatory to fill this text box in.

³ The purpose is, where relevant, to clarify for example severe unbalances in the report (for example, the report is good overall but recommendations very weak). Is optional to fill in.

⁴ E.g. this should only be used in case of significant events that has severely hampering the evaluation process like natural disasters, evaluators falling sick, unexpected significant travel restrictions, etc. More 'normal' limitations should be mentioned under relevant section above.