

Organizational unit: Evaluation Office **Year of report:** 2018

Title of evaluation report: Mid-term evaluation of the UNFPA Supplies Programme (2013-2020)

Overall quality of report: **Very Good** **Date of assessment:** 6 October 2018

Overall comments: The evaluation covers a complex programme that is central to UNFPA's functioning: how to ensure that the necessary supplies are in place to achieve family planning goals. The report can be used as a model for other evaluation reports in terms of design, structure and written presentation. While the evaluation focuses on procurement and the supply-chain, the connection with an improved enabling environment for population activities and the demand for services is also taken up. The evaluation includes a thorough theory of change analysis which is clearly presented in the text of the report and the annexes and the methods used, including the document review and key stakeholder interviews, allowed the evaluators to make extremely useful/relevant conclusions and recommendations, based on findings that were, on the whole, well evidenced/triangulated. Gender equality and the empowerment of women was integrated in the evaluation scope of analysis and the conclusions and recommendations, which were logically connected to/derived from the findings, reflect a gendered analysis.

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>		
1. Structure and Clarity of Reporting	Yes No Partial	Assessment Level:	Very good
<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 no issues of accessibility. It is easy to read, with large font text, pictures, and supporting figures improving readability.	
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)	Yes	At 78 pages, the full report (i.e. the running text of the report) is within a reasonable length for a major report on a key aspect of UNFPA actions.	
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 report has all of the key sections, except a distinct "lessons learned" section, although these can clearly be inferred from the findings and the conclusions.	
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 300 page volume on annexes contains all of the required information. The annexes contain the ToRs (annex 8); a bibliography (annex 5); a list of interviewees (annex 4); the evaluation matrix (annex 1); methodological tools used (annex 3, Part Two "Interview guides for field-based country case studies;" annex 3, Part Three "Online survey questionnaire (English Version).") Information on the stakeholder consultation process is presented in the methodology although there are not many details on the design of the participatory process.	
<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 thorough but concisely written and presented as a standalone section.	
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	Yes, there is a clear structure, with clearly delineated sections.	
7. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	Yes	The executive summary was five pages long.	

2. Design and Methodology	Yes No Partial	Assessment Level:	Fair
<i>To ensure that the evaluation is put within its context</i>			
1. Does the evaluation describe the target audience for the evaluation?	Yes	While the introduction in Part One does not specify the target audience as such, the recommendations clearly indicate to whom the evaluation is addressed (and the ToR of the evaluation - which is annexed to the evaluation report - also describes the target audience).	
2. Is the development and institutional context of the evaluation clearly described and constraints explained?	Yes	Sections 1 and 2 provide a thorough description of the context, which is itself complex. Chapter 2 "THE UNFPA SUPPLIES PROGRAMME 2013-2017" compares the UNFPA and global initiatives in reproductive health, as well as global partnerships and programmes for family planning. The evaluation explains the constraints of the program. For example, the evaluation refers to the DFID annual review (2015) of UNFPA Supplies which highlighted "A continuing gap between programme financial needs and resources..." among others. As another example, reflecting on the institutional context and potential constraints, the evaluation notes that while UNFPA initiated a change-making process and began to implement a new strategic approach, the changes "did not significantly influence the results observed during the evaluation data-collection phase. However, they have been noted in the evaluation, especially in the development of the evaluation's conclusions and recommendations."	
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 logic is well-described in the report, particularly in Figure 1, as well as in the extensive evaluation matrix in Annex I. In the methodology section the following is noted: the evaluation "reconstructed the programme's theory of change and, ultimately, developed key causal assumptions and related evaluation questions." Evidence of this analysis being undertaken can be found in the section 2.3.2 "Simplified theory of change for UNFPA supplies" and Annex 2 "Overall and pathway theories of change (ToC) for UNFPA Supplies."	
<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 matrix found in Annex I covers all of the evaluation questions in terms of indicators, data sources and methods in great detail. These are summarized in Chapter 3 of the main report.	
5. Are the tools for data collection described and their choice justified?	Yes	Yes. There are five types of tools (or methods) used: document review, key informant interviews and focus groups, four field-based country case studies and five desk-based case studies, and an on-line survey of key informants in all 46 countries in which the programme works.	
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	There is no comprehensive stakeholder map, but the consultants state in the annex 4 that "Key informants were first identified using stakeholder maps developed at global and country levels" (p. 189). It can be deduced that the evaluation involved stakeholders throughout the process, through for example, the Evaluation Reference Group which "provided substantive inputs, facilitating access to documents and informants, ensuring the high technical accuracy of the findings and co-authoring the recommendations to ensure their usefulness and feasibility" (xii).	
7. Are the methods for analysis clearly described for all types of data?	Partial	Data analysis is generally/broadly described in terms of quantitative and qualitative methods of analysis. Though it was clear that the evaluation applied contribution analysis to assess cause and effects in UNFPA Supplies (p. 9), the methodology does not include a section on data analysis. Chapter 3.2 "Data collection" includes 3.2.1 "Data-collection methods used" and 3.2.2 "Data-collection results."	
8. Are methodological limitations acknowledged and their effect on the evaluation described? (Does the report discuss how any bias has been overcome?)	Yes	Chapter 3 includes an extensive description of the limitation to the evaluation response.	
9. Is the sampling strategy described?	Yes	This is a purposive sample based on suggestions of who to interview in the 46 countries by the UNFPA country offices concerned. In the online survey, for example, the resulting sample frame from suggestions included 494 potential respondents "who were invited to complete the online survey. The evaluation was able to secure 134 completed responses from 39 of the 46 programme countries."	

10. Does the methodology enable the collection and analysis of disaggregated data?	Yes	Annex 4 shows that the methodology was appropriate for collection and analysis of disaggregated data. Data was collected from different types of stakeholder groups: international agencies, governments, and other public and private organizations. The evaluators interviewed both men and women.
11. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?	Yes	Yes, first there is a section that notes the intent of the evaluation to do this (i.e. section 3.1.3 Cross-Cutting Issues: Gender Equality and UNFPA Supplies In Humanitarian Contexts). It is evident in the evaluation itself that the methodology (including the methods chosen and sample selected) is able to take on cross-cutting issues, which are extensively reviewed in the report. For instance, the consultants explain that “The evaluation has taken a different approach to each of these cross-cutting issues” (pp. 9-10).

3. Reliability of Data	Yes No Partial	Assessment Level:	Very good
-------------------------------	----------------------	-------------------	------------------

To ensure quality of data and robust data collection processes

1. Did the evaluation triangulate data collected as appropriate?	Yes	Findings (and analysis) was supported by triangulated data (different data sources). The consultants explain that “The qualitative and quantitative evidence relevant to each evaluation question was summarized and triangulated in the completed evaluation matrix (Annex 1).” As can be seen from annex 1 (evaluation matrix), the results are drawn from different sources and forms of data: field and desk country case studies, global and regional interviews, national and global level document and data reviews and the online survey.
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Yes	The data is reliable and we can find the evidence of this in the evaluation (including in Annex 5 which shows different official documents used for the evaluation). The evaluation clearly identifies and makes use of quantitative data like expenditures and types of supply along with qualitative data from key stakeholders. The consultants were able “to access all the sources of information identified in the draft evaluation matrix” (p. 11).
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	The report clearly articulated limitations and described what was done, (i.e. additional key stakeholder interviews) to mitigate the issues. The evaluation discusses limitations in primary and secondary data sources and explain measures to mitigate them (pp. 13-14). For instance, the consultants acknowledge that “The desk-based country case studies provided the evaluation with a more limited body of evaluative information...” (p. 14). As such, the results of the field-based country studies were used to clarify data obtained from the desk review.
4. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?	Yes	Based on the way in which interviewees and survey participants were selected, as well as the questions being addressed, issues of discrimination and other ethical considerations were clearly considered in the data collection process.

4. Analysis and Findings	Yes No Partial	Assessment Level:	Very good
---------------------------------	----------------------	-------------------	------------------

To ensure sound analysis and credible findings

1. Are the findings substantiated by evidence?	Yes	Yes, findings are clearly substantiated by evidence. The evaluators provide references to the sources of information, for instance: “In the online survey, many respondents identified the support offered by UNFPA Supplies to developing national planning and strategy documents” (p. 17).
2. Is the basis for interpretations carefully described?	Yes	In each case, the basis for interpretation is shown. This is particularly the case with the thematic section dealing with procurement and supply-chain management.
3. Is the analysis presented against the evaluation questions?	Yes	Yes, as noted in Table 7, the findings respond to/presented against the evaluation questions.
4. Is the analysis transparent about the sources and quality of data?	Yes	The main sources of data are key stakeholder interviews and documents and the evaluation is clear about the source and quality.

5. Are cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Yes	The evaluators were careful to show the links and, particularly in terms of Chapter 6, several unintended outcomes were noted. As an example of cause-effect, the evaluators state that the outcome “demand is increasing” had happened due to the output “significant investment in a demand-generation strategy (mainly supported by the BMGF-funded Nigeria Urban Reproductive Health Initiative (NURHI) project)” (p. 25). The consultants point out the relation between key activities and output “This strategy was based on formative research and utilized a comprehensive range of mass media, community engagement and interpersonal communication interventions” (p. 25).
6. Does the analysis show different outcomes for different target groups, as relevant?	Yes	The analysis explains how different outcomes affect various groups differently (Government and non-governmental organizations (NGOs), development partners (WHO, the Clinton Health Access Initiative (CHAI), Population Services International (PSI) and the World Bank): the reasons why some groups have access while others do not as well as the factors which drive the outcomes and how they are different for different groups is taken up by the evaluation. The analysis discusses how demand is generated and how the approach needs to be tailored differently depending on for example the population served or their location. Particularly, in Chapters 4 and 5, demand creation is discussed and there is reference to different beneficiary groups and how access to supply differs/affects them differently.
7. Is the analysis presented against contextual factors?	Yes	Yes, the evaluation took context into account. This was particularly the case in issues of supply in situations of humanitarian problems. For instance, the evaluation states that “... in many countries, the advent of a humanitarian crisis brings into operation specific elements of UNFPA Supplies, such as the procurement and distribution of specialized kits. In some countries, a crisis which is national in scope (such as the Ebola virus disease (EVD) in Sierra Leone) can lead to a general shift in the focus and content of UNFPA Supplies” (p. 10).and national standards – p. 27), social (for instance, fears of increase in workload of service providers and provider’s attitude towards service integration – p. 36), and others.
8. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?	Yes	As noted, Chapter 7 of the report covered all of the cross-cutting issues and was able to develop findings on each.

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	Each conclusion, as presented, is linked to the findings that lead to it.	
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 were drafted to both summarize and expand upon the findings. This evaluation report is different from other UNFPA reports that have been assessed in terms of design of conclusions. The evaluation consultants highlighted strengths, challenges and provided narrative justification to each conclusion, including visual figures.	
3. Do the conclusions appear to convey the evaluators’ unbiased judgment?	Yes	Yes, they are unbiased and present a balanced picture (both positive and negative aspects are presented).	

6. Recommendations	Yes No Partial	Assessment Level:	Very good
<i>To ensure the usefulness and clarity of recommendations</i>			
1. Do recommendations flow logically from conclusions?	Yes	Each recommendation is grounded in conclusions to which it is addressed.	
2. Are the recommendations clearly written, targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?	Yes	Yes, they are clearly written and each recommendation includes a section called operational requirements or operational action provided to the entity to which the recommendation is addressed.	
3. Do recommendations appear balanced and impartial?	Yes	As drafted, the recommendations are impartial.	

4. Is a timeframe for implementation proposed?	Yes	Each operational requirement specifies a timeframe. For example recommendation one (asserting leadership) says "Requires UNFPA senior management to leverage UNFPA Supplies in global platforms for family planning, such as FP2020" which clearly means that leveraging needs to be built into the process of preparing the plan.
5. Are the recommendations prioritised and clearly presented to facilitate appropriate management response and follow up on each specific recommendation?	Yes	There are three levels of priority (very high, high, high/medium) and the recommendations are clearly presented.

7. Gender	0 1 2 3	Assessment Level:	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?	3	GEEW is integrated in the evaluation scope of analysis as one of the objectives of the evaluation is to assess "the extent to which issues of gender equality and social inclusion and equity have been taken into consideration" (p. 2). Indicators are designed in a way that ensures GEEW-related data to be collected, for instance: "National RH/FP strategies and plans (including in national health plans and reproductive health roadmaps) include focus on expanded access, including access for marginalized women and girls ..." (p. 157), "Relative priority given to improved access for marginalized women and girls in national programmes, policies and strategies" (p. 158).
2. Do evaluation criteria and evaluation questions specifically address how GEEW has been integrated into design, planning, implementation of the intervention and the results achieved?	2	Evaluation criteria and evaluation questions do not include specific GEEW statements, but the evaluators state that "Gender equality and social inclusion are particularly relevant when addressing evaluation questions two (increasing demand) and four (improving availability and access)... Gender equality, social inclusion and equity concerns are also linked to the programme's rights-based approach to the provision of reproductive health and family planning commodities and services (section 5.2.5)" (p. 10).
3. Have gender-responsive evaluation methodology, methods and tools, and data analysis techniques been selected?	2	The evaluation report states that "Issues of gender equality, social inclusion and equity have been "mainstreamed" during data collection, analysis and reporting" (p. 10). But, they do not provide specific details information on the extent to which tool and data analysis techniques are gender-responsive.
4. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?	3	Findings reflect GEEW as a cross-cutting issue in all sections of the report. For example, Chapter 4 notes: "UNFPA Supplies has supported countries to identify sound technical policies and build evidence-based programmatic approaches that reflect emerging priorities in reproductive health, particularly in support of marginalized women and girls" (p. 17) and Chapter 5 includes a section 5.2.5 "Gender equality, social inclusion and equity". Conclusions reflect a gender analysis, with conclusions under Cluster C addressing issues of "gender equality and social inclusion" (p. 69). Recommendations also reflect a gender analysis, for instance, recommendation 4 states that "UNFPA Supplies should ensure the systematic application of a human rights-based approach to the provision of family planning services. This should include specific guidance on how to improve gender equality and social inclusion..." (p. 76).

(*) This assessment criteria is fully based on the UN-SWAP Scoring Tool, see Annex 7. 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).

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	80	7	13	
Overall assessment level of evaluation report	Very 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?

The fair rating is based on unevenness in the evaluation design and in its applications in terms of findings. While the weaknesses of the methods and findings need to be considered, the fact that the conclusions are strong and clearly expressed, suggests that the evaluation can be used to help design the next country program.

- 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: