

**Organizational unit:** UNFPA Evaluation Office **Year of report:** 2019

**Title of evaluation report:** Evaluation of the UNFPA response to the Syria crisis (2011-2018)

**Overall quality of report:** **Very Good** **Date of assessment:** 15 May 2019

**Overall comments:** The evaluation evaluates a very complex programme concerned with humanitarian assistance in a complex political environment, involving multiple countries and contexts. Data collection was quite difficult, as a result, but the evaluation is itself well-defined in scope and approach, including by developing/agreeing on a theory of change at the onset (from which the evaluation questions emerged). There was a thorough consultation process with stakeholders and the data collection, within the limits of what was possible, was comprehensive and undertaken in accordance with good ethical practice. The findings were based on the data and were well-balanced between those that were positive and those that indicated areas for improvement. These were translated into conclusions and recommendations that were clear and targeted, although their time-frame and resource implications were less clear.

**Assessment Levels**

- Very Good** (blue box) strong, above average, best practice
- Good** (green box) satisfactory, respectable
- Fair** (yellow box) with some weaknesses, still acceptable
- Unsatisfactory** (red box) 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)*

<b>1. Structure and Clarity of Reporting</b>	Yes No Partial	Assessment Level:	<b>Very good</b>
--	----------------------	-------------------	------------------

*To ensure the report is comprehensive and user-friendly*

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	There are some minor errors with punctuation and run-on sentences, however the report is largely well-written and easy to understand.
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	The main report is 84 pages, which slightly exceeds the maximum length for a thematic evaluation, however, several of those additional pages are filled with images and several are used to separate sections and, therefore, did not compromise what is considered a reasonable length.
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 is structured in a logical way with a clear distinction made between sections. While there is not a specific section on "lessons learned", the evaluation does seek to draw lessons from UNFPA's response through the recommendations as well as boxes integrated throughout the findings on 'Good Practice.' The colored text boxes are effectively used to highlight good practices, key learnings, and areas for improvement.
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	There is a separate document containing the annexes, and it includes all of the required elements: approach and methodology, research tools, overview of UNFPA response, TOR, bibliography, results of the survey, key informants interviewed, predicted and actual limitations, financial information, presentation of strength of evidence of findings, evaluation matrix, and the evaluation inception report.

*Executive summary*

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 summary is clear and written as a stand-alone document with conclusions clearly linked to recommendations and providing an overview of the evaluation 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	There is a clear structure to the executive summary. Starting with a clear description of the purpose and structure of UNFPA's response, as well as an overview of activities, the report then provides a summary of the evaluation motivation, primary and secondary purpose, objectives, and intended audience. The methodology is also presented, including a brief discussion of limitations, noting the lack of quantitative outcome-related data as a significant limitation. Main conclusions and recommendations are succinct and numbered.
7. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	Yes	The summary is within the 5 pages maximum length.

<b>2. Design and Methodology</b>	Yes No Partial	Assessment Level:	<b>Fair</b>
----------------------------------	----------------------	-------------------	-------------

*To ensure that the evaluation is put within its context*

1. Does the evaluation describe the target audience for the evaluation?	Yes	The intended users of the evaluation are specified within the 'Introduction' section and primarily include stakeholders internal to UNFPA (i.e. UNFPA country offices, regional response hub, regional offices, leadership and special branches) as well as other organizations working in the region.
2. Is the development and institutional context of the evaluation clearly described and constraints explained?	Yes	The context is briefly explained in the report itself, primarily providing numbers on the scale of the crisis and details on the UN's coordinated approach and UNFPA's response. A brief situation analysis or findings from a rapid needs assessment would have strengthened this section, providing context and justification to UNFPA's interventions in the body of the report. However, more details on country contexts and constraints are provided in an annex, therefore satisfying this criteria.

3. Does the evaluation report describe the reconstruction of the intervention logic and/or theory of change, and assess the adequacy of these?	Yes	Yes, the evaluation presented a reconstructed theory of change, which guided the evaluation methodology.
<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 briefly described in the text in terms of methodological approach, the standards and guidelines to which it conformed, and its alignment with the ToC. It is covered in more depth within the annexes on the methodological approach (Annex I) and evaluation matrix (Annex X). The annexed evaluation matrix includes the questions, assumptions, indicators, data sources and methods. How the two case studies identified in the ToR and the Inception Report were specifically integrated into the evaluation process is not well-described. While referenced in several footnotes (# 23, #243, #257), they were not included in the annexes. As a key data source and method, we would expect to see them as part of the evaluation matrix.
5. Are the tools for data collection described and their choice justified?	Yes	The data collections tools are described. The templates for the KIIs, FGDs, clinic rapid assessment / service provider questionnaires, and online survey are provided. In terms of justification, it is noted that mixed methods are used to triangulate findings, and enhance their quality and credibility. The reason for the online survey for stakeholders in Syria is identified (the evaluators were not able to get visas).
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)?	Yes	A stakeholder map is provided in the annexed Inception Report - it is not referenced in the main report. The stakeholder consultation process including the timeline is outlined in Annex 1c (stakeholder consultation process). It is noted that the recommendations were co-developed with the ERG and other UNFPA business units at a final stakeholder workshop in NY.
7. Are the methods for analysis clearly described for all types of data?	Yes	The methods of analysis are described in detail in Annex I, pages 9-10, and include descriptive analysis, content analysis, and comparative 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 and mitigation strategies are briefly summarized in the main text and are described in more detail in Annex VII. The section on Quality Assurance (Annex I, p. 11) provides information on how bias would be overcome/minimized.
9. Is the sampling strategy described?	Partial	The annex includes a section on sampling. Sampling was primarily purposive with snowball sampling being used to identify additional potential interviewees. The criteria for the selection of site visits is provided. Any limitations of the sample and sample approach are not explicitly addressed as such, although the challenges of/mitigation strategy for not being able to enter Syria are described in the overall limitations section. Data from stakeholders in Syria was collected through an online survey and there were 28 responses but no discussion on the response rate, including the extent to which the responses are illustrative (or any limitations related to this).
10. Does the methodology enable the collection and analysis of disaggregated data?	Yes	The methodology enabled the collection and analysis of gender disaggregated data as focus group discussions were conducted separately with sex and age disaggregated groups. The online survey template allows for disaggregation by stakeholder group but not by gender.
11. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?	Yes	Yes, the design/methodology allow for evaluation to address cross-cutting issues, through for example, evaluation questions, regular consultation with the evaluation reference group, and through the use of multiple data collection methods (including focus group discussions with service recipients).
3. Reliability of Data	Yes No Partial	Assessment Level: <b>Very good</b>
<i>To ensure quality of data and robust data collection processes</i>		
1. Did the evaluation triangulate data collected as appropriate?	Yes	Triangulation was achieved through use of multiple data sources and methods.
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Yes	Qualitative and quantitative data sources are used and consistently referenced within the report, though the analysis heavily draws on qualitative data. However, this was noted as an anticipated limitation in the evaluation. The evaluation team sought to overcome this barrier through the use of survey data (which was cited on occasion in the report and presented in full within the annexes). In addition, the evaluation team sought transparency by identifying the reliability and strength of evidence for each finding (Annex IX). In most cases, key informant interview data on findings was seen as reliable, however, the lack of documentation on outcome-level results severely affected the reliability of some findings as the evaluators themselves noted (13, 20, 21, 22).
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	Data gaps and mitigation strategies, for example, are presented in the report as well as Annex IX and Annex VII.
4. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?	Yes	There is evidence that data has been collected with sensitivity to issues of discrimination. For example, citations made in the report text were anonymous, generally referencing the source but not identifying the specific name. In addition, the evaluation approach was described as gender and human rights responsive; and what this meant in practice was further described in the approach to focus group discussions (Annex I, p7) and the overview of quality assurance mechanisms applied (Annex I, p11-12).
4. Analysis and Findings	Yes No Partial	Assessment Level: <b>Very good</b>
<i>To ensure sound analysis and credible findings</i>		
1. Are the findings substantiated by evidence?	Yes	Findings are consistently substantiated by evidence, frequently citing both primary and secondary sources.

2. Is the basis for interpretations carefully described?	Yes	The rationale for the findings is thoroughly explained; in some cases this is somewhat overly descriptive for an evaluation report, unnecessarily adding to the length of the report (for example the findings under Efficiency related to the regional response hub, pp 55-57)
3. Is the analysis presented against the evaluation questions?	Yes	The section on findings is organized according to the evaluation questions.
4. Is the analysis transparent about the sources and quality of data?	Yes	The sources of data are consistently referenced in footnotes. Where data is missing or inadequate, the evaluation team mentions this and subsequently notes that in the conclusions and recommendations to be considered by UNFPA to ensure the quality of future evaluations.
<b>5. Are cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?</b>	Yes	The evaluators are able to show plausible cause-effect linkages and assess shortcomings, even with the constraints of outcome-level results not being systematically collected. These links are most clearly shown in the annexed evaluation matrix. Unintended outcomes were not explicitly framed as such, but there are cases of these been addressed, such as the involvement of males in GBV work.
6. Does the analysis show different outcomes for different target groups, as relevant?	Yes	The findings and analysis describe outcomes, such as service access, and ways that the program accommodated the needs of different groups, such as youth and persons with disabilities. In addition, the report consistently describes differences in the achievement of outcomes across countries.
<b>7. Is the analysis presented against contextual factors?</b>	Yes	This is particularly well done; for example, under Relevance, the ways that different countries responded to changes in the dynamics of the conflict are examined.
8. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?	Yes	The analysis includes extensive coverage of cross-cutting issues, including findings which focus on UNFPA's capacity and success in assessing and integrating persons with diverse gender identities and disabilities into activities. The analysis also focuses on access for the most marginalized groups (e.g. female-headed households and widows), including an assessment of methods for incorporating men/boys in GBV programming without hindering access for the most vulnerable. In addition, the report did not use the word 'beneficiary', citing instead participants, persons receiving services, etc. which are seen as aligning more with language preferred when using a human-rights based approach (HRBA encourages use of the words 'rights-holder' and 'duty-bearers' when referring to stakeholders).
<b>5. Conclusions</b>	Yes No Partial	Assessment Level: <b>Very good</b>
<i>To assess the validity of conclusions</i>		
<b>1. Do the conclusions flow clearly from the findings?</b>	Yes	The conclusions flow clearly from the findings, with direct links made between the finding number and conclusion number.
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 demonstrate a thorough understanding of the underlying issues of the Syria response, connecting multiple challenges raised within the findings to one, succinct conclusion. The conclusions are helpfully grouped into categories: Overall, Programming, Coordination and Leadership, Systems and Structures.
3. Do the conclusions appear to convey the evaluators' unbiased judgment?	Yes	The conclusions are clearly based on evidence from the findings and therefore do not convey bias.
<b>6. Recommendations</b>	Yes No Partial	Assessment Level: <b>Fair</b>
<i>To ensure the usefulness and clarity of recommendations</i>		
<b>1. Do recommendations flow logically from conclusions?</b>	Yes	Yes, there are direct links between the conclusion number and recommendation number.
<b>2. Are the recommendations clearly written, targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?</b>	Partial	The recommendations are clear and action-oriented, explicitly stating 'operational actions' and defining the intended users (which also align with the intended audience of the evaluation). While the technical implications of the recommendations are implicit, no specific reference to the human and financial implications is made.
3. Do recommendations appear balanced and impartial?	Yes	The recommendations flow clearly from the findings and conclusions without any evidence of bias.
4. Is a timeframe for implementation proposed?	Yes	Recommendations are prioritized as high, medium and low, thus suggesting a timeframe for implementation. Within some recommendations, short-term, medium-term and long-term strategies are also proposed (see recommendations 7 and 9, for example).
5. Are the recommendations prioritized and clearly presented to facilitate appropriate management response and follow up on each specific recommendation?	Yes	Recommendations are clearly prioritized and phrased so as to facilitate management response.
<b>7. Gender</b>	0 1 2 3 (**)	Assessment Level: <b>Very good</b>

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?		The evaluation was conducted for the purpose of accountability to stakeholders, implying the integration of GEEW and human rights. Specific criteria on gender and human rights were incorporated in the evaluation framework (e.g. To what extent were the objectives of the UNFPA humanitarian response responsive to refugees' identified needs, and to what extent did interventions reach the most vulnerable populations with the greatest need for services). EQ5 looks at UNFPA response's alignment with UNEG gender equality principles and EQ10 looks at UNFPA's contribution to the quality of SRH and GBV services (and, within that, evidence of gender equality). The evaluation highlights the lack of monitoring data on results in youth and SRH, while also identifying strengths in collecting data on GBV.
2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques?		The annex on quality assurance mechanisms describes gender considerations and most data collection methods were able to collect disaggregated data by sex. Diverse data sources and multiple methods are used, including focus groups, observations, document review, surveys and key informant interviews. Data is triangulated. While a mixed methods approach is used, and both qualitative and quantitative data was collected and used, there is limited disaggregated, quantitative data and overall limited use of gender disaggregated data (though the methods chosen allow for the collection and use). The perspectives of various stakeholders, including diverse vulnerable groups, were obtained through focus group discussions. Ethical standards, including WHO guidelines for researching sexual violence, were integrated into the evaluation methodology.
3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?		Overall, the findings, conclusions and recommendations reflect a gender analysis. The subsection on context does not specifically articulate how some groups are more marginalized as part of the crisis, but it does reference violations of international humanitarian and human rights law. This is then taken up in the analysis with Finding 3 addressing issues of consistency in referencing such laws within the programme. A weakness in the report is related to the disaggregation of quantitative data, but the perspectives of different groups are included in the text (for example, as part of Finding 7 on the engagement of males in GBV work (pp 30-31). The evaluators did not specifically identify any findings as being unanticipated, however they did address how the inclusion of boys and men appeared to prevent some women from accessing services in some cases.(3)There is a recommendation specific to the need to improve gender analysis in the planning of UNFPA interventions. (3)

(\*) 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).

(\*\*) 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

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			
<b>Total scoring points</b>	<b>76</b>		<b>24</b>	
<b>Overall assessment level of evaluation report</b>	<b>Very Good</b>			
	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**

The evaluation was carefully designed to measure a very complex programme that was affected by the conflict, and the data, to the extent possible, were collected carefully and reliably. The conclusions were sound and the inclusion of GEEW throughout the evaluation was excellent.

**Consideration of significant constraints**

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

Yes

No

If yes, please explain: