

Title of evaluation report: Evaluation of UNFPA's Country Programme 3 (CP3) in Afghanistan 2010 – 2013

OVERALL QUALITY RATING: Poor

Summary: The evaluation took place in a very difficult context: the evaluators were only able to conduct one site visit and for a large number of the CPAP indicators lacked data on baseline and/or endline results. One of the strengths of the report is the use of qualitative data to illustrate the activities and outputs of UNFPA's programmes. Nevertheless, the evaluation has major flaws. The evaluators used two ineffectual evaluation criteria and failed to sufficiently address the necessary criteria of Coordination and Added Value. They did not develop a manageable list of fewer than ten evaluation questions. The structure of the report is not logical and hinders understanding. Sources are insufficiently cited and there is no evidence of effort to validate information or assess the credibility of sources. Last, although most of the analysis does not go beyond documenting programme activities, in cases where outputs and outcomes are discussed (such as in the results matrix) cause and effect relationships are insufficiently analyzed.

Quality Assessment criteria	Assessment Levels			
	Very good	Good	Poor	Unsatisfactory
<p>1. Structure and Clarity of Reporting <i>To ensure report is user-friendly, comprehensive, logically structured and drafted in accordance with international standards.</i> Checklist of minimum content and sequence required for structure:</p> <ul style="list-style-type: none"> i) Acronyms; ii) Exec Summary; iii) Introduction; iv) Methodology including Approach and Limitations; v) Context; vi) Findings/Analysis; vii) Conclusions; viii) Recommendations; ix) Transferable Lessons Learned (where applicable) Minimum requirements for Annexes: ToRs; Bibliography; List of interviewees; Methodological instruments used. 	<p>Poor</p> <p>The report includes all required sections and annexes.</p> <p>The organization of the evaluation is not logical. First, there are multiple chapters with the title "Analysis of Programmatic Areas." It appears that these findings and analysis chapters are organized by programme area, but there is some overlap across chapters such that Adolescent Sexual and Reproductive Health is include in a chapter with Humanitarian Aid rather than the Reproductive Health chapter without any justification.</p> <p>Second, within these findings and analysis chapters there is</p>			

	<p>not a uniform structure for chapter sub-headings: some chapters are primarily organized by evaluation criteria (Relevance, Effectiveness, etc) while others are organized by programme type (Population and Development, Gender Equality). This lack of systematic structure makes the report difficult to follow.</p> <p>Third, the chapter on “Strategic Position and Future Directions” mixes findings from multiple evaluation criteria and recommendations.</p> <p>The pages of the annexes are incorrectly numbered.</p> <p>Acronyms are usually but not always written out in full when used for the first time (ex. “CSO” on p. 6).</p>
<p>2. Executive Summary <i>To provide an overview of the evaluation, written as a stand-alone section and presenting main results of the evaluation.</i> Structure (paragraph equates to half page max):</p> <ul style="list-style-type: none"> i) Purpose, including intended audience(s); ii) Objectives and Brief description of intervention (1 para); iii) Methodology (1 para); iv) Main Conclusions (1 para); v) Recommendations (1 para). Maximum length 3-4 page. 	<p>Poor</p> <p>The Executive Summary includes all the required components, although the discussion of the methodology is quite general and does not mention country-specific constraints that are apparent in other parts of the report.</p> <p>Findings and conclusions are presented together and the evaluators do not sufficiently distinguish the two. Most of the statements presented in the Findings and Conclusions sub-section of the Executive Summary are in fact findings: for example, “Training of midwives since 2005 by the MoPH, supported by UNFPA among others, is widely believed to have contributed to the drop in maternal mortality measured in 2010” (p. 9). Conclusions that are based on findings but address a higher level of analysis are lacking in the Executive Summary.</p>

3. Design and Methodology

To provide a clear explanation of the following elements/tools

Minimum content and sequence:

- Explanation of methodological choice, including constraints and limitations;
- Techniques and Tools for data collection provided in a detailed manner;
- Triangulation systematically applied throughout the evaluation;
- Details of participatory stakeholders' consultation process are provided;
- Details on how cross-cutting issues (vulnerable groups, youth, gender, equality) were addressed in the design and the conduct of the evaluation.

Poor

Methodological choices are clearly explained, with different types of data collection described and the tools for data collection are presented in the annexes (pp. 118-135). Methodological constraints are discussed (p. 16) but their severity and implications are not clearly communicated.

Although triangulation is mentioned (p. 16) and multiple sources of data were collected, there is little evidence in the findings and analysis chapters that it was used. For instance, in the discussion of the Socio-Demographic and Economic Survey in Bamiyan, which the evaluators note as a point of conflict between different stakeholders, triangulation should have been used and clearly explained to allow the reader to judge the veracity of different statements about the survey (p. 60). Instead, most information is not cited. To give one example, the evaluators claim that "Numerous studies prove that gender based violence has implications for every aspect of health policy and programming, from primary care to reproductive health programmes" (p. 64) yet do not provide any citations.

The list of stakeholders consulted and feedback from participatory workshops is included, although the structure of these documents is not explained (pp. 128-135).

The methodology section does not discuss how cross-cutting issues were taken into account. It also provides insufficient information on how sites and stakeholders were selected for visits or interviews.

One major deficiency in the evaluation is the lack of a small

	<p>number of targeted evaluation questions. The evaluation questions are only listed in the interview guides (pp. 121-125) and include all potential evaluation questions from the ToR.</p> <p>Furthermore, the evaluators use two ineffectual evaluation criteria, “Country Programme Management, Institutional Partnerships and Governance Arrangements” and “Strategic Positioning. As a result, the key criteria of Coordination and Added Value are insufficiently addressed. The Strategic Positioning criteria mixes findings and recommendations, and the criteria on management, partnerships, and governance is confusing and inconsistently applied: for example, on p. 51 there is a discussion of coordination, partnerships, and M&E all under the heading of “4.1.6 Management: Partnerships.” This is a major deficiency in the chosen methodology.</p>
<p>4. Reliability of Data <i>To clarify data collection processes and data quality</i></p> <ul style="list-style-type: none"> • Sources of qualitative and quantitative data have been identified; • Credibility of primary (e.g. interviews and focus groups) and secondary (e.g. reports) data established and limitations made explicit; • Disaggregated data by gender has been utilized where necessary. 	<p>Poor</p> <p>The evaluators use a number of qualitative and quantitative sources, although they were constrained by the security context in Afghanistan and by a lack of baseline data in many cases. For example, the evaluators draw in quotes from interviewed health workers that clearly show how UNFPA’s training enabled them to reduce maternal mortality (p. 40), and interviewed trainers and participants of gender trainings (p. 71). They do point out when appropriate the lack of baseline data (p. 16, 47) and the need to resort to national statistics to measure progress against outcomes that should ideally be measured only for UNFPA intervention provinces (p. 93).</p> <p>The sources of statistics are not always cited in footnotes,</p>

	<p>making it impossible to judge the credibility of the information. For example, but the source of a statistic on fertility rates presented as a 2013 Estimate in the section on context is not referenced (p. 17).</p> <p>There is little evidence that the evaluators carefully considered the veracity of information they were given, although they point out repeatedly that the lack of credible data is a problem in Afghanistan (p. 14). There is no discussion, for instance, of the generalizability or credibility of participant testimonials (for an example of where this would have been appropriate, see pp. 69-70).</p> <p>Disaggregated data by gender has been used in several places (for example, see pp. 69, 71).</p>
<p>5. Findings and Analysis <i>To ensure sound analysis and credible findings</i> <u>Findings</u></p> <ul style="list-style-type: none"> • Findings stem from rigorous data analysis; • Findings are substantiated by evidence; • Findings are presented in a clear manner <p><u>Analysis</u></p> <ul style="list-style-type: none"> • Interpretations are based on carefully described assumptions; • Contextual factors are identified. • Cause and effect links between an intervention and its end results (including unintended results) are explained. 	<p>Poor</p> <p>Findings often focus on activities of the UNFPA programs rather than outcomes or even outputs. For example, for Output 1.2 under Gender Equality, which focuses on “enhanced capacity of target communities to... prevent, respond to and monitor gender-based violence” the discussion of the health sector response focuses on the creation of UNFPA centers, provision of GBV assistance by UNFPA, and training rather than health sector capacity (p. 66).</p> <p>The outcomes of the program are discussed in the CP3 results matrix on pp. 93-98, but in these results matrix outcomes are not systematically linked to UNFPA’s interventions. In much of the results matrix UNFPA activities are listed as outputs (Output 2.1 of Population and Development on p. 96), or activities and outputs are listed</p>

	<p>in lieu of data on outcomes (Outcome 1 of Gender Equality on p. 97).</p> <p>Findings are not clearly presented. The evaluators should have included a highlights box with the main findings at the beginning of the sections on each programme component. Furthermore, it is not clear how programme components are split across chapters (ex. Adolescent Reproductive Health is with Humanitarian Aid rather than Reproductive Health) and the organization of each chapter is not systematic.</p> <p>In most cases the evaluators do not attempt to link the UNFPA activities to the outputs and outcomes in the CPAP. In many cases, this is because the data to measure progress against outputs does not exist, often because of the poor formulation of the CPAP 3, which in many cases chose indicators for which there was no baseline data or potential for collecting detailed follow-up data (p. 47). However, in the results matrix, data on the output indicators are presented but no attempt is made to link changes in the indicators to UNFPA interventions (for example, Output 1.1 of Gender Equality on p. 98).</p>
<p>6. Conclusions <i>To assess the validity of conclusions</i></p> <ul style="list-style-type: none"> • Conclusions are based on credible findings; • Conclusions are organized in priority order; • Conclusions must convey evaluators' unbiased judgment of the intervention. 	<p>Poor</p> <p>Conclusions are organized based on evaluation criteria rather than priority. They are not clearly linked to findings or to recommendations. Some conclusions are in fact <u>recommendations</u> (ex. “these [RH information] activities should be intensified and effectively extended to youth” on p. 80). There is no evidence of bias in the conclusions.</p>
<p>7. Recommendations</p>	<p>Poor</p>

<p><i>To assess the usefulness and clarity of recommendations</i></p> <ul style="list-style-type: none"> • Recommendations flow logically from conclusions; • Recommendations must be strategic, targeted and operationally-feasible; • Recommendations must take into account stakeholders' consultations whilst remaining impartial; • Recommendations should be presented in priority order 	<p>Recommendations are not presented in a way that enables them to be logically linked to conclusions. Recommendations are not presented in priority order, but are rather organized by evaluation criteria. Recommendations are accompanied by a longer list of "Suggestions for Programmatic Improvement" which provides more information on how the recommendations should be implemented.</p>
<p>8. Meeting Needs</p> <p>To ensure that Evaluation Report responds to requirements (scope & evaluation questions/issues/DAC criteria) stated in the ToR (ToR must be annexed to the report).In the event that the ToR do not conform with commonly agreed quality standards, assess if evaluators have highlighted the deficiencies with the ToR.</p>	<p>Poor</p> <p>The evaluators failed to adapt the list of suggested evaluation questions in the ToR into a reasonable list of fewer than ten specific evaluation questions, although this is required by the ToR (p. 110) and recommended by the Evaluation Handbook. These evaluation questions are not discussed in the methodology, although they are listed in the interview guides used with stakeholders (pp. 121-125).</p> <p>The ToR requests two unusual and ineffective evaluation criteria: "Country Programme Management, Institutional Partnerships and Governance Arrangements" and "Strategic Positioning." The evaluators should have pointed out this deficiency in the ToR and adapted their methodology to meet accepted standards.</p>

Quality assessment criteria (and Multiplying factor *)	Assessment Levels (*)			
	Very good	Good	Poor	Unsatisfactory

1. Structure and clarity of reporting (2)			2	
2. Executive summary (2)			<u>2</u>	
3. Design and methodology (5)			5	
4. Reliability of data (5)			5	
5. Findings and analysis (50)			50	
6. Conclusions (12)			12	
7. Recommendations (12)			12	
8. Meeting needs (12)			12	
TOTAL			<u>100</u>	

(*) Insert the multiplying factor associated with the criteria in the corresponding column e.g. - if “Finding and Analysis” has been assessed as “good”, please enter the number 50 into the “Good” column. The Assessment level scoring the higher number of points will determine the overall quality of the Report