

Title of evaluation report: UNFPA Country Programme Evaluation 2012-2016 (Albania)

OVERALL QUALITY RATING: Good

Summary: The report is clearly presented with the required structure and content as per the evaluation criteria. The executive summary is standalone and presents the objectives, methodology, and findings. The design and methodology are explained clearly, with limitations defined and methodological choices outlined. However, the non-random sampling used in data collection undermines the credibility of findings and conclusions of the report. The evaluators conducted rigorous data analysis, including documents review and interviews with stakeholders. Triangulation was used. Recommendations are strategic, targeted and operationally feasible but are also undermined by the non-random and non-representative sampling used in data collection, especially given the tone used in the report which implies causation and strong evidence.

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>Good</p> <p>The evaluation report includes a list of abbreviations/acronyms, a stand-alone Executive Summary, Introduction, Methodology, Context, and Findings. The Conclusions and Recommendations are batched into one chapter. There is no separate chapter “Transferable Lessons Learned.”</p> <p>In addition, the evaluation report includes a chapter dedicated to the ‘UN/UNFPA Strategic Response and Programme’ (p29). Furthermore, there is an additional chapter dedicated to ‘UNCT Cooperation and Value Added’ (p67). These sections respond to the specific requirements of the ToR outlined in the Annex. The consultants say in the report that “This evaluation examines factors that have facilitated or hindered achievements, and documents the lessons learned...” (p. 11). Lessons learned include the importance of the commitment of the UN Country Teams/HQ, Government and donors (p. 29). Thus, it is not clear why the consultants do not summarize lessons learned in the separate chapter.</p> <p>The Bibliography (References) is included in the Annex and detailed in the Table of Contents. However, the ToR, list of interviewees, and reproductions of the methodological instruments used are included in the Annexes, but is not detailed in the Table of Contents.</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>Good</p> <p>Executive summary is written as a standalone document and presents the main results of the evaluation. This includes a discussion of the overview and scope of the evaluation and an overview of the intervention. The executive summary also includes a description of the evaluation methodology (“approach” and “methodology” discussions), detail of the key findings/ “achieved results”, and an overview of recommendations organized according to programmatic area.</p> <p>However, the titling structure (font/style) used in the executive summary does not clearly distinguish between “Executive Summary” (p11) and “Key Findings” (p13). Moreover, the Executive Summary exceeds the maximum length of 3-4 pages specified in the evaluation criteria (totaling 6 pages).</p>
<p>3. Design and Methodology <i>To provide a clear explanation of the following elements/tools</i> Minimum content and sequence:</p> <ul style="list-style-type: none"> 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. 	<p>Good</p> <p>Methodological choices were sufficiently explained, including a description of non-random sampling and all data collection methods used (desk review, site visits, semi-structured group and individual interviews, group and individual follow-up interviews, and focus group discussions (p18). A clear description is provided for each method of data collection (on p18-20).</p> <p>This explanation of methodological choice includes a discussion of subsequent limitations and constraints, including the “non-representative and qualitative nature of the primary data that results from small, non-random samples and low response rates” (p12, p20). The explanation given for the non-random sampling methodology includes the lack of time and resources (p20). Other potential biases were discussed, including the selection of respondents, with mitigating strategies introduced (e.g. private interviews without UNFPA staff present) (p20).</p> <p>Triangulation is systematically applied throughout the evaluation. The consultants state that “The analysis is based on a synthesis and triangulation of information obtained from the above-mentioned five evaluation activities” (p. 12), but does not provide sufficient detail on the nature of triangulation. What limited discussion there is regards the use of triangulation to mitigate limitations and biases resulting from the non-random sampling methodology (p20). This limited discussion includes details on cross-cutting issues (e.g. vulnerable groups). There is also some discussion of cross-cutting issues in the Country Context chapter (p25).</p>

	<p>Details of participatory stakeholders' consultation process are provided in the paragraphs called "Stakeholder Involvement" and "Stakeholder Interviews with semi-structured questionnaire" (p. 18), a Table 1 "Achieved Stakeholder Interviews by Region and Type of Stakeholder" (p. 19). Findings are also based on stakeholders' consultations, for instance, the consultants explain that facts on Outcome 1 are based on "review of program data, site visits, stakeholder interviews, and training follow up interviews..." (p. 36). Details are also provided regarding stakeholder involvement in data collection; e.g. the report describes the use of an "evaluation reference group" of state/entity level ministers, civil society organizations, and other major stakeholders (p18).</p> <p>Youth and Adolescents as well as Gender Equality are the Program Areas. The methodology does not provide specific approaches for fostering participation and inclusiveness. But, it is said that "the team was able collect pertinent client/beneficiary feedback using focus group discussions with key populations including Roma, women in prisons, out of school youth, CSWs and IVDUs" (p. 20). In the methodology the consultants do not classify stakeholders with regards to gender or vulnerable groups. The evaluators say that "Per the Design Report, the target was for a total of 70 interviews, but only a total of 44 were conducted. Because several interviews had more than one respondent present, there were 74 respondents."</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>Good</p> <p>The sources of qualitative/quantitative data presented across the evaluation report are clearly identified; for example, regression statistics presented on contraceptive prevalence among married Albanian women is identified (p24).</p> <p>The methodology doesn't explain how the evaluators established credibility of primary and secondary data. Limitations are made explicit (p. 20). There is an effort to tie findings to specific data collection methods (p35).</p> <p>Disaggregated data by gender are utilized. For instance, it is said in the description of the country context that "While men are overall more likely to leave the country than women, the higher ratio of male to female emigration was more typical of the period from 1989 to 2001, with a more equal distribution as of men and women emigrants from" (p. 21). Findings section has disaggregated data by gender too, for instance, the Table 8 "UNFPA Outcome and Output Indicators" includes such indicator as Contraceptive Prevalence Rate for Modern Methods among married women age 15-44 - CPR (total), and others.</p>

5. Findings and Analysis

To ensure sound analysis and credible findings

Findings

- Findings stem from rigorous data analysis;
- Findings are substantiated by evidence;
- Findings are presented in a clear manner

Analysis

- 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.

Good

Interpretations are based on carefully described assumptions, presented in a clear manner and substantiated by evidence. For example there is a clear presentation of changes in output indicators according to a baseline and target, with actual/current progress defined and explained (T9; p40). Contextual factors are explained. Findings stem from rigorous data analysis based on the “review of program data, site visits, stakeholder interviews, and training follow up interviews” (p. 36). For instance, the evaluators refer to the interviews to support their analysis “People used to think that the Health Departments are only there for the sick people... I have seen the health personnel move more, go around meeting with the community” (p. 37). Examples of credible and rigorous assessments are given in the Evaluation Matrix.

Findings are substantiated by evidence. The report provide arguments and examples, for instance, such discussion is evidence-based “At the long-term strategic level, UNFPA Albania has been consistently successful in advocacy and support for the development and update of national strategies and guidelines. For example, the new National Strategy of RH (2016 – 2020) that is under development with leadership from the Ministry of Health as well as the development of an important Basic Package of PHC in 2014...” (p. 37). However, the non-random sampling used in data collection does not permit rigorous data analysis; therefore findings do not always stem from rigorous data analysis. The report attempts to show cause and effect links between the intervention and desired outcomes; however, the language employed is occasionally biased (e.g. “The CP SRH program has clearly contributed... [p36]). Additionally, given the non-random sampling used in data collection it is difficult to argue that the intervention can robustly be shown to have caused outcome changes.

Cause and effect links between an intervention and its end results (including unintended results) are explained. There are two tables which represent findings per outcome’ and output’ indicators (Table 8 UNFPA Outcome and Output Indicators as well as Table 9 UNFPA Albania SRH Output Indicators). However, when the report argues that “Based on review of program data, site visits, stakeholder interviews, and training follow up interviews there is strong evidence of progress in the three outputs in support of Outcome 1” (p36); given the non-random sampling used in data collection, this type of causal inference is not appropriate or justified. In addition, further statements of conclusion/causation are inappropriate given the underlying limitations of data collection: e.g. “it can be concluded that the SRH program...” (p37).

	<p>Contextual factors are identified. Interpretations are based on examples and statistics where necessary, for instance, “As shown by the dramatic 50% decline from US\$3.2 million in 2012 to US\$1.5 million in 2014, UNFPA/Albania was faced with a huge constraint to adjust its budget to the unexpectedly tight restrictions on its budget,” “As one respondent stated, “Apart from UNFPA, few donors are maintaining support for SRH. Almost everyone else has given up” (p. 41). Assumptions are discussed in the Annex 2. Albania CPE Evaluation Matrix.</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 divided into two parts “Strategic Conclusions” and “Program Related Conclusions.” There are five strategic conclusions. There are 13 program conclusions according to the program areas: 5 SRH, 2 Youth, 2 GBV, 4 PD.</p> <p>Most conclusions lack of references to findings. Whereas the conclusions are not specific, it is not clear if they convey evaluators’ unbiased judgment of the intervention. Also, many conclusions include recommendations. However, conclusions are presented clearly.</p> <p>Strategic conclusions 4 and 5 sound specific and reflect findings. Other strategic conclusions are not specific enough to be easily converted to further actions, for instance, such statements lack of clarity “trainings took place before the conditions were conducive” (conditions are not clear), “small and uncoordinated sub-activities” (it is not clear which activities are considered to be small and how the evaluators came to conclusion that they were uncoordinated). Another example “Albania does not have a centralized budget monitoring system...UNFPA Albania needs to improve control of the budgets of sub-activities in order to ensure that outputs are achieved as per annual working plan” (p. 71). It is not explained how the control of the budgets influenced achieving activities per annual working plan.</p> <p>Program Related Conclusions have both specific statements and not. It is not clear how the evaluators came to these conclusions: “...has demonstrated potential to improve,” (SRH 1) “...is needed to increase demand for effective methods of contraception... The UNFPA CP supported Social marketing effort in 2012 was inadequate in scope and duration” (SRH 2).</p> <p>In the SRH 3 it is not clear what “useful basis” mean: “The LMIS and abortion surveillance programs provide a useful basis for monitoring long-term progress with these initiatives.”</p> <p>SRH 4 in general is specific enough, except just one phrase; it is not clear what does “some effectiveness” mean.</p> <p>SRH 5 is formulated clear and specific based on credible findings.</p>

	<p>Youth 1 is specific. But, Youth 2 has general statement such as “has established a basis for,” but it is not clear what kind of “basis” has been established.</p> <p>GBV 1 should have more references on evidences to claim why the evaluators consider “further revisions of the GBV Training manual” not likely to be effective.”</p> <p>GBV 2 does not have references on findings explaining why “programs for men and boys to address gender violence” have both short- and long-term potential.</p> <p>PD conclusions are specific and based on credible findings, except one statement which is “There is a need for greater UNFPA Albania leadership, visibility and staff support for on PD issues” (p. 73). It is not clear how much leadership, visibility and staff support is needed. But this should be reflected in the recommendations.</p>
<p>7. Recommendations</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>Good</p> <p>Recommendations are divided into two parts “Strategic” and “Programme.” Recommendations flow logically from conclusions, for instance, it is said after the conclusion “(Linked to Strategic Conclusion 1, Program Areas SRH,GBV).” The recommendations flow logically from conclusions. Recommendations are strategic, targeted and operationally feasible: e.g. SRH R.5 provides options to assist the cervical cancer program effort with an “international consultant for additional short-term targeted support to re-invigorate the cervical cancer screening program” (p75). In this example, despite the use of broad terminology like “re-invigorate”, the report specifies a strategic recommendation and provides a targeted and operationally feasible course of action.</p> <p>Nothing is said in the report if the recommendations took into account stakeholders’ consultations. It may be this is because TOR did not require to answer this question. Finally, a minor issue is that all recommendations have “High” priority, so there is no priority order.</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>Good</p> <p>The Terms of Reference are provided in the annex of the report. Overall, the evaluation report attempts to respond to the scope and evaluation questions outlined in the ToR across the evaluation criteria (Relevance, Effectiveness, Efficiency, Sustainability) but this is undermined by the non-random sampling used in data collection. However, the report does respond to the evaluation criteria for UNCT Coordination and Added Value, providing additional sections for discussion on these topics. There is no annex “Methodological instruments used.</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)		2		
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		88	12	

(*) 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