

Organizational unit: **Papua New Guinea** Year of report: **2022**

Title of evaluation report: **UNFPA Papua New Guinea Country Programme Evaluation Sixth Programme Cycle, 2018 - 2022**

Overall quality of report: **Very good** Date of assessment: **12 December 2022**

Overall comments: Overall, this is a very good evaluation report that can be used by decision-makers with confidence. The sections on analysis and findings, conclusions, recommendations, and integration of gender were remarkably strong. The evaluation design was based on the Theory of Change; the results framework and the description of the data collection activities were provided in the annex, with shortcomings reported. The mixed methods approach enabled adequate triangulation of evaluative data in addition to validation through stakeholder consultations. One example of good practice is the very comprehensive table of UNFPA supported interventions, the implementing partners, other partners, and right holders for each strategic outcome presented in Annex 11. The evaluation findings systematically responded to all evaluation questions in a balanced way, with rigour and extensive triangulation of qualitative and quantitative data. In addition, the linkages between outputs and outcomes were demonstrated in the evaluation findings. The factors facilitating and/or hindering the achievement of results were reported. Furthermore, the findings, conclusions, and recommendations fully integrated human rights, gender equality, women's empowerment, and disability. The report clearly notes the lack of inclusion of persons with disabilities in the programme implementation despite being included in the programme design. However, it is not clear if persons with disabilities were consulted during data collection. The recommendations could have been more sharply formulated but are targeted and actionable with very clear prioritization. The methodology section could also have been sharpened by being organized by subsections and by being more explicit about the sampling plan, the rationale for data collection choices, and the reliability of data.

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 *Insert assessment level followed by main comments. (use 'shading' function to give cells corresponding colour)*

1. Structure and Clarity of Reporting Yes No Partial Assessment Level: **Fair**

To ensure the report is comprehensive and user-friendly

1. Is the report structured in a logical way? 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? Is there a clear distinction made between analysis/findings, conclusions, recommendations and lessons learned (where applicable)? Partial Overall, the report is logically structured and easy to navigate, although the Methodology section would benefit from having subsections. There are no major grammatical or spelling errors. The report uses tables and graphs effectively, and includes a few maps and key contextual facts of Papua New Guinea upfront to provide information about the country. However, to maximize the usefulness of the report, the recommendations could be more clearly stated, and the topic of both the recommendations and lessons highlighted.

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) No At 80 pages excluding the annexes, the report is longer than the 70-page requirements for CPEs.

3. Do the annexes contain – at a minimum – the ToRs; a bibliography; a list of interviewees; the evaluation matrix; methodological and data collection tools used (e.g. interview guides; focus group notes, outline of surveys)? Yes The report contains 16 annexes, including the TOR, the evaluation matrix, theory of change, list of consulted stakeholders, data collection tools, consent forms, and additional narrative information on evaluative data.

4. Is an executive summary written as a stand-alone section, presenting the i) Purpose; ii) Objectives, scope and brief description of interventions; iii) intended audience; iv) Methodology; v) Main results; Vi) Conclusions and Recommendations? Partial The executive summary mostly reads as a standalone document and includes all of the required subsections. However, it could be more clear about the outcome-level results of the Country Programme. Results are the topic of Conclusion 3 but are only discussed in terms of output achievement and restraining factors. It would be helpful if the reader was provided with more insight on the CP's main accomplishments, successes and shortcomings.

5. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)? Yes The executive summary is adequately concise at 5 pages in length.

2. Design and Methodology Yes No Partial Assessment Level: **Very good**

To ensure that the evaluation is put within its context

1. Is the development and institutional context of the evaluation clearly described and constraints explained? Yes The context section clearly describes the physical, political and socio-economic situation in Papua New Guinea including the physical topography, UN Human Development Index, the development challenges and national strategies. In addition, the context narrative provides a good overview of the national institutional and legislative framework related to sexual and reproductive health and Papua New Guinea's progress on achievement of Sustainable Development Goals.

2. Does the evaluation report discuss and assess the intervention logic and/or theory of change? Yes The report clearly outlines the three major outcomes of the sixth UNFPA Country Programme in Papua New Guinea 2018-2022 based on the TOC. In addition, the TOC in Annex 7 shows the interventions logic, depicting the results chain between outputs and outcomes, including assumptions and risks per outcome. The evaluators modified the TOC so they could use it to assess the causal linkages amongst output and outcome level changes.

To ensure a rigorous design and methodology

<p>3. 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?</p>	Yes	<p>The evaluation criteria and evaluation questions are clearly outlined. The evaluators presented the Country Programme results framework and overview of achievements in Annex 4, in addition to the overview of results achieved at outcome and output levels of CP results framework. The evaluation matrix presented in Annex 5 is organised by the evaluation questions and the relevant indicators, assumptions and methods for data collection and sources of information.</p>
<p>4. Are the tools for data collection described and their choice justified?</p>	Yes	<p>The methodology section in the report is very brief but with the methodological details being presented in Annexes 5 and 9. Annex 5 includes the methods (desk review, focus groups, interviews, and field observation), the type of sources associated with each method, and data analysis strategy. Although the evaluators do not clearly state their rationale for data collection choices besides stating the objectives for each method, they explain the use of multiple approaches such as theory-based, rights-based, and appreciative inquiry.</p>
<p>5. 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)?</p>	Partial	<p>Annex 11 provides a very comprehensive table of UNFPA supported interventions, the implementing partners, other partners, and right holders for each strategic outcome. The report provides general remarks about the roles of stakeholders in implementation. It is stated that the evaluation reference group were consulted at each of the stages of the evaluation process, but the report does not state explicitly how vulnerable groups, such as persons with disabilities or living with HIV, or their representatives were involved in the evaluation or the extent they have been included in the sample. It is stated in the report that the evaluation team presented and discussed the draft findings, conclusions and recommendations with UNFPA staff and ERG members.</p>
<p>6. Are the methods for analysis clearly described for all types of data?</p>	Yes	<p>The report indicates that the various data analyses methods were qualitative content analysis, context analysis, analysis of the TOC and results chain of the programme, SWOT analysis, contributions analysis, timeline and policy analysis. The evaluators provided the definition of what these methods are and a brief description of how each method was applied.</p>
<p>7. Are methodological limitations acknowledged and their effect on the evaluation described? Does the report discuss what was done to minimize such issues?</p>	Partial	<p>Some limitations such as the impact of the Covid pandemic on data collection were mentioned in the introduction, but the report could benefit from having a more structured section on limitations and mitigation strategies.</p>
<p>8. Is the sampling strategy described?</p>	Partial	<p>The sampling strategy could be more complete. The report states that sampling took place at the national and the sub-national levels and provides some further details on the sample strategy in Annex 9. Although the basis for selecting locations is noted - such as a variety of socio-geographic areas, and a mix of successful and challenging aspects of programming - it is not clear how people within those areas were selected as participants, how representative the sample was of different stakeholders, or how vulnerable stakeholder groups were represented in the sample.</p>
<p>9. Does the methodology enable the collection and analysis of disaggregated data?</p>	Partial	<p>Although the TOR stated clearly that the evaluation should disaggregate data by age, gender, vulnerable groups, etc., the report only mentions disaggregating data along gender: the other vulnerability criteria were not addressed. With respect to qualitative data collection, it is stated in the Introduction Section that of the 125 people interviewed, 53 percent were women. The gender disaggregated data are presented in the additional contextual details in Annex 8.</p>
<p>10. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, disability inclusion, gender equality and human rights)?</p>	Yes	<p>The report indicates that the evaluation followed a gender-sensitive approach to ensure equal representation of women, men girls and boys in all data gathering activities. It was stated that separate meetings were organized with male and female participants in focus group discussions. The evaluation matrix includes questions and indicators aimed at collecting disaggregated data, mostly by gender on GEWE outputs contributing to the achievement of the outcome level change, i.e. gender equality, the empowerment of all women and girls, and reproductive rights are advanced in development and humanitarian settings.</p>
<p>3. Reliability of Data</p>		
<p>Yes No Partial</p>		<p>Assessment Level: Fair</p>
<p><i>To ensure quality of data and robust data collection processes</i></p>		
<p>1. Did the evaluation triangulate data collected as appropriate?</p>	Partial	<p>Even though the data collection was described in the annex and the evaluators indicated that they used mixed methods and a participatory approach to allow triangulation of data across methods and respondents, there is no description on how data from the different sources were triangulated. It is not clear how the inclusion of multiple stakeholders allowed triangulation of data across the various methods and respondents to enhance validation of findings.</p>
<p>2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?</p>	Partial	<p>The report clearly stated that sources for both qualitative and quantitative data were field observations, focus groups and interview of evaluation participants and secondary data primarily from secondary data. However, the report is not explicit regarding the reliability of data besides information that data was validated through the various processes of the mixed methods approaches. The report would benefit from having a sub-section in evaluation methodology in which information on reliability and validation of data are more fully discussed.</p>
<p>3. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?</p>	Yes	<p>There is information in the evaluation matrix that the evaluation addressed collection of data with sensitivity to issues of discrimination. The evaluation matrix has questions and indicators that addressed issues of discrimination and there is evidence in the report that these issues were addressed throughout the evaluation. It is stated in the report that the evaluation matrix guided data gathering, analysis and reporting in the various phases of the evaluation process. Regarding ethical considerations, evaluators describe in Annex 9 how participants were informed about the confidentiality of information and their rights to respond or not respond to the interview questions. Adherence to UNEG ethical principles is also discussed.</p>
<p></p>		

4. Analysis and Findings	Yes No Partial	Assessment Level:	Very good
<i>To ensure sound analysis and credible findings</i>			
1. Are the findings substantiated by evidence?	Yes	The evaluation findings present a coherent and complete narrative that responds to each evaluation question with sufficient depth. Findings also provide a balanced picture of the programme's strengths and weaknesses. The report provides evaluative evidence in boxes, graphs, tables, and footnotes.	
2. Is the basis for interpretations carefully described?	Yes	There is evidence that the evaluation used the results framework of the CPD, the modified Theory of Change (TOC) presented in Annex 7, and the UNFPA global results framework as a framework for the interpretation of evaluation findings.	
3. Is the analysis presented against the evaluation questions?	Yes	The findings are structured along the evaluation questions and a brief summary of the findings for each evaluation question, making it easy for the reader to understand how these have been answered.	
4. Are the cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Yes	The findings are premised on the results framework and TOC in order to assess the causal linkages amongst output and outcome level changes. The unintended results of the intervention are discussed in paragraph 172 and throughout the findings sections.	
5. Does the analysis show different outcomes for different target groups, as relevant?	Yes	The findings clearly discuss the extent to which the programme has reached vulnerable groups, including women, youth, and persons living with HIV. Specifically, Evaluation Question #2 addressed the extent to which the programme integrated human rights, gender equality, women's empowerment, and disability inclusion. While gender equality and women's empowerment concerns are integrated across outcomes and emergency responses, the report is critical of insufficient inclusion of disability beyond programme design into implementation.	
6. Is the analysis presented against contextual factors?	Yes	Factors hindering and enabling the achievement of results are clearly discussed. For example, Box 2 includes factors such as cultural norms and beliefs, community stigma, and discrimination amongst barriers and facilitators for women's access and use of family planning products and services. Other factors affecting these services were identified in the report as geographical location (urban vs rural), marital status, and age.	
7. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, disability inclusion, gender equality and human rights?	Yes	As noted above in section 4.5, the evaluation findings elaborated on cross-cutting issues of equity and vulnerability, gender equality and human rights, in particular the vulnerability of young adolescent women. It is noted that persons with disabilities were mentioned in the design of interventions as well as in the CO Work Plans, but this did not translate into their inclusion in the implementation of the programme.	
5. Conclusions			
	Yes No Partial	Assessment Level:	Good
<i>To assess the validity of conclusions</i>			
1. Do the conclusions flow clearly from the findings?	Yes	The report has a standalone section on conclusions. All 12 conclusions are explicitly linked to their corresponding findings in the corresponding evaluation questions.	
2. Do the conclusions go beyond the findings and provide a thorough understanding of the underlying issues of the programme/initiative/system being evaluated and reflect as appropriate cross-cutting issues such as equality and vulnerability, disability inclusion, gender equality and human rights?	Partial	The conclusions are analytical and demonstrate a strong understanding of the key issues underlying the country programme in Papua New Guinea. The underlying issues related to the country context, the UNFPA strategic directions, and the cross-cutting issues - including gender equality, equity and vulnerability, human rights, and disability inclusion - are all addressed. However, the conclusions span six pages and include some findings-level information (i.e., the conclusions on staffing and monitoring) which somewhat diminishes the usefulness of this section. A more concise presentation would be appropriate.	
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	The conclusions highlight both the strengths and weaknesses of the programme and marshal evidence from the quantitative and qualitative findings.	
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 of the eight main recommendations is linked explicitly to its corresponding finding and the relevant evaluation conclusions. The recommendations are categorised as strategic or programmatic.	
2. Are the recommendations targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?	Yes	All the recommendations appear actionable and useful. Each clearly identifies the users responsible for implementation. The recommendations propose clear actions and also discuss implications for their implementation.	
3. Do recommendations appear balanced and impartial and address, as relevant, key cross cutting issues such as equity and vulnerability, disability-inclusion, gender equality and human rights?	Yes	Recommendations are balanced and clearly address key issues raised by the evaluation. Recommendations also address cross-cutting issues, including gender equality, equity and vulnerability.	
4. Are the recommendations prioritized?	Yes	The level of priority is identified for each recommendation was identified as ranging from intermediate to high.	

7. Gender	0	Assessment Level:	Very good
	1		
	2		
	3 (**)		

To assess the integration of Gender Equality and Empowerment of Women (GEEW) (*)

<p>1. Is GEEW integrated in the evaluation scope of analysis and indicators designed in a way that ensures GEEW-related data to be collected?</p>	<p>a. Does the evaluation include an objective specific to assessment of human rights and gender equality considerations or was it mainstreamed in other objectives? (Score: 0-3) The scope of the evaluation includes a focus on GEEW. (3)</p> <p>b. Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria? (Score: 0-3) The evaluation does not include a separate criterion on gender and human rights, however, GEWE is adequately mainstreamed in the evaluation matrix and findings. (3)</p> <p>c. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation? (Score: 0-3) The evaluation questions #1, 2 and 3 adequately integrate GEWE considerations. (3)</p> <p>d. Does the evaluation assess whether sufficient information was collected during the implementation period on specific result indicators to measure progress on human rights and gender equality results?(Score: 0-3) The use of the results framework and indicators provided a framework for assessing whether or not sufficient information was collected human rights and gender equality results. Data was disaggregated by gender, when possible. (3)</p>
<p>2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques?</p>	<p>a. Does the evaluation specify how gender issues are addressed in the methodology, including: how data collection and analysis methods integrate gender considerations and ensure data collected is disaggregated by sex? (Score: 0-3) The evaluation uses appropriate gender-responsive approaches in the methodology. (3)</p> <p>b. Does the evaluation methodology employ a mixed-methods approach, appropriate to evaluating GEEW considerations (collecting and analyzing both quantitative and qualitative data, and ensuring the appropriate sample size)? (Score: 0-3) The methodology uses a mixed-methods approach. However, evaluation participants are not disaggregated by sex. The evaluators held separate focus groups with men and women. (2)</p> <p>c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility? (Score: 0-3) The evaluation consulted a range of data sources which allowed for data to be triangulated and validated in stakeholder meetings. (3)</p> <p>d. Do the evaluation methods and sampling frame address the diversity of stakeholders affected by the intervention, particularly the most vulnerable, where appropriate? (Score: 0-3) The sampling frame includes a wide range of stakeholders. However, it is unclear the extent to which particularly vulnerable groups were consulted since the report does not include the list of direct beneficiaries consulted during data collection and analysis phases. (1)</p> <p>e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality? (Score: 0-3) The evaluation described the ethical approaches used. (3)</p>
<p>3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?</p>	<p>a. Does the evaluation have a background section that includes an intersectional analysis of the specific social groups affected by the issue or spell out the relevant normative instruments or policies related to human rights and gender equality? (Score: 0-3) The background section includes a good gender and intersectional analysis. (3)</p> <p>b. Do the findings include data analysis that explicitly and transparently triangulates the voices of different social role groups, and/or disaggregates quantitative data, where applicable? (Score: 0-3) The findings provide a good gender analysis and presents disaggregated data where possible. However, the voices of different groups could have been conveyed more clearly through direct quotes. (2)</p> <p>c. Are unanticipated effects of the intervention on human rights and gender equality described? (Score: 0-3) The evaluation does include the unanticipated effects of the intervention. (3)</p> <p>d. Does the evaluation report provide specific recommendations addressing GEEW issues, and priorities for action to improve GEEW or the intervention or future initiatives in this area? (Score: 0-3) Both the strategic and programmatic recommendations specifically address GEWE issues. (3)</p>

(*) This assessment criteria is fully based on the UN-SWAP Scoring Tool. Each sub-criteria shall be equally weighted.

(**) Scoring uses a four point scale (0-3).

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	71	11	18	
Overall assessment level of evaluation report	Very good			

The evaluation integrates adequately cross cutting issues of gender equality, human rights and disability inclusion, even though is not included as part of the evaluation objective.

Consideration of significant constraints (e.g. COVID-19 or civil unrest)

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

Yes

No

If yes, please explain: