

Organizational unit: **UNFPA Indonesia Country Office** Year of report: **2020**

Title of evaluation report: **Evaluation of the UNFPA 9th Country Programme of Assistance to the Government of Indonesia**

Overall quality of report: **Very Good**

Overall comments: This report provides a very comprehensive evaluation of UNFPA Indonesia's 9th Country Programme with a adequate integration of ethical standards and GEEW considerations within the scope, methodology, findings and recommendations. The strengths of the evaluation were in the analysis and findings and conclusions, which presented a clear analysis of CP results aligned with the United Nations Partnership for Development Framework, drawing cause-and-effect links and presenting data which was well validated, considering the apparent time constraints. The evaluation reached 185 stakeholders in an evaluation lasting only 25 days through focus groups, interviews and surveys. The report provides findings that reflect good analysis of the programme's strengths, achievements and challenges from a human rights and gender equality (GEEW) lens. The conclusions add additional value to the findings, clearly bringing the findings together to present an in-depth understanding of the country programme and context. Recommendations are coherently linked to findings and conclusions, and are clearly actionable, prioritized, and target users. However, more attention could be given to strengthening the evaluation design and methodology, more clearly incorporating GEEW considerations, and to the structure and clarity of reporting by correcting grammatical errors and more clearly drawing on stories and quotes from stakeholders, particularly rights holders engaged in the focus groups.

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)*

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 easy to read and understand (i.e. written in an accessible language appropriate for the intended audience) with minimal grammatical, spelling or punctuation errors?	Partial	The report is largely accessible and easy to understand, however, there are grammatical, punctuation and other editing errors, such as inconsistent use of capitalization and sentences with missing verbs or modifying words, which limit its readability. In addition, some acronyms are used which have not been first introduced: for example, the tables defining level of achievement across outputs (p20 and p27) mention MHH, RFP, YPD, GEN, PD, which can be inferred to align with the programme areas but are not explicitly explained or referenced.
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)	Partial	At 72 pages, excluding annexes, the report is slightly over the maximum page length for country programme evaluations.
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 logically structured, first presenting the methodology and country programme background and context, before highlighting findings, conclusions and recommendations. The findings are organized by the evaluation criteria and questions, including text boxes to highlight key findings under each criterion.
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 annexes contain all the required information, in addition to descriptive summaries of survey results as well as more detailed supporting evidence of key findings.

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?	Partial	The executive summary presents the main results of the evaluation, and highlights the recommendations drawn as a result. The usability of the section could have been improved for its intended users if summarized conclusions were more clearly organized around evaluation criteria. In addition, there are some grammatical and spelling errors within the executive summary, which also limits its clear presentation of 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	The executive summary includes all of the required information and follows a clear structure.
7. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	Yes	The executive summary is within the required page limits.

2. Design and Methodology	Yes No Partial	Assessment Level:	Fair
----------------------------------	----------------------	-------------------	-------------

To ensure that the evaluation is put within its context

1. Does the evaluation describe the target audience for the evaluation?	Yes	The target audience is described and drawn directly from the ToR.
2. Is the development and institutional context of the evaluation clearly described and constraints explained?	Yes	The development and institutional context of the evaluation is clearly described within the chapter on the country context, and more specifically in the section on 'development challenges and limitations', which highlights gaps in laws and services affecting population dynamics and sexual and reproductive health for the population.
3. Does the evaluation report describe the reconstruction of the intervention logic and/or theory of change, and assess the adequacy of these?	Partial	The evaluation assesses the adequacy of the theory of change, and since a clear theory of change was included during a recent mid-term review, the evaluators determined it sufficient and clear enough for use within the evaluation of the country programme considering existing time constraints. However, while the evaluators note gaps in the ToC that were not filled (risks, assumptions and limitations), the theory of change and overall vision (One Voice diagram) which the evaluation draws on, is not included but would have been useful as a reference and for providing additional clarity.

<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 not fully described in the text, but the questions are presented on page 3 with a rationale for streamlining questions from those previously identified in the ToR. The evaluation matrix provided in Annex 5 includes the evaluation questions, assumptions, indicators, data sources and methods for data collection.	
5. Are the tools for data collection described and their choice justified?	Partial	The data collections tools and rationale for their selection are not clearly described, though the protocols are provided in the annex. The description of data collection tools is folded into the discussion of sampling, without clearly defining the justification for each tool, and then again folded into the 'Process Overview'. The justification for the tools can be gleaned from the discussion of limitations and risks as well as the process description, though overall, this component of the evaluation could have been made more clear. For example, the number of stakeholders engaged within the evaluation is mentioned, though the data is not disaggregated by evaluation methodology (i.e. numbers completing surveys, interviews, focus groups).	
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	The stakeholder map's use in determining the sample selection is described and the map is presented as Annex 4, which organizes types of stakeholders in each of the four core programme areas. The process for incorporating stakeholder feedback on the report findings and recommendations is also mentioned. While the total universe is not described in terms of numbers, the evaluation notes reaching a total of 184 stakeholders identified in the mapping process.	
7. Are the methods for analysis clearly described for all types of data?	Yes	The methods of data analysis noted are context analysis and statistical analysis to generate descriptive statistics.	
8. Are methodological limitations acknowledged and their effect on the evaluation described? (Does the report discuss how any bias has been overcome?)	Yes	The methodological limitations are clearly articulated and actions to mitigate and reduce these limitations are also provided.	
9. Is the sampling strategy described?	Yes	The evaluation team used a purposive sampling approach to select sites for visits and participant focus groups/interviews, as well as stakeholder interviews. The criteria for selection include areas with pilot or high impact interventions, presence of interventions across all programme areas, the humanitarian situation and existence of some monitoring data, as well as individuals' depth of involvement in activities. Convenience sampling is specifically noted for the selection of participants for focus groups.	
10. Does the methodology enable the collection and analysis of disaggregated data?	Yes	The report is explicit about how the methodology will enable data collection and analysis of disaggregated data. The report specifically notes efforts to disaggregate data by sex and age, where feasible, which is also reflected in the presentation of findings.	
11. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?	Yes	The evaluation framework explicitly includes indicators for assessing cross-cutting issues, primarily in the assessment of relevance. Furthermore, the annexed list of stakeholders presents an extensive and diverse list of stakeholders (both rights holders and duty bearers) consulted as part of the evaluation. The evaluation also notes that: "Special consideration was made, where feasible, to include and reflect how boys, girls, men and women, and those belonging to marginalized groups, are affected differently in the CP9 design and implementation" (p. 6).	
3. Reliability of Data			
	Yes No Partial	Assessment Level:	Fair
<i>To ensure quality of data and robust data collection processes are used in the findings</i>			
1. Did the evaluation triangulate data collected as appropriate?	Yes	The evaluation notes the use of triangulation as a means to ensure data validity and quality. The triangulation process also included the validation of preliminary findings with key stakeholders to ensure that no factual or interpretative errors were present that would change the findings.	
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Partial	The evaluators are explicit about the source of qualitative data coming mainly from interviews and focus groups, and frequently noted limitations to data reliability (and associated mitigation strategies) as a result of time constraints. The discussion of efficiency and sustainability also draws quantitative data from the survey, and the evaluators note where clear findings could not be drawn as a result of missing or insufficient baseline information. Based on the list of stakeholders included in the annex and available citations in the report, it appears the evaluators were able to reach and access data from an indicative sample. However, the extent to which the evaluators make use of reliable data across qualitative and quantitative sources is not always clear as quotes are not consistently introduced or cited (see page 30, paragraph 2 as an example). In addition, there are few references to the focus groups conducted with participants/rights holders, and so it is not clear the extent to which data from these sources was incorporated. However, for the most part, it is clear that reliable qualitative and quantitative data sources were used.	
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 primary limitation described in the report is the availability of time to conduct the evaluation. The time constraint subsequently limited the evaluation team's ability to reconstruct the ToC and reach a larger sample (both in terms of stakeholders and geographic locations). Mitigation methods were primarily to use existing data, such as the existing ToC and programme reports, as well as to conduct an online survey to reach a larger group within the time constraints.	
4. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?	Yes	It is noted in the report that the evaluators followed UNEG guidelines and standards and ensured respondents' rights through informed consent, confidentiality and anonymity, and clear communication about the evaluation purpose. In addition, the evaluators mentioned how the purposive and convenience sampling methodologies may have limited the data collected. Since there was a relatively short time to collect data from quite an extensive group of stakeholders, the team only visited one site but made efforts to engage several small groups in different locations within the Palu area.	

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	As previously mentioned, the report does reference data sources in footnotes in most cases, and the use of quotes, statistics, and citations in most instances shows that the evaluators substantiated findings with evidence.	
2. Is the basis for interpretations carefully described?	Yes	This is well done throughout the Findings section across evaluation criteria. Findings are clearly drawn from the data collected and carefully interpreted.	
3. Is the analysis presented against the evaluation questions?	Yes	The relevant evaluation questions are listed at the beginning of the discussion of each criterion, and the analysis is structured accordingly.	
4. Is the analysis transparent about the sources and quality of data?	Yes	As noted earlier, the evaluation team referenced sources in footnotes and in-text, although the more frequent use and citation of direct quotes from focus groups could improve the strength of the analysis, particularly in the effectiveness section. The analysis is transparent about the sources and quality of data.	
5. Are cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Yes	While the evaluation did not present a reconstructed theory of change, the findings are framed around the cause and effect links between outputs and outcomes drawn from the United Nations Partnership for Development Framework within discussion of CP effectiveness. Unintended effects are also highlighted in a separate section (p. 62).	
6. Does the analysis show different outcomes for different target groups, as relevant?	Yes	The evaluation findings present different outcomes for different target groups, providing a clear gender analysis where possible and analyzing the impacts across diverse groups (for example, female sex workers, persons living with HIV, survivors of gender violence and youth across age groups).	
7. Is the analysis presented against contextual factors?	Yes	Findings are analyzed against relevant contextual factors, including changing government objectives, available funding, challenges with coordination and humanitarian crises/disasters.	
8. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?	Yes	As mentioned previously, the analysis elaborates on cross-cutting issues through the report, including noting where the needs of specific groups are not sufficiently addressed (such as with LGBT populations). Gender disaggregated data is included, where available and relevant.	
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	The conclusions are clearly drawn from the findings. The evaluators have specified the respective evaluation questions linked to each conclusion.	
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 are well-written and clearly bring together the findings across evaluation questions and criteria to present an overarching situation analysis for the country programme. The conclusions aggregate information across programme components (for example, defining the total number of knowledge products developed), which provides an added level of depth and clarity, and demonstrates a thorough understanding of the CP. The conclusions are categorized as strategic or programmatic.	
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	The conclusions are clearly based on evidence from the findings and therefore do not convey bias.	
6. Recommendations	Yes No Partial	Assessment Level:	Good
<i>To ensure the usefulness and clarity of recommendations</i>			
1. Do recommendations flow logically from conclusions?	Yes	The recommendations flow logically from the conclusions and are directly linked to conclusions by the evaluators. In most cases, the evaluators link the recommendations to six or more conclusions, though the rationale for this is presented within a summary at the start of the section which calls for more systematic approaches and thinking (as opposed to narrow and siloed thinking).	
2. Are the recommendations clearly written, targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?	Yes	The recommendations are very clearly targeted and action-oriented, explicitly defining the overall responsible institutional body or department as well as an action plan, which in some cases also defined more explicitly the persons or organizational levels responsible for each action item. The human and technical implications of recommendations are analyzed, though the evaluation notes that information was not available on resource allocation for the action plans in order to define the financial implications.	
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?	Partial	The recommendations did not clearly propose a timeframe for implementation. The evaluation only notes that: "the time period is for CPI0 and some design and HR related recommendations may have to be implemented during CP9 (2020) in preparation for CPI0." As such, it can be seen that the time period extends beyond one year into CPI0 implementation.	
5. Are the recommendations prioritized and clearly presented to facilitate appropriate management response and follow up on each specific recommendation?	Yes	The classification into high and medium seems appropriate to facilitate management response, considering that the evaluators got feedback from the CP stakeholders on the recommendations. Like the conclusions, the recommendations are also organized around the strategic and programmatic.	

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?			<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 evaluation does not define a specific objective related to human rights and gender equality, and it is not clearly mainstreamed into other objectives. (Score = 1).</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) While there is no standalone criterion on gender and/or human rights, it is mainstreamed into other evaluation criteria through the assumptions and indicators defined within the evaluation framework, notably within the assessment of relevance. As a result of the programme area on 'gender equality and the empowerment of women', GEEW is comprehensively addressed within its unique section in the evaluation framework. (Score = 3).</p> <p>2 c. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation? (Score: 0-3) Yes, as noted above, GEEW was integrated into evaluation questions, notably question one under relevance. (Score = 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) While the efficiency of monitoring and evaluation systems in turning data to support decision-making is assessed, the evaluation does not clearly or systematically assess whether sufficient information was collected during the implementation period to measure human rights and gender equality results. (Score = 1).</p>
2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques?	2		<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 methodology was gender responsive with regards to the methods and tools for data collection. For example, it is noted in the report that evaluators made conscious effort to select a fair representation of males, females and youth. In addition, a gender disaggregated list of persons consulted is included as an annex. How data collection efforts integrated gender considerations beyond representation/participation is not made clear. (Score=2).</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) Considering the clear time constraints, the evaluation did well to include a diverse sample of stakeholders through a mixed-methods approach. Associated questions on gender are clearly integrated into data collection tools, and list of interview respondents is sex disaggregated. (Score = 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 methodology notes that triangulation is applied through inclusion of diverse methods and sources, and that findings were validated by the CO. However, data drawn from focus groups with participants is infrequently referenced. (Score = 2)</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) Yes, this was ensured with samples representing the primary beneficiaries as noted above. However, since the sampling process engaged participants based on convenience, it is not made clear how the most vulnerable were included in the evaluation design. (Score = 2).</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 notes that stakeholders were treated with respect to their rights, integrity and confidentiality through data collection and analysis. (Score = 3).</p>
3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?	2		<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 evaluation report has a background section that includes an intersectional analysis of specific social groups affected and highlights normative instruments, frameworks and policies in Indonesia. The background section notes development challenges and national strategies aligned with each of the four core programme areas. (Score = 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) While the findings include data analysis that explicitly triangulates the voices of different groups, this is not consistently done. Quantitative data is used especially in the tables that show achievements against baselines and targets, and gender disaggregated data is presented, where available. However, there is minimal use of quotes or stories to highlight the voices of different groups, particularly participants and rights holders. The findings draw extensively on secondary data sources and interviews with duty bearers and programme staff (Score = 2).</p> <p>c. Are unanticipated effects of the intervention on human rights and gender equality described? (Score: 0-3) The unanticipated effects of the intervention are described in its own section, and include some assessment of impact on human rights and gender equality. For example, a government decree required VAW-related services to be covered within the local health insurance system, increasing affordability and access of survivor care. (Score = 2).</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) The primary recommendation highlighting GEEW issues relates to developing a male engagement strategy to more systematically address harmful gender norms and inequalities. (Score = 2).</p>

(*) 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 totaling 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		
Total scoring points	51	18	31	
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?

• 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: