



Organizational unit: **Pakistan** Year of report: **2022**

Title of evaluation report: **Country Programme Evaluation of the Ninth Country Programme (2018-2020) of Assistance to Government of Pakistan**

Overall quality of report: **Very good** Date of assessment: **20 May 2022**

Overall comments: This is a thorough and clear assessment of Pakistan's 9th Country Programme. Evaluators used a mixed-method approach to obtain qualitative and quantitative data from document review and remotely-conducted key informant interviews. The selection of respondents for interviews is based on a detailed mapping process. The methodology utilized is well-explained in the main report and backed up by annexes. To acquire qualitative data from primary sources, a content and thematic analysis framework is used, whereas descriptive statistical approaches such as tabulations and graphing are employed to analyze secondary data. Along with data triangulation, contribution analysis is used to analyze the coherence of the results chain and intervention logic, although shortcomings of the ToC could have been more thoroughly addressed. The findings are backed up by evidence derived from both primary and secondary sources. The quantitative data is gathered based on sex, region, and stakeholder type, and it is utilized to counteract any potential bias from key informant interviews. Cross-cutting issues of human rights, vulnerability and gender equality are incorporated in evaluation question 1 and are well addressed in findings, conclusions and recommendations. Conclusions and recommendations are clear and appear useful for decision-making. The evaluation addresses the success of the CP in meeting the needs of people with disabilities (PWD). An organization representing PWD is highlighted as being part of the ERG. This population is mostly covered under the umbrella of vulnerable groups although the analysis is more specific in some cases such as the analysis of the SRH component which highlights young people living with disabilities. Conclusions also specifically mention PWD; recommendations refer to marginalized groups more generally. It is noted that as part of the participatory approach adopted by the evaluation, attention was paid to ensuring the participation of women with disabilities in the evaluation, although it is not clear if this was actually achieved given the limitations of having to rely on remote data collection methods.

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		<i>Insert assessment level followed by main comments - (use 'shading' function to give cells corresponding colour)</i>	
1. Structure and Clarity of Reporting	Yes No Partial	Assessment Level:	Good
<i>To ensure the report is comprehensive and user-friendly</i>			
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)?	Yes	The report is well-organized. The language is simple and straightforward, with little jargon. Some minor formatting and typing errors can be identified in both the report and the annexes. The analysis/findings, conclusions, recommendations, and lessons learned are stated in a logical order, with the emphasized idea at the top and supporting details below. Evaluators should be encouraged to use rights-based terminology (referring to 'duty-bearers' and 'rights holders') in future reports.	
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 approximately 75 pages including the executive summary, the report exceeds the permitted length by 5 pages.	
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 annexes include all required elements in addition to a stakeholder map, the CPE Agenda, and more information on sampling and programme performance.	
<i>Executive summary</i>			
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 is well written and mostly serves as a stand-alone section. It includes all required components plus a lessons learned section. The main shortcoming is that the methodology does not indicate the number of people that participated in the evaluation process or whether data collection relied on remote processes - this is useful information for conveying the credibility and breadth of the exercise.	
5. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	Yes	The executive summary is written in exactly five pages with reasonable information provided in each part.	
2. Design and Methodology			
<i>To ensure that the evaluation is put within its context</i>			
1. Is the development and institutional context of the evaluation clearly described and constraints explained?	Yes	Population, poverty, the influence of COVID-19 on poverty, the Human Development Index, gender equality, economics, and security and humanitarian challenges are all part of the general development environment described in Chapter 2. Institutional difficulties are addressed as part of national policies that incorporate the role of external assistance.	
2. Does the evaluation report discuss and assess the intervention logic and/or theory of change?	Partial	Chapter 3 contains thorough information on both the theory of change and intervention logic. The theory of change was revised, but only by adding dates to when outputs and outcomes should be achieved. The evaluators did not note that the output-level results are actually outcomes. The ToC is included as figure 9 but the text is too small to easily read.	

<i>To ensure a rigorous design and methodology</i>			
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?	Partial	The evaluation framework is briefly described in the methodology section of the report, and the annex contains the complete matrix with specifics on evaluation questions, assumptions, indicators, data sources, and data collection methods. The evaluation matrix does not include summarized findings although some findings information can be found in Annex 10: Performance against CPD Indicators.	
4. Are the tools for data collection described and their choice justified?	Yes	The data collection tools are extensively detailed, including the justifications for choosing them. These include key informant interviews and document review.	
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)?	Partial	The annex contains a detailed stakeholder map. Key stakeholders, including UNFPA CO provided comments and feedback on conclusions and recommendations. Recommendations were finalized in consultation with the UNFPA CO after follow-up rounds of validation with the ERG. A shortcoming is that some parts of the methodology section appear to not have been updated from the Inception Report - there is mention of the involvement of marginalized groups and that permission was sought for interviewing minors under Ethical Considerations - but the limitations clearly state that community level beneficiaries were not engaged.	
6. Are the methods for analysis clearly described for all types of data?	Yes	For quantitative data, descriptive statistics are utilized; for qualitative data, content and thematic analysis are applied.	
7. Are methodological limitations acknowledged and their effect on the evaluation described? Does the report discuss what was done to minimize such issues?	Yes	Methodological limitations including their mitigation strategies are comprehensively described. They are related to COVID-19, gender bias, and other data limitations. The effect of the Covid-19 restriction on key informant interviews is discussed, as well as the methods that were taken to address the problem. Gender bias was said to be addressed by a balanced analysis that highlights female voices. To counteract potential biases from unheard recipients, quantitative data was employed.	
8. Is the sampling strategy described?	Yes	The sampling approach is explained as being based on the existing stakeholder map and desk reviews. The sampling strategy is clearly described under section 1.3.5, with the evaluators noting they adhered to guidelines outlined in the UNFPA Handbook recommendation on illustrative indicators of sample to select purposeful sampling. The evaluators incorporated a non-random purposive sampling strategy for the selection of sample of stakeholders at national and sub-national levels. The universe for the evaluation included all stakeholders engaged in the implementation of UNFPA interventions.	
9. Does the methodology enable the collection and analysis of disaggregated data?	Yes	Evaluators used an appropriate methodology to enable collection and analysis of disaggregated data. EQ6 assesses the extent to which UNFPA support contributed to improved disaggregation of data to ensure that "evidence-based development and implementation of plans, programmes and policies reflects needs of variety of stakeholders, including those furthest behind". Indicators in the evaluation matrix were also designed to capture disaggregated data. Additionally, it is noted under methodology that "descriptive data on indicators was disaggregated by key variables such as region, sex and stakeholder type (e.g., government, NGO, academia)	
10. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, disability inclusion, gender equality and human rights)?	Yes	EQ 1 particularly focuses on the program's adaptation to the needs of the marginalized and vulnerable, as well as issues regarding human rights and gender equality. The remaining evaluation questions focus on the cross-cutting concerns that each topic highlighted. In addition, evaluators note that "most of the UNFPA interventions were implemented at national and sub-national levels, which made it challenging to identify the direct beneficiaries of the interventions", however "rightsholders of humanitarian assistance from KP province were included in the sample".	
3. Reliability of Data			
	Yes No Partial	Assessment Level:	Very good
<i>To ensure quality of data and robust data collection processes</i>			
1. Did the evaluation triangulate data collected as appropriate?	Yes	Evaluators note that, "Data quality was maintained by triangulating the data sources and methods of collection and analyses". Additionally, the evaluation matrix shows the various data sources from which data was collected.	
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Yes	This criteria was fully addressed. Evaluators clearly discuss data quality assurance processes under section 1.3.7, specifying the type of data they were attempting to obtain. To mitigate possible biases from qualitative data gathered through interviews, they leveraged secondary data from internal UNFPA reports for quantitative data. They then validated any preliminary findings and recommendations with key stakeholders to ensure quality and reliability.	
3. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?	Yes	Referencing the UNEG Ethical Guidelines and Code of Conduct, the evaluation team listed several good practices that they followed including ensuring confidentiality and seeking consent from all respondents before conducting interviews. They note the potential for bias related to gender differential was limited through "a balanced analysis (gender equity perspectives) to bring out the voices of females (the under-represented category)".	
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	Each finding is well supported. Summary findings and topic headings are provided for each evaluation criterion, together with supporting facts and cited evidence. The findings are backed by adequate primary as well as secondary data. There are a total of 510 footnotes specifying the source from which the data was obtained.	
2. Is the basis for interpretations carefully described?	Yes	Where applicable, the foundation for interpretations is carefully described by quoting KIs as well as secondary data. The evidence for the UNFPA's contribution, for example under relevance criteria, it is noted that "the national wide lockdown in Pakistan resulted in the closure of some government services but UNFPA's support for operationalizing essential services and servicing hotlines for GBV survivors, catered directly to the needs of vulnerable women and girls. It was established that domestic violence (DV) increased due to the amplified exposure of women and children to the perpetrators, making them more vulnerable".	
3. Is the analysis presented against the evaluation questions?	Yes	The presented analysis provides comprehensive responses to the evaluation questions. Where applicable, each aspect of the UNFPA strategic plan is explained in detail. Findings are organized by criteria and evaluation questions.	
4. Are the cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Yes	Each question includes a thorough description of the cause-and-effect link. For example, under effectiveness criteria, evaluators assess contribution of outputs towards achievement of outcomes under a dedicated paragraph also containing a table showing performance achievement of output indicators. This is a good practice. Similarly, causal linkages are also noted under other criteria. For example, under relevance criteria, it is noted that "UNFPA's advocacy efforts and technical support resulted into the Sindh Reproductive Health Rights Bill 2019 being endorsed by the Sindh Assembly and the legislation on Reproductive Health Rights bill and Early Marriage Restraint Act being finalized and duly vetted by Law department in KP Province". Additionally, unintended outcomes were considered, and the evaluators state that no unintended outcomes were identified after in-depth assessment.	
5. Does the analysis show different outcomes for different target groups, as relevant?	Yes	Where applicable, the analysis highlights any differences in outcomes for distinct target groups. For example, under effectiveness criteria, evaluators note that marginalised groups such as women with disabilities, from religious/ethnic minorities, and rights-holders from urban slum/ remote rural areas were not targeted by the programme. Additionally, regional disparities are also described. For example under section 4.3.3.4 Leaving No One Behind, it is noted that "Evidence exists on regional disparities, where in KP/ NMD, and Balochistan, women suffer from GBV far more than in other areas, and in the case of the NMD, it was observed that a whole 95 percent of women and 75 percent of men believed that wife beating was justified. However, the evaluation observed that this segment of the women population remained untargeted".	
6. Is the analysis presented against contextual factors?	Yes	The evaluation takes into account the social and legal background for the CP. For example, the study highlights the COVID-19 context after Pakistan adopted a COVID-19 Response Framework based on WHO guidelines and national priorities. Under the connectedness criteria, evaluators note that "Drought in some districts of Sindh and Balochistan provinces was an emergency/ humanitarian need. Timely response from UNFPA was coherent and well connected with the population needs. The community resilience was enhanced among beneficiaries (including people living with disabilities) through the provision of SRH and GBV services."	
7. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, disability inclusion, gender equality and human rights?	Yes	Findings provide adequate analysis on cross-cutting issue such as gender, vulnerability, and disability inclusion. This is particularly evident under the relevance, effectiveness, and coverage criteria. For example, under effectiveness criteria, it is noted that "Gaps were identified and new areas were highlighted to be incorporated in the policies including the support for persons with disabilities and additional focus on adolescents and youth SRHR".	
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	Conclusions directly refer to the findings and are divided into strategic and programmatic level conclusions. Each specifies which evaluation question from which it was derived.	
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?	Yes	The conclusions provide a thorough and higher-level understanding of the program being evaluated. Conclusion 5 and Conclusion 7 particularly focus on cross cutting issues. For example, conclusion # 7 states "UNFPA was committed to the approach of Leaving No One Behind. However, not all vulnerable and marginalised groups were reached. There is need for novel ways and models of reaching the furthest behind first with quality information and services related to UNFPA's mandate areas".	
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	There is no indication of bias as there are direct linkages to each finding.	
6. Recommendations	Yes No Partial	Assessment Level:	Very good

To ensure the usefulness and clarity of recommendations			
1. Do recommendations flow logically from conclusions?	Yes	The recommendations are presented both at strategic and program levels. They flow logically from conclusions as reflected by their links to specific conclusions.	
2. Are the recommendations targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?	Yes	Each recommendation has adequate guidance for its implementation, citing financial, human resource, and technical implications as relevant. Target users are clearly shown.	
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?	Partial	The recommendations appear balanced and impartial. Cross-cutting issues of gender, equity and vulnerability are well addressed. However, recommendations do not explicitly cover disability.	
4. Are the recommendations prioritized?	Yes	They are divided into two levels: strategic and programmatic, with priority levels embedded within each.	
7. Gender			
	0 1 2 3 (**)	Assessment Level:	Very 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 objectives address assessing the topic of gender/GBV along with other programme areas but could be more specific about assessing whether HRGE was integrated across the CP. (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) HRGE is mainstreamed into the criteria as a cross-cutting issue, and is covered under relevance, effectiveness, and coverage criteria. (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) There are three EQs that focus on GE and vulnerable populations (EQ1, EQ5, and EQ9, respectively, under relevance and effectiveness and coverage criteria). (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 evaluation assesses whether sufficient information (to check the extent of UNFPA support) was collected during the implementation period on human rights and gender equality results as reflected by evaluation question 1. (3)</p>	
2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques?		<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) Evaluators note that given the high proportion of male evaluation participants (67% male and 33% female), the team adopted gender equity perspective) to bring out the voices of females (the under-represented category). It is also noted that the evaluators conducted debriefing sessions with the team members to dissect the analysis from the data to reduce bias. However, it would have been useful to provide a subsection explicitly describing the process of gender-equity perspectives and specific steps taken by the team to bring out gender perspectives. Evaluation participants are disaggregated by gender and stakeholder group. (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) The evaluation used an appropriate mixed-method approach. Data was disaggregated by sex, region, among other attributes. Data was collected from a wide variety of stakeholders including representatives of beneficiaries, however, beneficiaries themselves did not participate in the data collection (3)</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) This was adequate given the constraints of covid. There was a diversity of partners and implementing agencies participating, although not beneficiaries. Triangulation was used by combining data from document analysis and primary research. Data quality assurance used validation by key stakeholders to ensure the absence of factual and interpretation errors. (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) Evaluators note that they did not interact directly with community level beneficiaries but instead interviewed NGO representatives of beneficiaries. To mitigate this limitation, the Evaluation Team notes that they complemented the qualitative data with quantitative data from the KILs. (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 report makes reference to UNFPA Evaluation Policy, UNEG Ethical Guideline and Code of Conduct for Evaluations and the United Nations Norms and Standards for evaluation. Additionally, it is noted that the team followed accepted codes of conduct including seeking consent from respondents, confidentiality, keeping sensitive information, avoiding bias, being sensitive to issues of discrimination, avoidance of harm, and respect for dignity and diversity. (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 country context section discusses gender equality, security and humanitarian challenges, and challenges of Covid-19 towards SRH, ASRHR, and family planning program. (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) Findings clearly show the perspectives of different stakeholder groups. (3)</p> <p>c. Are unanticipated effects of the intervention on human rights and gender equality described? (Score: 0-3) It is noted that the evaluators observed no unintended consequences of the intervention in terms of human rights or gender equality. (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) Recommendation # 9 "UNFPA should develop a strategy and plan for addressing gender barriers through male engagement and select the appropriate male engagement partners with the right expertise and experiences.". (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	80	20		
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.

There is good integration of cross-cutting themes. The evaluation is disability inclusive to some extent as described in the summary box.

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: