



Organizational unit:

[Redacted]

Year of report:

2019

Title of evaluation report:

UNFPA Country Programme Evaluation Turkmenistan 2016-2020

Overall quality of report:

Good

Date of assessment:

19 November 2019

Overall comments:

This is a comprehensive evaluation of the country programme. The strengths of the evaluation's methodology include the diversity of stakeholders consulted, as well as the presentation of a clear and actionable set of recommendations. Potential areas for improvement include the need for an explicit description of the extent to which the evaluation was conducted in a gender responsive manner, as well as more concise conclusions and a clearer, more succinct Executive Summary. The report is notable for linking each recommendation to the relevant SDGs.

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)

I. Structure and Clarity of Reporting

Yes
No
Partial

Assessment Level:

Fair

To ensure the report is comprehensive and user-friendly

I. 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

There are a number of minor grammatical errors, but these do not detract from the overall readability of the report. Some of the visual aids (such as figures 5 and 6) have text that is too small to easily read. The report is, at times, choppy as it appears that different authors wrote different sections and these were not pulled together with final editing.

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)	Yes	The report is 72 pages, only slightly exceeding the maximum for a CPE.
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	It is logically structured with sections clearly delineated..
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	All required annexes are included with the exception of the stakeholder consultation process (although this is, to some extent, covered in the body of the report).
<i>Executive summary</i>		
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 summary provides an overview of the main results, however it would better serve as a stand-alone document for all intended users if the acronyms were defined. The summary is, at times, written in abrupt sentences, missing articles, or using run-on sentences which are difficult to understand at first glance.
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	There is a clear structure to the executive summary, presenting the evaluation purpose and objectives, intended audience, as well as the methodology, findings, main conclusions, and recommendations. While a brief description of the interventions under evaluation is not included, UNFPA interventions and expected outcomes are discussed in the findings.
7. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	Yes	The summary is 5.25 pages, and could be presented more concisely. The findings under effectiveness, for example, take up one page.
2. Design and Methodology		
	Yes No Partial	Assessment Level: Good
<i>To ensure that the evaluation is put within its context</i>		
1. Does the evaluation describe the target audience for the evaluation?	Yes	The primary users of the evaluation are decision-makers within the UNFPA Country Office and the Turkmenistan government. The UNFPA Executive Board, Regional Office for Eastern Europe and Central Asia, and specific HQ divisions were also highlighted as the target audience.
2. Is the development and institutional context of the evaluation clearly described and constraints explained?	Yes	<p>The development and institutional context in which the evaluation was conducted were described in several sections of the report. This includes a key facts table (p. 8-9) and an individual section on country context (p. 24). The country context details key statistics about sexual and reproductive health, gender equality, and population dynamics mostly using World Bank data and noting a limitation in the government's production of national statistics (information from the last census, which was conducted in 2012, has not yet been published). Missing data constrained both the country programme planning as well as the current evaluation.</p> <p>The report also considered the way in which context constrained the evaluation process and results itself. In the methodological constraints and limitations, the evaluation report mentions limitations based on missing national-level statistical data as well as the potential for positive bias as a result of the Turkmen culture (p. 23). Furthermore, the report notes that UNFPA's small physical presence in the country (i.e. few staff) as a challenge to the evaluation as well as the programme's efficiency, generally.</p>

<p>3. Does the evaluation report describe the reconstruction of the intervention logic and/or theory of change, and assess the adequacy of these?</p>	<p>Yes</p>	<p>The report includes a thorough assessment of the theory of change and proposes a revised version. The theory of change was described as inadequate for capturing the complexity of institutional change within governments, with only a "two-step (output-outcome) Theory of Change", and a detailed description of its reconstruction was provided and validated by the Evaluation Reference Group (p.34-38). A simple graphic depiction is also included but the typeface is too small to easily read.</p>
<p><i>To ensure a rigorous design and methodology</i></p>		
<p>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?</p>	<p>Yes</p>	<p>The evaluation matrix is included in annex 4 and includes all fields (evaluation questions, assumptions, indicators, data sources and data collection methods).</p>
<p>5. Are the tools for data collection described and their choice justified?</p>	<p>Yes</p>	<p>The data collection tools used included document review, semi-structured interviews and focus group discussions. The evaluation team highlighted the use of primarily qualitative approaches to get an in-depth understanding of qualitative change to government systems as a result of UNFPA support. The evaluation team also intended to integrate quantitative data through national health statistics, but noted an inability to access the data, a limitation of the evaluation that was mostly left unmitigated.</p>
<p>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)?</p>	<p>Partial</p>	<p>While the ToR calls for a stakeholder mapping (annex 1, page 84), the report does not include a specific stakeholder map or section on stakeholders, however Table 4 does list the types of stakeholders interviewed. There is a list of Evaluation Reference Group members (at the front of the report and on p.23) and it is noted that they provided input into the evaluation ToRs and were asked to validate the results of data analysis, however there is no specific mention of stakeholder participation or consultation specifically in drafting recommendations. National partners were also consulted on sampling of districts for field visits, and the stakeholder consultation process in the field is also described (i.e. the approach taken in interviews, focus groups, and observations).</p>
<p>7. Are the methods for analysis clearly described for all types of data?</p>	<p>Yes</p>	<p>The data analysis process is described, and although not stated as such, it primarily involved contribution analysis. There is a section on methods for data analysis, but no clear method of analysis is identified. In this section triangulation and other validation methods are primarily described (e.g. day-long analysis sessions for reviewing and validating collected evidence), though qualitative evidence was coded to align with the evaluation questions and criteria.</p>
<p>8. Are methodological limitations acknowledged and their effect on the evaluation described? (Does the report discuss how any bias has been overcome?)</p>	<p>Yes</p>	<p>Four limitations and the mitigation measures are described: an over-simplified theory of change, misalignment between activities reported in standard progress reports and annual work plans, missing national statistics, and possible positive bias in qualitative data. While the mitigation measures are mentioned, the measures taken could have been more proactive at the onset of the evaluation so as to improve the evaluation design and reliability of findings. For example, it is not mentioned how the application of a purposive sampling approach and use of primarily qualitative data could affect the reliability of the evaluation's results, especially if the evaluators found statistical data missing within the desk review. Therefore, the strategy for mitigating the challenge of inadequate statistical data, though included, could have been improved. While continuously asking for data is useful, the integration of some quantitative methodologies into the design could have also helped fill some of these anticipated data gaps.</p>

9. Is the sampling strategy described?	Yes	The evaluators provide a detailed explanation of the sampling strategy. The sampling strategy was purposeful in the selection of national, provincial and district-level government institutions targeted through country programme activities, as well as in the selection of specialists within these institutions. Purposeful sampling was also used in the selection of interventions within the 2016-2020 country programme framework and Annual Work Plans, identifying nine interventions across UNFPA's core intervention areas (reproductive health and rights, adolescents and youth, gender equality, population and development) and with a high potential for immediate impact on people's lives in terms of reproductive health, a high potential to advance human rights, and a high potential to advance gender equality. The selection of interventions for review informed the selection of sites for field visits, as most had a national focus, while few were located in specific provinces (Dashoguz and Mary). While the process for defining the sample is documented in detail, as described above, it is not clear how illustrative this sample is in reference to the full scope of UNFPA's work and stakeholders in the country (e.g. universe is not adequately defined as there is no stakeholder map to understand how many programs and stakeholders UNFPA has in the country from which they selected the sample from).
10. Does the methodology enable the collection and analysis of disaggregated data?	Yes	Although the process for doing this is not identified, the list of stakeholders consulted is disaggregated by gender and age (youth/adult) and much of the information in the Country Context section includes gender disaggregated data, suggesting that the methodology allowed, to an extent, the collection of disaggregated data.
11. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?	Yes	Although the methodology does not specifically articulate how human rights and gender equality are taken up, the evaluation design allows for the assessment of cross-cutting issues of vulnerability, human rights and gender equality. There are several evaluation questions that address the advancement of gender equality (i.e. EQ2, EQ3iii, EQ4) and beneficiaries are targeted in the data collection processes. For example, evaluation questions on relevance include an assessment of the extent of the country programme's responsiveness to the most vulnerable populations and the extent to which UNFPA support responds to the principle of leaving no one behind and reaching the furthest behind. In addition, diverse stakeholders were consulted within the evaluation process across administrative levels (from district to central government) and vulnerable groups were consulted, including primarily youth.
3. Reliability of Data	Yes No Partial	Assessment Level: Good
<i>To ensure quality of data and robust data collection processes</i>		
1. Did the evaluation triangulate data collected as appropriate?	Yes	The evaluators were careful to triangulate data, including by having the Evaluation Reference Group comment on the findings.

<p>2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?</p>	<p>Yes</p>	<p>The evaluators identify the primary sources of qualitative and quantitative data and sources appear to be reliable. Much of the quantitative data, mostly presented in Country Context, comes from World Bank sources, while qualitative data is primarily from interviews, which are triangulated with other data sources. The limited use of quantitative data, both as a result of missing national-level data from household surveys and the use of primarily qualitative data collection methods, does limit the reliability of findings, though the evaluation team did well to triangulate data, and draw on as many secondary quantitative sources as available during the desk review.</p>	
<p>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?</p>	<p>Yes</p>	<p>As mentioned, the methodological constraints and limitations are described and mention limitations based on missing national-level statistical data as well as the potential for positive bias as a result of a "Turkmen cultural imperative to preserve the positive image of the family, community and their country in the eyes of outsiders" (p23). Considering that most persons interviewed were members of government or other national stakeholders, it is assumed that this positive bias applies to these sources, but that these could be triangulated with the responses of others interviewed, including seven UNFPA staff interviewed. Triangulation was the primary mitigation method applied.</p>	
<p>4. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?</p>	<p>Partial</p>	<p>The evaluation report mentions receiving informed consent prior to each interview, as well as reviewing participants rights to confidentiality and ability to withdraw consent at any time, reflecting the application of evaluation ethics. However, the report does not mention whether consent was received within focus groups. In addition, it is not clear how issues of discrimination were addressed/the evaluation does not describe how data was collected to ensure non-discrimination. For example, the evaluators conducted focus group discussions with Y-PEER volunteers on reproductive health issues amongst their peers in school, their experience of reproductive health education in schools, and use of reproductive health services. While the evaluation team reached equal numbers of female and male volunteers, it was not defined whether groups were mixed or separated by sex, a common strategy in evaluation for ensuring that males and females feel comfortable, open, and un-influenced by power dynamics when talking about sensitive issues like reproductive health care practices.</p>	
<p>4. Analysis and Findings</p>	<p>Yes No Partial</p>	<p>Assessment Level:</p>	<p>Good</p>
<p><i>To ensure sound analysis and credible findings</i></p>			
<p>1. Are the findings substantiated by evidence?</p>	<p>Yes</p>	<p>Evidence is systematically used to back up findings, primarily using supportive quotes from interviews and focus groups, though occasionally utilizing data from multiple indicator cluster surveys or voluntary national reviews from previous years. As an example of substantiated findings, the government agreed in 2014 to assume full financial responsibility for the provision of free contraceptives by 2017, and made adapted policies and procedures to prepare for this transition. The adaptation of these policies was substantiated by quotes from national specialists interviewed and also available statistics from the 2019 VNR which showed an increase in use of contraceptives by women with specific health risks from 21% in 2013 to 71% in 2018. The finding was further substantiated by observations, noting that while 2018 data had stated that 94% of reproductive health service delivery points had stock of IUDs and condoms and 82% had oral contraceptives, no service points had oral contraceptives at the time of the evaluation.</p>	
<p>2. Is the basis for interpretations carefully described?</p>	<p>Yes</p>	<p>The basis for interpretation can be traced back to the evaluation matrix.</p>	

3. Is the analysis presented against the evaluation questions?	Yes	This section is well organized with the findings for each question clearly stated, bolded and numbered. The supporting evidence and analysis is then provided below each finding.	
4. Is the analysis transparent about the sources and quality of data?	Yes	Sources of data - specific documents, specific stakeholder groups, etc. - are frequently referenced. Where evaluators had a concern about data quality, they made note and then used other data sources to triangulate findings and provide additional analysis. For example, national stakeholders had noted that contraceptives were affordable, though the evaluators assessed the price against the average income to provide additional context and analysis on actual affordability.	
5. Are cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Yes	Causal links between outputs and strategic plan outcomes are well addressed under Effectiveness EQ 3 for each of the four outcome areas of the country programme. Unintended outcomes are not clearly mentioned within the report, and part of this is explained by the mismatch between Annual Work Plans and Standard Progress Reports where the report notes "templates for AWP and Standard Progress Reports on their implementation don't require [an explanation] of linkages between activities and country programme outputs and outcomes which also contributes towards losing focus on intended results."	
6. Does the analysis show different outcomes for different target groups, as relevant?	Partial	This is constrained by the limitations of national statistical data, and improving the production of data is one of the focus areas of the country programme. However, the opportunity to do so from primary evaluation data sources could have included comparing results from different sites visited (noting any differences) or among stakeholder groups consulted (though results were at times disaggregated by age when noting findings regarding rights holders - youth - versus duty bearers like national specialists or government officials).	
7. Is the analysis presented against contextual factors?	Yes	Context is regularly provided. Examples include how the status of GBV prevention legislation moving through Parliament affects the intended outcomes of the UNFPA country programme.	
8. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?	Yes	The analysis does well to consistently reflect on cross-cutting issues of vulnerability and human rights. For example, the first finding notes how the lack of national level "medical and population data" has hindered the CP's ability to develop programming that considers vulnerable groups specifically in access to reproductive health services. The analysis continues to look at adolescent's access to reproductive health services and analyzes reasons for fragmentation in the system. The evaluation team also consulted youth as stakeholders within the evaluation.	
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 lists the respective criteria and evaluation questions for 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?	Partial	This section includes 10 conclusions with supporting text and is quite lengthy (4 pages). Although some are clearly stated and add value, others (3, 6, 7) are less clearly formulated. In the case of 1 and 2, the explanatory text is at the level of findings.	
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	There is no indication of bias.	

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 recommendation references the number of the conclusion(s) on which they are based.	
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 clear and action oriented. Each states the intended user and then lists operational implications (suggested action items). They then go beyond most recommendations by also including the relevant SDG target(s) to which they respond.	
3. Do recommendations appear balanced and impartial?	Yes	There is no indication of bias.	
4. Is a timeframe for implementation proposed?	Yes	Although no timeframes are specified in this section, it was clear at the beginning of the report that the purpose of the evaluation was to provide information to inform the next country programme.	
5. Are the recommendations prioritized and clearly presented to facilitate appropriate management response and follow up on each specific recommendation?	Yes	They are all prioritized as either high or medium. All are presented in a manner that enables a management response.	
7. Gender	0 1 2 3 (**)	Assessment Level:	Good
<i>To assess the integration of Gender Equality and Empowerment of Women (GEEW) (*)</i>			
1. Is GEEW integrated in the evaluation scope of analysis and indicators designed in a way that ensures GEEW-related data to be collected?		Gender equality/women's empowerment is not included as a standalone objective of the evaluation, nor is there specific mention of GEEW being mainstreamed into other objectives of the evaluation. While the evaluation does not have a standalone criteria on gender equality, it is addressed under Relevance and Effectiveness. Additionally, there are, under evaluation questions, sub-questions that address the country programme's work on gender equality, including in EQA 1, which includes a question on the extent to which the country programme is adapted to the needs of the population with emphasis to the most vulnerable population. Finally, although the evaluators reference the weakness of existing national data as a limitation to tracking outcomes, there is no discussion about the country programme's monitoring system's ability to capture information on gender equality results.	

2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques?		<p>While the evaluation does not explicitly specify how gender equality is addressed in the methodology, it is noted that the methodology we guided by the UNEG Guidance on integrating human rights and gender equality. Moreover, the methodology itself is, to an extent, gender responsive, as it employs a mixed methods approach. Additionally, the approach to sampling ensured that a plurality of stakeholders were consulted, including UNFPA beneficiaries and rights holders, and that interventions and stakeholders were included that had a high potential to advance and protect human rights, advance gender equality, and increase access to reproductive health services and education for all. Finally, a range of sources was used, and triangulation was employed throughout.</p> <p>Ethical safeguards were noted for interview participants but not for focus group participants. All interviews began with presentation of the purpose of evaluation and obtaining informed consent of a respondent. Each respondent was informed that his/her contribution was anonymous. A person was also informed that he/she could decline to answer any of the questions and to stop interview at any time at his/her discretion.</p>
3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?		<p>In the background section of the evaluation, the evaluation makes reference to CEDAW, the National Action Plan on Gender Equality and various human rights instruments, and, the lack of focus on vulnerable groups in the national health system is noted. While the findings are based on triangulated data, and disaggregated data was collected by the evaluation, the specific perspectives of different stakeholders on various issues related to the country programme are not explicitly brought out. Relatedly, while data is disaggregated in at least one case (numbers of youth reached by the Y-Peer program is gender disaggregated), disaggregated data was not systematically used to assess differences in outcomes/results for different groups. Unintended or unanticipated effects of UNFPA support is not discussed (and it's unclear whether this was because there weren't any found, or if the evaluators did not consider this question). The evaluation does provide specific recommendations (namely recommendation #9) on addressing GEEW issues.</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 totalling 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

	Assessment Levels (*)			
Quality assessment criteria (scoring points*)	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	11	82	7	
Overall assessment level of evaluation report		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:

