

Organizational unit: _____ **Year of report:** **2017**

Title of evaluation report: **Evaluation of the 2nd UNFPA Country Programme for Ukraine (2012-2017). FINAL EVALUATION REPORT**

Overall quality of report: **Good** **Date of assessment:** **6 September 2017**

Overall comments: The report is logically organized and easy to read. It is structured according to the UNFPA guidelines, but it is too long. The annexes do not include methodological tools used and information on the stakeholder consultation process. The Executive Summary presents the main results of the evaluation. The final report describes the reconstruction of the logic. The evaluation framework is designed in accordance with the UNFPA requirements. The tools for data collection are clearly explained and the methods for analysis are generally described, but not for each type of data. Methodological limitations and bias are clearly explained. The evaluators developed stakeholder selection criteria and the methodology is appropriate for assessing the cross-cutting issues. The evaluation consultants triangulated data appropriately. The data sources, mostly a combination of desk reviews and interviews, are justified by the annexes. The findings are supported by evidence: sources of data like interviews or desk studies are often mentioned. Interpretations are carefully described throughout the analysis. The evaluation consultants explored causal links between outputs, outcomes, and impacts. In the meantime, major explanations are given at outputs level. Also, there are some logical inconsistencies in outputs-outcomes pathways analysis. Gender and youth were effectively covered in analyzing all program components. The consultants succeeded in presenting conclusions clearly and briefly (1 paragraph for one conclusion). Conclusions provide a thorough understanding of the underlying issues of the programme. The recommendations are clearly written and are impartial. While some recommendations have a time horizon or it can be easily understood from the context, some are general. GEEW was included into the evaluation scope: reproductive health and rights, youth, population and development, gender equality and humanitarian response. Evaluation questions include GEEW dimensions: women, youth, gender equality, and GBV. Evaluation findings, conclusions, and recommendations reflect a gender analysis.

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 <u>assessment level</u> followed by main <u>comments</u> . (use 'shading' function to give cells corresponding colour)		
I. Structure and Clarity of Reporting	Yes No Partial	Assessment Level:	Fair
<i>To ensure the report is comprehensive and user-friendly</i>			
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?	Yes	This report has a clear structure. It is logically organized and easy to read. However, is not consistent in its use of terms which makes it more difficult to read (e.g., it speaks of five focus areas - SRH, Youth, Gender, P&D and HR - then subsequently refers to them as Program Areas (pp 6, 7, 9) and lacks depth on key points (how it deals with the intervention logic in the report, for example).	
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	The report is 176 pages in total (114 pages without the annexes and therefore 40 pages longer than desirable).	
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 structured according to the UNFPA guidelines, lessons learned are described in the text and only in one case they are separated from the other text as a section "Lessons Learnt from the Humanitarian Response project include" (p. 92).	
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?	Partial	Most of the required items are included in the report with the exception of the methodological tools which are not.	

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?	Yes	The Executive Summary is written as a stand-alone document and it presents the main results of the evaluation.
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 follows the required structure, in addition, it briefly describes the key findings "Overview of Achieved Results" (2 pages).
7. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	Yes	The Executive Summary is reasonably concise, being little over 5 pages in length.

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 evaluation describes the target audience for the evaluation (p. 12).
---	-----	--

2. Is the development and institutional context of the evaluation clearly described and constraints explained?	Yes	Annex 8 presents the Key facts table. The Chapter 2 explains the country context, for instance, it is said that "Since mid-1990s Ukraine has suffered through a demographic crisis, a combination of rapid depopulation and deterioration of people's health and wellbeing" (p. 19). There are examples in the text such as "There were several major context issues that influence the achievement of results. Unfortunately there is a high level of stigma in Ukraine..." (p. 58). Constraints are explained on the p. 17.
--	-----	---

3. Does the evaluation report describe the reconstruction of the intervention logic and/or theory of change, and assess the adequacy of these?	Yes	The final report describes the reconstruction of the logic: Figure 7 "Logic Model of SRH Component", Figure 9 "Logic Model of the Youth Programme, 2012 – 1017", Figure 11 "Logic Model of Gender Programme, 2012 – 1017", Figure 13 "Logic Model of P&D Programme, 2012 – 1017", and Figure 15 "Logic Model of the Humanitarian Response Programme." The evaluators assessed the adequacy of the theory of change "based on the Draft country programme document for Ukraine; the Country Programme Action Plan between the Government of Ukraine and the UNFPA for 2012-2016; the UNFPA Strategic Plan for 2014-2017; annual program reports and interviews with stakeholders" (p. 32).
--	-----	---

To ensure a rigorous design and methodology

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 designed in accordance with the UNFPA requirements: the evaluation questions are based on standard four OECD-DAC criteria and two criteria specific to UNFPA (pp. 12-14). The evaluation questions are clear and correspond to the evaluation criteria. For instance, EQ 1 and EQ 2 relate to the relevance, whereas EQ 3 - EQ 7 describe the effectiveness. Evaluation methods are discussed in text of Chapter 1 The evaluation matrix is found in the annex 4, it is designed in accordance with the UNFPA Handbook for evaluation at UNFPA (pp. 109-110). The matrix has five columns: evaluation questions; assumptions to be assessed; indicators; sources of information; methods and tools for data collection; description of the data and information gathered during the field phase (pp. 139-169).
--	-----	---

5. Are the tools for data collection described and their choice justified?	Yes	The tools for data collection are clearly explained on the pp. 14-16. The choice of site visits, through a purposive sample that would illustrate results, is described.
--	-----	--

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	While no stakeholder map is presented, Section 1.3.4 "Evaluation Process" mentions stakeholder mapping on the stages 1. Preparation and 2. Design (p. 16). The stakeholder consultation process is clearly described in the section 1.3.3 "An Evaluation Reference Group (ERG) was established by the UNFPA Country Office in Ukraine comprising key programme stakeholders... The ERG reviewed and provided inputs to the CPE terms of reference, including list of evaluation questions ... in particular the final report at the draft stage." (p. 16).
7. Are the methods for analysis clearly described for all types of data?	Partial	The methods for analysis are generally described, but not for each type of data, for instance, it is said that "By conducting retrospective assessments for the most part, analysing what has happened and the reasons why, evaluators used prospective assessments as well. The evaluators use a variety of methods to ensure the validity of the data collected.... Counterfactual analysis was applied wherever possible to explore the cause-to-effect relationships within the programme being evaluated" (p. 15).
8. Are methodological limitations acknowledged and their effect on the evaluation described? (Does the report discuss how any bias has been overcome?)	Yes	Methodological limitations and bias are clearly explained, mitigation measures are described, for instance, the evaluators mention "careful planning and good logistical support and division of labour among the evaluation team members ... (p. 17).
9. Is the sampling strategy described?	Yes	The evaluators developed stakeholder selection criteria to prepare a list of persons/institutions to interview (p. 16). The sampling strategy is described, but the consultants do not explain how they chose the optimal size for individual and group interviews, they don't provide any numbers in the methodology: "The selection of sites outside of Kyiv were based on purposive sampling combined with the knowledge of groups, their characteristics, and the purpose of the study (p. 15), "...the evaluation team used secondary data and purposive sampling approach to identify 'right' key informants (p. 17).
10. Does the methodology enable the collection and analysis of disaggregated data?	Partial	The methodology enables the collection and analysis of disaggregated data, although there was not extensive use made of it due to the team's reliance on secondary data (p 17). The methodology does not explain how the consultants disaggregated data, but EQ6 includes assessment of "...disaggregated population data for policy making, programming and public use..." (p. 13). The report shows that the evaluators disaggregated data, for instance, by source of funding (p. 29).
11. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?	Yes	The methodology is appropriate for assessing the cross-cutting issues: "the collection of evaluation data was carried out through a variety of techniques ranging from direct observation to informal and semi-structured interviews and groups discussions" (p. 14).

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 evaluation consultants triangulated data appropriately: “Besides a systematic triangulation of data sources and data collection methods and tools, the validation of data was sought through regular exchanges with the UNFPA programme staff” (p. 15). It is possible to find examples of triangulation in the text: “The interviews and desk studies show...” (p. 49), “Based on interviews with stakeholders as well as desk reviews and analysis of secondary data...” (p. 55).
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Yes	The data sources, mostly a combination of desk reviews and interviews, are justified by the Annex 2 “List of persons / institutions met” and the Annex 3 “List of documents consulted.” Data sources look reliable. Titles and names are mentioned. Documents are official and cited in such a way that they can be consulted if necessary.
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 data limitations are clearly indicated and the mitigating measures for each described (p. 17).
4. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?	Partial	There is evidence that UNFPA and UNEG ethical standards were used such as, for example, interviewees were kept confidential: “All interviewees were assured by the evaluation team of the confidentiality of their responses. The team closely adhere to the UN Evaluation Group Code of Conduct and Ethical Guidelines for Evaluations (2008)” (pp. 14-15). Although, the evaluators do not explain in detail how exactly they applied the guidelines.

4. Analysis and Findings	Yes No Partial	Assessment Level:	Good
---------------------------------	----------------------	-------------------	-------------

To ensure sound analysis and credible findings

1. Are the findings substantiated by evidence?	Yes	The findings are supported by evidence: sources of data like interviews or desk studies are often mentioned, for instance, it is said that “The analysis of the interviews with the respondents and desk studies conducted within the framework of the Evaluation, demonstrate...” (p. 48). But, in some cases the evaluators are too general, for instance: “There is a lot of evidence that outputs of the SRH component contributed to SP Outcome 1...” (p. 36), “Over the evaluation period there was evidence of a growing recognition of the importance of adolescents and youth...” (p. 52), “...supported by evidence” (p. 53).
2. Is the basis for interpretations carefully described?	Yes	The basis for interpretations is carefully described. Evaluators interpret the data accurately explaining the logic of their analysis. The following example shows how the consultants analyzed different views and came to a conclusion: “Most respondents mentioned little cooperation among the agencies... However, ... there were several respondents that think that cooperation and coordination among the UN organizations in Ukraine is good... The evaluation has demonstrated that the UN agencies have better cooperation at the local level rather than at the national level” (p. 98).
3. Is the analysis presented against the evaluation questions?	Yes	The analysis is presented according to the evaluation questions which correspond to the evaluation criteria.
4. Is the analysis transparent about the sources and quality of data?	Yes	The analysis is transparent about the sources and quality of data, for instance, the consultants explain that “Changes of key personnel in some national counterpart institutions ... made it sometimes difficult to obtain accurate and representative data for the period under evaluation”(p. 17), “The lack of evidence, including population data, compromises the quality of humanitarian response” (p. 23).

<p>5. Are cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?</p>	<p>Partial</p>	<p>The causal links are discussed in the text. But, mainly, the explanations are related to the outputs, and not outcomes. For instance, the report states "Further analysis ... shows that humanitarian response programme contributes to different extent to all SP outcomes and six outputs" (p. 87); "The UNFPA partners and target population groups do not necessary attribute the interventions to the UNFPA and/or know about them..." (p. 100). The evaluators discuss the outcomes in the text, for instance, improvements in national policy and programmes, but without establishing cause-effect links (p. 67, 102).</p>
<p>6. Does the analysis show different outcomes for different target groups, as relevant?</p>	<p>Yes</p>	<p>The analysis examines whether outcomes are different for different groups: by program component (youth, women, older people, and others), type of organization (government, civil society organizations, and others), by regions/provinces/districts (county capital vs regions) and by types of outcomes/ outputs/ activities. For instance, the consultants found out that "155 PHC institutions upgraded their capacity to provide youth-friendly and integrated SRH/FP and STI/HIV services according to the national standards and protocols" (p. 39). Target groups here are the institutions and youth.</p>
<p>7. Is the analysis presented against contextual factors?</p>	<p>Yes</p>	<p>The analysis is presented against contextual factors, which is evident from the text of the report. For instance, the consultants explain that "The NMCR was successfully piloted in 4 oblasts and then expanded to 15 centers, although 4 facilities had to postpone implementation after annexation of Crimea by the Russian Federation and the armed conflict in Donetsk and Luhansk" (p. 38).</p>
<p>8. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?</p>	<p>Yes</p>	<p>Gender and youth were effectively covered in analyzing all program components: Sexual and Reproductive Health, Adolescents and Youth, Gender Equality, Population and Development, and Humanitarian response: gender equality (p. 63), women who use injectable drugs (p. 35), pregnant women from rural areas (p. 39), vulnerable women (p. 40), youth-friendly clinics (p. 31), youth policy (51), human rights (58), and other examples.</p>

<p>5. Conclusions</p>	<p>Yes No Partial</p>	<p>Assessment Level:</p>	<p>Good</p>
------------------------------	-------------------------------	--------------------------	--------------------

To assess the validity of conclusions

<p>1. Do the conclusions flow clearly from the findings?</p>	<p>Yes</p>	<p>The evaluators relate conclusions to the findings. For instance, gender conclusion 1 states that "No visible progress was made in the improvement of the national policy in the area of gender equality promotion." The consultants support their statement with description of the key findings like "...key political and regulatory papers were not adopted or advanced...The Istanbul Convention was not ratified..." (p. 105).</p>
<p>2. Do the conclusions go beyond the findings and provide a thorough understanding of the underlying issues of the programme/initiative/system being evaluated?</p>	<p>Partial</p>	<p>Conclusions go beyond the findings and provide a thorough understanding of the underlying issues of the programme, for instance, the evaluators explain that "The planned results were achieved to some extent due to the lack of policy preconditions and gender-mainstreaming in the national policy" (p. 102), "The advocacy work aimed at improving existing policies and legislation has not achieved its planned results to a great extent due to the lack of good understanding of the essence of gender equality and the necessity to promote it among the policy-makers, public servants and in the society." (p. 105). At the same time, the level and clarity of conclusions are mixed. Conclusions are so numerous and often a summary of findings without major conclusions, making it hard to understand the underlying issues being drawn.</p>

3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	Conclusions appear to convey the evaluators' unbiased judgement. Even though the conclusions lack of references to the Findings section, the evaluators succeeded in clear explanation of the issues. For instance, they reveal that "Support rendered in conducting a number of studies on gender issues was very instrumental as it enabled to receive substantial evidence-based data that could be used in the development and adoption of policies and programmes" (p. 105).
--	-----	---

6. Recommendations	Yes No Partial	Assessment Level:	Fair
---------------------------	----------------------	-------------------	-------------

To ensure the usefulness and clarity of recommendations

1. Do recommendations flow logically from conclusions?	Yes	Recommendations are divided into two categories: there are 12 Strategic and 13 Programmatic recommendations. They all flow from the conclusions with references to specific conclusions.
--	-----	--

2. Are the recommendations clearly written, targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?	Yes	<p>The recommendations are clearly written. The consultants specify the intended user which is primarily the Country office. No other users are mentioned. It is unclear from the recommendations if it is enough to note that all recommendations are targeted to country office (rather than, for example, specific programme staff, or function etc.).</p> <p>Recommendations are action-oriented (with information on their human, financial and technical implications), for instance, the consultants advise "both expert and technical support in order to improve the national mechanism for ensuring equal rights of men and women..." (p. 109), "...consider careful planning of timeframe, human resources and 'right' selection of targets, partners, and allies..." (p. 110).</p> <p>Minor issue is that some recommendations lack financial and technical implications, for instance, the consultants advise the UNFPA to "extend activities at regional and local levels more," but they do not suggest how to allocate UNFPA resources to do so.</p>
---	-----	--

3. Do recommendations appear balanced and impartial?	Yes	The recommendations are balanced and impartial. The consultants support their statements with arguments, for instance, they advise "to improve coordination and joint funding programming of the UN agencies ... for better synergy and in order to avoid overlapping and duplication of efforts" (p. 110).
--	-----	---

4. Is a timeframe for implementation proposed?	Partial	While some recommendations have a time horizon or it can be easily understood from the context (e.g. "The next UNFPA National programme for Ukraine should" – p. 109), some are general (e.g. "Capacity of Government officials and local civil servants in developing gender sensitive programs and plans still needs to be strengthened" – p. 112).
--	---------	---

5. Are the recommendations prioritised and clearly presented to facilitate appropriate management response and follow up on each specific recommendation?	Partial	Recommendations are clearly presented to facilitate appropriate management response and follow up on each specific recommendation. Some recommendations include operational suggestions. Thus, the consultants clarify that "It is recommended to focus on local authorities when planning future interventions ... to help them build skills in programmes' preparation and development of programme performance indicators and M&E of these programmes. These activities will have to be accompanied by capacity building of local public servants ..." (p. 112). While each recommendation is given a priority level, all are designated as high level priority. The total number of recommendations (25, 12 strategic and 13 programmatic) would appear to make preparation of a management response and follow-up difficult.
---	---------	---

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?		GEEW was included into the evaluation scope “The evaluation covers all activities planned and/or implemented ... (reproductive health and rights, youth, population and development, gender equality and humanitarian response)” (p. 12). 2 The evaluation matrix has GEEW indicators, for instance: “Number of women and/or girls reached with sexual reproductive health and/or gender-based violence services in humanitarian settings... (p. 156).	
2. Do evaluation criteria and evaluation questions specifically address how GEEW has been integrated into design, planning, implementation of the intervention and the results achieved?		Evaluation criteria are based on OECD-DAC criteria: relevance, effectiveness, efficiency, and sustainability. Two criteria specific to UNFPA (UNCT and HCT Coordination, Added value) did not address GEEW. But, the evaluation questions 1,4, and 7 include GEEW dimensions: women, youth, gender equality, GBV (p. 13).	
3. Have gender-responsive evaluation methodology, methods and tools, and data analysis techniques been selected?		Evaluation consultants claim that they considered ethical issues (like confidentiality) when collecting information (p. 14). The annexes do not have interview guides or focus group notes to check this. Even though, the methodology does not provide specific approaches for fostering participation and inclusiveness (recommended by the table 6.7. in the in the UNEG Guidance Integrating Human Rights and Gender Equality in Evaluations), the evaluation consultants tried to reach a wide range of stakeholders, including women: “The evaluation adopted an inclusive approach, involving a broad range of partners and stakeholders” (p. 15). 2 Desk review and data analysis stages included gender analysis as gender has been mainstreamed into almost all program components.	
4. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?		Evaluation findings reflect a gender analysis which is evident from the text: Sexual and reproductive health component had an outcome “integrated sexual and reproductive health services ... are gender-responsive and meet human rights standards” (p. 31), Youth component included young adolescent girls (p. 51, pp. 54-56), Gender equality component considered gender equality, women’s and girls’ empowerment and reproductive rights (p. 63), Population and development component included gender equality issues in the outcome (p. 73), and humanitarian response component investigated gender based violence (p. 91). 2 Evaluation conclusions reflect a gender analysis, for instance: Strategic Conclusions 3, 8, and 11; programmatic conclusions: gender conclusions 1 – 5. Evaluation recommendations reflect a gender analysis: strategic recommendations 1, 3, and 4, programmatic recommendations: SRH 1; gender Recommendation 1 – 3; and HR recommendation 1.	

(*) This assessment criteria is fully based on the UN-SWAP Scoring Tool, see Annex 7. 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).

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		69	31	
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: