

UNFPA Evaluation Quality Assessment Grid

Version: May 2024

REPORT RATING SUMMARY		
Overall Rating	87%	Satisfactory
••••• Excellent	5	
•••• Highly Satisfactory	4	
•••• Satisfactory	3	The report meets UNFPA/UNEG standards for evaluation reports, but some indicators are inadequately addressed or missing. Decision makers may use the evaluation with some confidence.
••• Fair	2	
•• Unsatisfactory	1	

REPORT DETAILS	
Title of the evaluation report	Country Programme Evaluation of the United Nations Population Fund (UNFPA) Serbia 2nd Country Programme
Region	Eastern Europe & Central Asia
Country	Serbia
Year of report	2025
Business Unit/programme country (managing evaluation)	Serbia CO
Date of assessment review (dd/mm/yyyy)	4/7/2025
Primary SDG(s) covered (list provided below)	IOD PARC
CLASSIFICATION OF EVALUATION REPORT	
Primary SDG(s) covered (list provided below)	
UNFPA Strategic Plan areas covered (lists provided below)	
Three transformative results	(a) an end to preventable maternal deaths; (b) an end to the unmet need for family planning; and (c) an end to gender-based violence and all harmful practices, including female genital mutilation and child, early and forced marriage
Six outputs	(a) policy and accountability, (b) quality of care and services, (c) gender and norms, (d) data and evidence, including on population changes, (e) humanitarian action, and (f) adolescent and youth.
Six accelerators	(a) Human rights-based and gender transformative approaches (b) Innovation and digitalization (c) Partnership, South-South and triangular cooperation, and financing; (d) Data and evidence; (e) "Leaving no one behind" and "reaching the furthest left behind first"; (f) Resilience and adaptation,
Organizational effectiveness and efficiency	yes
Humanitarian evaluation	no
Evaluation evaluand (e.g. country programme/intervention/policy/thematic area)	Country Programme Evaluation
Evaluation type (e.g. formative, summative, developmental)	Formative and summative
Geographic scope (e.g. global, regional, national)	National
EQA Summary: The rater will provide top line issues for this evaluation relevant for feedback to senior management (strengths and weaknesses), summarizing how the evaluation report meets or fails to meet all criteria. As relevant, the rater will highlight good practice/added value elements and the level of complexity of the evaluation. The rater should also highlight how cross-cutting issues were addressed in the report. Considerations of significant constraints (e.g. humanitarian crisis or political turmoil) should also be highlighted here.	
<p>The Country Programme Evaluation (CPE) of UNFPA Serbia's 2nd Country Programme (2021–2025) is rated satisfactory, offering a thorough and well-structured analysis aligned with UNFPA expectations. Key strengths include:</p> <ul style="list-style-type: none"> • Clearly defined purpose, objectives, scope, and evaluation questions aligned with OECD-DAC criteria. • Well-presented stakeholder mapping and logical frameworks/Theories of Change. • Systematic presentation of findings and well-evidenced conclusions and recommendations. • Active engagement of the Serbia Country Office and ERG in formulating recommendations. • Recommendations are clear, prioritised and directly linked to the evidence and conclusions. <p>However, several areas for improvement are noted:</p> <ul style="list-style-type: none"> • Limited integration of Leave No One Behind (LNOB), disability, and intersectional issues in context. • Inadequate detail on sampling, data analysis, ethical considerations, and data collection tools (e.g. small text, lack of consent forms). • Weak articulation of gender and human rights integration across findings and conclusions. • Limited description and contextualization of ethical guidelines, including explanation of safeguarding and "do no harm" practices. • GEEW (Gender Equality and Women's Empowerment) considerations could be better mainstreamed throughout the evaluation design and analysis. <p>Suggestions for future evaluators: The rater will identify key suggestions to improve the evaluation, and be specific to the sections of the report where shortcomings were found. As relevant, examples will be cited to assist evaluation managers in overseeing future evaluations.</p>	

Areas for improvement include:

- The country context would benefit from explicit inclusion/ analysis of the principles of LNOB and issues such as disability to be included, and contextualised to the work of UNFPA in the Serbian context, ideally with an intersectional analysis included.
- A clear and explicit description and outline of ICPD linked SDG indicators should be included in the background section
- In the methodology section, tables outlining numbers of stakeholders interviewed should include disaggregation of those stakeholders, by gender at a minimum, particularly as this data is available in Annex 3. It would be helpful to include in these tables in the main report, or to make specific reference to the fact disaggregated primary data is provided in the annexes. Overall the methodology did not explain how it would take a gendered approach or adequately include LNOB considerations, while data collection tools did not allow for different methods of data collection for different groups.
- With regard to the methodology and the principles sections the evaluation does note that it intended to capture data from vulnerable groups, including Roma women, it could do a better job of explicitly stating how this will be accomplished in terms of sampling, practical conduct of FGDs or KIIs and analysis of disaggregated data. Methods of data analysis would also benefit from more detail on how information was analysed and interpreted, for example through a gender lens or taking LNOB into account.
- Ethical considerations in collection tools are minimal with brief mention of consent and data protection, but the text is very small and difficult to read. Overall the data collection tools should be rewritten and presented in a clear and legible format, and ideally there should be a dedicated ethical section in the report, along with accompanying detailed data collection tools and consent forms
- GEEW, and other issues around equality, and marginalised groups could have been more fully integrated into the design of the evaluation if it has contained more specific information on how gender would be integrated across the evaluation questions, or there had been a separate GEEW evaluation question. With regard to the methodology, detail on how women's voices would be gathered and included in practice, and an analysis of the availability of disaggregated data should have been provided, and/or noted as a limitation. With regard to findings, conclusion and analysis, the application of an intersectional lens and consistent gendered analysis would have been expected with explicit use of disaggregated data.

SECTION RATINGS		
SECTION A: EXECUTIVE SUMMARY (weight 5%)	100%	Comments on Rating
Question 1. Can the executive summary inform decision-making?		
i	Is a clear, standalone document useful for informing decision making, (a minimum of 5 pages, up to a maximum of 7 pages).	Yes The executive summary is 6 pages in length, is a clear and stand alone document and is useful for decision-making.
ii	Includes all necessary components of the evaluation report, including: (1) overview of the context and intervention, (2) evaluation purpose, objectives and intended users, 3) scope and evaluation methodology, (4) summary of most significant findings, (5) main conclusions and (6) key recommendations	Yes The executive summary contains all relevant and expected information, including: (1) overview of the context and intervention, (2) evaluation purpose, objectives and intended users, 3) scope and evaluation methodology, (4) summary of findings, (5) main conclusions and (6) key recommendations.
iii	Includes all significant information in a concise yet clear manner to understand the theme, intervention, programme, project and the evaluation.	Yes The executive is concisely written, contains all significant information and enables a clear understanding of the CPE.
SECTION B: BACKGROUND (weight 5%)	70%	Comments on Rating
Question 2. Is the evaluand (i.e. intervention/policy/thematic area etc. that is to be evaluated) and context of the evaluation clearly described?		
i	Clear description of the evaluand (e.g. intervention), including: geographic coverage, implementation period, main partners, cost/budget, and implementation status.	Yes There is a clear description of the CP, its implementation status, main partners, geographic coverage, implementation period and cost/ budget.
ii	Clear description of the context of the evaluand (e.g. economic, social and political context, relevant aspects of UNFPA's institutional, normative and strategic framework, cross-cutting issues such as gender equality and human rights, disability and LNOB dimensions) and how the context relates to the evaluand (e.g. key drivers and challenges that affect the implementation of the intervention/policy/thematic area.	Partially There is a clear description of the context of the evaluation, from an economic, social and political perspective, and detailed analysis of how some of these factors currently influence (or could influence) UNFPA's work in Serbia. There is explicit reference to UNFPA's institutional and normative frameworks. However, issues relating LNOB or disability are not discussed in detail, with limited contextualisation to UNFPAs work in Serbia. For example, issues related to geographic disparities and the situation of migrants are not discussed as part of LNOB considerations. While there is reference to the status of Roma populations, this would benefit from greater detail, particularly as it relates to the work of UNFPA, and LNOB principles. EU accession is noted as being a primary driver, and challenge.
iii	Linkages drawn between the evaluand and ICPD-related SDGs outlined. This may include reference to ICPD benchmarks and relevant SDG targets and indicators.	Partially ICPD 30 is noted and mentioned in the report, but not clearly linked to SDG indicators and targets. This section could be improved if ICPD - related SDGS were included/or the background section were clearer about which specific SDGs were applicable to UNFPA interventions. There is however clear reference to the status of UNFPA's 3 TRs and their alignment with the UNSDCF and Serbian national priorities in Table 6. Overall SDG reference is not consistent or coherent, and this could be improved if the issues/ areas of intervention listed make reference to, and explicitly identify, relevant SDGs or indicators.
Question 3. Are key stakeholders clearly identified and analysed?		
i	Clear identification of key stakeholders which should include implementing partner(s), development partners, rights holders, and duty bearers among others; (e.g., stakeholder map).	Yes There is a clear identification of key stakeholders, and their role in the evaluation (primarily relating to data collection) which is explained narratively and in stakeholder mapping, set out in Tables 3 and 4.
ii	Stakeholders are analysed to understand their interests and needs, power and influence, and potential impact on the evaluand.	Partially Stakeholder analysis covers the interests and needs of stakeholders, from development partners to end level beneficiaries/ users, and outlines some of the concerns and issues which may impact on them. It would however, benefit from an analysis of power and influence, and how different stakeholders may be affected differently by the intervention, whether this be positive or negative. For example support to government ministries is very different to support to Roma women on issues of SRH or GBV, but the different types of programmes and their potential impacts are not explored as part of the stakeholder analysis, not their different roles in the evaluation.
SECTION C: EVALUATION PURPOSE, OBJECTIVES AND SCOPE (weight 5%)	100%	Comments on Rating
Question 4. Is the purpose of the evaluation clearly described?		

	i Purpose of evaluation is clearly defined, including why it was needed at that point in time, its intended use, and key intended users.	Yes	<p>The evaluation has a clear distinction between purpose, objectives and scope. The purpose is clearly linked to the UNFPA Evaluation Policy (2024) and outlined as being to:</p> <ul style="list-style-type: none"> • provide evidence to inform development, humanitarian response and peace responsive programming; • provide an independent, impartial perspective on the work of UNFPA and entail management accountability to act on recommendations; • aggregate and share good practices and credible evaluative evidence to support organisational learning; and • empower community, national and regional levels by providing them with access to information and skills to interpret and interrogate policies and programmes affecting their lives <p>The evaluation's main audience and primary intended users are described as being the UNFPA Serbia country office, the Government of Serbia, UNFPA Implementing Partners (IPs), the UNFPA Regional Office for Eastern Europe and Central Asia (EECARO), rights-holders involved in UNFPA interventions and the organisations that represent them, the United Nations country team (UNCT) and other development partners. It is also noted that evaluation results may also be of interest to a broader group of stakeholders.</p>
Question 5. Are the objectives and scope of the evaluation clear and realistic?			
	i Clear and complete description of the objectives of the evaluation, including reference to any changes made to the objectives included in the ToR (if applicable).	Yes	<p>The evaluation notes its objectives clearly, with overarching objectives being to provide an independent assessment of the UNFPA CP and to broaden the evidence base to inform the design of the for Serbia CP 2026-2030, which is the yet to be finalized.</p> <p>The specific objectives of this CPE are also provided, they are:</p> <ul style="list-style-type: none"> (i) provide an independent assessment of the relevance, effectiveness and sustainability of UNFPA support for Serbia; (ii) provide an assessment of the role played by the UNFPA Serbia country office in United Nations country team (UNCT) coordination, with a view to enhancing the United Nations collective contribution to national development results; (iii) draw key conclusions from past and current cooperation and provide clear, forward-looking, actionable recommendations for the next CP cycle <p>No changes from the ToR are noted.</p>
	ii Clear and relevant description of the scope (e.g. thematic, geographic, and temporal) of the evaluation, covering what will and will not be covered, as well as, if applicable, the reasons for this scope (e.g., specifications by the ToRs, lack of access to particular geographic areas for political, humanitarian or safety reasons at the time of the evaluation, lack of data/evidence on particular elements of the intervention).	Yes	<p>Section 1.2 outlines the scope of the evaluation. Temporally it covers CP interventions planned and implemented from January 2021 to the end of evaluation data collection in mid October 2024. Geographically it assesses interventions planned and implemented at the national and municipality levels, with detail of regions / areas provided in Annex 7 and summarised in the map at the start of the evaluation. Thematically the CPE covers all outputs and expected CPD outcomes in the areas of sexual and reproductive health (SRH), including gender-based violence (GBV), adolescents and youth (A&Y) and population dynamics (PD), including interventions added in response to changing contexts, priorities and needs.</p> <p>It is also well noted that the evaluation will explore the CP's focus on and application of the principle of leaving no one behind (LNOB), gender equality and humanitarian preparedness and response, though how it will do this is not specified.</p>
SECTION D: EVALUATION DESIGN AND METHODOLOGY (weight 20%)		73%	Comments on Rating
Question 6. Are the selected evaluation questions and evaluation criteria appropriate for the purpose of the evaluation and is there clear justification for their use?			
	i Evaluation questions and sub-questions are appropriate for meeting the objectives and purpose of the evaluation. The relevant criteria are specified and are aligned with the questions.	Yes	<p>Evaluation questions are appropriate for meeting the objectives and purpose of the evaluation. Three OECD DAC criteria have been selected and aligned with questions. These are Relevance, Effectiveness and Sustainability, and they align with specifications in the ToR. Each question also includes assumptions to be tested, which are clear and appropriate.</p>
	ii Evaluation matrix clearly presents the evaluation criteria used as well as the corresponding criteria, evaluation questions, assumptions, indicators/ lines of inquiry, and methods for data collection and analysis, including relevant data sources.	Yes	<p>The evaluation matrix is included at Annex and includes key components expected of UNFPA evaluation matrix, including:</p> <ul style="list-style-type: none"> (i) evaluation criteria (ii) evaluation questions (iii) assumptions associated with each evaluation question; (iv) indicators to be empirically verified for each assumption; (v) data collection methods to measure the indicators (though it is noted these are listed next to each piece of data collected) <p>Sources from which data will be collected is not clearly outlined in the evaluation matrix, though it can be inferred from the list of FGDs and KIIs, but the sources could be clearer.</p>
Question 7. Is the theory of change, results chain, logical framework, or equivalent framework well-articulated?			
	i Clear description of the intervention's intended results, or of the parts of the results chain that are applicable to, or are being tested by, the evaluation.	Yes	<p>There is a clear description of the interventions intended results and the result chain and Theory of change is summarised in Figure 3, with individual ToCs also provide in Annex 5. Overall there is a clear narrative and figurative description of the logic chain and the elements of the CP and its ToCs being tested for different elements of the work of UNFPA in Serbia.</p>

	ii	Causal relationships between the various elements (e.g. outcomes, including the three or relevant Transformative Results, outputs) of the theory of change, results chain or logical framework are presented in narrative and/or graphic form).	Yes	There is a clear narrative analysis of the ToCs that exist for the three main programmes in the Serbia CP, attached as annexes and discussed in the main body of the text, the update of these ToCs by the evaluation team is also discussed in relation to outputs and outcomes. The evaluators also provide a useful summary of all three ToCs in narrative text, as well as Figure 3. Overall casual relationships and results chains are clearly presented and explained. Usefully risks and assumptions with regard to the Serbia CP and its pathways of change have also been provided in Table 2, which particularly given political instability is a relevant inclusion, in its own right, as well as being part of the ToCs.
	iii	Comprehensive analysis and assessment of the theory of change, results chain or logical framework, and if requested in the ToR, it is retrofitted/reconstructed by the evaluators.	Yes	The CP has three ToCs which guide its work, these are provided at Annex 5. However, the evaluation does a good job of narratively explaining and analysing the theories of change in section 1.3.1, as well as producing a combined logical framework linking the outputs and outcomes across all streams of the CP in a clear logical framework (Figure 3). It is clear that this is to aid in understanding, not replace or duplicate already existing ToCs, which have not been reconstructed, but have been updated.
Question 8. Does the report specify adequate methods for data collection, analysis, and sampling?				
	i	Evaluation design and set of methods are clearly described, and are relevant and robust for the evaluation's purpose, objectives and scope, including the use of AI in the evaluation process if applicable.	Partially	<p>It is not noted if the evaluation is formative or summative (it is assumed both), however this should be clarified. It would also be useful to note if other approaches, such as participatory or utilization focused approaches were applied.</p> <p>The evaluation questions align with the purpose and objectives of the evaluation and data collection methods are explicitly mentioned and include document review, KIs and FGDs. While it is clear that vulnerable people will be involved, and the principles of LNOB will be applied, it is not clear in practice how diverse voices of stakeholders, including vulnerable and marginalised people will be captured. Additionally the focus on Roma women, as explained in the context section and discussed in the findings section, is not matched with a methodology to support data collection from vulnerable and marginalised people or groups. The almost complete lack of information on how the views of these women will be captured is a significant gap in the methodology.</p> <p>Data collection tools are minimal and consist of a list of questions for each stakeholder, with minimal preamble and no consent forms are included.</p>
	ii	Data sources are all clearly described and are relevant and robust; these would normally include qualitative and quantitative sources (unless otherwise specified in the ToR).	Yes	<p>A mixed methods approach has been applied with a range of qualitative and quantitative data sources which are listed in the report, and include:</p> <ul style="list-style-type: none"> • Detailed document review • KIs and group interviews with a range of stakeholders which included government officials, service providers and representatives of IPs, CSOs, private sector, academia, other United Nations organisations and development partners • FGDs with beneficiaries, including young people and older people. <p>While it is noted that data will be disaggregated, it would have been helpful/ expected to include the gender and type of people (women from Roma community, young people etc) participating in KIs or FGDs in tables 3 and 4 which outline number of people interviewed broken down by stakeholder group; but not by gender of stakeholders themselves, or whether they come from marginalised/ excluded groups. The percentage of women interviewed at the national level is noted, but without further context as to type of stakeholder this information is not well contextualised in the main body of the report. Given that Annex 3 contains detailed lists, disaggregated by gender of people interviewed it would be helpful to note this in the methodology section itself, and expand on how the voices of the most marginalised will be captured as a source of data.</p>
	iii	Sampling strategy is provided - it should include a description of how diverse perspectives are captured (or if not, provide reasons for this).	Partially	<p>The evaluation notes that the sampling for stakeholders was based on a review of UNFPA programmes in different geographic regions, and as representative a sample as possible chosen from that initial analysis, with support from the CO. It is clear that stakeholders were chosen based on their ability to provide information on each of the three outputs, which is appropriate. However, this section could be improved if more detail were provided on what measures were employed to ensure that bias was minimised, and detail of the original sample size provided to enable an analysis as to whether the sample size for this evaluation is appropriate, and if/how it represents diverse views.</p> <p>It is noted that geographic sampling is well described, and that including certain municipalities will enable sampling of Roma women, and other vulnerable groups, but this is not described further either in terms of numbers or how Roma many Roma women were identified and interviewed.</p>
	iv	Methods allow for rigorous testing of the theory of change, results chain or logical framework (e.g. methods help to understand the causal connections, if any, between outputs and expected outcomes (3TRs).	Yes	Methods allow for rigorous testing of the ToC and results chains to identify causal connections between outputs and expected outcomes

	v	Clear and complete description of the methods of analysis, including explainability and full disclosure of the use of AI in the evaluation process, if applicable.	Partially	The evaluation notes that all evidence was inputted into an Excel-based evidence database and the evidence matrix was used to extract and analyse all data by indicator. While it is noted that this was process was used systematically to identify common themes and patterns, more detail could be provided on how exactly this was accomplished, for example detail on e.g. contribution analysis (as noted earlier in the report), or a content analysis. Additionally it is not noted how disaggregated data was analysed, whether a gendered lens was applied, or how LNOB were integrated across the analysis, as was stated earlier in the report. AI is not mentioned.
	vi	Clear and complete description of limitations and constraints faced by the evaluation in its data collection and analysis, along with the mitigation measures implemented by the evaluation to address these limitations, where feasible.	Partially	Limitations and mitigation measures are present in the analysis, however one mitigation states that the methodology does not apply an intersectional lens, but the mitigation for this is to engage with a wide range of stakeholder - it is not explained why the methodology was not designed from the outset with an intersectional lens, or how a wide range of stakeholders will be engaged. This is particularly relevant given that the cross-cutting issues of LNOB and gender equality are not well attended to in the approach and design of the evaluation. It might also be expected that a limitation could be included regarding the number and range of rightsholders who were able to be interviewed as a separate category, not an overarching category relating to all stakeholders. An additional limitation to include might also relate that the availability of disaggregated data, relating to secondary data as well as primary, by gender as well as other marginalised groups.
Question 9.		Are ethical issues and considerations described? The evaluation should be guided by the UNEG ethical standards for evaluation. As such, the evaluation report should include:		
	i	Explicit and contextualized reference to the UNEG obligations of evaluators (independence, impartiality, credibility, conflicts of interest, accountability) and/or UNEG Ethical Principles.	Partially	The evaluation report contains brief mention that UNEG ethical standards were applied, but nothing further. There is limited mention of what the UNEG standards are, but no contextualisation.
	ii	Clear description of ethical issues and considerations (e.g. respect for dignity and diversity, fair representation, confidentiality, and avoidance of harm) that may arise in the evaluation, safeguard mechanisms for respondents (e.g. parental consent forms for adolescents, compliance with codes for vulnerable groups; WHO standards of safe data collection on GBV) and ethical considerations in the use of AI as applicable (e.g., transparency of use, explainability, privacy, data protection, accuracy, human rights). If AI is used in the evaluation, there should be transparency and disclosure on the ethical and responsible use of AI in the report.	No	The evaluation does not explain how ethical issues were addressed at each phase of the evaluation, and during data collection in particular, other than to note that UNEG ethical standards were applied. There is no mention of safeguarding or do no harm, or detail of how confidentiality and management of data would be managed. Consent, as an issue, is not explained nor is it present in the data collection tools (Annex 4), which is not a list of data collection tools, but a list of questions to be asked of stakeholders. As such the data collection tools contain brief mention of consent, and other ethical issues. Additionally the data collection annex (Annex 4) has text which is very small and hard to read/understand. The data collection tools should be rewritten and presented in a clear and legible format, with consent forms included.
Question 10.		Does the evaluation incorporate innovative practice that adds value to the evaluation process?		
	i	Innovation practice is used to improve the quality of evaluation process. This could include efforts to optimize the evaluation process (e.g., use of AI or new technology for data gathering, content analysis, outcome harvesting among others), or components introduced to enhance inclusion and participation in the evaluation processes (e.g. a youth steering committee), or ways of sharing of evaluation results.	Not Rated	Not rated as not a requirement of a UNFPA evaluation and ToR.
SECTION E: EVALUATION FINDINGS (weight 25%)			100%	Comments on Rating
Question 11.		Do the findings clearly and adequately address all evaluation questions and sub-questions?		
	i	Findings are presented clearly and provide sufficient levels of evidence to systematically address all the evaluation's questions	Yes	Each finding correlates with each evaluation question, and contains a summary of key findings, followed by analysis of assumptions or sub questions. Relevant criteria are specified for each EQ. Findings are then listed, with analysis noting where evidence has been gathered, i.e. from KIIs, FGDS, or from documents/ statistical data. Overall, findings are presented clearly and provide sufficient levels of evidence to systematically address all of the evaluation questions.
	ii	Explicit use of the evaluand's theory of change, results chain, logical framework in the formulation of the findings.	Yes	There is explicit reference to, and use of, the CP ToCs and logical frameworks in the formulation of the findings against each output area.
Question 12.		Are evaluation findings derived from credible data sources as well as a rigorous data analysis?		
	i	Evaluation uses credible forms of qualitative and quantitative data. It presents both output and outcome-level data as relevant to the evaluation framework. Triangulation is evident using multiple data sources.	Yes	The evaluation uses credible forms of qualitative and quantitative data with evidence footnoted against KIIs, FGDS, or other sources so that there is clear traceability between data collection, and findings. Both output and outcome level data are presented, as relevant to the evaluation framework/question. For example, outcome level data is present in Finding 9 relating to reducing adolescent birth rates, and the evaluation uses both output and outcome level data to analyse the reduction in adolescent birth rates for the population overall, but not within the Roma community, where it has increased. There is also evidence of triangulation in the text itself, and clearly demonstrated in the footnoted data sources, where FGDS, KIIs and document review are all referenced.

	ii	Findings are clearly supported by the evidence presented, both positive and negative. Findings are based on clear performance indicators, standards, benchmarks, or other means of comparison as relevant for each question.	Yes	Findings are supported by evidence, and both positive and negative evidence is clearly presented, for example on EQ1, there is mixed evidence on UNFPA's role in supporting Serbia on specific issues relating to EU Accession. For example, UNFPA has played a significant role in support the government to meet demographic and statistical requirements (among others) to meet EU requirements, but overall, UNFPA is not perceived as a significant player in advancing EU accession reform requirements in Serbia. Findings are based on standards indicators and benchmarks, or other means of comparison where outcome indicators are not available. For example, it is also well noted that outcome indicators are not available for key issues, such as maternal health or GBV rates, but data has been used from other sources, such as SDG reporting, wider UNSDCF reporting and government statistics. However, the lack of accurate and targeted survey data, particularly for vulnerable groups such as the Roma population are noted as challenging to report on accurately. UNFPA is therefore involved in improving reporting and results, and population and demographic data (as noted previously).
	iii	Causal factors (contextual, organizational, managerial, etc.) leading to achievement or non-achievement of results are clearly identified. For theory-based evaluations, findings analyse the logical chain (progression -or not- from outputs to high level results).	Yes	Causal factors are noted throughout the findings section, for example in relation to challenges in reaching the Roma community, low health literacy and challenging social norms are outlined as factors contributing to the achievement of results. The relevance of UNFPA's work in terms of supporting the government in gathering accurate demographic and population data is well evidenced. However, due to the size of the CO and its relatively small budget, its role is not always visible. Additionally, for some issues, such as GBV rates, there is no baseline data against which to measure changes. Given this is a theory based evaluation there is clear analysis of logical chain progression from outputs to outcomes (where possible), and, where appropriate, results.
Question 13.		Does the evaluation assess and use the intervention's Results Based Management elements?		
	i	Assessment of the adequacy of the intervention's planning, monitoring, and reporting system (including completeness and appropriateness of results/performance framework - including vertical and horizontal logic, M&E tools and their usage) to support decision-making.	Not Rated	Not rated as not a requirement of a UNFPA evaluation and ToR.
SECTION F: EVALUATION CONCLUSIONS (weight 10%)			100%	Comments on Rating
Question 14.		Do the conclusions clearly present an unbiased overall assessment of the evaluand?		
	i	Conclusions are clearly formulated and present unbiased summative statements that respond to the evaluation questions.	Yes	Conclusions are clearly formulated and present unbiased summative statements that respond to the evaluation questions, and are forward-looking in nature. Each conclusion presents a concise statement, followed by a more detailed description, which synthesis the findings. all conclusions clearly indicate the evaluation finding to which they are linked, and from that it is possible to link them to the appropriate EQ. They do not link to the recommendations, but the recommendations do provide a link back to each relevant conclusion.
	ii	Conclusions are well substantiated and derived from findings and add deeper insight and analysis beyond the findings.	Yes	Conclusions are well supported and substantiated, it is clear they are derived from findings, while adding deeper insight and analysis, with a particular view to trends already evidenced to support the development of the next CP.
Question 15.		Are lessons learned identified?		
	i	Lessons learned are derived from the findings and are well substantiated with practical, illustrative examples.	Not Rated	Not rated as not a requirement of a UNFPA evaluation and ToR.
	ii	Lessons learned are clearly presented and provide actionable insights on the positive aspects of the evaluand as well as any areas of improvement.	Not Rated	Not rated as not a requirement of a UNFPA evaluation and ToR.
SECTION G: EVALUATION RECOMMENDATIONS (weight 15%)			100%	Comments on Rating
Question 16.		Are recommendations well-grounded and articulated?		
	i	Recommendations are clearly formulated and logically derived from the conclusions.	Yes	Recommendations are clearly formatted and logically derived from findings, with a clear link to the conclusion from which each recommendation was derived.
	ii	Recommendations are useful and actionable for primary intended users. Specific guidance is provided for its implementation (e.g. strategic or programmatic level, suggested actions/operational implications, responsible actors, as appropriate).	Yes	Recommendations are useful and actionable for primary intended users and guidance is provided for implementation. It is clear who is responsible for each recommendation to the level of units, teams and individuals within the Serbia CO. This is an example of good practice.
	iii	Process for developing the recommendations is described, and includes the involvement of key stakeholders (e.g. evaluation reference group members), including those who will be affected by the recommendations.	Yes	The process for developing the recommendations is clearly described, and the role of the UNFPA Serbia CO, as well the role of the ERG to review and provide feedback and inputs is clearly outlined. This is an example of good practice.
	iv	Recommendations are prioritized based on their importance, urgency, and potential impact.	Yes	Recommendations are prioritised based on importance, and include a separate section on rationale, as well as operational implications, provided for each recommendation.
SECTION H: REPORT STRUCTURE AND PRESENTATION (weight 5%)			83%	Comments on Rating
Question 17.		Does the evaluation report include all required information?		

	i Opening pages include: Name of evaluation and/title of evaluation, timeframe of the evaluation, date of report, location of evaluation, names and/or organization(s) of the evaluator(s), name of organization commissioning the evaluation, list of the evaluation reference group members, acknowledgements, table of contents (including, as relevant, tables, graphs, figures, annexes), and list of acronyms/abbreviations.	Yes	Opening pages contain all expected information including key facts table, and list of ERG members.
	ii Annexes include, if not in body of report: terms of reference, evaluation matrix, stakeholder map, list of respondents, results chain/ToC/logical framework, list of site visits, data collection instruments (such as survey or interview questionnaires), list of documentary evidence. Other appropriate annexes could include: additional details on methodology (e.g. inception report), additional details on ethical considerations, (country) case study reports, thematic papers or "deep dives."	Partially	The annexes include all expected annexes and also include a stakeholder map. Areas for improvement could include more detail on ethical considerations, greater detail on methodology, especially sampling and analysis. Additionally the annex on data collection tools is difficult to read and presented in a format that is not appropriately tailored for each different type of stakeholder group/ method of data collection, this could be improved for understanding and readability. Consent forms should also be included
Question 18.	Is the report logically structured and of reasonable length?		
	i The report has a logical structure that is easy to identify and navigate (for instance, with numbered sections, clear titles, well formatted).	Yes	The report has a logical structure and accords with UNFPA expectations for a CPE.
	ii Structure and length accords to UNFPA guidelines for evaluation reports; it does not exceed number of pages that may be specified in ToR.	Yes	At 70 pages in length the report accords with UNFPA guidelines, and structure.
Question 19.	Is the report well presented?		
	i Report is easy to understand (written in an accessible way for the intended audience) and generally free from grammar, spelling and punctuation errors.	Yes	The report is easy to understand and written in an accessible way for the intended audience. It is generally free from grammar, spelling and punctuation errors.
	ii Frequent use of visual aids (such as infographics, maps, tables, figures, photos) to convey key information. These are clearly presented, labelled, and referenced in text.	Partially	There is frequent use of visual aids in the background context and methodology sections, and these are clearly presented labelled and referenced in the text. However, there are more limited figures and data visualisation present in the findings section, which would aid in understanding of complex information.
SECTION I: CROSS-CUTTING ISSUES (weight 10%)		46%	Comments on Rating
Question 20.	Are cross cutting issues - in particular, human rights-based approach, gender equality, disability inclusion, LNOB - integrated in the core elements of the evaluation (e.g. evaluation design, methodology, findings, conclusions and recommendations)?		
	i Evaluation's data collection methods designed to capture the voices/perspectives of a wide range of stakeholders including rights holders, marginalized and vulnerable persons, young people, people with disabilities, migrants or refugee populations, indigenous communities, and other persons that are often left behind.	Partially	Data collection methods were designed to capture voices from a wide range of stakeholders, but as programme participants/ beneficiaries were not disaggregated in the methodology section it is difficult to assess what types of programme participants/beneficiaries specifically were consulted, and how left-behind groups and vulnerable people were explicitly included. The evaluation does note that it intended to capture data from vulnerable groups, including Roma women, but could do a better job of explicitly stating how this was accomplished both in terms of sampling, practical conduct of FGDs or KIs and analysis of disaggregated data. Data collection tools are minimal and contain a list of questions to be asked of each stakeholder, they are not tailored for different groups, or different types of data collection, they also contain no mention of consent, or accessibility, including language barriers or others.
	ii Evaluation questions address cross-cutting issues, such as human rights-based approach, gender equality, disability inclusion, LNOB, social and environmental standards as appropriate.	No	The evaluation states that the fundamental principles of gender equality, human rights and LNOB were mainstreamed throughout the evaluation questions. However these principles were not systematically or clearly integrated throughout the evaluation. A key factor which contributed to this was that none of the three evaluation questions included questions which related to vulnerable or marginalised groups, or to gender equality, or the principles of LNOB. It would be expected that these explored under both relevance and effectiveness criteria to explore the relevance of UNFPA programming for vulnerable and marginalised groups, and its effectiveness with regard to what results/ outcomes were achieved, and for whom.
	iii Data is disaggregated by population groups (e.g. persons with disability, age, gender, etc.) where there are implications related to UNFPA's portfolio/interventions for these population groups; differential results are assessed (distribution of results across different groups), as feasible.	Partially	Some data is disaggregated by population groups, particularly for Roma women, for young people and for older people - the focus on programming in Serbia. Some data and evidence is therefore presented by group, but within each group there is a lack for further disaggregation, or marginalisation, for example women with disabilities. This lack of a consistent intersectional analysis, or analysis across groups, is noted as a limitation, however it does not appear to have been appropriately addressed by the listed mitigation measures.
	iv Intersectional lens is applied in the data analysis, looking at various and multiple forms of exclusion and discrimination (and how they overlap with each other) and how this may impact the performance or results of the evaluation.	No	An intersectional lens has not been applied, this is noted as a limitation in the methodology section under limitations and mitigations. There is also little evidence of how mitigation measures to counter this were applied.
	v Findings, conclusions and recommendations, address cross-cutting issues such as equality and vulnerability, disability inclusion, leave no-one behind, social and environmental as relevant.	Partially	Cross-cutting issues are explored across some of the findings, conclusions and recommendations, but not all. LNOB and gender are mentioned the most; however, there is little mention of disability inclusion and vulnerable groups more widely.

	vi Inclusion of young people in the evaluation team and/or Reference Group.	Yes	Although not requested in the ToR a young evaluator was included in the evaluation team.
Question 21.	Does the evaluation meet UN SWAP evaluation performance indicators?	5	Comments on Rating
	i GEEW is integrated in the evaluation scope of analysis, and evaluation criteria and questions are designed in a way that ensures GEEW-related data will be collected.	Partially integrated	The evaluation includes analyses of GEEW considerations in the design of the project and gender and LNOB principles were stated to be integrated across evaluation questions, though it is unclear in practice how this was accomplished. There is not a dedicated evaluation question on GEEW, though the evaluation framework does include gender criterion. The disaggregation of gender could have been better explained in the design, and both human rights and gender equality could have been integrated into the design to enable an assessment of HR and GE results. There is also little information provided in relation to indicators relating to HR or GE and, as noted previously, no standalone question on gender equality is included.
	ii A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Satisfactorily integrated	<p>The evaluation employs a mixed methods approach and gender considerations are considered, some data was disaggregated by sex, though it would have been helpful to conclude the availability of disaggregated data, for secondary data collection as well as primary.</p> <p>It is unclear if the methods enabled/ ensured meaningful participation of women's voices. Women were certainly well represented in the primary data collection, but other vulnerable groups are not mentioned, nor is the role of women as rights holders discussed.</p> <p>The sampling framework does not explicitly explain how the most vulnerable groups were reached and addressed by the evaluation. While ethical standards were applied, detail of how this was conducted in practice is scant.</p>
	iii The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.	Satisfactorily integrated	<p>The findings do include an analysis of vulnerable and marginalised groups affected by issues in this evaluation, though it is not an intersectional analysis. The findings also analyse data that explicitly references the voices of different groups, such as Roma women, young people and older people, though disaggregated data in terms of primary and secondary data is not clearly described.</p> <p>Unanticipated effects of the intervention are not noted, and while gender issues are addressed by the findings, conclusions and recommendations human rights is not explicitly mentioned, and gender is primarily noted with regard to programmes such as SRH and GBV.</p> <p>Conclusions do mainstream gender across them, and there are specific recommendations addressing GEWE issues and priorities for action relating to LNOB and intersectionality.</p>

List of SDGs

1. No Poverty
2. Zero Hunger
3. Good Health and Well-being
4. Quality Education
5. Gender Equality
6. Clean Water and Sanitation
7. Affordable and Clean Energy
8. Decent Work and Economic Growth
9. Industry, Innovation and Infrastructure
10. Reduced Inequality
11. Sustainable Cities and Communities
12. Responsible Consumption and Production
13. Climate Action
14. Life Below Water
15. Life on Land
16. Peace, Justice and Strong Institutions
17. Partnerships for the Goals

Three transformative results

1. Ending unmet need for family planning
2. Ending preventable maternal deaths
3. Ending gender-based violence and harmful practices

Six outputs

1. Policy and accountability
2. Quality of care and services
3. Gender and social norms
4. Population change and data
5. Humanitarian action
6. Adolescents and youth

Six accelerators

1. Human rights-based and gender-transformative approaches
2. Innovation and digitalization
3. Partnerships, South-South and triangular cooperation, and financing
4. Data and evidence
5. Leaving no one behind and reaching the furthest behind first
6. Resilience and adaptation, and complementarity among development, humanitarian and peace-responsive efforts