EQA for UNFPA China Country Programme
Evaluation (2011-2014)

Title of Evaluation Report: UNFPA Country Programme Evaluation: CHINA – Final Evaluation Report

Overall Quality Rating: Poor

Overall Assessment: The evaluation covers a programme in a large and complex country, but one which is relatively limited in scope since China is a middle-income country. While it uses an appropriate methodology and rigorous analysis of data to assess the extent to which the UNFPA country programme in China produced the outputs and outcomes outlined in the CPAP, and provides good background on what has been happening in China regarding population, it does not, in most cases, show a causal connection between what UNFPA produces itself and the results in terms of country programme outputs and outcomes. The data collection from documents is thorough, and the many interviews are indicated in the report. However, the selection of sites to visit to observe project results was a limitation on what could reasonably be shown and the findings in terms of effectiveness, as a result, are not well supported. The conclusions and recommendations are rather general and the recommendations are not sufficiently strategic and operationally feasible.

Quality Assessment criteria	Assessment Levels			
	Very good	**Good**	**Poor**	**Unsatisfactory**
1. Structure and Clarity of Reporting *To ensure report is user-friendly, comprehensive, logically structured and drafted in accordance with international standards.* Checklist of minimum content and sequence required for structure: • Acronyms; ii) Exec Summary; iii) Introduction; iv) Methodology including Approach and Limitations; v) Context; vi) Findings/Analysis; vii) Conclusions; viii) Recommendations; ix) Transferable Lessons Learned (where applicable) • Minimum requirements for Annexes: ToRs; Bibliography; List of interviewees; Methodological instruments used.	**Good** All of the required sections are included. The annexes include the ToR, a list of documents consulted, and a list of interviewees. They do not include all of the methodological instruments used: the stakeholder mapping is included by interview guides are not. The summary boxes at the beginning of the findings subsections are useful previews of the main findings for each evaluation question (e.g. on p48). The report's length (73 pages without annexes) is in line with the recommended length.			

Quality Assessment criteria	Assessment Levels			
	Very good	Good	Poor	Unsatisfactory
2. Executive Summary *To provide an overview of the evaluation, written as a stand-alone section and presenting main results of the evaluation.* Structure (paragraph equates to half page max): • i) Purpose, including intended audience(s); ii) Objectives and Brief description of intervention (1 para); iii) Methodology (1 para); iv) Main Conclusions (1 para); v) Recommendations (1 para). Maximum length 3-4 page.	**Poor** While the summary contains all of the required sections, it is too long (6 pages) and has too much text on conclusions and recommendations that make it difficult to read. The section on purpose is missing (no indication of the intended audience). There is no description of what the UNFPA programme intended to do, which undermines the standalone quality of this section. The description of methodology has too many paragraphs and focuses more on the limitations of the evaluation than on the methodology.			
3. Design and Methodology *To provide a clear explanation of the following elements/tools* Minimum content and sequence: • Explanation of methodological choice, including constraints and limitations; • Techniques and Tools for data collection provided in a detailed manner; • Triangulation systematically applied throughout the evaluation; • Details of participatory stakeholders' consultation process are provided; • Details on how cross-cutting issues (vulnerable groups, youth, gender, equality) were addressed in the design and the conduct of the evaluation.	**Poor** The evaluation had too many questions to answer (12) and, while this number was found in the ToR, the evaluators added many sub-questions and as a result, the data collection was complicated. More importantly, the evaluators took the output in the country programme as the basis for the evaluation (shown in detail in Annex 3) without indicating the expected causal connection with what UNFPA would be providing to obtain those outputs (which, for UNFPA, would be outcomes). While the data collection techniques were a standard menu, there were some issues. For document analysis, the evaluators state that they reviewed 2000 documents but this is not fully presented in Annex 7. Many were studies that were related to population, but not necessarily to UNFPA's role. There were over 300 interviews done, but over half were with officials (UNFPA, UNCT and government) concerned with implementing programmes and the remainder were with local level administrators of projects and beneficiaries. Another issue is the selection of ten municipalities selected for site visits which were close to Beijing and were purposively selected.			

Quality Assessment criteria	Assessment Levels			
	Very good	**Good**	**Poor**	**Unsatisfactory**
	Although work with minorities (as a vulnerable group) was noted as a key niche for UNFPA, the evaluators did not go to areas with minority populations, on the grounds that time did not allow that. This was listed as a limitation, one of several that were carefully set out in the methodology section. Another was that the team leader left half way through the evaluation and was replaced by one of the team members and that this affected the ability to draw conclusions; however, the specific effect on data collection and finding is not discussed. The methodology included an effort at triangulation throughout the analysis, although this would be affected by the limitations in data collection. The evaluators describe that they used triangulation "to test preliminary findings and to improve on validity" (p26). There is evidence that triangulation was used to validate data. For example, the evaluators cite evidence gathered in field visits that is in line with the decline in C-section rates reported in national statistics (p57). Here also, the causal connection is not clear. There was evidence of considerable consultation with stakeholders, particularly at the central level. The methodology does not describe how cross-cutting issues were incorporated into the design and conduct of the evaluation. Cross-cutting issues are mentioned later (p38) and the cross-cutting issue of gender equality is discussed in the analysis section in depth (p63).			
4. Reliability of Data *To clarify data collection processes and data quality* • Sources of qualitative and quantitative data have been identified; • Credibility of primary (e.g. interviews and focus groups) and secondary (e.g. reports) data established and limitations made explicit;	**Poor** Sources of data are identified in the report, but there are gaps. The quantitative data came primarily from government sources and the qualitative from interviews, focus groups and field visits. The limitations in these were noted under design and methodology. The major weakness is the lack of data showing a causal connection between what UNFPA produces and its effect on CP7 output and			

Quality Assessment criteria	Assessment Levels			
	Very good	**Good**	**Poor**	**Unsatisfactory**
• Disaggregated data by gender has been utilized where necessary.	outcomes, partly because of the lack of geographic coverage in the qualitative data. Data is appropriately referenced either in the text (ex. "…the CPE's visit to the street sex worker (SW) project in Donghu District of Nanchang city, Jiangxi province showed that…" on p57) or in footnotes but there are instances where references are vague ("this was demonstrated in a presentation to the CPE", p45). There is some usage of data disaggregated by gender (ex. in the discussion of the UN regional study on violence against women on p47, statistics for men and women are cited). However, in many cases quantitative figures are cited that are not disaggregated. In some cases aggregated figures are presented for UNFPA outputs where the Country Office is likely to have collected data disaggregated by gender: "In 2013 the project covered a total of 8600 migrant workers with 111 received training as peer educators" (p57) and "Nearly 100 peer educators were mobilized- they provided information to about 9,000 young people at project sites" (p58).			
5. Findings and Analysis *To ensure sound analysis and credible findings* <u>Findings</u> • Findings stem from rigorous data analysis; • Findings are substantiated by evidence; • Findings are presented in a clear manner. <u>Analysis</u> • Interpretations are based on carefully described assumptions; • Contextual factors are identified.	**Poor** Each of the 12 questions has findings that contain considerable information. Some of the information is about what the government and other partners have done, but not what was causally connected with UNFPA products. In fact, other than listing UNFPA projects, what UNFPA has done is not well described, so the findings, particularly in terms of effectiveness, are not well substantiated. For example, the evaluators state that "The reforms at Liuyang, which seek to reduce VAW, have contributed to changing the attitude to GBV, while strengthening action to reduce it" but provide insufficient supporting evidence to link the cause and effect (p62). Questions			

Quality Assessment criteria	Assessment Levels			
	Very good	**Good**	**Poor**	**Unsatisfactory**
• Cause and effect links between an intervention and its end results (including unintended results) are explained.	related to relevance are supported by evidence that population issues are important, but this is not exceptional. Findings on efficiency are mostly about expenditure and are supported by data. Under effectiveness, the only area where clear causal connections are show is Violence Against Women where the evaluators find an important role by UNFPA. Findings about UNFPA's role in the UN Country team are based on comments made by persons interviewed, and there is no clear reference to supporting documentation (e.g. minutes). There is some discussion of contextual factors that influence the effectiveness of the UNFPA programmes, but it is insufficient. For example, the ownership and capacity built in programme areas cannot, however, be solely attributed to UNFPA support. For example, in Yongchuan, Chongqing, the RH center has long been a recipient of programme support from other agencies, including national and international" (p68). However, this kind of logic is missing from most of the discussion. The evaluation produces considerable information about context and assumptions, although again, without showing the causal connection with UNFPA products, it is difficult to see whether contextual factors were important or not. The problems with site selection are manifest in the findings, many of which are based on reviews of other studies rather than direct data collection.			
6. Conclusions *To assess the validity of conclusions* • Conclusions are based on credible findings; • Conclusions are organized in priority order; • Conclusions must convey evaluators' unbiased judgment of the intervention.	**Poor** The conclusions are short, not prioritized, and do not summarise effectively the data presented in considerable detail in the findings section. The confusion between UNFPA's output and the country programme output continues in the conclusions. The strategic conclusions are all very general (and based on interviews			

Quality Assessment criteria	Assessment Levels			
	Very good	**Good**	**Poor**	**Unsatisfactory**
	that stated that UNFPA was well-regarded in China). The programmatic conclusions are not in priority order, but rather by question and some conclusions (relevance) are not sufficiently substantive.			
7. Recommendations *To assess the usefulness and clarity of recommendations* • Recommendations flow logically from conclusions; • Recommendations must be strategic, targeted and operationally-feasible; • Recommendations must take into account stakeholders' consultations whilst remaining impartial; • Recommendations should be presented in priority order	**Poor** While the recommendations are connected to the conclusions, the fact that the conclusions are very general leads to very general recommendations. All of the recommendations are given high priority, which means that in effect there is no prioritisation. All recommendations are targeted at the UNFPA Country Office. While some recommendations are strategic and operationally feasible (ex. "strengthen ownership and coherence of next CP by establishing a twice yearly strategic management policy dialogue with key leading partners in GoC and enhance project coordination mechanisms at sub-national level by an annual IP's forum for discussions of programme results" on p90), others are too vague (ex. "It is important for the CO to stimulate population analyses in its support to address societal impact of complex population dynamics China is facing" on p89). By being very general, the recommendations cannot be said to be operationally feasible (most, in fact, say keep on doing what you are doing.)			
8. Meeting Needs To ensure that Evaluation Report responds to requirements (scope & evaluation questions/issues/DAC criteria) stated in the ToR (ToR must be annexed to the report). In the event that the ToR do not conform with commonly agreed quality standards, assess if evaluators have highlighted the deficiencies with the ToR.	**Good** The report follows the ToR, including the longer than normal list of questions. The evaluators have been very systematic in describing data sources and approach, but did not question the lack of information on what UNFPA actually produced to achieve country programme outputs.			

Quality assessment criteria (and Multiplying factor *)	Assessment Levels (*)			
	Very good	Good	Poor	Unsatisfactory
1. Structure and clarity of reporting (2)		2		
2. Executive summary (2)			2	
3. Design and methodology (5)			5	
4. Reliability of data (5)			5	
5. Findings and analysis (50)			50	
6. Conclusions (12)			12	
7. Recommendations (12)			12	
8. Meeting needs (12)		12		
TOTAL		14	86	

(*) Insert the multiplying factor associated with the criteria in the corresponding column e.g. - if "Finding and Analysis" has been assessed as "good", please enter the number 50 into the "Good" column. The Assessment level scoring the higher number of points will determine the overall quality of the Report

OVERALL QUALITY OF REPORT: Poor