Organizational unit:	Evaluation Office, UNFPA	2016
Title of evaluation report:	Evaluation of UNFPA support to population and housing census data to inform decision-making and policy formulation 2005-2014	

Overall quality of report: | Very Good | **Date of assessment:** 12 December 2016

Overall comments: The Census Data Evaluation is a very comprehensive, well written user-friendly report that deals with a critical aspect of UNFPA's activity of supporting population and housing census data that underpins its development policies and programs. In spite of the difficult task of undertaking a balanced and objective evaluation over a multi-year period on a global basis, the report provides a plethora of information, insights and suggestions for one of UNFPA's most fundamental world-wide core activities The heart of the evaluation is an expansive, in-depth presentation of the findings and analysis in Section 3 which, as supported by the various annexes, provides a sterling accounting of UNFPAs census data efforts for 2010 as they measured up – what worked well, what didn't and what might be done to improve them in the next (2020) cycle. The other sections of the report are of comparable caliber – all of which provides a strong basis for accepting the Conclusions and Recommendations in Section 4. The evaluation is an exemplar for other thematic evaluation in terms of how the evaluation results of thematic evaluations can be optimized so they are reported in an objective and credible manner.

Assessment Levels

Very good: strong, above average, best practice	**Good:** satisfactory, respectable	**Fair:** with some weaknesses, still acceptable	**Unsatis-factory:** 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)	
1. Structure and Clarity of Reporting	Assessment Level:	**Very Good**
To ensure the report is comprehensive and user-friendly • Is the report easy to read and understand (i.e. written in an accessible non-technical language appropriate for the intended audience)? • Is the report focused and to the point (e.g. not too lengthy)? • Is the report structured in a logical way? Is there a clear distinction made between analysis/findings, conclusions, recommendations and lessons learned (where applicable)? • Do the annexes contain – at a minimum – the ToRs; a bibliography, a list of interviewees, the evaluation matrix and methodological tools used (e.g. interview guides; focus group notes, outline of surveys)? *Executive summary* • Is an executive summary included in the report, written as a stand-alone section and presenting the main results of the evaluation? • 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)? • Is the executive summary reasonably concise (e.g. with a maximum length of 5-10 pages)?	Comment: The report is clearly organized and well structured. It adheres to the sequence of the established report outline and includes all required sections and components. The supporting evidence and analysis was excellent both in terms of creative use of displays in graphs, tables and boxes provided throughout the body of the report as well as in terms of the annexes provided in the two supporting volumes. The subject matter (censuses) is very broad, but the report while thorough is still of reasonable length. The executive summary is a stand-alone presentation of the evaluation findings, conclusions and recommendations. At eight pages it is within the concise standards used by UNFPA and covers the expected content.	

2. Design and Methodology	Assessment Level:	**Very good**

To ensure that the evaluation is put within its context
- Does the evaluation describe whether the evaluation is for accountability and/or learning purposes?
- Does the evaluation describe the target audience for the evaluation?
- Is the development and institutional context of the evaluation clearly described?
- Does the evaluation report describe the reconstruction of the intervention logic and/or theory of change?
- Does the evaluation explain any constraints and/or general limitations?

To ensure a rigorous design and methodology
- Is the evaluation approach and framework clearly described? Does it establish the evaluation questions, assumptions, indicators, data sources and methods for data collection?
- Were the methods chosen appropriate for addressing the evaluation questions? Are the tools for data collection described and justified?
- Is the methods for analysis clearly described?
- Are methodological limitations acknowledged and their impact on the evaluation described? (Does it discuss how any bias has been overcome?)
- Is the sampling strategy described? Does the design include validation techniques?
- Is there evidence of involvement of stakeholders in the evaluation design? (Is there a comprehensive/credible stakeholder map?)
- Does the methodology enable the collection and analysis of disaggregated data?
- Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?

Comment: The design began with a careful determination of expected results, based on a clear theory of change that is reflected in a solid results matrix. Based on this, the evaluator selected data acquisition tools that allowed them to measure results. The context for the evaluation, centring on the 2010 round of censuses, was clearly expressed that put the results matrix into context.

The evaluation design is based on a theory of change that is clearly established in Section 2: Context. Although the evaluation questions were not explicitly designated early in this section, they were subsequently linked to specific parts of the methodology. Each of the six evaluation components that are the lines of evidence used to informed the analysis are described and consist of a good mix of tools and techniques and are tied to an explanation of the sampling criteria. The design and methodology are based on a solid Evaluation Matrix in Annex 5 which addresses each of the respective Evaluation Questions.

The use of multiple methods to obtain data was particularly good. The use of document review, visits to a purposive sample of countries, where the basis for selection (region, extent of UNFPA involvement) was clear, and three surveys that were undertaken with particular care and yield a reasonably high rate of return, all contributed to obtaining the information necessary to measure results and why or why not they

	had been obtained. This permitted obtaining data that could be disaggregated. Cross-cutting issues, including particularly gender, were addressed. There was extensive involvement of stakeholders in both design and interpretation phases.

3. Reliability of Data	Assessment Level:	**Very Good**
To ensure quality of data and robust data collection processes • Did the evaluation triangulate all data collected? • Did the evaluation clearly identify and make use of qualitative and quantitative data sources? • Did the evaluation make explicit any possible issues (bias, data gaps etc.) in primary and secondary data sources and if relevant, explained what was done to minimize such issues? I.e. did the evaluation make explicit possible limitations of the data collected? • Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations? • Is there adequate gender disaggregation of data? And if this has not been possible, is it explained? • Does the evaluation make explicit the level of involvement of different stakeholders in the different phases of the evaluation process?	Comment: Together the text of the report and the Annexes provide a credible and firmly established basis for the report. Data sources for both qualitative and quantitative data are identified in the Annexes as well as footnoted appropriately throughout the body of the text and primary and secondary data are established, with their limitations noted. The constraints and limitations were noted and treated in terms of mitigating actions taken to offset them which accentuated the proactive approach the evaluation team followed in the course of the evaluation. The three tiered manner in which country data serves as the building block for regional and global data serves well the multiple purposes for the evaluation at the country, regional and global levels. Limitations were shown, including most importantly, how they were overcome. Ethical issues and how they were addressed is clearly shown. This is an evaluation whose data are exceptionally reliable.	

4. Analysis and Findings	Assessment Level:	**Very Good**

To ensure sound analysis

- Is information analysed and interpreted systematically and logically?
- Are the interpretations based on carefully described assumptions?
- Is the analysis presented against the evaluation questions?
- Is the analysis transparent about the sources and quality of data?
- Are possible cause and effect links between an intervention and its end results explained?
- Where possible, is the analysis disaggregated to show different outcomes between different target groups?
- Are unintended results identified?
- Is the analysis presented against contextual factors?
- Does the analysis include reflection of the views of different stakeholders (reflecting diverse interests)? E.g. how were possible divergent opinions treated in the analysis?
- Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?

To ensure credible findings

- Can evidence be traced through the analysis into findings? E.g. are the findings substantiated by evidence?
- Do findings follow logically from the analysis?
- Is the analysis of cross-cutting issues integrated in the findings?

Comment: Section 3 of the evaluation report is clearly organized and well documented. It presents a sound analysis and credible findings based on several key factors: First, all the basic elements are met for both the Findings and Analysis. Second, the manner in which the section is organized is consistent and thorough: the findings address each of the respective Evaluation Questions starting with a short summary of findings followed by detailed findings of the evaluation question and sub-issues, and includes an explanation of factors affecting the performance. For each of the evaluation questions, the analysis is tiered to include the three levels of analysis with national-level analysis providing information for the regional and global levels of analysis which, in turn, provided a solid basis for ensuring sound analysis and clear findings. Third, a discerning approach to "results" was taken whereby the evaluation team notes the "contribution" that UNFPA made as part of a broader effort to improve census data as opposed to claiming credit for specific results, thereby avoiding any effort for UNFP to take credit for results.

The evaluation focused on the production of census data as well as its dissemination and utilization, noting the latter as an area that came up noticeably short and that requires much more attention and resources in the coming cycle to optimize the census data endeavor. The subsequent section of the report, both the conclusions and recommendations, pick up on this short-fall.

	The evaluators do an outstanding job throughout this section of the report (indeed throughout the full report itself) in providing data in various presentational modes (tables, charts, graphs) to complement the narrative text in a very illustrative and visual manner. A few of the more prominent examples, drawn from numerous ones, include: Figure 7: The outcomes' contribution line of the theory change of UNFPA support to census; Table 17: Types of uses of consensus data (at the country level); Box 19: The ability to deliver under stringent deadlines, in challenging contexts and despite delays; Figure 19 A zoom-in into the outcomes of the ex-post theory of change in the conclusions section that shows a version of the theory of change with revisions based on evaluation findings.It includes views from stakeholders aspart of the evidence for the findings and included a section on cross-cutting issues, especially gender. This is an exceptionally thorough analysis and the findings are clear.

5. Conclusions	Assessment Level: **Very Good**
To assess the validity of conclusions • Are conclusions credible and clearly related to the findings? • Are the conclusions demonstrating an appropriate level of analytical abstraction? • Are conclusions conveying the evaluators' unbiased judgement of the intervention?	Comment: The conclusions flow from the findings, but have been organized to show their implication for UNFPA's assistance to censuses. The source in findings for each conclusion is shown clearly.

6. Recommendations	Assessment Level: **Very Good**
To ensure the usefulness and clarity of recommendations • Do recommendations flow logically from conclusions? • Are the recommendations sufficiently clear, targeted at the intended users and operationally-feasible? • Do recommendations reflect stakeholders' consultations whilst remaining balanced and impartial? • Is the number of recommendations manageable? • Are the recommendations prioritised and clearly presented to facilitate appropriate management response and follow up on each specific recommendation?	Comment: Recommendations are well presented and explicitly linked to the conclusions on a one-for-one basis by following a consistent and effective sequence: a summary of the recommendation with a brief elaboration, the priority given it, the target unit responsible for follow-through, and the cross-referenced to the conclusions. Each recommendation has a rational stating the basis for the needed action followed by operational suggestions and implementation considerations. Taken together, this provides for a consistent and effective set of recommendations that serves the underlying purpose of the report, improving UNFPA support to the next round of census process and most notably improving the use of data for decision- and policymaking at the country level. In short, the report provides a basis for making progress on the 2020 census by building on the lessons learned from the 2010 census.

7. Gender	Assessment Level:	**Fair**

To assess the integration of Gender Equality and Empowerment of Women (GEEW)[1]

- Is GEEW integrated in the evaluation scope of analysis and indicators designed in a way that ensures GEEW-related data to be collected?
- Do evaluation criteria and evaluation questions specifically address how GEEW has been integrated into design, planning, implementation of the intervention and the results achieved?
- Have gender-responsive evaluation methodology, methods and tools, and data analysis techniques been selected?
- Do the evaluation findings, conclusions and recommendations reflect a gender analysis?

Comment:

Gender equality is integrated in the evaluation scope and in selected indicators.

One evaluation question (EQ7) specifically addresses the integration of gender equality (together with human rights). However, there is no reference to GEEW in the other evaluation questions. In particular, dimensions of GEEW could have been integrated in the relevance and the effectiveness criteria (respectively under EQ1, EQ2 and EQ3), which prevents the evaluators from assessing how GEEW had actually been integrated in the planning and the implementation of UNFPA support to census data and in the results achieved through this support.

The evaluators have selected and used gender-responsive evaluation methods and tools, ensuring, in particular, the avoidance of gender biases and the reinforcement of gender discriminations. Attention to the gender equality dimension is particularly visible in the interview protocols provided in annex 7.

Findings do reflect a gender analysis, mainly under EQ7. However, this analysis does not lead to specific conclusions and/or recommendations pertaining to gender equality (in the conclusions and recommendations section, gender is referred to as one

---

[1] This assessment criteria is fully based on the UN-SWAP Scoring Tool. Each sub-criteria shall be equally weighted (in correlation with the calculation in the tool and totalling the scores 11-12 = very good, 8-10 = good, 4-7 = Fair, 0-3=unsatisfactory).

	UNFPA mandate area among others, with no specific analytical lens).

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	93		7	
Overall assessment level of evaluation report	**Very 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 'finding and analysis' 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[2]:

- How it can be used?

- What aspects to be cautious about?

Where relevant, please explain the overall assessment Very good, Good or Unsatisfactory[3]:

Consideration of significant constraints[4]

The quality of this evaluation report has been hampered by exceptionally difficult circumstances: ☐ yes X no

If yes, please explain:

---

[2] The purpose here is to clarify in what way the report can be used. This in order to assist the elaboration of a relevant Management Response and the wider use of the evaluation findings back into programming. When a report has been assessed as Fair, it is obligatory to fill this text box in.

[3] The purpose is, where relevant, to clarify for example severe unbalances in the report (for example, the report is good overall but recommendations very weak). Is optional to fill in.

[4] E.g. this should only be used in case of significant events that has severely hampering the evaluation process like natural disasters, evaluators falling sick, unexpected significant travel restrictions, etc. More 'normal' limitations should be mentioned under relevant section above.