

GEROS Evaluation Quality Assurance Tool

Version: September 2021

Federal Government of Nigeria–UNICEF Health and HIV/AIDS Country Programme 2018–2022 - End-Cycle Evaluation

REPORT RATING SUMMARY			
Overall Rating		68%	Satisfactory
●●●●●	Exceptional (96% - 100%)	5	
●●●●●	Highly Satisfactory (87.5% - 95.99%)	4	
●●●●●	Satisfactory (62.5% - 87.49%)	3	Meets UNICEF/UNEG standards for evaluation reports and decision makers may use the evaluation with confidence
●●●●●	Fair (35% - 62.49%)	2	
●●●●●	Unsatisfactory (0% - 34.99%)	1	

REPORT DETAILS	
Title of the evaluation report	Federal Government of Nigeria–UNICEF Health and HIV/AIDS Country Programme 2018–2022 - End-Cycle
Report sequence number	Nigeria/61/2024/18723
Region	WCAR
Year of report	2024
Office	Nigeria Country Office
Coverage (countries)	Nigeria
ToRs present	Yes
Date of review (dd/mmm/yyyy)	February 3, 2024
Name of review firm	IOD PARC

CLASSIFICATION OF EVALUATION REPORT	
Management of evaluation (Managerial control and oversight of evaluation)	UNICEF managed
Unicef goal areas (Alignment with strategic plan priorities)	
Every child survives and thrives	Yes
Every child learns	No
Every child is protected from violence and exploitation	No
Every child lives in a safe and clean environment	Yes
Every child has an equitable chance in life	No
Gender equality (cross-cutting)	Yes
Humanitarian action (cross-cutting)	No
Evaluation object	Programme
Evaluation type	Summative and formative
Evaluation strategy	Mixed methods
Evaluation design (primary method used)	Quasi-experimental
Evaluation level	Output & Outcome
Geographic scope	National
Primary SDG(s) covered (number)	3

EQA Summary: *The rater will provide top line issues for this evaluation relevant for feedback to senior management (positive and negative), summarizing how the evaluation report meets or fails*

This is, overall, a good quality evaluation that is well-designed, well-executed (with one significant caveat, discussed below) and benefits from a very polished presentation. It covers a very wide scope and has engaged with a substantial number of different sources of data in order to generate analysis and findings. It could easily have been rated higher according to the Geros criteria had the evaluators been more diligent in following some of the standard good practices in evaluations that this review assesses against. This said, the major issue with the evaluation is its length. While the scope and analysis required to meet it are complex and broad, the evaluators should have made greater efforts to rein in the scope of work committed to (in the evaluation matrix) and to summarize, avoid duplication and leverage annexes for longer analyses to ensure a tighter and more accessible report. Some of the major strengths and added value elements are as follows:

- Although long, the executive summary is clear, structured, and reflects the comprehensive analysis of the main report.
- The programme context is well-described, with detailed demographic and socio-economic data relevant to Nigeria and a comprehensively described theory of change.
- The design is strong, with the evaluation questions fully aligned with DAC criteria (including additional focus on equity, gender equality, and resilience), well-documented ethical considerations, including formal ethical clearance, and a robust sampling strategy.
- The analysis and findings are extensive, with robust quantitative (secondary) and qualitative (primary) data, ensuring triangulated findings presenting both descriptive data and causal analysis, especially under effectiveness and impact.
- Conclusions are clearly formulated and logically linked to the findings and evaluation objectives.
- Recommendations are mostly actionable, prioritized, and logically linked to the evaluation conclusions.
- The evaluation incorporates gender-disaggregated data and includes considerations of equity and inclusion.
- Visual aids, such as charts and tables, enhance readability and support data presentation.
- The report is well-written, with minimal grammatical or typographical errors.

Recommendations for Improvement: *The rater will identify topline recommendations to improve the evaluation, and be specific to the sections of the report where shortcomings were found. As*

As noted, the major area for improvement in this evaluation is the overall length of the report, which makes engaging with the specifics of analyses and findings challenging. Further (and despite the existing length and complexity of the report), the scope of data collection and analysis committed to (as documented in the evaluation matrix) was excessive, and could not reasonably be fully met. Hence, the evaluators appear to have largely set the evaluation matrix aside and focused on a selection of evaluation question - this is not good practice and the evaluation design (which is otherwise very robust) should have been better aligned with what was ultimately feasible. Other topline areas for improvement are:

- Both the executive summary and the overall report are considerably longer than what is considered good practice in evaluations, reducing accessibility for decision-makers to the main report in particular. There is some repetitive narrative and insufficient summarization that could have alleviated this, although the extensive analysis does still justify a longer-than normal report.
- The evaluation matrix includes numerous sub-questions and indicators, many of which are not fully addressed. The extensive number of these means the matrix is unsystematically applied in the findings.
- Headline 'overall conclusions' at the beginning of findings subsections are more analogous to findings.
- Some of the analysis lacks specific references to data sources, with some limited causal analysis, especially under coherence and sustainability.
- Lessons learned are few in number and overly generic, limiting practical applicability.
- Recommendations lack explicit linkages to specific findings, reducing clarity and traceability and are inconsistent in presentation, with some not having sub-recommendations or actions that would assist in implementation.
- The recommendation development process is not described, notably with respect to any stakeholder involvement.
- Use of the term "beneficiaries" instead of "rights-holders" is inconsistent with rights-based language.
- The monitoring and evaluation systems of the programme are only superficially assessed.

SECTION RATINGS		
SECTION A: EXECUTIVE SUMMARY (weight 5%)	83%	Comments on Rating
Question 1: Can the executive summary inform decision-making?		
Is clearly presented, serves as a standalone document useful for informing decision making, and is of relevant conciseness and depth for key users (Maximum of 5 pages unless otherwise specified in ToR).	Partially	The executive summary is ten pages, whereas the TOR specifies that it should be five pages, so is therefore considerably longer than directed (as is the main report). The length issue aside, it is clearly presented and reflects the considerable amount of analysis and discussion in the main report.

	ii	Includes all necessary elements (overview of the intervention, evaluation purpose, objectives and intended audience, evaluation methodology, key conclusions on findings, lessons learned if requested, key recommendations) as per the ToR.	Yes	All of the required elements are presented in the summary, although the findings and conclusions are presented together - the headline finding for each criterion following by a narrative justification for the finding is useful, although the narrative is essentially the text from the conclusions section that has been somewhat augmented. Nonetheless, the structure is sound.
	iii	Includes all significant information needed to understand the intervention and the evaluation AND does not introduce new information from what is presented in the rest of the report.	Yes	Notwithstanding the excessive length of the summary, the information it presents is good, although the background, context and methods could have been condensed, as could many of the findings/conclusions (which are themselves more-or-less a restating of the 'overall conclusions' at the beginning of each findings subsection). In this, it does indeed present sufficient information to understand the programme and the evaluation, with no additional information included.
SECTION B: BACKGROUND (weight 5%)			64%	Comments on Rating
Question 2.		Is the object of the evaluation clearly described?		
	i	Clear and relevant description of the intervention, including: location(s), timelines, cost/budget, and implementation status.	Yes	Section 2, 'Object of the Evaluation' describes the programme well. There is a very useful summary table (table 1) presented initially that provides all of the essential information relating to the programme in one place. The remainder of the section goes on to provide extensive narrative detail of the programme itself (including the finances, locations, status) and its place within the context of the wider UNICEF Country Programme in Nigeria. One minor issue is that the approximate exchange rate for Naira to USD (in the background/context section 1.2) could be provided in a footnote.
	ii	Clear and relevant description of intended rightsholders (beneficiaries) and duty bearers (state and non-state actors with responsibilities regarding the intervention) by type (i.e., institutions/organizations; communities; individuals...), by geographic location(s) (i.e., urban, rural, particular neighbourhoods, town/cities, sub-regions...) and in terms of numbers reached, with disaggregation by gender, age, disability... (as appropriate to the purpose of the evaluation).	Partially	The intended beneficiaries (this term is used in place of rights-holders throughout) are described in general terms with respect to major health indicators and needs across section 1.2 and 2. Key demographic or geographic factors (such as the differences between north and south of Nigeria) are noted in relation to the performance against these macro-level indicators. The programme descriptions do not include numbers of people targeted or reached, rather the programme was aimed at system and services support at a broader scale (nationwide, but a focus on 18 states). The TORs, however, do present specific quantitative population targets - these are present in the annexes, but not referenced in this section. Duty-bearer stakeholders are described in more detail, with the narrative in sections 1 and 2, and particularly table 1, providing a list of various institutional stakeholders (although not fully comprehensive).
Question 3.		Is the context of the intervention clearly described?		
	i	Clear and relevant description of the context of the intervention (i.e. relevant policy, socio-economic, political, cultural, power/privilege, institutional, international factors) and how context relates to the implementation of the intervention.	Partially	Section 1.1. covers the general background to UNICEF operations in Nigeria, noting the extant CPD, the UNDAF and the main national stakeholders in health and HIV and linking this clearly to the programme under evaluation. Taken with section 1.2 (which describes the specific health context relevant to the programme), it provides a reasonable clear description of the context. What might have been noted in this section, or the next, are other initiatives that seek to address the issues of health/HIV that overlap or are complementary to this programme (covered under the Coherence criterion in the findings section, but not referenced here).
	ii	Linkages drawn to the SDGs and relevant targets and indicators for the area being evaluated.	Partially	The context and background clearly notes and relates the programme to SDG 3 (health and wellbeing) and highlights work done in Nigeria to evaluate progress on this SDG. The sections do not, however, explicitly detail any of the specific indicators and/or targets related to this SDG and the programme, notably targets 3.2 and 3.3 which specifically relate to HIV and children.
	iii	Clear and relevant description (where appropriate) of the status and needs of the rightsholders/beneficiaries of the intervention.	Yes	Section 1.2 provides a comprehensive and relevant description of the background to child health and HIV in Nigeria, breaking it down across different dimensions (maternal, neonatal, >5) and the geographical context, which is very relevant in Nigeria, given the clear disparities between the north and south of the country. It provides well-referenced and recent data on the major health indicators for the intended rights-holders, supported by useful graphics/charts where necessary.
Question 4.		Are key stakeholders, their relationships and contributions clearly identified?		
	i	Identification of implementing agency(ies), development partners, right holders, and additional duty bearers and other stakeholders; and of linkages between them (e.g., stakeholder map) (if relevant).	Partially	As noted above, there is a narrative description and tabular summary listing of the main implementing agencies and development partners (as well as rights-holders) provided in sections 1 and 2 (and in table 1). This appears to be the only explicit listing of stakeholders provided in these sections (there are more details of the evaluation targets provided in the methods, but that is not relevant to this criterion). It is clear that the listing is not comprehensive as table 1 notes "other international and local non-governmental organizations (NGOs)" under the "partners" description. A more comprehensive map of the stakeholders (perhaps in the annexes) would have been useful.
	ii	Identification of the specific contributions and roles of key stakeholders (financial or otherwise), including UNICEF.	Partially	As above, the narrative and tables in sections 1 and 2 provide good general descriptions of the roles and contributions of key stakeholders, with table 1 providing a separate row for the sources of funding (although the specific contribution of each of the ten donors is not provided) and another row detailing the 'institutional/implementing' partners, though, again, not fully specifying each nor differentiating between these two categories.
SECTION C: EVALUATION PURPOSE, OBJECTIVES AND SCOPE (weight 5%)			92%	Comments on Rating
Question 5.		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	The purpose of the evaluation, presented specifically in section 3, is well-described and as per the TORs. The overall purpose, the accountability and learning functions, and the three broad intended uses of the evaluation (objective performance assessment, identify data gaps and identify future engagement) are fully and clearly described. The coinciding of the evaluation with the conclusion of this programme, the UNICEF country programme and various national strategies is also well-noted, highlighting its likely contribution to the development of successor strategies. A full list of the intended primary and secondary audiences is presented in table 2, with the rationale for their inclusion well-explained. The section is strong.
Question 6.		Are the objectives and scope of the evaluation clear and realistic?		
	i	Clear and complete description of what the evaluation seeks to achieve by the end of the process with reference to any changes made to the objectives included in the ToR (if applicable).	Yes	The five objectives of the evaluation are presented in their own section (4, which is possibly unnecessary - they could have been incorporated into section 3). They are clearly stated, complete and are as per the TORs, with no deviations.

	ii	Clear and relevant description of the scope of the evaluation: what will and will not be covered (thematically, chronologically, geographically with key terms defined), 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 or safety reasons at the time of the evaluation, lack of data/evidence on particular elements of the intervention).	Partially	Scope of the evaluation is described in section 5. The section covers thematic, geographic and chronological dimensions and notes that it is in line with the TOR and subsequent consultations during the inception phase. The thematic elements are as per the TOR, and are clear, as is the chronological scope. The geographic scope is somewhat less clear, however - the subsection (5.2) notes that there are four dimensions (as does the TOR), but the text has three numbered items. It appears that the cholera/COVID response element of the evaluation - noted separately in the TOR - has been incorporated into the first point in this section. While there is some justification for this approach, given that it aligns with the nationwide scope of the overall programme, the inconsistency and deviation from the ToR should be acknowledged. Additionally, the third point references "focus states with high HIV burdens" but does not specify which states. This should have been clearly identified.
Question 7.		Is the theory of change, results chain or logic 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	There is an extensive narrative description of the theory of change of the programme provided in section 2.3, covering the four health outputs and the two HIV outputs. This is supported by a graphical representation of the theory of change provided in Annex 1 (although the formatting of the graphic is poor - it is split across two pages incorrectly). The description is detailed and comprehensive; however, given the length of the report, consideration should have been given to moving the narrative to the Annex and presenting the graphic in the main report on a single page or summarizing the programme in one concise page. While this is more relevant to the overall presentation of the report rather than this specific criterion, it remains an important consideration.
	ii	Causal relationship between outputs and outcomes is presented in narrative and/or graphic form (e.g., results chain, logic model, theory of change, evaluation matrix).	Yes	As noted above, the causal relationship between the programme outcomes in Health and HIV and the relevant outputs, activities, risks and overarching sectoral/cross-cutting themes is described in detail in the narrative in section 2 and represented graphically in the theory of change in Annex 1.
	iii	For theory-based evaluations, the theory of change or results framework is assessed, and if requested in the ToR, it is reformulated/improved by the evaluators.	Yes	Although not explicitly specified by the TOR, the theory of change for the programme was assessed by the evaluation for plausibility and relevance (covered under relevance, section 8.1.1) and was deemed to be robust, with no changes recommended.
SECTION D:		EVALUATION DESIGN AND METHODOLOGY (weight 20%)	64%	Comments on Rating
Question 8.		Does the evaluation use questions and the relevant evaluation criteria that are explicitly justified as appropriate for the purpose of the evaluation? <i>UNICEF evaluation standards refer to the OECD/DAC criteria - Relevance; Coherence; Effectiveness; Efficiency; Sustainability; Impact (not all are necessarily relevant for all evaluations). Evaluations should also consider equity and leaving no-one behind, gender and human rights based approach (these can be mainstreamed into other criteria). Humanitarian evaluations should also consider Coverage; Connectedness; Coordination; Protection; Security.</i>		
	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.	Partially	The evaluation questions are presented in section 6, in accordance with standard DAC evaluation criteria, with some additional criteria included per the TOR (equity, gender equality and resilience). While the questions are generally appropriate for meeting the objectives and purposes of the evaluation, they are somewhat unclear in places (e.g. Relevance EQ2 notes that it is subject to an "evaluability assessment" - this is typically undertaken before any evaluation) and could have been revised or clarified. Further, there are some additions/changes to the evaluation questions from the TOR that are not noted in the report text, specifically the addition of a question related to commodities under Efficiency. It is good practice to note any deviation from the TOR.
	ii	In addition to the questions and sub-questions, the evaluation matrix includes indicators, benchmarks, assumptions and/or other processes from which the analysis can be based and conclusions drawn.	Partially	There is an evaluation matrix included in Annex 2 that breaks the evaluation questions into subquestions and links them to indicators and means of verification/sources and analysis methods. This is a very systematic and appropriate approach to the evaluation, though the large number of subquestions (51) means that responding to each appropriately and concisely (and with sufficient rigour) presents challenges. In addition to the large number of subquestions, there are many (160+) indicators to be assessed (or at least considered), from 1-8 for each subquestion, which again commits the evaluators to a substantial amount of analysis, or be found wanting in their treatment of the framework.
Question 9.		Does the report specify adequate methods for data collection, analysis, and sampling?		
	i	Evaluation design and set of methods are relevant and adequately robust for the evaluation's purpose, objectives and scope; and are fully and clearly described.	Partially	The evaluation methods are explained in considerable detail in section 7. The evaluation combines both quantitative (secondary) research with qualitative (primary and secondary) approaches from a variety of stakeholders and sources that should result in a robust evaluation analysis. The description of the methods is complete, with no apparent missing elements in the section. However, the section is repetitive and over-reliant on narrative descriptions not only makes it challenging for the reader to clearly and quickly grasp the suite of methods, it has clearly increased the propensity for errors in the description. For example, under 7.4.2, the "interviews and discussions" subsection (pg. 53) notes that 58 KIIs were held, yet in the following paragraph that 59 were held. The summary table (#3, pg. 55) then has a total of 70 KIIs. While almost all of the required information on the approach is present, the report would have benefited from summarising all of the approaches and tools into a single table (as was done with the sampling) for quick reference, and consideration to moving much of the narrative description to an annex. Finally, what appears absent is the timeline - when the data was collected etc. The inception report for the evaluation has a detailed workplan, but this is a plan - a timeline for the evaluation should have been annexed.
	ii	Data sources are appropriate - these would normally include qualitative and quantitative sources (unless otherwise specified in the ToR) - and are all clearly described.	Yes	The evaluators have chosen data sources well. For the quantitative secondary analysis, a mix of robust and varied health data have been used, balancing robustness and acceptability (e.g. the MICS from 2011, 2016 and 2021) with more recent sources of data (e.g. the NACA from 2022) to enhance comparability and the ability to triangulate data from different sources. Similarly for the qualitative primary data, the evaluators combine feedback from institutional duty-bearers at different levels with primary data from rights-holders at community level (from FGDs) and observations of health facilities directly. This is well structured and comprehensively described.

	iii	Sampling strategy is provided - it should include a description of how diverse perspectives are captured (or, if not, provide reasons for this).	Yes	The description of the overall methods and the specific data collection approaches includes details on the various sampling methods applied (primarily a mix of purposive and convenience for primary data). The narrative (summarised in section 7.5) is supported by a useful table (pg. 55 that notes the specific stakeholders sampled, the technique used and the total of each sample. This is adequately described and should have ensured diversity of perspectives across the evaluation in combination with the secondary data.
	iv	Clear and complete description of the methods of analysis.	Partially	There is extensive description of the analytical approaches used for treatment of the primary and secondary data. While this is difficult to parse (and the overall criticism of the report relating to the heavy narrative applies), in particular for the secondary quantitative data, this is a reflection of the complex nature of the datasets and the evaluation requirement to apply detailed statistical analyses (such as propensity-score-matching, difference-in-difference etc.) and thus the detail is warranted. A definite issue, however, is that the description of the statistical analysis is not provided in a single place - the methods overview (7.2.1) includes considerable detail around this that should have been moved to the relevant analysis subsection (7.6.1). Description of the analytical approaches to the qualitative primary data is considerably more briefly described (in 7.6.4) but is nonetheless complete and adequate.
	v	Methodology allows for drawing causal connections between outputs and expected outcomes.	Yes	Considerable efforts on the part of the evaluators to ensure robustness of the (quasi-experimental and theory-based) approaches to the data collection and analysis have been made, and are well-described. The choice of statistical analysis techniques to the secondary health datasets appears sound, and the qualitative data collection methods are equally robust, with the specific tools provided in the annexes and the questions linked to the specific evaluation criteria. Correct and diligent application of these approaches to the data collection should have led to the determination of causal links between outputs and outcomes, per the purpose and objectives of the evaluation.
	vi	Clear and complete description of limitations and constraints faced by the evaluation, including gaps in the evidence that was generated and mitigation of bias, and how these were addressed by the evaluators (as feasible).	Partially	There are two lists of limitations, in different locations. The first relates specifically to the limitations on the quantitative secondary dataset analysis (section 7.6.1) with a more detailed and comprehensive list of limitations, risks and constraints presented in section 7.8. While, taken together, they present a clear and complete description, both sections should have been combined for a single presentation of the issues and appropriate mitigation strategies, where relevant. Beyond this, the section is complete.
Question 10.		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 the principles in the 2020 revised UNEG Ethical Guidelines (integrity, accountability, respect, beneficence).	Partially	Section 6.1, which discusses the evaluation criteria, provides something of a preamble to the ethical issues by referencing a number of guiding principles for the evaluation (independence, technical rigour, transparency, validity, reliability, partnership and usability) and contextualising these. While this is positive, the source of these principles is unclear (they are not UNEG) and they are somewhat out of place in this section. Further, the evaluators claim that one of the measures taken to ensure application of the principles is to make a distinction between "facts (findings) and opinions of the evaluation team" (bullet 3). Good practice evaluation does not permit the opinions of evaluators - it relies on interpretation of evidence only. This said, the evaluation design section (7) does make explicit reference to the UNEG guidelines and standards and contextualises a range of these in section 7.7. There is some repetition in the section which could have been eliminated to reduce the overall report length.
	ii	Description of ethical safeguards for participants appropriate for the issues relevant to methodology and how they are applied (respect for dignity and diversity, right to self-determination, fair representation, compliance with codes for vulnerable groups, confidentiality, and avoidance of harm). For those cases where the evaluation involves interviewing children, explicit reference is made to the UNICEF procedures for Ethical Research Involving Children.	Yes	The evaluators provide a very complete description of the specific ethical safeguards and principles followed during the evaluation. This includes both general principles and also specific provisions for ethical approaches in dealing with the adolescents that were included in the data collection. The data collection tools in the annexes include all appropriate consent/assent processes (including assurance of the right to withdraw) and explicit reference to UNICEF's Ethical Research Involving Children guidelines was made. Finally, ethical clearance for the research was obtained from the National Health Research Ethics Committee of Nigeria, provided with the report (but separately, not annexed, which it might have been).
Question 11.		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 be evident in several ways such as the design of the methodology (i.e. use of technology for data gathering, extensive participatory processes, systematic analysis processes such as collaborative outcomes reporting and incorporation of big data, specific strategies to address complexity such as outcome harvesting, strong child rights focus), or ways of sharing of evaluation results.	No	There is no evidence of any particularly innovative practices being applied to the evaluation.
SECTION E:		EVALUATION FINDINGS (weight 25%)	75%	Comments on Rating
Question 12.		Do the findings clearly address all evaluation objectives and scope?		

	i Findings marshal sufficient levels of evidence to systematically address all of the evaluation's questions, sub-questions and criteria.	Partially	Assessing the systematic approach of the findings and analysis, it is apparent that the evaluators only partially used the evaluation matrix as the basis for generating findings. While the evaluation criteria have been followed somewhat consistently (although not in the same order - Efficiency has been swapped with Impact), there are significant inconsistencies between what is presented in the matrix (with the many questions, sub-questions and indicators) and what is presented in the report. For example, both Relevance and Coherence have some, but little overlap with the questions or subquestion areas in the matrix (one question from Coherence). The Effectiveness criterion follows more closely the questions, but adds an extra question related to the social mobilization/behaviour change strategy effectiveness (Effectiveness EQ4) that is not in the matrix. Under Efficiency, a subquestion from the matrix has been promoted to the status of a full evaluation question. The criteria of Sustainability (8.6) and Gender Equality/Equity (8.7) do not even mention the related evaluation questions nor follow the subquestions while the final criterion (Resilience) is the only one that is in line with the questions in the matrix. Essentially, the evaluators have largely ignored the evaluation matrix as a guide to the analysis and have picked and chosen specific evaluation questions and/or subquestions to answer in accordance with a logic that has not been explained in the report. This may be understandable given the very substantial number of subquestions present, but it undermines the purpose of an evaluation matrix and is quite unsystematic. From an evidential perspective, those questions that have been answered, are done so with, by and large, a robust level of analysis that is drawn from the considerable amount of evidence that it is clear has been gathered. There are extensive analyses under Effectiveness and Impact, though less so under Coherence, which covers less than two pages. The voluminous nature of the report may have served to obfuscate some of these deficiencies among earlier reviewers.
	ii Explicit use of the intervention's results framework/ToC in the formulation of the findings.	Yes	The evaluators explicitly and clearly analyse the results of the programme in line with the Theory of Change as articulated in the background/context section and in annex 1. Although the theory of change is referenced in several places in the findings, the analysis specifically addresses its relevance (section 8.1.1) and under 8.3 (Effectiveness) where the outputs and outcomes of the results chain are analysed.
Question 13.	Are evaluation findings derived from the conscientious, explicit and judicious use of the best available, objective, reliable and valid data and by accurate quantitative and qualitative analysis of evidence.		
	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 through the use of multiple data sources.	Yes	The report is clearly based on a robust analysis of a comprehensive suite of both quantitative and qualitative data, the sources having been itemised in the methods and the data sources noted in the evaluation matrix. Different subsections (based on the evaluation criteria) use a different mix of evaluation data and triangulation is evident.
	ii Findings are clearly supported by, and respond to, the evidence presented, including both positive and negative. Findings are based on clear performance indicators, standards, benchmarks, or other means of comparison as relevant for each question.	Partially	Each subsection presents a headline finding with a rating of the quality of evidence against it, which is a useful and visually engaging rubric for readers to gauge the strength of the finding. For some subsections (Effectiveness, Impact, Efficiency), the individual questions are also itemized with the strength of evidence presented for each again - why this is not done for the other subsections (some of which have multiple evaluation questions associated) is unclear. Beyond these headline findings, the report does not present any itemized findings, rather it includes (at the start of each subsection, after the headline finding) a boxed list of 'overall conclusions' - these are directly analogous to findings, even to the extent of including a reference to the specific paragraphs on which the conclusion is based. As discussed below, having conclusions within the analysis as well as at the end of the report is duplicative and somewhat confusing. A quick and effective solution would be to rename the 'overall conclusions' to 'findings'. Further, interpolating these throughout the analysis rather than as a block at the beginning of each subsection would enhance clarity. Beyond this, the findings are overall well-supported by the evidence, in line with the summary ratings. Some elements of the report are more deeply analysed than others - for example, much of the effectiveness and impact subsections (which comprise 45 pages of analysis alone) are presented with a wealth of quantitative and qualitative analysis that is based on output and outcome indicators (covered extensively in 8.3.1, 8.3.2 and 8.4.1) related to children's health. The analysis in other subsections is relatively superficial by comparison. For instance, the Coherence subsection (8.2) acknowledges alignment with various strategies and policies and lists them, but it lacks substantive evidence to support these positive findings. A more systematic explanation of each point of alignment, with appropriate references to sources, would have been expected.
	iii Causal factors (contextual, organizational, managerial, etc.) leading to achievement or non-achievement of results are clearly identified. For theory-based evaluations, findings also analyse the logical chain (progression -or not- from implementation to results).	Yes	The findings sections are very focused on presentation of descriptive analysis of data, i.e. what was seen or found by the evaluators from the research, but there is also consideration of some of the causal factors that may have driven the observed changes. The extensive nature of the analysis in some subsections (particularly Effectiveness and Impact) can obscure some of the causal analysis, but it is still present. For example, under Impact (subsection 8.4.1.5, para 119 onwards, table 17), the evaluators discuss some of the potential reasons that underlie observed changes in <5 mortality and seek to establish factors common to areas reporting positive change. This is also reflected in the unintended effects analysis in 8.4.2, which brings in useful evidence from national stakeholders and rights-holders to explore the nature, causes and consequences of changes observed by the evaluators. This is very much in line with the logical chain of results, partially articulated in the theory of change, but also with recourse to the programme results framework, which included elements of impact, whereas the theory of change only goes to the level of outcomes (noted by the evaluators in their analysis in section 8.4.1). Even in the subsections where the analysis has been more cursory, some evidence of causal elements is presented, e.g. under Coherence, which (as noted above) has limited concrete evidence to support the findings assertions, but includes testimony (quotes) from stakeholders that highlight the reasons why the programme was aligned well with global and national strategies.
Question 14.	Does the evaluation assess and use the intervention's Results Based Management elements?		
	i Assessment of the adequacy of the intervention's monitoring system (including completeness and appropriateness of results/performance framework - including vertical and horizontal logic, M&E tools and their usage) to support decision-making.	Partially	The monitoring of the programme is assessed, though not to any significant depth. It is noted under the Effectiveness criterion (overall conclusion EFFE10) and more particularly under Efficiency (8.5.3, para 158 onwards) where the weakness of the HMIS and internal monitoring systems are flagged, although not analysed in any depth, merely the testimony of stakeholders cited.

SECTION F: EVALUATION CONCLUSIONS & LESSONS LEARNED (weight 10%)		75%	Comments on Rating
Question 15. Do the conclusions clearly present an objective overall assessment of the intervention?			
i	Conclusions are clearly formulated and reflect the purpose and objectives of the evaluation. They are sufficiently forward looking (if a formative evaluation or if the implementation is expected to continue or have additional phase).	Yes	One of the features of the report is the presentation of 4-10 "overall conclusions" before each evaluation criterion in the findings sections. The report also includes the standard final conclusions section towards the end of the report (section 10). This section comprises a selection of the interim conclusions from earlier in the report, mostly copied word-for-word, but with some conclusions omitted, and others summarized. The final section has clearly formulated conclusions, presented well under each of the evaluation criteria, thus reflecting the purpose and objectives of the evaluation well, although they are more summative than forward looking. However, the logic of duplicating these conclusions in the findings sections (with additional conclusions) in a report that is already overly-long, is not clear. The evaluators would have been better served by removing the conclusions from the findings section and moving all to section 10 (or renaming them "findings", which they appear to be more akin to, as noted above).
ii	Conclusions are derived appropriately from findings, and present a picture of the strengths and limitations of the intervention that adds insight and analysis beyond the findings.	Partially	As noted, the <i>final conclusions</i> are derived from the conclusions statements in the preamble to each findings subsection, which in themselves are clearly and logically linked to the analysis. As noted, no itemised findings are presented in the main report, with the overall conclusions in this section taking the place of findings. The latter are clearly linked to the analysis (via numbered source paragraphs), and thus the final conclusions are derived well from the analysis, if somewhat confusingly. This said, the conclusions statements are quite summative for the most part - noting what was and was not achieved, with only partial insights into the performance of the programme beyond this. For example, under the Effectiveness conclusion, the evaluators note simply where the programme did or did not achieve its targets, without presenting potentially useful information on the underlying causes of this that would usefully link to the recommendations. There are some instances where some deeper insights are presented, such as around the effectiveness (and drivers) of social mobilization and behaviour change, or on linking state health expenditure to programme impacts - more attention to these factors and drivers would have been expected in this section, particularly for the other criteria.
Question 16. Are logical and informative lessons learned identified? [N/A if lessons are not presented and not requested in ToR]			
i	Identified lessons stem logically from the findings, have wider applicability and relevance beyond the object of the evaluation.	Yes	Inasmuch as they are presented, the lessons are logically derived from the findings - each lesson discusses from where it emerged and uses examples to highlight how it was derived and impacted the programme. They are also more broadly applicable to other programming, <u>which is positive, but are somewhat generic and few in number (discussed below).</u>
ii	Lessons are clearly and concisely presented, yet have sufficient detail to be useful for intended audience.	Partially	While the lessons presented (in section 10) are clear and concise, there are few of them when the extensive amount of analysis presented in the preceding sections is taken into consideration. The report includes only four (of the 'most pertinent') lessons on a single page. This might be useful in an executive summary, but for a report of 150+ pages it seems very thin, particularly given the emphasis placed in the TOR on learning lessons (the TOR scope notes some key areas where lessons are expected to be learned). For example, the presumption that the HMIS would be adequate for monitoring of results in real-time was not valid, leading to deficits in this area - this is a clear lesson that could be learned for future programming in Nigeria and elsewhere (as HMIS' are used worldwide). The evaluators should have paid greater attention to this element of the evaluation.
SECTION G: RECOMMENDATIONS (weight 15%)		38%	Comments on Rating
Question 17. Are recommendations well grounded in the evaluation?			
i	Recommendations align with the evaluation purpose, are clearly formulated and logically derived from the findings and/or conclusions.	Partially	The recommendations are organized into strategic and operational categories (Although the two tables could have more prominent headings) and are generally well-aligned with the evaluation purpose. The strategic recommendations are specifically aligned with the evaluation criteria, which is useful, but the operational recommendations are not, which is an omission - some are relevant to effectiveness, others efficiency, some gender/equity - so there is definite scope for this. While a review of the recommendations indicates that they are derived indeed from the evaluation findings and conclusions, there is no explicit linkage to specific findings or conclusions, allowing the logical inferences to be clearly noted - readers need to be familiar with the extensive analysis or spend time cross-referencing the recommendations with the findings section - <u>this is not good practice, which is to clearly indicate the linkages.</u>
ii	Recommendations are useful and actionable for primary intended users and uses (relevant to the intervention); guidance is given for implementation, as appropriate.	Partially	All of the recommendations make sense and are useful, and the priority that they are assigned is reflective of the analysis/findings and conclusions. Most (but not all) of the strategic recommendations have a number of suggested actions and opportunities that can help guide their implementation, which are very useful (though individual actions are not assigned to stakeholders, rather the overall recommendation is). However, this is reserved for the high-priority strategic recommendations - the medium priority recommendations have no action points, and none of the operational recommendations are broken into actions - each of the five operational recommendations is a paragraph of narrative that combines a short sentence (which is essentially the recommendation) and a narrative justification for this. Again, this is not standard good practice in evaluations.
iii	Process for developing the recommendations is described, and includes the Involvement of duty-bearers, as well as rights holders when feasible (or explanation given for why they were not involved).	No	There is no description of how the recommendations were developed other than as part of the overall evaluation analysis and report-writing in the methodology section. The development of the recommendations as evaluation deliverables is noted in section 6, as are the intended audiences for them, but the actual process and any participation of key stakeholders is absent.
Question 18. Are recommendations clearly presented?			
i	Clear identification of groups or duty-bearers responsible for action for each recommendation (or clearly clustered group of recommendations). Clear prioritization and/or classification of recommendations to support use.	Partially	As noted above, each of the ten recommendations is assigned to a group of stakeholders, with a series individual actions associated with the first two that are not assigned individually, so it is not fully clear which stakeholder is responsible for each action. The remaining eight recommendations do not have action points so the target stakeholders are clearer. The recommendations are divided into high and medium priorities (although there are only two medium-priority recommendations - the remainder are high). This is minimally adequate (although not required, good practice frequently includes a time horizon for individual recommendations, as well as a priority-level).
SECTION H: REPORT STRUCTURE AND PRESENTATION (weight 5%)		79%	Comments on Rating

Question 19.	Does the evaluation report include all relevant information?		
	i Opening pages include: Name of evaluated object, timeframe of the evaluation, date of report, location of evaluated object, names and/or organization(s) of the evaluator(s), name of organization commissioning the evaluation, table of contents -including, as relevant, tables, graphs, figures, annexes; list of acronyms/abbreviations, page numbers.	Yes	All of the required elements are present, with no omissions apparent. All page numbers line up, as do the lists of figures/tables/boxes which also have page numbers and links assigned, which is useful to navigate quickly to items of interest.
	ii Annexes include: terms of reference, evaluation matrix, list of interviewees, results chain/ToC/logical framework (unless included in report body), 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, information about the evaluator(s), etc.	Partially	The annexes include most of the required elements that are not otherwise in the main report. However, an important omission is the observational survey tool (observation of facilities was noted as part of the methodology).
Question 20.	Is the report logically structured?		
	i Structure is easy to identify and navigate (for instance, with numbered sections, clear titles and sub-titles, well formatted).	Yes	While the structure is generally straightforward, there is considerable narrative and many subheadings under the findings section. A minor issue is that the recommendations are not numbered, but this is a minor stylistic issue that does not go against formal guidance. The report is otherwise very well formatted and polished.
	ii Structure accords to UNICEF guidelines for evaluation reports: context, purpose and methodology would normally precede findings, which would normally be followed by conclusions, lessons learned and recommendations.	Yes	The structure is in accordance with the relevant guidelines.
Question 21.	Is the report well presented?		
	i Report is of reasonable length; it does not exceed number of pages that may be specified in ToR.	Partially	This is an exceptionally long report. The entire main report (not including annexes) is 163 pages, with 88 pages of findings and analysis alone. The scope - both thematic and geographical - is extensive, with a very substantial commitment to analysis made, via the evaluation matrix and related sub-questions and indicators (of which there are many, as noted under 8ii above), which necessitates a considerable amount of work to do justice to. However, the evaluators have compounded this by making few apparent efforts to compress the report length. The writing - while being of good quality - is long-winded and repetitive in many places, with many pages of introductory, summary or 'filler' text that a judicious editing could have removed and led to a more to-the-point and shorter report. Some of these have been noted in rating reviews above. This said, there is no guidance to page length provided in the TORs, and given the depth and breadth of the subject matter, the length is partially justified, per the rating, but, as the analysis is largely sound and well-undertaken more could and should have been done to tighten the report to a more manageable length for the reader.
	ii Report is easy to understand (written in accessible way for intended audience) and generally free from grammar, spelling and punctuation errors.	Yes	Despite the length, the report is very well-written. There were very few grammar, spelling or punctuation errors noted.
	iii 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	Overall, the report makes extensive and good use of visual graphics, which serve to break up the very substantial size of the narrative and illustrate the findings and analysis. Individual findings subsections also benefit from visually engaging boxes, quotations, well-formatted headline findings (with the strength of evidence ratings color-coded). This all adds up to a visually engaging report. While all graphics are well-presented/labelled, some of the graphics could be better formatted - e.g. figure 1 (which is actually two figures, but labelled as one), figure 2, figure 3 are presented across 1.5 pages, and could be shrunk to a considerably smaller size in line with the text explaining them (which follows the charts, rather than before or contiguous with them). Further, some of the graphics, while interesting, are unnecessary in the context of an already overly-long report. For example, figures 5, 6 and 7 are illustrative of generic approaches, and unnecessary to the report. Efforts to reduce the overall length of the report could have included removal of some of these to annexes.
SECTION I:	EVALUATION PRINCIPLES (weight 10%)	71%	Comments on Rating
Question 22.	Did the evaluation design and style consider incorporation of the UN and UNICEF's commitment to a human rights-based approach to programming, to gender equality, and to equity?		
	i Reference and use of rights-based framework, and/or CRC, CCC, CEDAW and/or other rights related benchmarks in the design of the evaluation.	Partially	Although the design of evaluation was well-oriented towards a rights-based approach (both for child rights, gender equality and wide considerations of vulnerability inclusion), evident in the overall purpose, the evaluation questions and the analytical approach, there is little reference to normative rights-based frameworks in the overall design. The only framework specifically cited in the design aspects is that of Leave No-One Behind, although the language of a rights-based approach is used throughout (a positive aspect). The findings/analysis section incorporates this language of rights-based approaches (and gender equity and LNOB) and does note both the CRC and CEDAW, but only briefly - "UNICEF's efforts" are noted to "support the principles" of the CRC and CEDAW, with no further explanation or evidence that supports this finding. There is another brief reference to <u>the CRC under a partnership activity of the programme (section 8.3.3, para 74) but no more</u>
	ii Clear description of the level of participation of key rights holders and duty bearers in the conduct of the evaluation (for example, a reference group is established, stakeholders are involved as informants or in data gathering).	Partially	It is clear that duty-bearers and rights-holders have been involved in the evaluation in terms of their participation as informants or contributors to data collection - this is well described in the methods and their inputs are reflected in the analysis. The report also notes the presence of an evaluation reference group, but the membership of this is not clearly specified. The acknowledgement also note the existence of a "UNICEF Nigeria Evaluation Review Committee" which was an internal management group. A list of the full membership of the <u>reference group should have been included in the annexes or with the cover pages.</u>
	iii Language is empowering and inclusive, avoiding gender, heterosexual, age, cultural and religious bias, among others; use terminology of rights holders and duty bearers; data is disaggregated by marginalized group; differential results are assessed (distribution of results across different groups).	Partially	There is a brief mention of rights-holders in the design of the report (section 8.1.5), but otherwise the evaluation refers to "beneficiaries". Similarly, duty-bearers are referred to in the same section and the evaluation matrix, but not elsewhere. Notwithstanding this, there is no particular evidence of bias towards or against any groups, and the evaluation makes particular efforts to explore the gender equity and equality dimensions. There is extensive gender-disaggregation of data throughout the analysis, and qualitative analysis also includes gender dimensions and analysis by various factors of vulnerability.

iv	Evaluation assesses the extent to which the implementation of the intervention addresses child rights and Leave No-one Behind (gender and other excluded and marginalized groups). It is disability inclusive.	Partially	The dimensions of equity and gender equality (which included disability and other vulnerability dimensions) were explicitly included as an evaluation criterion, and thus received good attention in the report. The evaluators present specific findings related to this under section 8.7, as well as discussion the LNOB aspects of the programme. Disability is noted in this section, but is not given extensive treatment.
Question 23.	Does the evaluation meet UN SWAP evaluation performance indicators? (Note: this question will be rated according to UN SWAP standards with detail provided below)	9	
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.	Fully integrated	Equity and Gender Equality are included as a specific additional evaluation criterion for this evaluation, which is a positive and proactive step to ensure full integration of this consideration.
ii	A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Fully integrated	In line with the overall scope and evaluation approach, the evaluators incorporated gender dimensions into the evaluation matrix and is well reflected in the primary data collection tools, which have specific sections (and associated questions) related to gender equity/equality.
iii	The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.	Fully integrated	In line with the overall scope and evaluation approach, the evaluators incorporated gender dimensions into the evaluation matrix and is well reflected in the primary data collection tools, which have specific sections (and associated questions) related to gender equity/equality. Data collection was undertaken with gender considerations in mind (e.g. FGDs conducted with gender disaggregated groups) and analyses disaggregated by/focused on gender and vulnerability.

SWAP Rating Guidance

<p>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.</p> <p>a. Does the evaluation assess whether sufficient information was collected during the implementation period on specific result indicators to measure progress on human rights and gender equality results?</p> <p>b. Does the evaluation include an objective specific to assessment of human rights and gender equality considerations or was it mainstreamed in other objectives?</p> <p>c. Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria?</p> <p>d. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation?</p>
<p>ii A gender-responsive methodology, methods and tools, and data analysis techniques are selected.</p> <p>a. Does the evaluation specify how gender issues are addressed in the methodology, including: how data collection and analysis methods integrate gender considerations and ensure data collected is disaggregated by sex?</p> <p>b. Does the evaluation methodology employ a mixed-methods approach, appropriate to evaluating GEWE considerations?</p> <p>c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility?</p> <p>d. Does the evaluation methods and sampling frame address the diversity of stakeholders affected by the intervention, particularly the most vulnerable, where appropriate?</p> <p>e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality?</p>
<p>iii The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.</p> <p>a. Does the evaluation have a background section that includes an intersectional analysis of the specific social groups affected by the issue or spell out the relevant normative instruments or policies related to human rights and gender equality?</p> <p>b. Do the findings include data analysis that explicitly and transparently triangulates the voices of different social role groups, and/or disaggregates quantitative data, where applicable?</p> <p>c. Are unanticipated effects of the intervention on human rights and gender equality described?</p> <p>d. Does the evaluation report provide specific recommendations addressing GEWE issues, and priorities for action to improve GEWE or the intervention or future initiatives in this area?</p>