GEROS Evaluation Quality Assurance Tool

Version: September 2021

Summative evaluation of the CRVS programme in Mozambique

REPORT RATING SUMMARY			
Overall Rating		58%	Fair
● ● ● ● ●	Exceptional (96% - 100%)	5	
● ● ● ●	Highly Satisfactory (87.5% - 95.99%)	4	
● ● ● -	Satisfactory (62.5% - 87.49%)	3	
● ● - -	**Fair (35% - 62.49%)**	**2**	**Meets UNICEF/UNEG standards for evaluation reports in some regards, but not all. Decision makers may use the evaluation with caution. Substantive improvements in some areas are needed**
● - - -	Unsatisfactory (0% - 34.99%)	1	

REPORT DETAILS	
Title of the evaluation report	Summative evaluation of the CRVS programme in Mozambique
Report sequence number	Republic of Mozambique 63-2024-19602
Region	ESAR
Year of report	2024
Office	Mozambique Country Office
Coverage (countries)	Mozambique
ToRs present	Yes
Date of review (dd/mmm/yyyy)	December 9, 2024
Name of review firm	IODPARC

CLASSIFICATION OF EVALUATION REPORT	
Management of evaluation *(Managerial control and oversight of evaluation*	Jointly managed with organizations outside the UN system
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	Yes
Every child lives in a safe and clean environment	No
Every child has an equitable chance in life	No
Gender equality (cross-cutting)	Yes
Humanitarian action (cross-cutting)	No
Evaluation object	Joint Programme
Evaluation type	Summative and formative
Evaluation strategy	Mixed methods
Evaluation design *(primary method used)*	Theory-based
Evaluation level	Output & Outcome
Geographic scope	National
Primary SDG(s) covered *(number)*	16, 17

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 to meet all criteria. As relevant, the rater will highlight best practice/added value elements and the level of complexity of the evaluation.*

Overall, the evaluation was well-designed, with good analysis that relies on a solid mix of triangulated primary and secondary data. The analysis is relatively sound, and thus the evaluation report has a good foundation. It is undermined by some key issues, however. Firstly, the absence of headline findings under each analysis area in the findings section is a gap - it is difficult for readers to parse the analysis and grasp what the analysis - otherwise sound - is pointing towards in many areas. While a sound conclusions section bridges this gap somewhat, it comes through again in the recommendations, which are not linked to specific findings or conclusions (or evaluation criteria), and some resultingly appear to not be rooted in the analysis. These are key issues in the report. Some key strengths of the evaluation are:
- The executive summary provides a clear and concise background to the evaluation and the programme.
- The purpose and objectives of the evaluation are clearly outlined and faithfully reflect the ToR.
- The scope of the evaluation is clearly delineated, covering programme duration and geographic reach.
- The reconstructed Theory of Change is generally well-constructed and logically articulated.
- Stakeholder roles, relationships, and contributions are extensively described with a useful stakeholder map.
- The evaluation design is good, with comprehensive primary and secondary data sources for analysis, ethical considerations explicitly addressed, gender-responsivity and inclusivity integrated.
- The analysis is well-laid out and seeks to systematically address all of the evaluation questions.
- The conclusions section provides a structured summary of findings aligned with the evaluation criteria.

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 relevant, resources will be cited to assist evaluation managers in overseeing future evaluations.*

As noted above, articulation of key findings and more explicit linkages to recommendations are significant areas for improvement. Other key areas to address are as follows:
- Key demographic and programme activity details are insufficiently described in the context and programme sections.
- Stakeholder information contains discrepancies between narrative and tabulated data, as well as in the stakeholder map.
- The sampling strategy description lacks clarity and contains conflated terms.
- Limitations are presented without sufficient mitigation strategies or detail.
- There is some inconsistency between presentation of evaluation questions vs. sub questions in the analysis section, leading to some missed analysis.
- Analysis by evaluation criterion overlaps, with themes like effectiveness discussed under relevance (although this is more an issue with the evaluation design).
- Gender-disaggregated data is limited (although mostly an external limitation), and vulnerable groups are not consistently detailed.
- The statistical analysis in the impact section is overly-technical and detailed - it should be summarised in the main report and much detail removed to the annexes.
- The report's language and formatting require some editing for clarity and consistency.
- Some recommendations are overly generic, duplicative, or insufficiently supported by analysis.

SECTION RATINGS			
SECTION A:	**EXECUTIVE SUMMARY (weight 5%)**	50%	Comments on Rating
Question 1.	Can the executive summary inform decision-making?		
i	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 relatively clearly presented, with the background to the evaluation and programme itself clear and concise. It is seven pages, which is quite long, but not excessively so. Notwithstanding some of the issues articulated below, it should be of use to decision makers and can serve as a standalone document.
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.	Partially	The summary has a good synopsis of the programme context and evaluation background, methods etc. It does not mention the intended audience for the evaluation, which should be included. It includes the evaluation conclusions and recommendations, but not the lessons learned, which are not well presented in the main report either. Some work is needed here.

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.	Partially	Aside from the context and background/approach to the evaluation, the executive summary essentially reproduces the evaluation conclusions and recommendations found at the end of the report. In this, it replicates one of the major issues with the report itself, which is the absence of articulation of key headline findings with the analysis. While the conclusions go some way to synthesising the analysis into a concise and coherent summary, this is a significant gap. The second issue is that the recommendations are poorly linked to either the analysis or conclusions - as with the main report, this is a feature of the summary. This issue should have been addressed to ensure recommendations adequately reflect the evidence and data gathered and the analysis undertaken. No new information is presented in the executive summary.
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.	Partially	Section 3.1 provides a description of the programme - largely via Table 1, which presents all of the relevant details such as the budget size, location, implementation status, implementing agencies (and other stakeholders), intended outcomes, components and key milestones (as well as dates). There is some superfluous information in this table (e.g. # of steering committee meetings is not relevant). Importantly, while the table is a concise way to present the key data, it lacks important details on the nature of the programme activities beyond the list of three programme components in the table. Some narrative description of what the key activities were (and who undertook them) would be useful in this section. Much of the description of the programme is covered in the Theory of Change subsection (particularly section 3.2.4) - this is not the optimal place for these details to be included and the report could have been organised more clearly. Finally, there appears to have been various changes to the foundational programme documents over time. This is alluded to in the introductory sections but not explained fully - the report should describe the when, what and why of such changes as they have clear implications on the ultimate evaluand.
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/cites, sub-regions…) and in terms of numbers reached, with disaggregation by gender, age, disability . . . (as appropriate to the purpose of the evaluation).	Partially	There is very brief description of the intended rightsholders targeted by the programme but not of duty bearers. Basic demographic information on population is presented in a table (table 2 - unreferenced) and supported by two maps (referenced) and the context section notes the proportion of Mozambicans that are not registered at birth. The Problem Statement section (3.2.1) does provide additional narrative information, although this is not linked to any specific data or statistics that quantify or specify the issues further. This section contains a considerable amount of information that could be moved to the context section.
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	The first section of the report is a description of the context. While it has much useful information, it would benefit from some additional inputs where some obvious gaps have been left. For example, the description states that "significant strides" have been made in birth registration in Mozambique, but does not note what these might be. Further, the drivers behind the limited extent of birth registration are not explained in the section. The issues are more clearly explained in the Problem Statement subsection under the Theory of Change description (3.2.1), but these should be in the previous section. Further, there is no description of the overall context of Mozambique in this section - its population, economy etc are not described - the single page only provides some brief details and then moves straight to a description of the intervention.
ii	Linkages drawn to the SDGs and relevant targets and indicators for the area being evaluated.	Yes	The context section clearly notes the SDG targets relevant to birth registration (SDG targets 16.9 and 17.19.02b) in the context section - both of which pertain to birth registration. This is fully clear. There is also a link to these and more indirect SDGs in Table 1 which describes the intervention.
iii	Clear and relevant description (where appropriate) of the status and needs of the rightsholders/beneficiaries of the intervention.	Partially	As discussed above, the report provides some, albeit brief, details on the status and needs of the rights holders in the programme, specifically the prevalence of birth registration (in the context section) and death registration (in the problem statement section). While the importance of birth and death registration is made clear in the problem statement section, this is more high-level/theoretical and does not provide any specifics around how the issues affect the population in Mozambique, and thus is minimally adequate.
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	The context and intervention description sections provide some information on the stakeholders to the programme, which is brief (Table 1 simply lists the stakeholders). However, there is an extensive description of stakeholder roles, and a diagrammatic stakeholder map (figure 3) provided in the section relating to the Theory of Change (section 3.2.4). Although not necessarily in the more appropriate section, this detail is useful and relevant. One issue is the membership of GITEV - the stakeholder map in Figure 3 includes only government ministries and an academic institution, whereas the narrative in many places in the report notes a much more extensive membership of government bodies, UN agencies, international NGOs and others - this should be clarified. A further (minor) discrepancy exists between the narrative in section 3.2.4 (pg. 24) which notes eleven local partners, whereas there are only eight "local" implementing partners listed in Table 1, of which two (SCF and WVI) are international NGOs, not local. This should be addressed.
ii	Identification of the specific contributions and roles of key stakeholders (financial or otherwise), including UNICEF.	Yes	Section 3.2.4 clearly and concisely describes the inputs, actors and partnerships of the key stakeholders to the programme, including the budget of the programme.
SECTION C:	**EVALUATION PURPOSE, OBJECTIVES AND SCOPE (weight 5%)**	75%	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	Section 4.1 clearly outlines the purpose of the evaluation and the rationale for conducting it at the time (as the CRVS programme drew to a close in 2023). This text is drawn verbatim from the TOR and thus is a faithful representation of the desired purpose by UNICEF. It also notes the key primary stakeholders who will use the evaluation, although secondary stakeholders (wider UNICEF/WHO parties, implementing agencies etc. are not noted, which is sometimes desirable, albeit more often when an evaluation is formative (i.e. there is learning that should be disseminated).
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	As with the purpose, the objectives of the evaluation are reproduced verbatim from the TOR, without change (although this is not noted). The objectives are sound, although it is noted that objectives 2 and 3 are forward-looking, and thus this is not a fully summative evaluation, as the TOR specifies, but has a significant formative component. This should have been noted at TOR design.
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	The scope is concisely described in section 4.2. The scope itself appears to be straightforward, covering the entirety of the programme duration (2016-2023) and the entirety of Mozambique, as noted in the beginning of the section. However, the latter half of the section notes some amendments to this scope, both geographically and temporally, due to a phased approach of the programme. While a little unclear, this is acceptable. Programmatic scope is not explicitly mentioned in the section, though elements of programme design and implementation are referred to. A clearer delineation of the programmatic scope should be made.
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.	Partially	The report presents the programme logic/result chain that was articulated by the implementers / designers in section 3.2.2. This is useful and relevant. However, there are two relatively small issues: a. The footnote on page 21 (#8) notes that the programme logic pertains to the "latest project design document" - ideally, the evaluation should use the design document that governed implementation of the programme rather than a final version - or if the final version is used, the rationale for this explained. Further, if there were various design changes in the programme, these should be mapped and explained as they are very relevant to the evaluation. b. Section 3.2.3 on pg. 23 notes that the evaluators only accessed the programme monitoring framework (which sets out the programme logic) later in the evaluation (June 2024 - the evaluation started in January 2024). The reasons for this issue (which is of some significance to the robustness of the evaluation) is not explained.
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	The evaluators present a reconstructed Theory of Change in Section 3.2 (although it is preceded by a considerable amount of description of the programme itself that should be moved to the preceding sections). There is a reference to issues with the development of the TOC noted in the preamble to section 3.2 (pg. 18) that should be unpacked, and text between this and page 26 (where the actual TOC is presented diagrammatically) moved to the preceding sections. The TOC itself is generally well-constructed, following a logical path from inputs to outcomes and the overall goal, with the problem statement articulated well.
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.	Partially	The TOC as presented in the relevant section was reconstructed by the evaluators as part of the preparation process. It was intended to have been done so with the participation of various programme stakeholders, but this process appears to have been somewhat poorly supported by these stakeholders (noted in the limitations section). Nonetheless the TOC as reconstructed was still utilised by the evaluators in the design of the evaluation (explicitly noted in section 4.3) and forms the basis of some of the analysis, notably 5.1.3 - programme design logic, but this is quite brief, and no other reference to the TOC is made in the findings, which is inconsistent with what is stated in 4.3.
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? *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	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	Firstly, the evaluation questions are not discussed or presented in the main report, indeed not referred to at all in the methods/design section. This is a significant omission. The EQs are, however, presented in the evaluation matrix (in the annexes), discussed below. Review of the EQs in the matrix indicates that while they are substantially in line with the evaluation objectives (and are clearly aligned with OECD/DAC evaluation criteria, which are specified throughout), they are somewhat duplicative and could have been reduced at inception or TOR design stage (18 evaluation questions is very high). For example, the coherence questions (#s 4-7) could have been condensed to two questions, and under effectiveness, questions 9 and 10 are very similar.
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.	Yes	Notwithstanding the large (and duplicative) number of EQs, the matrix in annex 8.2 breaks sub questions down into specific and individual sources of information/judgement criteria/indicators (only one of these terms would have sufficed). This analytical process is sound, and the indicators are allocated against data collection methods, also sound.
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	Section 4.3 provides a brief overview of the data collection methods. While concise (and with no details provided on the "how" of data collection), the combination of methods is generally adequate and robust. There are some issues, however: The difference between KIIs and SSIs is not clear. KIIs can be (and generally should be) semi-structured. This should be explained further if there is indeed a significant difference between the two approaches. Further, the "Statistical Models" noted as a data collection method is more accurately a component of the desk review. Statistical modelling is an analysis technique, not a data collection method. The underlying data is (presumably) secondary quantitative data from government sources. This should be explained better and included under the correct data collection method. Given that analysis of secondary quantitative data appears to be a significant pillar of the evaluation, the "desk review" method could have been split into qualitative (with the existing text) and quantitative (incorporating the statistical models information) for better clarity.

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	Overall, the sources of the primary and secondary data for this evaluation appear to be sound - it combines primary data from a good cross-section of duty-bearers involved in implementation of the programme with qualitative data from rights-holders to explore effectiveness and impact, with secondary quantitative data from microdata and published statistics. While some issues with how the methods are presented are noted above, the requisite detail on the sources is all present.
iii	Sampling strategy is provided - it should include a description of how diverse perspectives are captured (or, if not, provide reasons for this).	Partially	The sampling strategy (4.4) provides some useful description of the approach used, though there is some confusion of terms ("purposive snowballing" is a conflation of two techniques) and the selection of locations is described as "non-probabilistic" - which purposive sampling is, by definition. The text notes the use of a "prior sample definition" in (presumably) the selection of locations for inclusion in the evaluation. This definition is not provided, although the explanation is not clear so the evaluators may be referring to the criteria which follow. Either way, the text should be clarified. Further, the sampling criteria on pg. 29 includes participation in "the pilot phase". This is the first reference to such a phase. This should be explained in the programme context/description. These criteria are also not fully clear in terms of how they were used to select provinces. The first two are clear, but the latter four are simply values, and how they are used to select a province (cut-off values, a representative mix...?) is not clear, other than #5. This contrasts with the sampling criteria for districts within provinces (pg. 30, which are clearly applied) The table which presents the total number of individuals sampled is useful, though it needs a total # reached.
iv	Clear and complete description of the methods of analysis.	Partially	There is very good description of how the evaluators will treat the quantitative secondary data (statistical model, section 4.5). However, this is not matched with a discussion on how qualitative primary and secondary data will be compiled, coded, analysed etc. This is a significant omission.
v	Methodology allows for drawing causal connections between outputs and expected outcomes.	Yes	The evaluation uses multiple overlapping sources of primary/secondary; quantitative and qualitative sources of information, combined and cross-referenced (although this is not described - see above). It also explores the underlying aspects of the programme outputs which lead to outcomes, This is articulated in the reconstructed TOC and also reflected in the data collection tools (the KII, semi-structured interviews and the FGD tools) which include questions on the drivers and barriers to registration of life events that underpin programme outcomes. If applied and analysed as per the evaluation plan (for example, the analysis notes the use of formal statistical regression - which isolates individual variables and allows causal inference), this should lead to good capacity to draw causal connection between outputs and outcomes.
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	The report includes a limitations/challenges section (4.8) which includes the major limitations. However, mitigation strategies were not well-described (or were indeed absent). For example, the first limitation articulated in the section relates to poor validation of the reconstructed TOC - the intended workshop was poorly attended and did not achieve its aim. The evaluators did not appear to mitigate this, for example by having the TOC validated by the ERG (as is common practice) or by evaluation managers (in UNICEF/WHO). Thus a foundational underpinning of the evaluation is apparently significantly flawed. Other limitations are not clearly described, for example the limitation relating to the duration of the programme and availability of documentation actually combines two separate limitations: 1. Staff turnover and loss of institutional memory and relationships 2. Poor knowledge management processes inherent to the design and management of the programme. The section could have included mitigation strategy descriptions for each limitation.
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	The ethical considerations section (4.6) explicitly notes the UNEG guidelines for evaluations, as well as various UNICEF normative standards for ethics, although none of the documents are referenced via footnotes. There is some contextualised description of measures taken (trainings, briefings, etc.) but the UNEG obligations are simply listed.
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	While the section is brief (and some of the references to the normative standards perfunctory, as noted above), there is good description of specific processes and safeguards followed to ensure an ethical approach. This includes ensuring of appropriate consent from participating rights-holders and others, the use of an external IRB, referral mechanisms in case of confidential disclosures and separation of male and female respondents to ensure good representation of the perspectives of both. No children under 17 years old were included in the research.
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 innovative practice used in the evaluation.
SECTION E:	**EVALUATION FINDINGS (weight 25%)**	50%	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	While the report does roughly follow the proposed analytical approach (i.e. providing the analysis under the evaluation criteria), the findings section has a somewhat inconsistent structure vis a vis the evaluation matrix, which should normally be explicitly and systematically followed. For example, the Relevance section starts with a subheading only partially related to the evaluation question, and the organisation of the section is not as per the matrix. The report follows the evaluation questions systematically for the next section (coherence), but uses sub-questions as headings for the next (effectiveness). Efficiency does not appear to follow the proposed analytical approach. Both the Sustainability and Impact sections mix evaluation questions and sub questions. In particular, for Efficiency, the report does a poor job of analysis, in that it focuses on programme spending. The fact that 99.7% of programme funding was spent is not a measure of efficiency. Evaluation of efficiency should look at what that spending was on - did it offer value for money ("were there cheaper alternatives" per the analytical framework). A key theme in the analysis is that costs of implementing activities (e.g. the eCVRS and the SMS pilot) or maintaining equipment on an ongoing basis are fundamental impediments - this should be reflected in the efficiency and sustainability sections. Being less than systematic means that either (a) some of the 'bigger picture' elements present in the overarching evaluation questions are missed, and/or (b) specific analysis to be captured via the sub-questions is missed. Despite this, the evaluators have made clear efforts to cover most of the ground in the (many) evaluation questions.
ii	Explicit use of the intervention's results framework/ToC in the formulation of the findings.	Yes	The report analyses the overall design of the programme via the originally stated logic and the reconstructed TOC - a high-level analysis of alignment of the design with needs is presented under the relevance criterion (section 5.1.3). This is positive. The evaluation also explicitly maps achievements against the intended programme outcomes (1 & 2) under the effectiveness criterion, thus looking at achievements against the intentions of the implementers. This is explicitly linked to the reconstructed theory of change in section 5.3.4, meeting this criterion well.
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.	Partially	Within the context of the report itself, it does seek to present qualitative and quantitative evidence against the evaluation questions/sub-questions for the most part and triangulation of data sources is evident, but in places the analysis appears to miss the theme intended by the evaluation framework. For example, there is some overlap in the analysis by evaluation criterion under Relevance (EQ1), where there is discussion of the effectiveness and perceived utility of training/capacity building and of the benefits of the programme to the population - this is effectiveness. Relevance should explore the extent to which such a need existed, by reference to both primary and secondary data. Similarly, under the relevance of the programme logic (5.1.3) the analysis presented (in the table) is not related to the sub question, which is about the internal logic of the programme design, not how it was adapted to the external context. To some extent, this is an issue with the evaluation analytical design. As noted above, duplicative evaluation questions mean that there is significant overlap between sections, for example a lot of the Relevance and Coherence sections discuss effectiveness or efficiency issues. This is partly design, but also the evaluators have not been sufficiently focused on the overall theme of each section, e.g. the table under 5.2.2 should present issues related to coherence/coordination, but it is largely a list of operational gaps and recommendations. Some good analysis relating to this aspect is presented in the following subsection (5.2.4).
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	Firstly, there is a considerable amount of analysis and evidence presented in the report from both primary and secondary sources, including data from interviews with duty-bearers, FGDs with rights-holders, analysis of published documents and data. This is a solid basis for the evaluation. However, the major weakness of the report is that it does not present specific and clear findings per se. Each subheading includes a block of analysis text under it, and readers must interpret this to determine the overall finding themselves, which is not standard practice in evaluations - and makes the report challenging to read. For example, under Coherence with other initiatives (5.2.1), the report discusses various coordination bodies, instruments and instances noted in primary/secondary data, and notes challenges with each but does not synthesise this into a clear finding, so the reader is not sure what to make of the analysis. Other good examples are in Effectiveness and Impact. For effectiveness, there is considerable discussion of the achievements, challenges and underperforming areas of the programme, but is not synthesised into a series of findings that are supported by the discussion. The reader must draw their own conclusions. Similarly for Impact, a complex series of statistical analyses are presented, with little in the way of interpretation - much of this could be removed to an annex, with the headline findings/analysis based on this presented in a straightforward manner. Specific (and ideally numbered) findings before each sub-section, supported appropriately by the evidence, could have been included here.
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).	Partially	Typically, the evaluators seek to ensure that the analysis discusses specific areas of strong or weak performance by the programme and assign this to underlying causes. For example, a significant challenge to programme effectiveness that is explored is the lack of decentralised authority among the government which means that local/provincial level authorities do not take the initiative on implementation decisions. One aspect that was not explored was the drivers of registration amongst rights-holders - an underlying causal factor of programme success. For example, a significant impediment to birth registration globally is indirect costs to rights holders (e.g. travel and time costs) - this could be analysed further in the efficiency section. The progression of the programme from implementation to results is explored via the programme logical frameworks and the reconstructed theory of change in the effectiveness section.
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.	No	The report notes the programme monitoring framework in the introduction and undertakes some analysis of this in the Effectiveness section (5.3). This involves some level of review of the M&E system (noting gaps in the development of the Performance Management Framework), but what is not made adequately clear is the evolution of the Performance Management Framework and related system. The evaluators note that the "latest" version of the Performance Management Framework (dating June 2024) was used for analysis, but what about earlier versions? There is something missing here - if the gap is on the part of the programme, that needs to be made clearer. As written, it appears that the gap is in the evaluation analysis. Beyond this, there is no specific analysis of the monitoring system of the programme, such as an analysis of the logic, appropriateness of indicators or data collection tools/processes.
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	The conclusions section is well laid out according to the evaluation criteria. It presents a good summary of findings under the subheadings. The conclusions are clearly stated, and indeed are reflective of the purpose of the evaluation and can be used for future iterations of the programme or help in the design of future support to CVRS in Mozambique, as they usefully lay out the opportunities, strengths and likely challenges.
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	The conclusions in a way address one of the key issues of the main analysis section of the report - the absence of higher-level findings that summarise the analysis. Their absence is one of the key flaws of the report. The conclusions section is therefore where the analysis is synthesised and can be grasped - but it should not be the case that the section is simply a summary of findings.
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.	Partially	There is a short lessons learned section, which does includes some useful lessons and good practices (although the heading is not correctly formatted). There are definitely a range of other lessons in the main analysis which have not been reflected in this section, however. For example, under Unintended Consequences (5.6.3) an interesting lesson on aligning campaigns with local schedules (specifically local harvest times in the example) is noted, but not in the lessons-learned. Also, the issue of internet connectivity in remote areas is an important lesson that should have been highlighted and teased out.
ii	Lessons are clearly and concisely presented, yet have sufficient detail to be useful for intended audience.	Yes	Insofar as the lessons in the section are presented, they are indeed clear, concise and well-presented, so they are useful. As noted above, however, more lessons could have been extracted from the analysis and presented here.
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.	No	As there are no specific and itemised findings in the report, it is challenging to link the recommendations to specific findings. Further, they are not organised according to the evaluation framework (relevance, coherence etc.) as are the analysis and the conclusions. They do, however, align with the overall report purpose reasonably well, but some of the recommendations do not appear to have significant analysis underpinning them. For example, Recommendation 6, in relation to elimination of birth registration costs does not appear to be derived from any analysis in the relevant section. There is only a single mention of the costs involved in registration, and this is mostly in relation to transportation costs of people to a registration point or the opportunity cost of time taken for registration - the report makes no other mention of "birth registration fees" other than this recommendation. The recommendations should have been reorganised to be more directly related to the conclusions and the evaluators should have ensured that any recommendation is underpinned by robust analysis in the finding section.
ii	Recommendations are useful and actionable for primary intended users and uses (relevant to the intervention); guidance is given for implementation, as appropriate.	Partially	The recommendations are somewhat useful, insofar as they present useful guidance for the implementers and funders, with some specific action points presented in narrative form under each one. However, some of the recommendations are quite generic and somewhat duplicative, for example, recommendations 1 and 2 are essentially the same - improving coordination. They could easily be combined and have specific action points related to GITEV (recommendation 1) and wider interagency coordination (recommendation 2). Recommendation 4 (improvement of M&E) is noted as strategic, but should be operational (or both) and could go a lot further in terms of specific action points - the narrative is very high-level and general.
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).	Partially	The preamble to the recommendations section notes that the process for their development was based on "input from respondents, the evaluation team's overall analysis, and feedback from the Evaluation Reference Group during a validation workshop held in August 2024". No further details were provided regarding this (e.g. the specific attendees, presence/absence of rights-holders or duty-bearers). This is therefore minimally adequate to meet the criterion.
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	Each recommendation includes identification of the specific duty-bearers to which they are targeted, and a timeline for implementation (immediate, short, medium or long-term), although the difference between "immediate" and "short-term" is not clear. They are not prioritised according to importance, and the vagueness of the discussion under each means that it is not clear who is responsible for what actions under the recommendations.
SECTION H:	**REPORT STRUCTURE AND PRESENTATION (weight 5%)**	64%	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.	Partially	Some of the required elements are present in the opening pages, but it misses the date range of the evaluation and the page numbering begins on the cover page (this is not normal practice). Further, the Table of Contents has incorrect page numbering in the table of tables - the table need to be updated and carefully checked. The table of figures has some formatting errors.
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.	Yes	There is a full suite of annexes with the report. Some additional analysis (i.e. the regression analyses used for the impact assessment and financial data) are presented as annexes rather than in the main report, which is positive - but more of the impact analysis should have been excerpted to the annex and a more simplified explanation provided in the main report. All other elements are present.
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).	Partially	The report is well-organised, with sections/subsections organised according to the standard template, and with the analysis largely according to the evaluation questions and response areas. There is a useful meta-summary of the contents of each section presented as a short preamble to each. The formatting needs some work, however - sections (Relevance, Effectiveness etc.) should each start on a fresh page. The Lessons Learned section has no heading (bottom of page 73).
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 of the report is mostly in accordance with this formulation, with the exception of the lessons learned being presented before the conclusions.
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	The report is on the long side (68 pages excluding the executive summary), but not unexpected due to the considerable amount of analysis required to answer all of the evaluation questions. Note that the TOR specifies 60 pages. Some of the programme description detail (of which a considerable amount is in the Theory of Change subsection - see above) could be reorganised to shorten the length, and as noted above, some of the details in the Impact section could be removed to the annexes.
ii	Report is easy to understand (written in accessible way for intended audience) and generally free from grammar, spelling and punctuation errors.	Partially	The formatting and writing are somewhat but not fully clear, with a variety of formulations that should have been revisited and there are a number of grammatical/syntax errors - the report was not edited for language and formatting in line with UN/UNICEF style guidelines. Some specific issues noted are (among others): Numbers should start after the cover page (no number on cover page). The list of tables has incorrect numbering of pages. Some of the language/writing style is not clear, starting with the very first sentence of the full report on page 15 which does not make sense. Mix of UK/US spellings ("program" vs. "programme"). Inconsistency of some terms/spelling e.g. COVID vs COVID-19. Some acronyms not spelled out at first usage (e.g. UNLIA). Some instances of unexplained jargon ("telessaude platform", "mobile brigade").
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	The tables, charts and figures are useful in most cases and add to the quality of the report. They are also clearly labelled and referenced in the text itself and via the table of contents. There are a couple of instances where the tables/charts are not clear or have errors in text - these are highlighted via in-line comments.
SECTION I:	**EVALUATION PRINCIPLES (weight 10%)**	69%	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.	Yes	The report explicitly notes gender-responsivity and inclusiveness as being integral elements of the evaluation design. Section 4.7 describes this, with clear reference to CEDAW and the CRC in design.
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 - section 4.7 notes that it had "diverse membership", including "UN agencies and ministries". A list of the full membership should be included in the annexes.
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 some mention of rights-holders in section 4.7 (on inclusivity), but otherwise the evaluation refers to "beneficiaries". Similarly, duty-bearers are referred to in section 4.7, 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 any gender dimension of birth registration and includes particular references to disabilities. There was, some, but limited gender-disaggregated (secondary) data available to the evaluators, which is noted as a limitation, but qualitative analysis did include gender dimensions.
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	There are various references to vulnerable groups in the evaluation design and programme description, however, they are not clearly listed in the design elements - there are references to women, girls, IDPs, PWDs, "people with low literacy" in different places but no specific discussion on them. Some greater details are required on these, particularly if they were to have participated in data collection (as noted in the design section - 4.2) and have been targeted by the programme as is claimed. This said, there is some reference to people with disabilities and gaps are briefly noted. There is a recommendation (#7) specifically focused on the rights-based approach, including disability, although there is little in the way of corresponding analysis to underpin this recommendation.
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)	7	
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	GEEW forms a robust part of the evaluation scope, as part of the aim of the overall programme was to improve gender-responsiveness of the CRVS system. The scope of the evaluation includes "a particular focus on equity/inclusion and gender equality" (section 4.2), with gender dimensions cutting through the analytical approach (specifically evaluation questions 4 and 8).
ii	A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Satisfactorily integrated	The evaluation includes a specific section (4.7) on measures to ensure GEEW and inclusivity in the evaluation (as was mandated by the TOR). This included description of the incorporation of explicit gender considerations into the reconstructed TOC and research tools and implementation of a gender-sensitive approach in data collection. Gender considerations were somewhat integrated into tools and analysis - each tool (KII, SSI, FGD) includes either explicit questions or probes that explore gender dimensions of CVRS - while not a comprehensive approach, it is adequate to provide some useful GEEW data - the evaluation is more gender sensitive than responsive.

iii	The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.	Satisfactorily integrated	The report analysis does include due regard to gender aspects around the programme, and around birth registration in Mozambique, discussing some of the achievements and gaps it exhibits. This is well reflected across several of the evaluation areas. The conclusions are somewhat more optimistic regarding GEEW (noting that the programme had a "gender-transformative approach") than the analysis (which suggests it was more gender-sensitive than anything else), but the specific recommendation on gender and inclusivity is warranted and positive to note.

SWAP Rating Guidance

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.
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?
b. Does the evaluation include an objective specific to assessment of human rights and gender equality considerations or was it mainstreamed in other objectives?
c. Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria?
d. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation?

ii A gender-responsive methodology, methods and tools, and data analysis techniques are selected.
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?
b. Does the evaluation methodology employ a mixed-methods approach, appropriate to evaluating GEWE considerations?
c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility?
d. Does the evaluation methods and sampling frame address the diversity of stakeholders affected by the intervention, particularly the most vulnerable, where appropriate?
e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality?

iii The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.
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?
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?
c. Are unanticipated effects of the intervention on human rights and gender equality described?
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?