


**IMPORTANT NOTICE**

Please do NOT express views on the proposals under evaluation or the experts' opinions on the proposals.

 If you notice shortcomings, please do NOT directly address the experts but liaise with the Commission/Agency staff involved in the evaluation (for instance, to suggest immediate and short term improvements).



**SENSITIVE**

## OBSERVER REPORT

CALL	
Call:	HORIZON-JU-GH-EDCTP3-2025-01-two-stage
Topic(s):	<p>HORIZON-JU-GH-EDCTP3-2025-01-TB-01-two-stage Global collaboration action for the development of vaccines for reducing the disease burden of Tuberculosis in sub-Saharan Africa</p> <p>HORIZON-JU-GH-EDCTP3-2025-01-MALARIA-02-two-stage Global collaboration action for research on existing Malaria therapeutics and clinical development of new antimalarial candidates</p> <p>HORIZON-JU-GH-EDCTP3-2025-01-NTD-03-two-stage Accelerating the development of prophylactic vaccines against Neglected Tropical Diseases (NTDs) in sub-Saharan Africa</p>
Type(s) of action:	HORIZON JU Research and Innovation Actions (RIA)
Service:	Global Health EDCTP3
Call deadline:	03/09/2025
Submission model:	stage 2

EVALUATION	
Evaluation model:	single
Panel(s):	HORIZON-JU-GH-EDCTP3-2025-01-TB-01-two-stage, HORIZON-JU-GH-EDCTP3-2025-01-MALARIA-02-two-stage, HORIZON-JU-GH-EDCTP3-2025-01-NTD-03-two-stage
Observer(s):	██████████

## TABLE OF CONTENTS

SUMMARY FOR PUBLICATION .....	3
1. BACKGROUND AND SCOPE .....	3
2. OBSERVER ASSESSMENT .....	4

## SUMMARY FOR PUBLICATION

### Summary for publication

*Include a summary of the main findings for publication in a call update in the Portal, including the overall assessment on the conduct and fairness of the evaluation sessions, and compliance with the applicable rules. (max 2000 characters)*

 The content of this section will be published in the Portal as a call update.

One Independent Observer assisted EDCTP3 in the evaluation of proposals submitted to 3 topics (TB, MALARIA, NTD) within the second stage of the two-stage call HORIZON-JU-GH-EDCTP3-2025-01-two-stage. They followed the briefings and Consensus meetings in order to assess and report on the implementation of the evaluation procedures, on the conduct and fairness of the evaluation process and on the application of the evaluation criteria. Based on the analysis conducted, the Observer gives independent advice for improvement of the evaluation process.

The evaluation process was fully transparent. Rules and guidelines were clearly communicated by documents provided to Experts, online and on-site briefings and instructions given and, where necessary, repeated by the Moderators during the Consensus meetings. This ensured a fair and transparent procedure. The evaluation was conducted in a fair and professional way, thanks to a thorough preparation and the helpful and competent staff involved, including the Quality Checkers and the assistant team.

The Independent Observer was impressed by the high quality of the evaluation, and made some recommendations for improving and optimally streamlining the process. The most important advice, as in the stage 1, is to improve the focus on excellence and innovative aspects and impacts, rather than focusing on shortcomings. Such a change would be in line with other international procedures and is likely to increase the quality and impact of the successful proposals.

## 1. BACKGROUND AND SCOPE

### Background and scope

This report describes the observer's assessment of the evaluation of the following call:

**Call for proposals:** HORIZON-JU-GH-EDCTP3-2025-01-two-stage stage 2

**Deadline:** 03.09.2025

**Budget:** EUR 122,700,000

This call covers the following topic(s)/type(s) of action:

HORIZON-JU-GH-EDCTP3-2025-01-TB-01-two-stage/HORIZON JU Research and Innovation Actions RIA

Global collaboration action for the development of vaccines for reducing the disease burden of Tuberculosis in sub-Saharan Africa

HORIZON-JU-GH-EDCTP3-2025-01-MALARIA-02-two-stage/HORIZON JU Research and Innovation Actions RIA

Global collaboration action for research on existing Malaria therapeutics and clinical development of new antimalarial candidates

HORIZON-JU-GH-EDCTP3-2025-01-NTD-03-two-stage//HORIZON JU Research and Innovation Actions RIA

Accelerating the development of prophylactic vaccines against Neglected Tropical Diseases (NTDs) in sub-Saharan Africa

The report analyses the efficiency of the procedures, usability of the instruments (including IT tools), conduct and fairness of the evaluation sessions, and compliance with the applicable rules.

The objective is to give independent advice for improving the evaluation processes for EU funding.

## 2. OBSERVER ASSESSMENT

### Methodology

#### Methodology

*Describe how you proceeded for observing the evaluation procedure (e.g. participation in briefing; present at evaluation session; analysis of IERs, CRs and panel report; comparison with similar procedures at national/international level, etc)*

As Independent Observer, I have followed the evaluation in order to assess and report on the implementation of the evaluation procedures, the conduct and fairness of the evaluation process and the application of the evaluation criteria. Based on the analysis conducted, the Observer gives independent advice for improving the evaluation process.

As methodology for observing the evaluation procedure, I have attended the preliminary briefings and received relevant information and documents before the starting of the evaluation, and I have attended the consensus meetings. I could partially follow only two of the topic-specific briefings (TB and NTD) at the beginning of the Consensus meetings, since the briefings for the three topics I observed were held all at the same time and because of some connection problems to one of the topics (MALARIA). I have followed the quality checking and the finalization of the CR text.

### Assessment

#### Assessment

##### Scale of complexity of the evaluation task

The evaluation task was well managed despite its complexity. The briefing for Experts held at the beginning of the evaluation was clear and informative. The number of proposals to be evaluated in each topic was relatively small (6 for TB, 9 for MALARIA and 12 for NTD), allowing for sufficient time for a thorough and constructive discussion during the Consensus meetings. The meetings were scheduled along a precise agenda, provided in advance to the participants, with a clear allocation of Experts' involvement and timing. The agenda greatly helped in the overall meeting organisation. Changes, mainly due to unforeseen needs of Experts and Recorders, were managed by appropriately re-scheduling discussions. Some confusion rarely arose from repeated agendas distributed at short distance one from the other, but rapidly solved.

The initial briefing aimed at reminding Experts of the actions to be taken during the Consensus meeting and at proving the indications for scoring, in particular the need of matching scores with the comments made in the CR. The general procedure for the discussion included a brief presentation of the proposal by the Recorder, who then started reading the text of each criterion, drafted on the basis of the IERs, drawing attention on the discrepancies in the Experts' opinions so as to discuss and find consensus. For the second stage, all three criteria were evaluated, i.e., Excellence, Impact and Implementation. Recorders kept note of all the comments and issues raised during the discussion and corrected/implemented the CR text and obtained the Experts' approval. Some Recorders (e.g., TB) did it during the Consensus meeting, in the presence of all Experts, so as to immediately obtain the approval, while in the other meeting this was done after the discussion. The CRs were then sent to the Quality Checkers, who made sure the text was in line with the guidelines and rules and, importantly, that all the CRs were harmonized and aligned in structure and evaluation criteria. In some cases, quality

checking resulted in a few rounds of re-writing and re-wording, with the draft CR sent back to the Experts for revision, before final approval.

Despite the limited number of proposals to be assessed, the Experts, Recorders and Quality Checkers worked full-time during the meetings, to ensure the best possible evaluation and the most complete and fair reports. Overall, the evaluation process worked well.

#### Transparency of the procedures

The evaluation process was fully transparent. The rules and guidelines to be followed were clearly communicated by documents provided to Experts, by online and on-site briefings and by instructions given initially by the moderators and often repeated during the discussions. All discussions took place online. The overall process took place along a fair and transparent procedure, with Moderators trying to include all involved experts in the discussion.

#### Throughput time of the evaluation and the efficiency of the procedures

At variance with Stage 1, none of the Experts complained about the time allocated to individual proposal evaluation and CR preparation.

During the Consensus meetings, time management was in general unproblematic, without serious deviations. The overall evaluation planning was distributed before the start of the meetings and adjusted during the meetings to meet the needs of Experts and Recorders.

Some proposals required additional time for discussion in order to reach consensus or because of conflicting opinions, but these were exceptions. No minority opinions had to be recorded.

Overall, the evaluation procedures were straightforward and efficient.

The style of discussion conduction was very different in the three panels I have followed. In TB, the Chair (Project Officer) directly moderated the discussion and very efficiently kept the time and managed to obtain the opinion of all involved Experts, and was supported by an excellent Recorder who prepared the shared CR during the discussion. In NTD, the moderation was initially left to the Recorders, who introduced the CR and asked Experts to intervene in the discussion, but without making sure that all involved Experts were included. It turned out that several discussions were almost monopolized by the most talkative and provocative Experts, and the Chair did not limit or restrict them (a choice that I frankly share, provided all others have the opportunity to speak their opinion). However, the Chair would intervene when no consensus was easily reached, to avoid full monopolisation by few experts, and when the discussion was at risk of derailing. In MALARIA, the conduction was almost entirely left to Recorders, with the Chair just making sure the Recorder had all the necessary information for implementing the CR. Despite the differences in the procedure (complete freedom to Experts vs. efficient steering of the discussion), eventually all three panels completed their work within the allocated time without major discrepancies.

#### Efficiency, reliability and usability of the procedures, including the IT-tools

The presence of Recorders greatly helped in speeding the CR finalisation, and the interaction between Chair and Recorder is essential in steering and streamlining the discussion. The majority of Recorders was very good, some exceptionally good. It should be noted that the draft reports produced by the Recorders are most useful and time-saving if they are close to a complete text, in which alternative phrasings are included only in places where Experts have expressed conflicting opinions. This approach reduces throughput time, as it supports clearly focused discussions.

All Experts were asked to be present at all times during the Consensus meetings, even when not directly involved in the discussion. This is excellent because Experts can greatly benefit from the discussion of other proposals and have less difficulties in streamlining CRs and scores. However, we have again observed a behaviour that needs to be corrected, i.e., some Experts offering strong opinions on proposals that they had not read. Non-reviewing Experts are welcome to ask relevant questions and provide relevant information related to their particular expertise, but should not actively offer opinions and suggest scoring. Without limiting the discussion, Chairs should find the way of reducing such interventions.

Although highlighted in briefings and infographic, the procedure for CR modification by the Recorder (including the QC comments) and following check and approval/further comments by Experts was unclear to several experts, leading to lengthy explanations, especially to those recorders and evaluators that were new to the process. This highlights some unclarity in the guidelines, in particular relative to the practical steps taking place in SEP (e.g., email communication of new task). Teams worked fine for online meetings. In the context of IT tools, Experts were again reminded that the use of generative AI tools in producing an evaluation report is not allowed.

### Impartiality, fairness and confidentiality of the evaluation

The evaluations were conducted with utmost fairness, allowing for extensive discussions on all proposals. Most of the proposals were, as expected in stage 2, of high level. Thus, discussion was thorough despite the general positive opinion of Experts, in order to be able to accurately rank the proposals eligible for funding.

For all topics, impartiality was easily ensured by the fact that all proposals in a topic were collegially discussed in a single panel, an excellent practice that should be maintained whenever possible.

No particular problem for confidentiality was identified. We have never observed a single instance where a non-impartial behaviour could be suspected. Experts with real or possible CoIs readily left the panel when requested.

### Conformity of the evaluation with the applicable rules (including guidance documents)

Guidance documents were provided to the Experts, and important points were repeated by the Chairs during the meetings. Although guidance was clear, interpretation was not always easy. Some challenge was again posed in defining the opinions provided in the CRs according to the score descriptors and the definitions of "minor shortcoming" and "shortcoming". It has been observed that, in several cases, Experts had difficulties in grasping the meaning and difference between "minor shortcoming" and "shortcoming", and in assessing how many of these you should count in order to decrease the score and how much should the score be decreased. Since this is actually open to interpretation and depends on the context, although not easy, the Chairs did a good job in explaining that scoring is not mathematically calculated based on the number of shortcoming and minor shortcomings, but that Experts should carefully assess how much these may actually impact on the feasibility and relevance of the project.

Another issue that may need correction is the fact that some Experts included in their IER recommendations for improving the proposed study. Although I may agree with the Experts that offering recommendations for improvement is a useful action, which could benefit the quality of the successful projects, the current rule is that recommendations should not be given, and the CRs should only impersonally state the evaluation of the proposal as it was written, in order to ensure impartiality. Therefore, in the current evaluation it was left to the Recorders' skills, when preparing the CR, to re-phase the experts' comments so as to eliminate recommendations and provide a formally impartial evaluation.

### Quality of the evaluation process in comparison with similar national/international evaluation procedures

The higher value assigned to Excellence and Impact criteria is good and in line with other international evaluation processes.

The evaluation process, as I have observed, placed emphasis on identifying problems, which is common practice. Many other international evaluation processes also put a strong emphasis on identifying outstanding strengths, which was less emphasized in the present process, as in all Horizon Europe evaluation processes. We again noted that CRs rarely contain descriptors like outstanding, exciting or ground-breaking, which are used elsewhere to identify special strengths and to take them fully into account for reaching a score.

### Overall quality of the evaluation

The evaluation quality was very good. All proposals were assessed in a fair and transparent way, with extensive discussions on all aspects. Evaluators were in general all very expert in their areas, and always willing to discuss and compare opinions in order to find consensus. Chairs and Recorders were able to direct the discussions very well. The general pitfall, as already mentioned above, is the tendency to pay more attention on identifying shortcomings and weaknesses rather than to the positive, innovative and exciting aspects. This may lead to score as "excellent" proposals that are formally flawless according to the listed evaluation criteria, although not providing any really outstanding ideas/outputs.

The evaluation was conducted in a highly fair and professional way, thanks to a thorough preparation and to the helpful and competent EDCTP3 staff involved, including the Quality Checkers and the assistant team.



**Other remarks** *(optional)*

The general online briefings and the on-site briefings were informative and useful. Nevertheless, Chairs should reiterate information as often as necessary during the Consensus meetings, to make sure that all Experts are fully aware of the applicable rules. This may have to be done repeatedly, in particular when considering that Experts may be unfamiliar with the many evaluation rules that must be applied.

Several Experts had problems understanding how to score, based on the definition and number of shortcomings and weaknesses and their impact on the project development and, consequently, on the scoring. The current Horizon Europe guidance is not prescriptive, thereby giving flexibility of interpretation. However, this may raise doubts in some experts, or generate different interpretations between experts. This should be changed in the next FP, to eliminate doubts, and to better highlight outstanding proposals.

Allocation of Experts was very well balanced in all aspects.

A particular case happened in the NTD panel, when the Chair became aware that several of the proposals included pre-clinical work, which is not allowed as indicated in the call text. This issue could not be spotted at Stage 1, since the Implementation details were not requested. Whether the call text was not sufficiently clear or whether the applicants tried anyway since there are no other funding opportunities for pre-clinical work is difficult to say. This problem was serious, because it regarded a substantial number of proposals, and resulted in a prolonged and sometimes heated discussion about whether these proposals should be all punished in the Excellence score because not adhering to the call requirements (considering this as a weakness and therefore putting the Excellence score below threshold, as indicated in the HE guidelines) or retained because of their truly excellent and innovative content. Both Experts and EDCTP3 officers sorely regretted to decrease the Excellence score of these proposals, but eventually all experts agreed that not respecting the call requirements should be considered a weakness. The Chair let the Experts discuss the issue without restrictions until they eventually reached the conclusion. They kept their decision even when given the chance to change it in the panel meeting, through benchmarking of the same issue.

**Recommendations****Recommendations**

1. **Engage all Experts in the discussion but prevent Experts who did not evaluate a given proposal to offer their opinion.** The non-evaluating Experts can ask questions and provide useful information, but should not give opinions or suggest scores. The Chairs can make sure to engage all the involved Experts and control the intervention of the non-involved ones.
2. **Discuss and amend the CR text in the meeting with the Experts.** It would be excellent if CR drafts discussed during the Consensus meetings could be corrected and implemented together with the Experts, so that they are practically ready for Quality Checking at the end of the discussion. This was done in one of the panels I have followed and worked beautifully, saving a lot of time. The discussion time was more than sufficient for doing so. The expertise of Chairs and recorders should be sufficient for obtaining a CR with comments that fully match the scores.
3. **Implement a thorough guidance of Experts on understanding the meaning and impact of "shortcomings".** This can be done by Chairs during the Consensus meetings, as the interpretation may vary depending on the context, which cannot be included in the guidance documents.
4. **Give more focus to the strengths of proposals.** It would also be good to find the way of prioritizing proposals that are particularly exciting/meaningful over those that are formally flawless but less innovative.
5. **Provide clearer/stronger information regarding the non-inclusion of pre-clinical studies.** The fact that several of the 12 NTD proposals in stage 2 had this problem may underlie lack of clarity in the call text, which should be made stronger next time.