

---

OCB META EVALUATION OF

---

SEU-MANAGED  
EVALUATIONS  
2017-2021  
*SUMMARY REPORT*

---

DECEMBER 2022

This publication was produced at the request of MSF-OCB under the management of the Stockholm Evaluation Unit. It was prepared independently by [Michael Harnar](#), [Zach Tilton](#) and [Tian K. Ford](#).

This summary report has been prepared by the SEU.

DISCLAIMER

The authors' views expressed in this publication do not necessarily reflect the views of Médecins sans Frontières and the Stockholm Evaluation Unit.

## ACRONYMS

MSF	Médecins Sans Frontières
OCB	Operational Centre Brussels
SEU	Stockholm Evaluation Unit
SC	(SEU) Steering Committee
PrgES	Program Evaluation Standards
ALNAP	Active Learning Network for Accountability and Performance
EMQF	SEU Evaluation Manifesto Quality Framework
CP	Checkpoint
TOR	Terms of Reference
IR	Inception Report
FR	Final Report
MR	Management Response
MEQ	Meta-evaluation Question
A8CP1	An example of a reference point within the PrgES checklist. The first letter, “A”, refers to the PrgES criteria, and “8” references the relevant sub-criteria. “CP1” refers to the first checkpoint of the sub-criteria.
AEA	American Evaluation Association
EHA	Evaluating Humanitarian Action

The following report presents findings, conclusions, and recommendations from an external independent meta-evaluation study commissioned by Médecins Sans Frontières (MSF) Operational Centre Brussels (OCB) and conducted from May to October 2022.

## PURPOSE

The goal of this meta-evaluation was to assess the quality and value of past evaluations and to produce useful processes and products to “maximiz[e] the strategic value of requesting, conducting and making use of [future] evaluations” at OCB.

### META-EVALUATION OBJECTIVES

1. Assess quality and value of past evaluations.
2. Establish understanding of evaluation quality at OCB.
3. Determine organizational factors of evaluation quality.

### META-EVALUATION QUESTIONS

1. How does the SEU and OCB define evaluation quality?
2. What was the quality of past evaluation performance?
3. What was the value of past evaluation performance?
4. What factors determined the quality of past evaluation performance?

## SCOPE

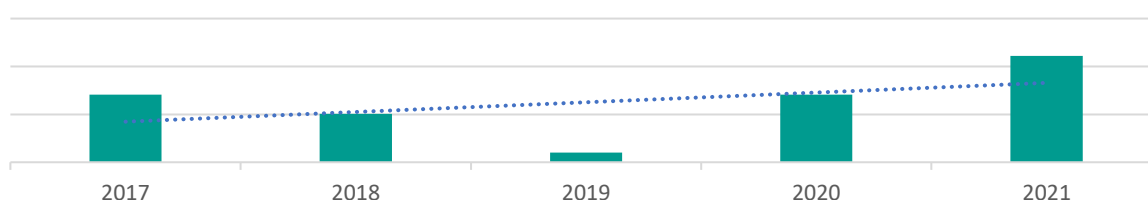
### OBJECT

The primary object of this meta-evaluation was the portfolio of 31 evaluations managed by the SEU from 2017-2021; the secondary object of this meta-evaluation was the evaluation system at OCB.

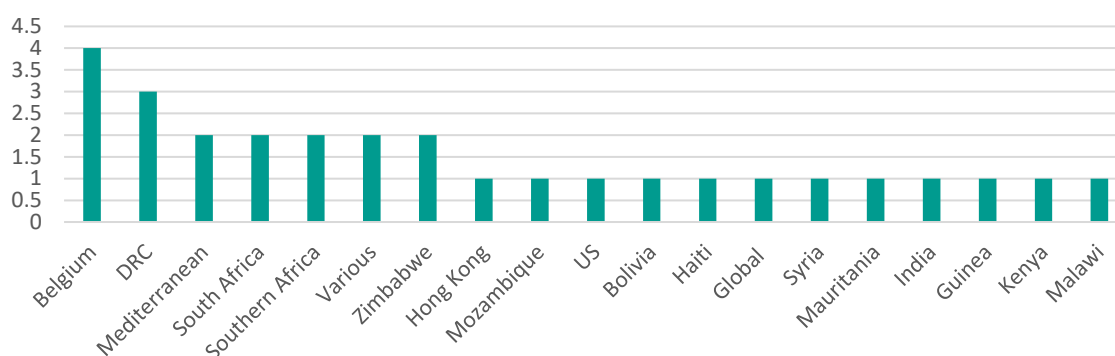
### THE EVALUATIONS

For this meta-evaluation, 31 individual evaluations, conducted between 2017 and 2021, were assessed. Below are graphs detailing some of the main characteristics of the total evaluation portfolio.

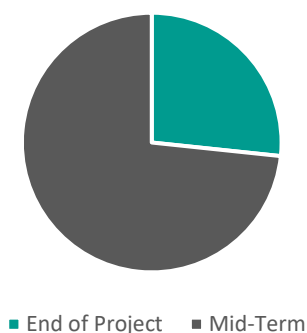
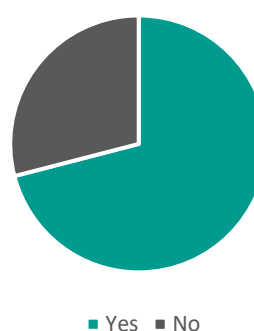
Total by year



By Country



End of project vs Mid-term

Operational project  
(vs non-operational project)

## OCB'S EVALUATION SYSTEM

The Stockholm Evaluation Unit (SEU), based in Sweden, is an operational support unit tasked to manage and guide evaluations of MSF's operational projects, with a focus on OCB. It is housed within OCB's Operations Support Team within the Operations Department. The functional management of the SEU is overseen by the SEU Steering Committee (SC) that ensures SEU independence and accountability for SEU quality of work and progress. At the time of this meta-evaluation, the evaluation function was set up as follows:

### SEU

- 3 evaluation managers
- 1 coordinator
- 1 head of unit

### SEU Steer Co

- OCB Medical Director
- OCB Deputy GD
- OCB Director of Operations
- OCB Board member
- MSF SE GD
- MSF SE Board member

## METHODS

The design of this meta-evaluation is an external and independent ex-post portfolio meta-evaluation. Methods used were:



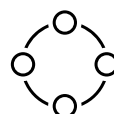
Document  
review



Key informant  
interviews



Online survey  
questionnaires



Focus group  
discussion



Participant  
observation

Checklist and numerical weight and sum methodologies were used for analysis and synthesis.

A synthesis of quality frameworks defines quality for this study, which includes the Program Evaluation Standards, the ALNAP Proforma, the UNEG Norms and Standards, and the SEU Evaluation Manifesto Quality Framework (EMQF). With the PrgES and ALNAP frameworks, evaluations were assessed individually and then the outcomes of the sub-criteria and indicators were averaged to provide analysis of the entire 2017-2021 dossier. The UNEG and EMQF norms, standards and values were used only to assess the dossier in its entirety (and not by individual evaluation), by applying the PrgES and ALNAP. Data from sources were also used to assess the evaluations and or dossier per framework.

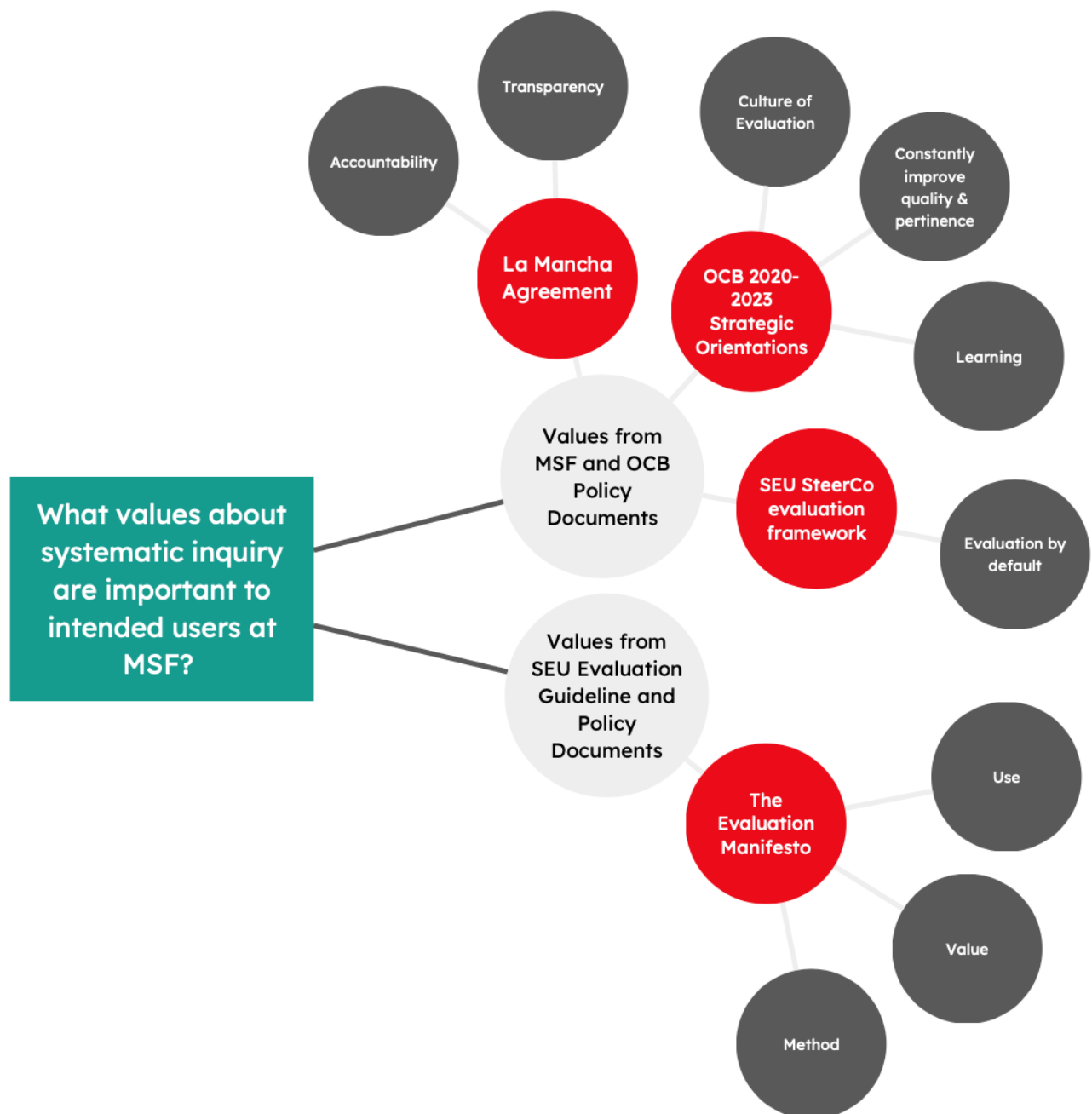
FRAMEWORK	CRITERIA	SUB-CRITERIA	INDICATORS	UNIT OF ANALYSIS	EVIDENCE SOURCES
PrgES	5	30	180	Evaluation	Documents; survey; interviews; participant observation
ALNAP	5	15	43	Evaluation	Documents
UNEG	19	24	168	Evaluation Portfolio	PrgES and ALNAP Ratings; survey; interviews; participant observation
EMQF	3	12	NA	Evaluation Portfolio	PrgES and ALNAP Ratings; survey; interviews; participant observation

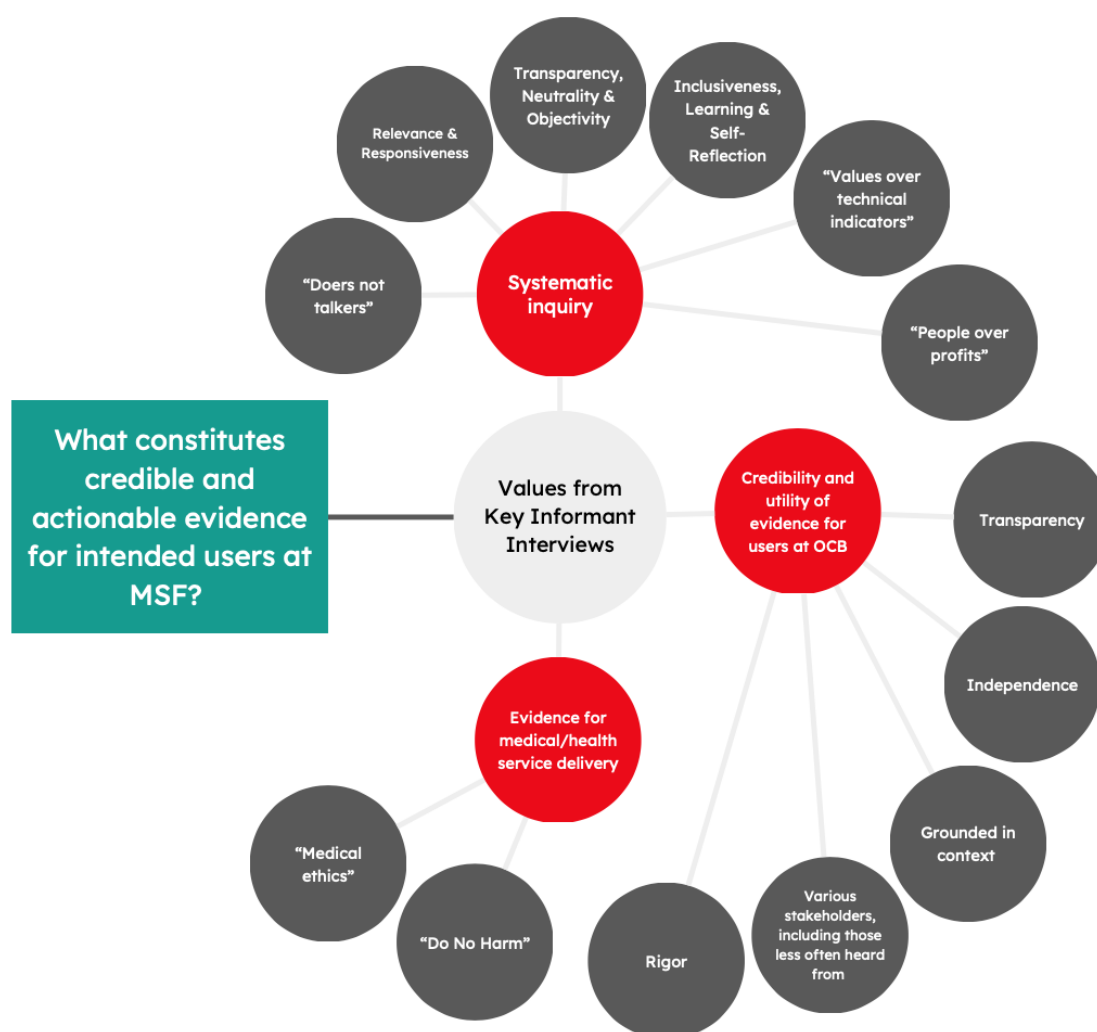
## FINDINGS

### MEQ #1: WHAT ARE THE MOST IMPORTANT, RELEVANT, AND USEFUL META-EVALUATION CRITERIA FOR MSF IN DETERMINING THE QUALITY AND VALUE OF PART AND FUTURE EVALUATIONS?

The meta-evaluation team reviewed existing documents, including agreements, strategies, and policies, and gathered input directly from key informants at OCB. Interviews confirmed ideas on evaluation quality differ across interviewees, but these differences are complimentary and not in

contradiction with an overall vision of evaluation quality. Meanwhile, document content analysis revealed the SEU does in fact manage an (un-stated) quality framework, dispersed across multiple guideline documents.





## MEQ #2: WHAT IS THE QUALITY OF EVALUATION AT OCB?

Program Evaluation Standards	Good	★★★
ALNAP Proforma	Very good	★★★★
UNEG Norms & Standards	Good	★★★
SEU Evaluation Manifesto	Very good	★★★★
Quality Framework		

### Program Evaluation Standards ★★★

The portfolio of evaluations managed by the SEU from 2017-2021 received a GOOD rating with a score of 60% when compared with the highest bar of transdisciplinary evaluation quality. To an uninformed reader these findings may seem underwhelming. Despite the room for improvement, this is a laudable result. This finding reveals OCB has a healthy and well-functioning evaluation institutional framework and emerging evaluation culture and suggests high value of evaluations for OCB.

CRITERION	DEFINITION	EXAMPLE(S) OF SUB-CRITERIA	SCORE	RATING
Utility	The extent to which evaluations are “aligned with stakeholders’ needs related to contribution towards use and influences are possible” but does not directly measure actual evaluation use and influence	Evaluations should be conducted by qualified people who establish and maintain credibility in the evaluation context; Evaluations should devote attention to the full range of individuals and groups invested in the program or affected by the evaluation.	68%	★★★★
Feasibility	The extent to which evaluations are “viable, realistic, contextually sensitive, responsive, prudent, diplomatic, politically viable, efficient, and cost effective.”	Evaluations should use effective project management strategies. The procedures should be practical and responsive to the way the program operates.	66%	★★★
Evaluation accountability	The extent to which an evaluation is systematically, thoroughly, and transparently documented and then assessed, both internally and externally for its utility, feasibility, propriety, and accuracy.	Evaluations should use these and other applicable standards to examine the accountability of the evaluation design, procedures employed, information collected, and outcomes.	63%	★★★
Propriety	The extent to which evaluation are conducted properly, fairly, legally, ethically, and justly with respect to (1) evaluators’ and stakeholders’ ethical rights, responsibilities, and duties; (2) systems of relevant laws, regulations, and rules; and (3) roles and duties of professional evaluators.	Evaluations should be responsive to stakeholders and their communities. Evaluations should be designed and conducted to protect human and legal rights and maintain the dignity of participants and other stakeholders.	53%	★★★
Accuracy	Examines the extent to which evaluations “[employ] sound theory, designs, methods, and reasoning in order to minimize inconsistencies, distortions, and misconceptions and produce and report truthful evaluation findings and conclusions.”	Evaluation conclusions and decisions should be explicitly justified in the cultures and contexts where they have consequences. Evaluation information should serve the intended purposes and support valid interpretations.	51%	★★★



**ALNAP: ★★★★★**

When compared with the best available and generally accepted sector-specific quality framework for humanitarian evaluation, strong evidence reveals the portfolio of past SEU managed evaluations is “Very Good” with a score of (80%). This framework makes direct claims about the quality of evaluation reports, which provide indirect claims about the quality of evaluation processes.

CRITERION	DEFINITION	SCORE	RATING
Terms of reference	Addresses the extent to which the terms of references include all relevant and comprehensive information deemed necessary to establish the scope of the evaluation and attract the proper talent to meet evaluation needs.	82%	★★★★★
Methods	This quality domain investigates the strength of the evaluation design and methodology as described in evaluation documentation.	72%	★★★★★
Contextual analysis	This quality domain addresses the extent to which the evaluation reports conduct adequate analysis of the humanitarian context.	80%	★★★★★
Intervention assessment	The extent to which the evaluation report describes and assesses the humanitarian intervention.	78%	★★★★★
Report assessment	Assesses specific dimensions of evaluation report comprehensiveness and quality.	88%	★★★★★

**UNEG Norms and Standards**

The SEU evaluation portfolio and evaluation system perform “Good” against one of the most generally accepted quality frameworks for evaluating international aid evaluation cases and systems, the *UNEG Norms and Standards*. Of all the quality frameworks, this one deals the most with the institutional framework and evaluation culture that backdrop evaluation case performance.

**Overall UNEG Norms Rating: ★★★**

NORM	DEFINITION EXCERPT	RATING
Utility	“...there should be a clear intention to use [the evaluation] to inform decisions and actions”	★★★★★
Credibility	“Credibility is grounded on independence, impartiality and a rigorous methodology.”	★★★★★

Independence	“...evaluators [should] be impartial and free from undue pressure throughout the evaluation process.”	★★★★
Transparency	“Evaluation products should be publicly accessible.”	★★★★
Professionalism	“Evaluations should be conducted with professionalism and integrity.”	★★★★
Evaluation Policy	“clear explanation[s] of the purpose, concepts, rules and use of evaluation within the organization”	★★★★
Responsibility for the Evaluations Function	“governing body [is] responsible for...independent, competent and adequately resourced evaluation [unit]”	★★★★
Ethics	“integrity and respect for [culture]...human rights... gender equality; and ‘do no harm’”	★★★
Enabling environment	“an organizational culture that values evaluation as a basis for accountability, learning and evidence-based decision-making”	★★★
Impartiality	“The key elements of impartiality are objectivity, professional integrity and absence of bias.”	★★
Human Rights and Gender Equality	“[integrate] principles of human rights and gender equality... into all stages of an evaluation.”	★★
Evaluation Use and Follow-Up	“...promote evaluation use and follow-up, using an interactive process that involves all stakeholders.”	★★

### UNEG Standards: ★★★

UNEG STANDARDS RATINGS FOR SEU EVALUATION PORTFOLIO AND SYSTEM	
Management of the evaluation function	★★★★
Evaluation competencies	★★★★
Institutional framework	★★★
Conduct of evaluations	★★★
Quality	★★★
OVERALL STANDARD RATING FOR EVAL PORTFOLIO AND SYSTEM	
	★★★

### Evaluation Manifesto Quality Framework: ★★★

These domains and sub-domains were drawn exclusively and without modification from the quality domains and sub-headings under the quality domains in the SEU’s Evaluation Manifesto. The SEU receives a “Good” rating suggesting it is working hard to live up to its own standards of quality, no matter how fixed or unfixed.

DOMAINS AND SUBDOMAINS	RATING
VALUE (domain)	★★★
Choosing Criteria (activity)	★★★★
Ask the right questions (principle)	★★★
Engagement and ownership (domain)	★★★
Languages (domain)	★★★
Ethics (domain)	★★★
Engage the voices of those less present (principle)	★
USE (domain)	★★★
Transversal Learning (domain)	★★★★★
Annual report (product)	★★★★★
Evaluation day (event)	★★★★★
External communication (domain)	★★★★
Annual presentation at OCB board (event)	★★★★
Real time learning (domain)	★★★
Link to strategic platforms and meetings (principle)	★★★
Communicate and disseminate findings (principle)	★★★
Learning (domain)	★★
Follow up on findings and recommendations (principle)	★
METHOD (domain)	★★★
Discuss Evaluator Competencies (principle)	★★★★
Data (domain)	★★★
Consider the Evaluability of the Project (principle)	★

**MEQ #3: WITH THESE META-EVALUATIVE CONCLUSIONS, TO WHAT EXTENT DO THESE COMPLETED EVALUATIONS PROVIDE VALUE TO OCB?**

This meta-evaluation makes the distinction between quality and value. Quality is the merit of something. Value is the worth of something. With that, the value of something is to some extent determined by the quality of that same thing. The following analysis uses these premises to investigate the value of evaluations and the evaluation function at OCB. This is accomplished by 1) describing and judging the use and use outcomes of evaluations at OCB; 2) conducting a cost utility analysis that compares evaluation quality (in terms of utility) with evaluation cost (in terms of external evaluation consultant budgets); and 3) making an overall claim about the value of evaluations at OCB.




The Meta-evaluation concludes that OCB is getting good value, or worth, from its evaluation function.

## MEQ #4: WHAT ARE THE FACTORS OF EVAL QUALITY AT OCB?

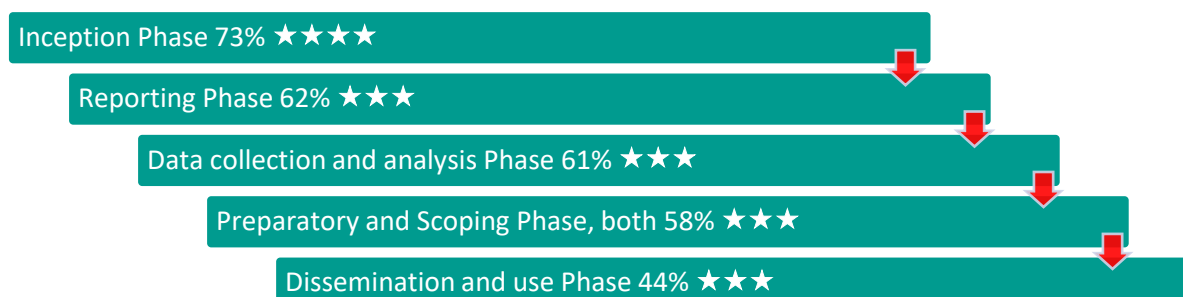
### Quality Gap list Represented by PrgES Indicators with Scores of 25% or Less

The simplest analytical approach to answering what factors mediate quality – or what contributed to the quality scores and ratings observed from the portfolio for evaluation performance criteria – is to investigate the ratings for the sub criteria and indicators. There are two reasons framework indicators could have received a “not met” rating: The first reason is due to systematic gaps in data availability, or data gaps. The second type of gaps are directly observable systematic quality gaps. These are sub-criteria and indicators that had sufficient evidence to demonstrate standards regularly not being met across the portfolio.

Score		Standard	Criteria
23%		Feasibility	Practical procedures
			Sound designs and analyses
15%		Accuracy	Reliable information
12%			Information management
10%		Utility	Attention to stakeholders
			Concern for consequences and influence
		Accuracy	Explicit evaluation reasoning
3%			Evaluation documentation
0%		Utility	Evaluator credibility
		Feasibility	Resource use
	Propriety	Clarity and fairness	
	Accuracy	Justified conclusions and decisions	
	Evaluation accountability	Internal meta-evaluation	

## Six-Step Evaluation Analysis

The next procedure to answer this question was to use the Six-Step Evaluation process as a diagnostic tool. Each of the 180 indicators for the PrgES were coded by the SEU six-step evaluation process and corresponding scores and ratings for the 31 evaluation cases were sorted by the new index of PrgES indicators by the six-step process codes.

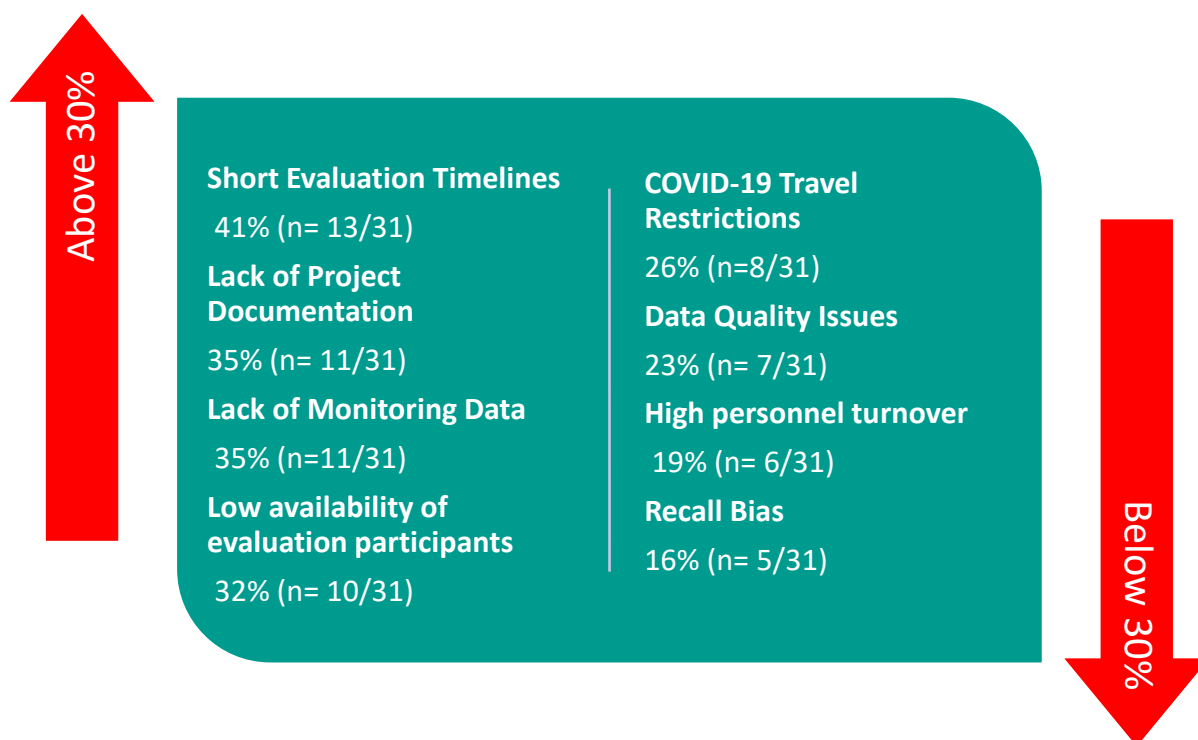


## Maximum Deviation Analysis for Outcome Factors

One of the questions posed to respondents of the meta-evaluation surveys was “What were the most important factors that determined the outcome of this evaluation?” In analysing survey responses from both the highest and lowest scoring evaluation, two clear trends seemed to dominate respondents' views of what determined a positive or negative outcome. First, was the competency of the contracted external evaluation consultant(s). The second most frequently reported factor said to determine the outcome of evaluations was evaluation participant engagement. This included high levels of engagement and collaboration from consultation groups, commissioners, and other participants. The presence or absence of this engagement was indicated as a clear factor towards the evaluation outcome.

## Limitation Analysis

Each evaluation report reviewed included a limitations section detailing potential issues that might influence the accuracy, feasibility, or other dimensions of quality for the given evaluation. A transversal analysis identified eight limitations that occurred frequently enough to suggest the evaluation system or environment could play a role in addressing these.



### Role and Responsibility Analysis

OCB staff (SEU, commissioners, and project contacts) were surveyed on “How are [their role] responsible for the success or failure of evaluations at OCB?” Together, all three see it as their shared responsibility to align expectations around scope, design, purpose, and use. Below are the most commonly cited responsibilities, shared perceptions and how the roles identify themselves in the process.

SEU Evaluation managers (n=24/29)	Commissioners (n= 9/39)	Project contacts (n= 6/39)
Planning for intended use (n= 7)	Promote engagement among participants (n= 5)	Aligning expectations (n= 1)
Aligning expectations (n= 8)	Align expectations (n= 4)	Helping evaluators access information (n= 1)
Regular communication with and between evaluation teams and intended users (n= 7)	Apply findings (n= 2)	Supporting the coordination for data collection or field visits (n= 1)
General management activities for adaptive accompaniment of the evaluation (n= 7)		Guide the process with local knowledge (n= 1)
Adequately identifying and mitigating risks to the evaluation (n= 6)		Applying findings (n= 1)

These findings reveal important distinctions between these roles and how they can support one another and the evaluation process, as well as signal a shared degree of accountability for ensuring the

quality of evaluations are high and that their potential value are realized with the alignment, planning, and application of findings for intended use.

Evaluation manager	Commissioner	Project contact
Facilitator, coordinator, and safeguard in terms of quality and risk management	Provide legitimacy to the evaluation process for operational staff	Granting access and contextual knowledge

## CONCLUSIONS

### Definitions of Evaluation Quality at OCB are Emerging and Defensible.

The SEU and OCB have a mature view of the nature of evaluation and an emerging quality framework informed by generally accepted evaluation quality frameworks. It is expected that the SEU and SEU steering committee will continue to refine the answer to this question with updates to and consolidation of evaluation policy following this study.

### Evaluation Quality at OCB is GOOD to VERY GOOD.

Using the definition of quality identified from this study, the portfolio of 31 evaluations managed by the SEU from 2017-2021 has been judged “Good” to “Very Good”. These ratings and historical analysis reveal evaluation quality at OCB is high and has improved over time.

### Evaluation Use at OCB is GOOD; Use Outcomes are FAIR; the Full Extent of Evaluation Use and Outcomes is still UNKNOWN.

Comparing the “Very Good” evaluation *Utility* rating with the “Good” *Evaluation Use* and “Fair” *Evaluation Use Outcome* ratings suggest a few conclusions. Evaluators and managers are fulfilling their responsibility to prepare the conditions for evaluation use. Commissioners and clients within cell offices and operations departments can do more to make use of evaluation. Lower ratings of use, use outcomes, and degree of use and influence may be under-representative due to lacking SEU evaluation follow-up procedures and limitations with primary data collection and analysis measures of *Evaluation Use* from this meta-evaluation study.

### The OCB is receiving GOOD Value from the Evaluation Function.

Factoring utility, use, use outcomes, and evaluation costs, the portfolio of evaluations have provided good value or worth to OCB. The majority of evaluations are used, used in multiple ways, and among multiple users. The best available evidence suggests use leads to positive outcomes for those involved and affected by evaluations. Cost Utility Analysis reveals good utilization for money.

### The Evaluation System at OCB is Well Functioning and Healthy.

Quality and value ratings for the 2017-2021 portfolio of evaluations is a barometer for the health of the evaluation system in which they were commissioned, managed, and used. Findings for evaluation quality and value indicate there is an enabling environment for useful evaluations at OCB. This is the

result of the intentional and concerted efforts by the head of the SEU, evaluation managers, qualified external evaluation consultants, invested consultation groups, supportive evaluation commissioners, and helpful project contacts. It is also a likely result of a long-standing ambition to foster a culture of evaluation at OCB.

### The OCB has a Roadmap for Sustaining and Improving Quality and Value.

The chosen methods of meta-evaluation checklists and rubrics have resulted in transparent and actionable findings and recommendations for the SEU to sustain and improve evaluation quality and value at OCB.

Strong quality dimensions to <b>sustain</b>	Include utility, Reporting and Communication, Management of the Evaluation Function, and Transversal Learning.
Weak quality dimensions to <b>improve</b>	Include accuracy, Evaluability Assessment, Explicit Evaluation Reasoning, Human Rights and Gender Equity, and Engage the Voices of Those Less Present
Significant factors of evaluation <b>quality</b>	Include evaluation competencies and evaluation participant engagement
Frequently occurring <b>limitations</b>	Include short evaluation timelines and lack of program documentation and monitoring data, among others.



## RECOMMENDATIONS

Following SEU guidance and tradition, the meta-evaluation team offers five main strategic recommendations deemed to have the most potential in assisting the SEU and OCB realize the value of this meta-evaluation through intentional and concerted follow-up.



### STRENGTHEN THE EVALUABILITY ASSESSMENT FUNCTION

- Apply existing scoping procedures consistently.
- Strengthen scoping procedures with evaluability assessment methods.



### RE-INVEST IN DOCUMENTING EVALUATION USE AND INFLUENCE

- Monitor of evaluation activity, output, and outcome-level data.
- Recommit to policies for Evaluation Use Follow up.
- Expect a more active role from evaluation commissioners and evaluation requesters.



### DEMAND STRONGER EVALUATIVE LOGIC, REASONING, AND VALUING

- Codify stronger expectations for evaluation consultants to be more explicit and transparent in their evaluative reasoning.
- Establish criteria standards for each evaluation criteria.
- Use evaluation rubrics to improve the transparency and credibility of evaluative judgments.



### FORMALISE THE INTERNAL META-EVALUATION FUNCTION

- Improve the evaluative reasoning of external evaluation consultants by practicing improved meta-evaluative reasoning themselves.
- Draw a more confident line in the sand about what constitutes evaluation quality.
- Identify and articulate how sub-sector specific contingencies for medical humanitarian evaluation.
- Use meta-evaluation findings and recommendations to realize an enhanced vision of quality.



### ADOPT TRANSFORMATIVE EVALUATION POLICIES

- Manage the evaluation about if OCB is doing things right, but also about if OCB is doing the right things.
- Give serious consideration to a new cluster of evaluation policies that extend cosmopolitan notions of ethics to include a vision of transformative evaluation practice that is culturally responsive, culturally specific, and equitable for those involved in and affected by the evaluation function.

Stockholm Evaluation Unit  
<http://evaluation.msf.org/>  
Médecins Sans Frontières

Independently written by  
Michael Harnar, Zach Tilton and Tian K. Ford.  
(December 2022)