SUPPLEMENTARY MATERIAL S6

Performance Monitoring and Evaluation Framework

Metrics, After-Action Review Standards, Reporting Templates, and 24-Month Evaluation Framework

Purpose and Audience

Purpose: Provide a practical monitoring package for Preparedness Intelligence Unit (PIU) pilots, including metric definitions, calculation methods, reporting templates, After-Action Review (AAR) standards, and an evaluation framework aligned with the 7-1-7 framework.

Audience: PIU teams, Public Health Emergency Operations Centre (PHEOC) leadership, monitoring and evaluation units, and partner evaluators. This document aligns with and cross-references Supplementary Box S1 (weekly risk brief template) and the implementation roadmap in Supplementary Material S2.

1. Core Metrics (Pilot Minimum)

1.1 Alert-to-Response Time (ART)

Definition: Time elapsed from the electronic Integrated Disease Surveillance and Response (eIDSR) alert timestamp in District Health Information Software 2 (DHIS2) to the first documented response action — Rapid Response Team (RRT) dispatch, public health advisory, supply activation, or field investigation start — as recorded in the incident log.

Unit: Hours (preferred) or days; report median and interquartile range (IQR).

Disaggregation: Disease, district, facility level (where available), and event type.

Target: ≤48 hours for ≥70% of priority disease alerts (country-set benchmark pending baseline measurement).

Calculation Method (ART)

1. Extract alert_time from eIDSR on DHIS2 event record.
2. Extract response_time from incident log or RRT dispatch log.
3. Compute ART = response_time − alert_time.
4. Exclude records with missing timestamps; report exclusion count and reasons.
5. Report ART distributions and compare against ≤48-hour target for priority diseases.

1.2 Forecast Accuracy (Brier Score)

Definition: Accuracy of probabilistic forecasts for defined threshold events at 14-day and 30-day horizons.

Primary score: Brier score for event E (e.g., cases exceed threshold T in district d during horizon h).

Interpretation: Lower Brier score is better. Scores <0.25 often indicate useful predictive skill; scores >0.50 often indicate poor performance. These thresholds should be interpreted in context of disease-specific baseline rates and forecast horizon.

Calculation Method (Brier Score)

6. Define event E: cases exceed threshold T within horizon h.
7. For each forecast instance i, record predicted probability p_i and observed outcome o_i (0 = event did not occur; 1 = event occurred).
8. Compute Brier = mean((p_i − o_i)^2) across all forecast instances.
9. Report by disease, horizon (14-day and 30-day), and district group (high burden vs low burden).
10. Track calibration: compare predicted probabilities vs observed frequencies across probability bins.

https://doi.org/10.37432/jieph-d-25-00266

Note: The Brier score is referenced in Supplementary Material S1 (Data Scientist KPI: Brier scores <0.25) and is the primary forecast accuracy metric used in PIU performance reviews.

1.3 AAR Closure Rate

Definition: Proportion of issues identified in exercises or activations that reach documented resolution within 90 days, with verification evidence.

Target: ≥70% closure within 90 days, with quarterly reporting and independent verification of completion evidence.

Calculation Method (AAR Closure Rate)

11. Log each issue identified in exercises or activations with: owner, due date, expected evidence of resolution, and current status.
12. Mark status: open, in progress, resolved, verified, or deferred.
13. Compute closure rate = verified resolutions within 90 days ÷ total issues logged.
14. Report quarterly; include exceptions with justification for any deferred items.

2. Reporting Templates

2.1 Weekly PIU Risk Brief (One-Page)

Use Supplementary Box S1 as the reference template. Minimum required fields:

- Priority alerts with disease, location, case counts, and risk level
- 7-1-7 performance summary (detection, notification, and response timelines for active events)
- 14-day probabilistic forecast summary with Brier score where available
- Recommended actions with named owners and deadlines
- Data quality flags for any reporting gaps

2.2 Monthly Performance Dashboard Report (2–3 Pages)

- ART summary by disease and district: median, IQR, and target tracking (≤48 hours)
- Forecast performance: Brier score by horizon and disease; calibration summary
- AAR issue tracker summary: open vs verified; overdue items highlighted
- Data quality Key Performance Indicators (KPIs) and planned corrective actions

2.3 Quarterly AAR and Learning Brief (3–5 Pages)

- What happened: timeline of key events, exercises, and activations
- What worked and why (evidence-based)
- What did not work and why (evidence-based)
- Corrective actions: owner, deadline, verification evidence
- Follow-up status from previous quarter corrective actions

Note: AARs should be written within 14 days of an exercise or activation debrief. Use a standard template across pilot sites for cross-country comparability (see Section 3 below).

3. AAR Standards

AARs should be written within 14 days of an exercise or activation debrief and follow these standards to ensure consistency and comparability across pilot sites.

- Use a standard template across pilot sites for comparability (template provided in Section 2.3 above)
- Include objective evidence: incident logs, dispatch logs, dashboard screenshots, and decision memos — not only narrative accounts

https://doi.org/10.37432/jieph-d-25-00266

- Separate findings into four categories: People, Process, Data, and Tools
- Record decisions on policy or Standard Operating Procedure (SOP) changes with their implementation timeline
- Every corrective action must have a named owner, deadline, and description of expected verification evidence
- AAR completion must be signed off by PIU Lead and PHEOC Director

Corrective action tracking: Each action identified in an AAR is entered into the AAR issue tracker (linked to the monthly dashboard report). Status is updated monthly. Actions marked 'verified' must include evidence (photograph, updated SOP document, test result, or equivalent). The quarterly AAR and learning brief must report follow-up on all open items from prior quarters.

4. Evaluation Framework (24-Month Pilots)

4.1 Evaluation Questions

- Does the PIU reduce alert-to-response time (ART) for priority diseases compared to pre-pilot baseline?
- Does the PIU improve data integration and routine use of analytics across surveillance, laboratory, and climate systems?
- Do forecasts provide actionable lead-time and improve decision readiness as measured by Brier scores over time?
- Do AAR loops close issues on schedule and does this improve subsequent performance over successive quarters?

4.2 Design Options

Design	Description
Before–after comparison	Within pilot districts, using pre-pilot baseline data. Requires baseline ART measurement before PIU launch. Simplest design; vulnerable to temporal confounding.
Stepped-wedge rollout	Across districts where feasible. Districts begin PIU at staggered intervals, allowing each to serve as control before rollout. Stronger causal inference; requires careful planning.
Comparator districts	Districts without PIU included as contemporaneous controls (if politically and ethically acceptable). Provides direct contemporaneous comparison.

Note: Given the nature of the PIU model (systems change, not individual intervention), no design will yield controlled experiment-level causal evidence. The evaluation should be designed to maximize learning and inform adaptation, not primarily to generate causal proof.

4.3 Data Sources

- DHIS2/eIDSR events and analytics extracts (primary source for ART calculation)
- Incident logs, dispatch logs, and supply activation logs (ART and AAR inputs)
- PIU weekly risk briefs and monthly dashboard reports (forecast performance and outputs)
- Exercise reports and AAR trackers (corrective action closure)
- Stakeholder surveys (PHEOC leadership and district health officers; minimum quarterly)

4.4 Reporting Schedule

https://doi.org/10.37432/jieph-d-25-00266

Frequency	Report
Weekly	Risk brief (see Supplementary Box S1 template)
Monthly	Performance dashboard report (2–3 pages; see Section 2.2)
Quarterly	AAR and learning brief (3–5 pages; see Section 2.3)
Month 12	Independent mid-term evaluation summary (external evaluators)
Month 24	Independent final evaluation summary (external evaluators); peer-reviewed publication

4.5 Phase Gate Performance Criteria

The following criteria align with the Phase Gate criteria defined in Supplementary Material S2 (Implementation Roadmap):

Phase Gate	Metric	Threshold
Gate 2 (Month 12)	Data quality (completeness)	≥80% completeness in pilot districts
Gate 2 (Month 12)	Forecast validation	At least 1 forecast validated against observed outcome
Gate 3 (Month 18)	Alert-to-response time (ART)	≤48 hours in pilot districts
Gate 3 (Month 18)	Stakeholder satisfaction	≥70% satisfaction from surveys
Gate 4 (Month 24)	7-1-7 framework adherence	≥70% of priority disease outbreaks meeting all three targets
Gate 4 (Month 24)	AAR closure rate	≥70% of issues verified within 90 days
Gate 4 (Month 24)	Annual performance report	Published and publicly available

5. Conclusion

The three core metrics — alert-to-response time (ART), Brier score forecast accuracy, and AAR closure rate — provide a performance profile covering timeliness, analytic accuracy, and organizational learning. Together, these are the dimensions on which PIU effectiveness can be most directly assessed using routinely available data.

Performance monitoring should begin from the first week of operations, even if initial measurements reflect a pre-improvement baseline. The value of the monitoring system lies in the learning trend over time, not only in meeting thresholds from the outset. All metrics should be reported transparently in public-facing annual performance reports to build accountability and stakeholder trust.

This document should be read alongside Supplementary Box S1 (weekly risk brief template), Supplementary Material S2 (implementation roadmap and phase gate criteria), and Supplementary Material S1 (KPIs for individual staff roles).

https://doi.org/10.37432/jieph-d-25-00266

List of Abbreviations

Abbreviation	Full Term
AAR	After-Action Review
ART	Alert-to-Response Time
Brier score	A metric for measuring forecast accuracy (mean squared difference between predicted probability and observed outcome)
DHIS2	District Health Information Software 2
eIDSR	Electronic Integrated Disease Surveillance and Response
IQR	Interquartile Range
KPI	Key Performance Indicator
PHEOC	Public Health Emergency Operations Centre
PIU	Preparedness Intelligence Unit
RRT	Rapid Response Team
SOP	Standard Operating Procedure
WHO	World Health Organization
7-1-7	Detect within 7 days, notify within 1 day, initiate early response within 7 days

https://doi.org/10.37432/jieph-d-25-00266