Supplementary File 8. Analysis plan for the whole protocol

Estimands, analysis of every outcome, integration, economics, equity, progression and reporting

Version 1.4 - 17 August 2026

Companion to: Implementing a competency-based Advanced Field Epidemiology Training Programme in Sierra Leone: protocol for a mixed-methods pilot evaluation (JIEPH-D-26-00266).

1. Status and governing principles

This plan covers every analysis the pilot will run. It is fixed before enrolment and versioned. Supplementary File 3 holds the technical statistical specification; this document is the plan for the study as a whole, including the qualitative, economic, equity and decision analyses, and it states who performs each analysis and when.

- Analysis is estimation focused. No p-value threshold is used to declare the pilot successful or unsuccessful.
- The primary fidelity outcome and six key secondary measures enter four graded progression criteria: field-practice attainment, retention, acceptability/appropriateness/perceived feasibility, and primary-data completeness. Safety, safeguarding and integrity are checked first as non-negotiable conditions. Every other analysis is secondary or exploratory and is named before data collection.
- This plan covers every analysis the pilot will run. It is fixed before enrolment and versioned. Supplementary File 3 holds the technical statistical specification; this document covers the study as a whole, including quantitative, qualitative, mixed-methods, economic, equity, safety and decision analyses, and states who performs each analysis and when.
- No analysis may be added, removed or reclassified after data collection begins without a dated protocol amendment approved before the relevant data are examined.
- All analyses run from version-controlled scripts against a locked data dictionary, with software versions, random seeds and an analysis decision log archived.
- Cells smaller than five are suppressed in any output that could identify a resident, a mentor or a site.
- Every analysis in this plan is conducted by the independent research team described in Supplementary File 7, Section 5. No person accountable for programme delivery scores a component, analyses an outcome, or holds authority to change a result.

2. Analysis populations

Population	Definition	Used for
Enrolled cohort	All residents who accept a place and start orientation	Reach, retention, field-practice attainment, graduation, all primary denominators
Active cohort at timepoint	Residents enrolled and not permanently withdrawn at the start of the measurement window	Dose, mentorship, milestone opportunity
Research-consenting cohort	Residents consenting to the relevant survey, interview or linkage	Individual-level research analyses
Programme actor population	Assigned mentors, supervisors, faculty, assessors and managers	AIM, IAM and FIM; delivery survey; qualitative sampling
Site population	The eight accredited sites, plus invited sites for adoption	Adoption, site-level description, accreditation reporting
Decision-user sample	Products with a named decision owner, purposively selected	Recommendation uptake and system contribution
Component set	The twenty fidelity components applicable at the assessment point	The primary outcome

3. Estimands and the analysis of each outcome

3.1 Primary outcome

Element	Specification
Outcome	Overall implementation fidelity at Month 24
Estimand	The proportion of applicable fidelity components meeting their operational standard in the enrolled programme at Month 24, with the separate condition that every applicable critical component is met
Population	Component set, all twenty applicable at Month 24
Measurement	Instrument I07, scored and double scored by the independent research team as set out in Supplementary File 7
Analysis	Count and exact proportion; due, applicable, met, partial, not met, unknown, disputed, not-due and not-applicable components reported; component-level and critical-component tables reported in full; quarterly trajectory shown; no sampling confidence interval for the 20-component census.
Handling of unknowns	Unknown and unresolved disputed components count as not met in the primary analysis. A sensitivity analysis may exclude them. Not-applicable exclusions require written IRET adjudication.
Missing data	No imputation. Observed evidence only
Who	Scored independently by two IRET fidelity assessors; unresolved cases adjudicated by a third reviewer; computed by the study statistician from the signed Month-24 lock.

3.2 Key secondary outcomes

Outcome	Estimand	Analysis	Timing
Field-practice attainment	Proportion of enrolled residents completing at least 68 qualifying weeks by Month 24 or an approved window	Proportion with Wilson 95% CI; distribution of weeks; reasons for shortfall; sensitivity to the completion window	Monthly; reported at Month 24
Retention	Proportion of enrolled residents active at Month 24	Proportion with Wilson 95% CI; cohort flow diagram; timing and stated reason for each interruption and withdrawal	Continuous; Months 12 and 24
Acceptability, appropriateness and perceived feasibility	Stakeholder-group mean AIM, IAM and FIM scores at each wave	Item and scale distributions, group mean with 95% CI, floor and ceiling inspection, internal consistency, response rate by role and wave; qualitative findings reported beside each	Months 6, 12, 18, 24
Primary-data completeness	Required fields available divided by required fields expected, by source and wave	Proportion by instrument and timepoint; lateness, invalid values, duplicates, chronology errors and query closure	Quarterly

3.3 Non-negotiable conditions

Safety, safeguarding and scientific integrity are not analysed as a rate. Every serious event is reported individually with type, severity, time to response, resolution status and corrective action taken. The analytic question is binary: does any serious event remain unresolved at the decision point. No aggregation, and no statistical test.

3.4 Secondary outcomes

Domain	Outcomes	Planned analysis
Reach and adoption	O01, O02	Enrolled divided by eligible and offered; sites accredited and active divided by invited; distribution by geography, profession, sector and gender; reasons for non-adoption
Delivery dose	O10 to O12	Sessions, exercises and mentor contacts delivered divided by planned, with numerator, denominator and trend by quarter; low-bandwidth substitutions counted separately
Portfolio	O13, O14	Completion proportion across the eight categories; rubric score distributions by product, site and review point; critical-criterion failure and remediation counts
Outbreak evidence	O15	Investigations per resident; proportion completing two; proportion leading an analytic investigation; proportion meeting the requirement through corrective placement; time from signal to resident deployment
Research and publication	O16 to O20	Milestone completion and cumulative incidence; median actual month against planned month; stated reasons for delay; outputs by channel and editorial status through Month 30
Uptake and contribution	O23, O24	Recommendation status proportions; verified changes by type; time from recommendation to decision; contribution-strength rubric with narrative and alternative contributors named
Sustainability	O28	Scorecard across domestic financing, local faculty share, mentor retention, governance integration, site continuation, curriculum maintenance and graduate deployment readiness

3.5 Exploratory analyses

The following are exploratory. They are reported with uncertainty, interpreted as signals for a larger study, and excluded from the progression decision. No progression classification may be altered on the basis of an exploratory result.

Analysis	Specification	Condition for running it
Competency change (O21)	Paired change from Month 0 to Months 12 and 24 with bootstrap percentile intervals, 2,000 resamples, seed reported	Always reported; model-based estimates only if the conditions below are met
Longitudinal models	Linear mixed model, random resident intercept, categorical time, unstructured covariance where estimable and compound symmetry otherwise; REML with Kenward-Roger degrees of freedom	Convergence achieved and cell sizes adequate. Random slopes are not fitted. Non-convergence is reported as not estimable
Ordinal ratings	Cumulative-link mixed model with a checked proportional-odds assumption	Cell sizes permit and the assumption holds
Assessor agreement (O22)	Weighted kappa with the full confusion matrix; two-way random-effects absolute-agreement ICC for composite scores	Always reported, with intervals, as a check on rating consistency rather than a reliability claim
Site variation	Descriptive counts and proportions by site	No multilevel model is fitted. Eight sites cannot support stable variance estimation

4. Qualitative analysis

- Framework Method, with a matrix informed by CFIR domains, implementation outcomes, mechanisms, adaptations, equity, burden, safety and sustainability, and open coding for emergent concepts.
- Sufficiency judged by information power against study aim, sample specificity, use of established theory, quality of dialogue and analysis strategy; the anticipated 40 to 65 stakeholders follow from those criteria and are revisable in the decision log.
- At least 20% of transcripts at each major wave independently coded; discrepancies refine the codebook rather than produce a reliability coefficient.
- Comparison by role, site, profession, progression status and timepoint; negative cases retained and reported.
- Trustworthiness procedures as specified in Supplementary File 2, Section 24.
- Analysts: two qualitative analysts from the independent research team, neither holding a teaching, mentoring or assessment role in the programme.

5. Mixed-methods integration

Integration occurs at design, sampling, analysis and interpretation. Joint displays align each quantitative outcome with the qualitative explanation, the adaptations recorded, the cost and the site context. Relationships are classified as convergence, complementarity, dissonance or silence. Dissonance triggers source review, stakeholder interpretation or a bounded additional analysis, and is never removed to produce a uniform account. Silence is reported as a finding about the evaluation rather than treated as absence of the phenomenon.

Joint display	Quantitative side	Qualitative side	Integrated question
Delivery	Fidelity and field-week estimates by quarter	How sites protected or lost field opportunity	Was the programme delivered as designed, and what determined that
Experience	AIM, IAM and FIM distributions by role and wave	Reasons for ratings; workload accounts	Which refinements are priorities, and for whom
Attainment	Outbreak, research and publication completion and timing	Opportunity, supervision and institutional mechanisms	Is shortfall recoverable, and what would recovery require
Contribution	Verified uptake and documented changes	Decision-maker explanation; alternative contributors	How strong is the contribution claim
Equity and safety	Exposure and outcome differences; incident counts	Experience of inclusion, coercion or protection	Did the programme distribute opportunity fairly and operate safely

6. Economic analysis

Provider perspective, time-driven activity-based costing, reported against CHEERS 2022 and the reference case for costing global health services. Programme delivery, implementation strategy and research-only evaluation costs are reported separately so that the recurrent cost of routine delivery can be estimated without the cost of evaluating it. Parameters, base-case values and sensitivity ranges are specified in Supplementary File 3, Section 21.

- Outputs: total and component costs; cost per enrolled resident, retained resident, graduate, completed field product, authentic investigation, internal publication, submitted manuscript and approved thesis.
- Discounting at 3% per year beyond twelve months, with the same rate used to annualize capital; sensitivity between 0% and 5%.
- One-way sensitivity analysis on every parameter in the table, with a tornado presentation of the parameters that move unit cost most.
- No cost-effectiveness ratio is computed. The pilot has no comparator and no health outcome measured, so an incremental ratio would be uninterpretable.
- Analyst: the independent health economist, working from the locked cost ledger and time logs supplied under the data-access agreement.

7. Equity analysis

Equity is analysed as exposure before outcome, so that unequal results can be traced to unequal opportunity rather than attributed to residents.

- Exposure measures compared across approved strata: outbreak opportunity offered, mentor contact received, specialist access, protected research time and editorial support.
- Outcome measures then compared: progression, product quality, interruption and completion.
- Strata may include geography, profession or sector, gender, disability or accommodation need, site type, and baseline digital or quantitative access, subject to ethics approval.
- Descriptive differences with confidence intervals. No inferential subgroup modelling, because 15 residents cannot support it and because disclosure risk rises with each stratum. Cells smaller than five are suppressed and categories are combined only where substantively defensible.
- Every equity finding is interpreted alongside qualitative evidence rather than reported as a ranking of individuals or sites.

8. The progression decision

The decision sequence is fixed and is applied in this order. Reordering it would allow a strong aggregate to mask a disqualifying failure.

Step	Action
1	Check the non-negotiable conditions. Any unresolved serious safety, safeguarding or integrity event stops the sequence and produces red
2	Check the critical fidelity components. Any unresolved critical failure produces red irrespective of the overall percentage
3	Classify the primary outcome and the key secondary progression criteria against the prespecified green, amber and red rules in Supplementary File 3.
4	Review source quality and uncertainty. Resident- or site-level proportions near a boundary are flagged for integrated review; the 20-component fidelity census is reviewed through status and evidence sensitivities rather than a sampling interval.
5	Integrate the qualitative evidence and the equity findings for every flagged or amber criterion
6	Estimate the feasibility and cost of the corrective action each amber classification would require
7	Record the decision, the reasoning, any minority view, and the evidence relied on

9. Analysis timetable and responsibilities

When	Analysis	Lead (independent research team)
Prelaunch	Baseline context, readiness, site accreditation, cost baseline; assessor calibration on archived records	Fidelity assessors and health economist
Quarterly	Fidelity scoring, double scoring, reconciliation, dose, data quality, cost tracking, adaptation log	Fidelity assessors, then study statistician
Month 6	First AIM, IAM and FIM wave; first qualitative wave	Study statistician and qualitative analysts
Month 12	Interim integration; retention; research governance milestone; competency change from	Study statistician

When	Analysis	Lead (independent research team)
	baseline	
Month 18	Third survey and qualitative wave; portfolio interim; field-week trajectory	Study statistician and qualitative analysts
Month 24	Primary outcome; all key secondary outcomes; full joint displays; economic analysis; equity analysis; evidence for the progression decision	Independent principal investigator, with statistician, health economist and qualitative analysts
Month 30	Follow-up: publication status, verified system contribution, sustainability scorecard, final integration	Independent principal investigator
Month 30 to 33	Final report, reporting checklists, repository deposit	Independent principal investigator

The progression decision itself is taken by the steering committee and the Agency executive sponsor on the evidence the research team presents. The research team produces and interprets the evidence and does not take the scale-up decision; the committee takes the decision and does not alter the evidence.

10. Reporting shells and reproducibility

Output	Content
Table A	Cohort, sites, programme actors and baseline context
Table B	Primary fidelity component matrix and exact aggregate score, every critical component, key secondary outcomes with appropriate resident/site confidence intervals, and progression classification.
Table C	All twenty fidelity components with standard, computed value, status and adaptation
Table D	Portfolio, outbreak, research, thesis and publication outcomes
Table E	Recommendation uptake, verified system contribution, cost, equity and safety
Figure A	Cohort flow and resident milestone completion
Figure B	Quarterly fidelity trajectory and field-practice accumulation
Figure C	AIM, IAM and FIM distributions with qualitative integration
Joint display	Determinants, outcomes, mechanisms, adaptations and scale-up implications

The analysis repository will hold the protocol and amendment log, the data dictionary, the de-identification specification, the data-quality report, this plan and the technical statistical plan, all analytic scripts, software and package versions, the output log, the joint displays and the disclosure review. Scripts use relative paths and regenerate every final table and figure from the approved analysis datasets. Any manual recoding appears in an auditable correction file.

11. Version record

1.4 - 17 August 2026: final coordinated analysis plan. The revision removes a sampling interval from the fidelity component census, counts unresolved evidence as failure in the primary analysis, aligns the key-secondary hierarchy, clarifies integrated boundary review and assigns all analyses, locks and scientific reporting to the independent research team.