SUPPLEMENTARY FILE

Search Strategy and Methodological Documentation

Section S1: Systematic Search Framework

S1.1: Detailed Database Specific Search Strings

1. PubMed / MEDLINE Search Matrix

- **Execution Date:** May 3, 2026
- **Search Interface:** National Center for Biotechnology Information (NCBI)
- **Query Formula:**

```
(("community health workers"[MeSH Terms] OR "community health workers"[All Fields] OR "village health workers"[Text Word] OR "community health volunteers"[Text Word] OR "health extension workers"[Text Word] OR "accredited social health activists"[Text Word] OR "community health extension workers"[Text Word] OR "relais communautaires"[Text Word] OR "CHW"[All Fields] OR "CHEW"[All Fields] OR "VHT"[All Fields] OR "HSA"[All Fields] OR "ASHA"[All Fields]) AND ("data quality"[Text Word] OR "reporting completeness"[Text Word] OR "data accuracy"[Text Word] OR "report submission"[Text Word] OR "recording completeness"[Text Word] OR "information quality"[Text Word] OR "data completeness"[Text Word] OR "routine health information systems"[Text Word] OR "RHIS"[All Fields] OR "HMIS"[All Fields]) AND ("maternal health"[MeSH Terms] OR "child health"[MeSH Terms] OR "maternal health"[All Fields] OR "child health"[All Fields] OR "newborn"[Text Word] OR "neonatal"[Text Word] OR "infant"[Text Word] OR "antenatal care"[Text Word] OR "immunization"[Text Word] OR "vaccination"[Text Word] OR "postnatal care"[Text Word] OR "MNCH"[All Fields]) AND ("developing countries"[MeSH Terms] OR "low- and middle-income countries"[Text Word] OR "LMICs"[All Fields] OR "low income countries"[Text Word] OR "middle income countries"[Text Word] OR "Africa"[MeSH Terms] OR "Asia"[MeSH Terms] OR "South America"[MeSH Terms]))
```

- **Filters Applied:** Publication Date: January 1, 2005 – December 31, 2025; Language: English.

Embase Search Matrix (Elsevier Interface)

```
( 'community health worker'/exp OR 'community health worker' OR 'village health worker':ti,ab OR 'community health volunteer':ti,ab OR 'health extension worker':ti,ab OR 'accredited social health activist':ti,ab OR 'relais communautaire':ti,ab ) AND ( 'data quality'/exp OR 'data quality' OR 'reporting completeness':ti,ab OR 'data accuracy':ti,ab OR 'data completeness':ti,ab OR 'health management information system':ti,ab ) AND ( 'maternal health'/exp OR 'child health'/exp OR 'newborn':ti,ab OR 'neonatal':ti,ab OR 'antenatal care':ti,ab OR 'postnatal care':ti,ab OR
```

```
'immunization':ti,ab OR 'mnch':ti,ab ) AND ( 'developing country'/exp
OR 'low and middle income country' OR 'lmic':ti,ab OR 'africa'/exp
OR 'asia'/exp ) AND [2005-2025]
```

CINAHL Search Matrix (EBSCOhost Interface)

```
(MH "Community Health Workers" OR "village health workers" OR
"community health volunteers" OR "health extension workers" OR
"accredited social health activists") AND (TX "data quality" OR TX
"reporting completeness" OR TX "data accuracy" OR TX "data
completeness" OR TX "HMIS") AND (MH "Maternal Health" OR MH "Child
Health" OR TX "newborn" OR TX "neonatal" OR TX "antenatal care" OR
TX "postnatal care" OR TX "immunization" OR TX "MNCH") AND (MH
"Developing Countries" OR TX "LMIC" OR TX "low and middle income
countries")
```

WHO IRIS and Grey Literature Target Directives

Manual search syntax used within the World Health Organization Institutional Repository for Information Sharing, UNICEF Research Centers, USAID Development Experience Clearinghouse (DEC), Jhpiego Technical Publications, and the Community Health Impact Coalition digital index:

- *Terms:* "community health worker data quality" / "CHW routine reporting completeness" / "mHealth reporting accuracy low- and middle-income countries".

Section S2: Data Elements and Variable Dictionary

The specific data extraction layout used during database mapping consists of the following parameters:

Extraction ID Element	Variable Name	Domain Specification & Coding Rules
S2.1	Study Identification	Primary author surname, peer-review publication year, publication journal source.
S2.2	Geographic Context	Country location, World Bank economic income class tier, and designated WHO Regional territory.
S2.3	CBHW Cohort Specs	Cadre denomination, educational minimum baseline, total operative sample size, country program type.
S2.4	Reporting Medium	Platform Type: Paper, Hybrid, DHIS2 link, native Mobile App.
S2.5	Health Indicators	Specific tracking targets grouped under the

		five primary review domains.
S2.6	Completeness Proportion	Numerical ratio: $\frac{\text{Filled/Submitted Data Fields}}{\text{Expected Total Row Counts}}$.
S2.7	Accuracy Performance	Percentage match rate from independent primary verification audits.
S2.8	Determinant Vectors	Multivariable adjusted Odds Ratio (aOR) arrays with 95% Confidence Intervals.
S2.9	Evaluated Interventions	Structural descriptions of programmatic data quality improvement strategies.
S2.10	Methodological Quality	RoB grading metric (Low Risk, Moderate Risk, High Risk).

Section S3: Methodological Quality Assessment Standards
S3.1: ROBINS-I Evaluation Criteria Matrix (Non-randomized Frameworks)

Each non-randomized study will be evaluated across the seven mandatory domains specified by the ROBINS-I guidelines:

1. **Bias due to Confounding:** Evaluation of background imbalances in CBHW education levels, regional funding variation, and localized mobile network infrastructure.
2. **Bias in Selection of Participants into the Study:** Assessment of selection mechanisms to ensure participants were not systematically selected based on historical data compliance performance.
3. **Bias in Classification of Interventions:** Verification that paper-based, hybrid, and digital formats were explicitly classified without overlap during the evaluation timeframe.
4. **Bias due to Deviations from Intended Interventions:** Assessment of whether high supervisor turnover or logistical disruptions affected the delivery of supervision or training.
5. **Bias due to Missing Data:** Evaluation of whether missing values in primary monthly records led to the exclusion of specific clinics or cadres, which could bias overall completeness metrics.
6. **Bias in Measurement of Outcomes:** Analysis of whether data quality audits were conducted using objective verification protocols or relied on subjective self-reporting by supervisors.
7. **Bias in Selection of the Reported Result:** Verification that analysis profiles match pre-specified institutional evaluation protocols to prevent selective reporting of positive outcomes.

S3.2: Cochrane Risk of Bias (RoB) Criteria Matrix (Randomized Trial Designs)

Randomized evaluations will be assessed across five core criteria:

1. **Selection Bias:** Methods used for random sequence generation and allocation concealment during cluster assignment of health posts or sub-districts.
2. **Performance Bias:** Assessment of deviations from intended intervention protocols due to contamination across study arms (e.g., control paper-tier sites adopting digital tools).
3. **Attrition Bias:** Evaluation of the volume, distribution, and handling of missing outcome data from dropped or unsubmitted monthly reports.
4. **Detection Bias:** Assessment of whether data quality auditors were blinded to intervention status during validation reviews.
5. **Reporting Bias:** Cross-referencing published results with prospective trial registries to ensure all secondary indicators are fully reported.

Section S4: Statistical Mathematical Foundations

S4.1: Logit Transformation and Back-Transformation Formulae

Because raw proportions (p) are naturally bounded between 0 and 1, a standard random-effects meta-analysis can generate confidence intervals that exceed these realistic thresholds when pooling high data completeness rates. To resolve this restriction, raw completeness proportions are transformed into an unconstrained logit scale before pooling:

$$y_i = \text{logit}(p_i) = \ln\left(\frac{p_i}{1 - p_i}\right)$$

The corresponding intra-study variance (v_i) for each logit-transformed proportion is estimated as follows:

$$v_i = \frac{1}{n_i \cdot p_i} + \frac{1}{n_i \cdot (1 - p_i)}$$

where n_i represents the total expected report entries or data fields evaluated within study *i*.
Following random-effects pooling of the logit-transformed values (*Y*), the global pooled estimate and its 95% confidence boundaries are back-transformed into standard proportions (*P*) for intuitive reporting:

$$\boldsymbol{P = \frac{\exp(Y)}{1 + \exp(Y)}}$$

S4.2: Random-Effects Weights Assignment Model

The global pooled effect size is computed using a random-effects model that weights individual studies by both their intra-study variance and across-study variance:

$$W_i = \frac{1}{v_i + \tau^2}$$

where v_i represents the intra-study variance, and τ^2 represents the between-study variance component. Between-study variance (τ^2) is calculated using the DerSimonian-Laird estimator:

$$\tau^2 = \max\left(0, \frac{Q - (k-1)}{\sum W_i^0 - \frac{\sum (W_i^0)^2}{\sum W_i^0}}\right)$$

where k represents the number of independent study arms, $W_i^0 = \frac{1}{v_i}$ represents the fixed-effect inverse-variance weight, and Q represents the overall Cochran's heterogeneity statistic:

$$Q = \sum W_i^0 \, (y_i - \bar{y}_{\text{fixed}})^2$$