

Cross-Outcome Associations of Candidate Neonatal Risk Predictors in Kenya: A Multisite Analysis of Early Neonatal Outcomes

Danny Nyatuka

Strathmore University

Paul Macharia

Strathmore University

Esther Kahata

Strathmore University

Faith Siva

Strathmore University

Betsy Muriithi

Strathmore University

Md Shafiqur Rahman Jabin

`mdshafiqur.rahmanjabin@lnu.se`

Linnaeus University

Research Article

Keywords: neonatal morbidity, risk stratification, maternal sepsis, gestational maturity, multiple gestation, antenatal corticosteroids, hypertensive disorders, Kenya

Posted Date: October 4th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10792831/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Machine learning (ML)–based neonatal risk prediction models may support earlier identification of newborns requiring clinical attention, but predictor–outcome relationships can vary across populations, clinical settings, and outcome definitions. Understanding how candidate predictors relate to different early neonatal outcomes is, therefore, important when evaluating the contextual relevance of ML-derived risk predictor sets in routine care.

Objective

This study aimed to examine the associations between candidate neonatal risk predictors and three early neonatal outcomes—birth weight, bag-and-mask resuscitation, and hospitalization by Day 2—in a multisite Kenyan cohort, and to determine whether associations were consistent across outcomes.

Methods

We conducted a multisite observational analysis of 394 maternal-neonatal observations from three Kenyan health facilities. Candidate predictors were derived from an existing ML-based neonatal risk prediction model and included maternal, pregnancy, and delivery characteristics. Birth weight was analyzed using Spearman rank correlation, whereas bag-and-mask resuscitation and hospitalization by Day 2 were examined using bivariate Wilcoxon rank-sum tests. False discovery rate (FDR) adjustment was applied to account for multiple comparisons, with an adjusted q value of $< .05$ considered statistically significant.

Results

Seven statistically significant predictor–outcome associations involving five candidate predictors remained after FDR adjustment. Birth weight was positively associated with gestational age at delivery ($\rho = 0.183$; $q = .002$) and negatively associated with multiple birth ($\rho = -0.202$; $q < .001$) and antenatal corticosteroid exposure ($\rho = -0.163$; $q = .006$). Bag-and-mask resuscitation was associated with maternal sepsis ($q = .030$) and antenatal hypertension/eclampsia ($q = .030$). Hospitalization by Day 2 was associated with maternal sepsis ($q < .001$) and gestational age at delivery ($q = .00018$). Maternal sepsis and gestational age at delivery, therefore, demonstrated associations across 2 outcomes, whereas multiple birth, antenatal corticosteroid exposure, and antenatal hypertension/eclampsia demonstrated outcome-specific associations. Other examined predictors did not retain statistically significant associations after FDR adjustment.

Conclusions

Candidate neonatal risk predictors showed heterogeneous, outcome-specific associations across early neonatal outcomes in this Kenyan multisite cohort. Maternal sepsis and gestational age at delivery showed cross-outcome signals, whereas other predictors were associated with individual outcomes only. These findings provide contextual evidence regarding the empirical behavior of ML-derived candidate predictors but should not be interpreted as evidence of causality or model performance. Multivariable analyses, external validation, calibration assessment, and prospective evaluation of clinical utility are warranted before translating these associations into clinical prediction or decision support.

Background

Neonatal Mortality and the Need for Earlier Risk Identification

Neonatal mortality remains a major global health challenge, with the burden disproportionately concentrated in low- and middle-income countries (LMICs) [1, 2]. In 2022, approximately 2.3 million newborns died within the first 28 days of life, and neonatal deaths continue to account for a substantial share of deaths among children younger than 5 years [1–3]. Sub-Saharan Africa bears a particularly high burden, and progress toward Sustainable Development Goal (SDG) Target 3.2, ending preventable newborn and child deaths, remains uneven across countries and health systems [1, 3].

Many neonatal deaths occur in the context of conditions for which earlier recognition and timely clinical management may improve outcomes. Prematurity, low birth weight, birth asphyxia, neonatal sepsis, and other maternal, obstetric, and neonatal complications can develop or become clinically significant during the immediate perinatal and early postnatal period. Consequently, timely identification of newborns at increased risk is an important component of efforts to strengthen neonatal care, particularly in settings where clinical resources and specialist capacity are limited [1, 4].

In Kenya and other resource-constrained settings, early risk identification occurs within health systems marked by substantial variation in staffing, clinical workload, infrastructure, documentation practices, and access to digital decision-support tools. These contextual conditions may hinder consistent identification and management of vulnerable newborns. The need for more systematic approaches to neonatal risk assessment has, therefore, fueled growing interest in data-driven and machine learning (ML)-based approaches that can use routinely available maternal, obstetric, delivery, neonatal, and early postnatal information to support clinical risk stratification [4–6].

Machine Learning for Neonatal Risk Prediction

ML-based prediction has emerged as a promising approach for identifying newborns at elevated risk of adverse outcomes. A systematic review by Mangold et al. (2021) documented the growing use of ML models for predicting neonatal mortality, underscoring the potential of computational approaches to

synthesize multiple clinical characteristics into individualized risk estimates [4]. However, the availability of a technically developed prediction model does not, by itself, establish that its underlying predictor set will demonstrate the same clinical relevance when transferred to another population or health-system context [7].

Shukla et al. (2020) developed a particularly relevant model for resource-constrained settings, using routinely collected maternal, obstetric, delivery, neonatal, and early postnatal variables to build predictive models for perinatal mortality in resource-limited settings [5]. This work provides an important foundation for using routinely collected clinical information for neonatal risk prediction and formed the basis for the candidate predictor set subsequently implemented in the Kenyan neonatal risk-prediction project.

The Kenyan implementation used a paper-based predictor set comprising 21 candidate variables, organized into maternal, pregnancy-related, delivery, neonatal, and early post-delivery domains. Importantly, the accompanying evaluation examined implementation-related characteristics of this predictor set rather than re-evaluating the original ML algorithm's predictive performance [8].

This distinction is important. A prediction model may demonstrate performance in its development or validation setting, while the empirical behavior of individual candidate predictors may vary across populations, facilities, and clinical contexts [7]. Thus, transferring a predictor set from one setting to another raises a question related to, but distinct from, formal model validation: do the candidate variables maintain observable relationships with clinically relevant outcomes in the population where they are implemented?

From Prediction Models to Contextual Clinical Signal

The growing use of ML in healthcare has increasingly highlighted the gap between technical model development and real-world clinical application. Reviews of ML implementation have emphasized that successful translation requires attention to the organizational and clinical context in which predictive technologies are introduced [9, 10]. However, implementation should not be understood solely as a question of whether a technology can be introduced into routine practice. We also need to understand the empirical clinical information generated within the implementation context and whether candidate predictors show meaningful relationships with observed clinical outcomes.

This issue is particularly relevant to neonatal risk prediction. A predictor may be routinely available and easy to record, yet its relationship to a specific outcome may vary by population, clinical environment, timing of measurement, or outcome definition. Conversely, variables that consistently associate with early neonatal outcomes may provide important empirical evidence supporting their continued consideration in context-specific risk stratification approaches.

Such evidence should not be confused with predictive model validation. Demonstrating an association between a candidate predictor and an outcome does not establish discrimination, calibration, clinical

utility, or causal influence, nor does it show that a particular combination of predictors constitutes an effective prediction model. Rather, association analysis provides an intermediate empirical layer between predictor availability and demonstrated predictive performance.

This intermediate layer is particularly relevant when an existing predictor set is contextualized within a new health system. The Kenyan implementation offers an opportunity to assess whether routinely captured candidate variables are associated with early neonatal outcomes in the local population, without making claims about the performance of the underlying ML algorithm. This analytical distinction also complements previous work from the project, which has focused separately on operational feasibility, user experience, and health-system determinants of implementation [8, 11, 12].

Early Neonatal Outcomes as Empirical Anchors

The immediate postnatal period offers several clinically observable outcomes that can be used to evaluate candidate risk indicators. Birth weight is an important neonatal characteristic associated with vulnerability during the neonatal period, while interventions such as bag-and-mask resuscitation reflect immediate clinical needs at birth. Hospitalization during the first days of life is an additional marker of early clinical concern and resource utilization.

These outcomes are particularly useful for contextual analysis because they occur during the same early period when the neonatal predictor set is captured. The Kenyan implementation collected information at delivery and during a post-delivery assessment, allowing candidate maternal, pregnancy, delivery, and neonatal characteristics to be examined alongside early neonatal outcomes [8, 11]. The protocol and subsequent implementation similarly describe the predictor set as spanning maternal characteristics, pregnancy factors, delivery characteristics, neonatal clinical indicators, and early post-delivery observations [8, 11].

Examining associations across multiple outcomes is preferable to treating a single outcome as a universal indicator of neonatal risk. Different outcomes may reflect distinct stages or mechanisms of neonatal vulnerability. For example, birth weight captures an important neonatal characteristic, whereas resuscitation reflects immediate physiological compromise, and early hospitalization may capture subsequent clinical concern. Examining candidate predictors across these distinct outcomes can, therefore, provide a more nuanced understanding of whether particular variables demonstrate consistent, outcome-specific, or limited clinical signal in the Kenyan setting.

Importantly, this approach does not assume that every candidate predictor should be associated with every outcome. A predictor originally included in a mortality prediction model may plausibly relate more strongly to one early outcome than to another. Identifying such patterns can help distinguish variables with recurrent empirical signals from those whose relevance may be outcome-specific or require further investigation.

Evidence Gap

Existing research has established the potential of ML-based approaches for predicting neonatal mortality [4, 5], while implementation research has highlighted the challenges of translating ML technologies into routine healthcare practice [9, 10]. The Kenyan neonatal risk-prediction project has subsequently generated evidence on the operational feasibility of capturing the candidate predictor set, frontline healthcare workers' usability and workflow experiences, and the governance and health-system conditions that influence longer-term scale-up [8, 11, 13].

However, these dimensions do not directly address a separate empirical question: whether the candidate predictors themselves demonstrate observable associations with early neonatal outcomes in the Kenyan implementation cohort.

This distinction is important for at least three reasons. First, successful routine capture of a variable establishes its operational availability but not its empirical relationship with an outcome. Second, favorable user perceptions establish usability and acceptability but do not provide evidence of predictor–outcome relationships. Third, organizational readiness and governance conditions determine the potential for implementation and scale-up but do not establish whether the underlying clinical variables carry locally observable risk information. These represent complementary yet analytically distinct layers of evidence.

Accordingly, a focused analysis of the empirical relationships between candidate neonatal risk predictors and early neonatal outcomes in the Kenyan context is needed. Such an analysis can provide contextual evidence on which candidate variables are associated with specific outcomes, whether some predictors show signals across multiple outcomes, and whether other candidate variables show limited evidence of association within this cohort.

Study Aim and Objectives

The aim of this study was to examine the empirical associations between candidate neonatal risk predictors and selected early neonatal outcomes among newborns receiving care across three Kenyan health facilities.

The specific objectives were to:

1. examine the associations between candidate maternal, pregnancy, delivery, and neonatal predictors and birth weight;
2. assess the associations between candidate predictors and the need for bag-and-mask resuscitation at birth;
3. examine the associations between candidate predictors and neonatal hospitalization by Day 2;

4. identify predictors demonstrating statistically significant associations with one or more early neonatal outcomes after accounting for multiple comparisons; and
5. characterize patterns of recurrent and outcome-specific predictor–outcome associations to inform contextual interpretation of the candidate risk-predictor set.

This study, therefore, occupies a distinct analytical position within the broader Kenyan neonatal risk-prediction program. Rather than reassessing the original ML model's technical performance, evaluating frontline usability, or examining governance and scale-up, it focuses specifically on the empirical clinical signal of the candidate predictors in the Kenyan population.

Methods

Study Design and Analytical Approach

This study used a quantitative, observational, analytical design to examine associations between candidate neonatal risk predictors and selected early neonatal outcomes among newborns receiving care at three Kenyan health facilities. The analysis was conducted as part of a broader prospective implementation project in which a paper-based neonatal risk predictor tool, derived from an existing machine learning model for perinatal mortality in resource-limited settings, was incorporated into routine neonatal documentation [5, 8].

The present analysis was deliberately distinct from the operational feasibility and usability evaluations conducted as part of the broader project [14, 15]. It did not assess the completeness of predictor documentation as an implementation outcome, frontline usability or user experience, or the organizational and governance determinants of scale-up. Instead, the analysis focused on the empirical relationship between candidate predictors and early neonatal outcomes observed in the implementation cohort.

The predictor set was derived from the model developed by Shukla et al. (2020), which used routinely collected maternal, obstetric, delivery, neonatal, and early postnatal data to predict perinatal mortality in resource-limited settings [5]. The Kenyan implementation incorporated 21 candidate predictors across five domains: maternal characteristics, pregnancy-related factors, delivery characteristics, neonatal clinical indicators, and early post-delivery observations. Figure 1 presents the analytical framework for the present study.

Figure 1. Analytical framework for examining candidate neonatal risk predictors and early neonatal outcomes

The framework outlines the analytical pathway from candidate predictors spanning maternal, pregnancy, delivery, neonatal, and early post-delivery domains to three selected early neonatal outcomes: birth weight, bag-and-mask resuscitation, and hospitalization by Day 2. The analysis evaluates outcome-

specific and cross-outcome statistical associations without claiming validation of the underlying machine learning model.

Study Setting

The study was conducted across three Kenyan health facilities representing different service-delivery contexts:

1. **Kibra**, Nairobi County, representing an urban informal-settlement context;
2. **Matuu**, Machakos County, representing a peri-urban public hospital context; and
3. **Nazareth**, Kiambu County, representing a faith-based referral hospital context.

These facilities were selected as the implementation sites for the broader neonatal risk-prediction project. The predictor tool was incorporated into routine neonatal documentation during a four-month implementation period. The implementation documentation reports 400 newborn cases across the three facilities: 100 at Kibra, 200 at Matuu, and 100 at Nazareth. For the present quantitative analysis, the project's analytical dataset comprised 394 observations, distributed as 197 from Matuu, 99 from Nazareth, and 98 from Kibra.

We retain the distinction between the 400 cases documented during implementation and the 394 observations in the analytical dataset because the underlying project documentation reports them as separate denominators. We do not infer or assign reasons for the six-record difference without source documentation that establishes the exclusions.

Study Population and Analytical Sample

The study population comprised newborns whose clinical information was captured using the paper-based neonatal risk predictor during routine care at the participating facilities.

The implementation tool was used prospectively throughout the four-month implementation period, with predictor information recorded at delivery/Day 1 and again during a post-delivery assessment on Day 2. The tool included 21 candidate predictors spanning maternal, pregnancy, delivery, neonatal, and early post-delivery domains.

The present analysis used the 394-record analytical dataset available from the project. The number of observations varied slightly across individual predictor–outcome analyses due to variable-specific availability, as reflected in the underlying statistical results. For example, the birth-weight analyses included between 375 and 391 observations, depending on the predictor, whereas the resuscitation and hospitalization analyses similarly used predictor-specific denominators.

We, therefore, report the analytical denominator for each association rather than assuming a single complete-case denominator across all analyses.

Candidate Neonatal Risk Predictors

The candidate predictor set was derived from the ML-based neonatal mortality prediction model reported by Shukla et al. (2020) [5]. In the Kenyan implementation, the predictor set comprised 21 candidate variables organized into five domains:

- maternal characteristics;
- pregnancy-related factors;
- delivery characteristics;
- neonatal clinical indicators; and
- early post-delivery observations.

The predictor set was incorporated into a paper-based clinical documentation tool and used during routine neonatal care.

In the present association analysis, candidate predictors were examined based on their available measurement type and their relationship with each prespecified early neonatal outcome. The project analysis included maternal and perinatal variables such as maternal age, parity, multiple birth, gestational age at enrollment, gestational age at delivery, antenatal corticosteroid exposure, antenatal hypertension/eclampsia, antepartum hemorrhage, maternal sepsis, obstructed labor, and delivery attendant. The available analyses did not necessarily evaluate every candidate predictor against every outcome; therefore, we retained outcome-specific analytical sets rather than imputing or inventing unreported comparisons.

This approach also avoids conflating the complete predictor set implemented in the clinical tool with the subset for which an inferential association was available in the current analysis.

Early Neonatal Outcomes

Three early neonatal outcomes were prespecified for the present analysis:

Birth weight

Birth weight was analyzed as a continuous neonatal outcome. It represents an immediately observable neonatal characteristic that can be examined in relation to maternal, pregnancy, and delivery predictors recorded during the same care episode.

Bag-and-mask resuscitation

Bag-and-mask resuscitation was analyzed as a binary early neonatal outcome indicating whether the newborn required immediate breathing support at birth. The project dataset recorded 42 resuscitation events in the corresponding bivariate analysis.

Hospitalization by Day 2

Neonatal hospitalization by Day 2 was analyzed as a binary outcome indicating whether the newborn required hospitalization by the second day after birth. The underlying analysis recorded 35 hospitalization events for the maternal sepsis comparison and 38 events for the gestational-age-at-delivery comparison, with the denominator varying by predictor availability.

Neonatal antibiotic administration was available as an additional descriptive Day-2 variable in the broader project dataset; however, because the supplied analytical report does not provide a corresponding inferential predictor–antibiotic analysis, it is not treated as a primary outcome in this manuscript. The project documentation confirms that antibiotic administration by Day 2 was fully recorded, but completeness alone does not establish a predictor–outcome association.

Data Preparation and Quality Control

Data preparation followed the project's existing analytical workflow. The broader dataset underwent completeness and plausibility checks before inferential analysis. The project report noted generally low missingness across the selected predictors, with somewhat higher missingness for gestational age at enrollment and antenatal corticosteroid exposure. It also flagged two records with gestational age at enrollment recorded as zero weeks for verification.

These checks are treated as data-preparation procedures rather than study outcomes. Therefore, this study does not present detailed completeness and plausibility findings as substantive results. For each inferential analysis, we used the available observations for the relevant predictor and outcome. Consequently, sample size varied across predictor–outcome comparisons.

Statistical Analysis

Descriptive analysis

Descriptive statistics were used to characterize the analytical cohort and the distribution of the selected early neonatal outcomes. Continuous variables were summarized using appropriate measures of central tendency and dispersion, while categorical outcomes were summarized using frequencies and percentages.

Because the purpose of this manuscript is inferential rather than operational, detailed site-level distributions of the 21 predictor variables are not reproduced here. Those descriptive distributions are reported separately in the project's operational analysis.

Birth-weight analysis

The relationship between candidate predictors and birth weight was assessed using Spearman rank correlation, consistent with the nonparametric bivariate analysis documented in the project report.

The resulting correlation coefficient (ρ), analytical sample size, nominal P value, and false-discovery-rate-adjusted P value were retained for interpretation. The existing analysis identified statistically significant

associations with multiple birth, gestational age at delivery, and antenatal corticosteroid exposure after multiple-comparison adjustment.

Bag-and-mask resuscitation analysis

Associations between candidate predictors and bag-and-mask resuscitation were assessed using the bivariate nonparametric comparison approach documented in the project analysis, with resuscitation status treated as the binary outcome.

The analysis compared the distributions of candidate predictors between newborns who did and did not require bag-and-mask resuscitation. The project report identifies Wilcoxon-based bivariate comparisons for this outcome and reports predictor-specific event counts, effect estimates, nominal P values, and FDR-adjusted P values.

Day-2 hospitalization analysis

The same bivariate analytical approach was used to assess hospitalization by Day 2. We compared predictor distributions between hospitalized and non-hospitalized newborns, retaining predictor-specific analytical denominators and event counts.

The existing analysis identified maternal sepsis and gestational age at delivery as the predictors demonstrating statistically clear associations after multiple-comparison adjustment.

Multiple-comparison adjustment

Because the project examined multiple candidate predictors across several outcomes, the analysis applied a false discovery rate (FDR) adjustment to control the expected proportion of false discoveries among statistically significant findings.

Both nominal and FDR-adjusted P values were retained in the analytical output. Statistical interpretation in this manuscript is based primarily on the FDR-adjusted results, rather than on nominal P values alone. A two-sided FDR-adjusted P value below 0.05 was considered statistically significant.

Interpretation of Associations

The analyses were designed to identify empirical associations, not to establish causality or validate the predictive performance of the underlying ML model.

Accordingly, we interpret statistically significant associations as evidence that a candidate predictor showed an observable relationship with a specific early neonatal outcome in this Kenyan cohort. They are not interpreted as evidence that the predictor independently causes the outcome or that including it necessarily improves model discrimination or calibration.

Similarly, we interpret the absence of a statistically significant association cautiously. A nonsignificant result may reflect a genuinely weak relationship, a limited number of outcome events, measurement

characteristics, or insufficient statistical precision in the available sample.

This distinction matters because the original ML model demonstrated predictive performance in its development context, whereas the present study examines the empirical behavior of candidate predictors within that context rather than externally validating model performance. The broader implementation manuscript similarly distinguishes evaluation of the predictor set from direct reassessment of the ML algorithm itself.

Cross-Outcome Analytical Synthesis

After outcome-specific analyses, statistically significant predictor–outcome relationships were considered together to identify whether particular predictors demonstrated:

1. **cross-outcome associations**, where a predictor showed statistically significant relationships with more than one outcome;
2. **outcome-specific associations**, where a predictor demonstrated evidence of association with only one outcome; or
3. **no statistically significant association** with the outcomes examined.

This synthesis is descriptive and interpretive, not a new statistical test. It is intended to identify patterns in the empirical signal across different early neonatal outcomes while avoiding the implication that effect estimates from different outcome types are directly interchangeable.

Reporting and Analytical Transparency

The manuscript reports the analytical denominators for each predictor–outcome comparison rather than assuming that all 394 records contributed to every analysis. This approach reflects the variable-specific sample sizes documented in the underlying project analysis.

The complete candidate predictor set and the implementation tool are described in the associated project documentation and the companion protocol/manuscript. The present manuscript focuses on the inferential association analysis and does not reproduce the operational feasibility, usability, or governance analyses reported elsewhere.

Results

Analytical cohort

The analytical dataset comprised 394 newborn observations from three participating Kenyan health facilities: Matuu (n = 197), Nazareth (n = 99), and Kibra (n = 98). The number of observations available for individual predictor–outcome comparisons varied by predictor-specific data availability. Overall, missingness was low across the variables included in the analysis.

The analysis examined three early neonatal outcomes: birth weight, bag-and-mask resuscitation at birth, and hospitalization by Day 2. Birth weight was treated as a continuous outcome, whereas bag-and-mask resuscitation and hospitalization were treated as binary outcomes. The underlying analysis applied FDR adjustment to account for multiple comparisons.

Table 1 summarizes the analytical sample and outcome-specific denominators.

The overall dataset contained 394 observations. Predictor-specific denominators were retained from the underlying inferential analyses rather than imposing a single complete-case denominator across all tests. The source analysis reports 42 resuscitation events and predictor-specific hospitalization counts ranging from 34 to 39.

Associations between candidate predictors and birth weight

The first outcome-specific analysis examined the relationship between candidate predictors and birth weight using Spearman rank correlation.

After FDR adjustment, three predictors were statistically significantly associated with birth weight: multiple birth, gestational age at delivery, and antenatal corticosteroid exposure. Multiple birth was negatively correlated with birth weight ($\rho = -0.202$, $n = 391$, FDR-adjusted $P < 0.001$), whereas gestational age at delivery was positively correlated with birth weight ($\rho = 0.183$, $n = 387$, FDR-adjusted $P = 0.002$). Antenatal corticosteroid exposure was also negatively correlated with birth weight ($\rho = -0.163$, $n = 379$, FDR-adjusted $P = 0.006$).

The remaining predictors showed small correlations that did not remain statistically significant following FDR adjustment. These included parity, maternal age, obstructed labor, teenage pregnancy, antenatal hypertension/eclampsia, antepartum hemorrhage, delivery attendant, maternal sepsis, and gestational age at enrollment. Their FDR-adjusted P values ranged from 0.239 to 0.829. Table 2 presents the complete birth-weight analysis.

Associations with bag-and-mask resuscitation

The second outcome analysis examined whether candidate predictors differed according to bag-and-mask resuscitation status. The corresponding analysis included 42 recorded resuscitation events.

Two predictors demonstrated statistically significant associations after FDR adjustment: maternal sepsis and antenatal hypertension/eclampsia. Maternal sepsis was associated with resuscitation among 386 observations (39 events; nominal $P = 0.004$; FDR-adjusted $P = 0.030$). Antenatal hypertension/eclampsia was also associated with resuscitation ($n = 391$; 42 events; nominal $P = 0.005$; FDR-adjusted $P = 0.030$).

Obstructed labor showed a nominal association ($P = 0.042$), but this did not remain statistically significant after FDR adjustment (adjusted $P = 0.168$). The other examined predictors likewise did not demonstrate statistically significant associations following adjustment. Table 3 presents the complete analysis.

This outcome-specific pattern indicates that maternal complication-related variables provided the strongest detectable associations with immediate resuscitation, while the broader demographic and delivery-related variables examined did not retain statistical significance after multiple-comparison adjustment.

Associations with hospitalization by Day 2

The third outcome analysis examined predictors associated with neonatal hospitalization by Day 2.

Two predictors demonstrated statistically significant associations after FDR adjustment: maternal sepsis and gestational age at delivery. Maternal sepsis showed a strong association with hospitalization ($n = 386$; 35 events; nominal $P < 0.001$; FDR-adjusted $P = 2.37 \times 10^{-5}$). Gestational age at delivery was also significantly associated with hospitalization ($n = 390$; 38 events; nominal $P < 0.001$; FDR-adjusted $P = 0.00018$). The source analysis reported a median gestational-age difference of approximately one week between the groups.

Antenatal corticosteroid exposure showed a nominal association ($P = 0.0265$), but this association did not remain statistically significant after FDR adjustment (adjusted $P = 0.106$). The remaining predictors also did not demonstrate statistically significant associations after adjustment. Table 4 presents the complete hospitalization analysis.

Integrated pattern of predictor–outcome associations

Taken together, the three outcome-specific analyses identified seven statistically significant predictor–outcome associations involving five distinct candidate predictors.

Two predictors demonstrated cross-outcome associations. Maternal sepsis was associated with both bag-and-mask resuscitation and hospitalization by Day 2, whereas gestational age at delivery was associated with both birth weight and hospitalization.

Three predictors demonstrated outcome-specific associations. Multiple birth and antenatal corticosteroid exposure were associated with birth weight but not with either of the two binary outcomes after FDR adjustment. Antenatal hypertension/eclampsia was associated with bag-and-mask resuscitation but not with birth weight or Day-2 hospitalization. Figure 2 shows the integrated pattern.

Figure 2. Cross-outcome pattern of statistically significant candidate predictor associations

The figure maps candidate predictors against the three early neonatal outcomes and identifies associations that remain statistically significant after false-discovery-rate adjustment. Predictors

associated with two outcomes are visually distinguished from those associated with only one outcome. The figure summarizes the pattern of empirical signal across outcomes rather than comparing effect magnitudes across different statistical tests.

The integrated pattern shows that maternal sepsis and gestational age at delivery were the only predictors associated with both early neonatal outcomes, whereas multiple birth, antenatal corticosteroid exposure, and antenatal hypertension/eclampsia showed outcome-specific associations.

Summary of Findings

Across the three early neonatal outcomes, seven statistically significant predictor–outcome associations involving five candidate predictors remained after FDR adjustment.

For birth weight, significant associations were observed with multiple birth, gestational age at delivery, and antenatal corticosteroid exposure. The associations were negative for multiple birth and antenatal corticosteroid exposure and positive for gestational age at delivery.

For bag-and-mask resuscitation, significant associations were observed with maternal sepsis and antenatal hypertension/eclampsia. No other examined predictor remained statistically significant after FDR adjustment.

For hospitalization by Day 2, significant associations were observed with maternal sepsis and gestational age at delivery. No other examined predictor remained statistically significant after FDR adjustment.

Overall, maternal sepsis and gestational age at delivery were each associated with two of the three outcomes, whereas multiple birth, antenatal corticosteroid exposure, and antenatal hypertension/eclampsia were each associated with one outcome only.

Discussion

Principal Findings

This study identified seven statistically significant predictor–outcome associations involving five candidate predictors across three early neonatal outcomes. After false discovery rate (FDR) adjustment, multiple birth, gestational age at delivery, and antenatal corticosteroid exposure were associated with birth weight; maternal sepsis and antenatal hypertension/eclampsia were associated with bag-and-mask resuscitation; and maternal sepsis and gestational age at delivery were associated with hospitalization by Day 2.

The pattern is notable because the statistically significant predictors were not uniformly associated with all outcomes. Maternal sepsis and gestational age at delivery demonstrated cross-outcome associations, whereas multiple birth, antenatal corticosteroid exposure, and antenatal

hypertension/eclampsia demonstrated outcome-specific associations. This suggests that the empirical signal in the candidate predictor set may vary by clinical outcome and time point.

This finding is important in the context of neonatal machine-learning research. The systematic review by Mangold et al. (2021) identified substantial heterogeneity in neonatal mortality prediction models with respect to the number and type of predictors, timing of prediction, algorithms, validation strategies, and performance measures [4]. Only a minority of the reviewed studies conducted external validation or reported calibration, highlighting the difficulty of translating apparently promising prediction models across populations and settings. Shukla et al. (2020), whose model underpins the predictor set examined here, similarly showed that prediction of perinatal mortality in resource-limited settings improves when prediction models incorporate information collected after delivery [5].

The present findings should, therefore, be interpreted as context-specific empirical evidence concerning candidate predictor–outcome relationships, rather than as validation of the original machine-learning model or evidence that any individual predictor is causally responsible for an outcome.

Gestational Age at Delivery and Birth Weight

Gestational age at delivery demonstrated a positive association with birth weight ($\rho = 0.183$; FDR-adjusted $P = 0.002$). This is biologically coherent and consistent with the established relationship between gestational maturity and fetal growth. The finding is also compatible with the broader neonatal prediction literature, in which gestational age is repeatedly incorporated among clinically relevant predictors of neonatal morbidity and mortality [4, 5].

Gestational age is particularly important in neonatal prediction because it provides information relatively early in the care pathway and can, therefore, inform risk assessment before subsequent neonatal complications become evident. Shukla et al. (2020) showed that maternal and post-delivery variables contribute differently to predicting neonatal death, emphasizing the importance of when predictor information becomes available [5]. The present finding extends this logic by showing that gestational age at delivery was not only statistically associated with birth weight but also independently associated with hospitalization by Day 2.

However, the correlation with birth weight was modest. This is important because statistical significance should not be equated with strong predictive power. The relatively large analytical sample makes it possible to detect associations that may still have limited explanatory power when considered individually. The result should, therefore, be understood as evidence of a consistent relationship rather than as evidence that gestational age alone is sufficient to characterize neonatal risk.

The association between gestational age at delivery and hospitalization further underscores the relevance of gestational maturity to early neonatal clinical needs. Newborns delivered at lower gestational ages often require closer monitoring and additional supportive care, although the present bivariate analysis cannot establish that gestational age independently caused hospitalization. The

observed association may also reflect correlated factors, such as respiratory adaptation, feeding difficulty, infection risk, or other clinical conditions associated with earlier delivery.

Multiple Birth and Birth Weight

Multiple birth was negatively associated with birth weight ($p = -0.202$; FDR-adjusted $P < 0.001$). Among the statistically significant associations identified in this study, this was one of the strongest correlations.

The direction of this association is clinically plausible. Multiple gestations are characterized by shared intrauterine resources and are more often associated with restricted fetal growth and lower birth weight than singleton pregnancies. Within a predictive framework, multiple birth can, therefore, inform assessments of neonatal vulnerability even before postnatal clinical deterioration becomes apparent.

The finding is also relevant to the distinction between risk prediction and causal explanation. Multiple birth is not a mechanism that explains every subsequent adverse neonatal outcome. Rather, it is a readily observable characteristic that can summarize several underlying pathways associated with neonatal vulnerability. This aligns with the rationale for incorporating routinely observable maternal and obstetric characteristics into prediction models for neonatal mortality in resource-limited settings [4, 5].

Importantly, multiple birth was not significantly associated with bag-and-mask resuscitation or hospitalization by Day 2 after FDR adjustment. This outcome-specific pattern illustrates why the present analysis should not be interpreted as an exercise in identifying universally important predictors. Instead, it demonstrates that the same predictor can carry detectable information for one neonatal outcome while showing insufficient evidence of association with another.

Antenatal Corticosteroid Exposure: Association Versus Causation

Antenatal corticosteroid exposure was negatively associated with birth weight ($p = -0.163$; FDR-adjusted $P = 0.006$). At first sight, this finding might appear counterintuitive because antenatal corticosteroids are an evidence-based intervention intended to improve outcomes among women at risk of preterm birth. However, the interpretation should account for the clinical indication for corticosteroid administration.

The most plausible explanation is confounding by indication. Antenatal corticosteroids are more likely to be administered when clinicians anticipate preterm delivery or another pregnancy complication associated with increased neonatal risk. Consequently, corticosteroid exposure may function as a marker of an underlying high-risk pregnancy rather than as an independent cause of lower birth weight. The project analysis itself recognized that corticosteroids were more frequently administered in pregnancies at risk of preterm delivery.

This interpretation is also supported by the distribution of corticosteroid use across the study sites, which varied substantially between facilities. The underlying survey material describes considerably greater use at Nazareth than at Kibra and Matuu and suggests that differences may reflect case mix, referral patterns, recognition of preterm risk, or clinical practice.

The finding, therefore, provides an important methodological lesson for interpreting candidate predictors in machine-learning systems: a statistically informative predictor does not necessarily represent a modifiable causal determinant. In predictive modeling, a variable may improve risk estimation precisely because it captures information about an underlying clinical process. Removing or disregarding such a predictor simply because its association appears counterintuitive could, therefore, reduce predictive information. Conversely, interpreting the association causally could lead to inappropriate clinical conclusions.

The fact that antenatal corticosteroid exposure was not significantly associated with resuscitation or hospitalization after FDR adjustment further supports an outcome-specific interpretation rather than a broad claim about neonatal morbidity.

Maternal Sepsis and Bag-and-Mask Resuscitation

Maternal sepsis was significantly associated with bag-and-mask resuscitation after FDR adjustment ($P = 0.030$). Importantly, this association occurred alongside a separate association between maternal sepsis and hospitalization by Day 2.

The recurrence of maternal sepsis across two clinically different outcomes provides stronger empirical evidence than an isolated association with a single endpoint. In this cohort, maternal sepsis was associated with an immediate neonatal intervention at birth and with subsequent inpatient care within the first two days of life.

The biological plausibility of this pattern is strong. Maternal infection and sepsis can be associated with fetal and neonatal compromise through inflammatory, infectious, and obstetric pathways. Neonates exposed to maternal infection may require closer observation and treatment because of concerns about infection, respiratory compromise, or other consequences of perinatal illness. The broader neonatal literature similarly recognizes maternal and obstetric complications as important sources of information for early neonatal risk assessment [4, 5].

Nevertheless, the present analysis cannot establish the pathway linking maternal sepsis to resuscitation. The association could reflect direct effects of maternal infection, associated obstetric complications, differences in gestational age, delivery circumstances, or combinations of these factors. Because the analysis was bivariate, it cannot determine whether maternal sepsis would remain independently associated with resuscitation after adjustment for correlated clinical variables.

The appropriate interpretation is, therefore, that maternal sepsis demonstrated a reproducible statistical association with immediate neonatal resuscitation in this cohort, rather than that maternal sepsis

independently caused the need for resuscitation.

Maternal Sepsis and Hospitalization by Day 2

Maternal sepsis showed the strongest statistical association observed in the study with hospitalization by Day 2 (FDR-adjusted $P = 2.37 \times 10^{-5}$).

This finding is particularly relevant because hospitalization represents a later early-neonatal outcome than resuscitation. The association, therefore, extends from immediate neonatal support to subsequent clinical care needs.

The consistency of the maternal-sepsis signal across these two outcomes is more informative than treating either result independently. Nevertheless, the strength of the statistical evidence should not be confused with the magnitude of an independent clinical effect. The current analysis does not provide an adjusted effect estimate or establish whether maternal sepsis independently predicts hospitalization after accounting for gestational age, delivery characteristics, or neonatal condition.

This distinction is especially important for interpreting machine-learning predictors. Shukla et al. (2020) showed that predictive performance can change with the timing and availability of clinical information, while Mangold et al. (2021) emphasized heterogeneity in predictor selection and validation among neonatal ML models [4]. A predictor that shows an association with an early outcome in one analysis should, therefore, not be interpreted as having the same predictive contribution to neonatal mortality.

Antenatal Hypertension/Eclampsia and Resuscitation

Antenatal hypertension/eclampsia was significantly associated with bag-and-mask resuscitation after FDR adjustment ($P = 0.030$), but not with birth weight or hospitalization by Day 2.

The outcome-specific nature of this association is important. Hypertensive disorders of pregnancy can be associated with placental dysfunction, impaired fetal growth, preterm delivery, and fetal compromise. These pathways could plausibly contribute to an increased need for immediate neonatal support, even when the same predictor does not demonstrate a detectable association with the other outcomes examined here.

The absence of a significant association with birth weight should not be interpreted as evidence that hypertensive disorders are unrelated to fetal growth. Rather, it indicates that within this cohort and under the present bivariate analysis, the association with birth weight did not remain significant after FDR adjustment. Similarly, the absence of an association with hospitalization may reflect the relatively small number of hospitalization events and the multiple factors influencing the decision to admit a newborn [6].

This illustrates one strength of examining several outcomes: the analysis can reveal where the empirical signal appears and where it does not, rather than collapsing all neonatal morbidity into a single endpoint.

Cross-Outcome Versus Outcome-Specific Signals

The integrated findings reveal two distinct patterns. First, maternal sepsis and gestational age at delivery were cross-outcome signals, each demonstrating significant associations with two different outcomes. Second, multiple birth, antenatal corticosteroid exposure, and antenatal hypertension/eclampsia were outcome-specific signals, each associated with only one outcome after FDR adjustment.

This distinction matters for interpreting candidate predictors. A conventional analysis focused on a single outcome could have missed that maternal sepsis and gestational age at delivery were associated across different stages of early neonatal care. Conversely, an analysis that assumed all predictors should behave similarly across outcomes could have incorrectly interpreted the outcome-specific associations as inconsistent or unimportant.

The cross-outcome pattern is particularly relevant to the original purpose of the predictor set. Shukla et al. (2020) developed prediction models for perinatal mortality using information available at different points in the maternal and neonatal care pathway [5]. The present findings suggest that individual candidate variables may likewise contain information that manifests differently depending on the outcome and time horizon.

However, cross-outcome association should not be interpreted as cross-outcome predictive validity. Demonstrating an association with two outcomes does not establish that a predictor improves discrimination, calibration, or clinical utility for either outcome. These questions require formal prediction-model evaluation.

Implications for Interpretation of ML-Derived Predictor Sets

The findings have an important implication for how ML-derived neonatal predictor sets should be evaluated after transfer into a new clinical population.

The original model was developed using a large multicountry dataset and showed that incorporating post-delivery information could improve neonatal mortality prediction. The systematic review by Mangold et al. (2021), however, showed that neonatal ML prediction studies vary substantially in their predictor sets, prediction time points, validation procedures, and reporting of calibration and external validation [4].

Our findings add an intermediate form of evidence: the empirical relationships of individual candidate predictors with clinically observable early neonatal outcomes in the implementation population.

This is valuable, but it is deliberately narrower than model validation. A predictor can demonstrate a statistically significant bivariate association without making a meaningful contribution to a multivariable prediction model. Conversely, a predictor can contribute to prediction through interactions, nonlinear relationships, or correlations with other variables even when its simple bivariate association with an individual outcome is weak.

Therefore, these findings should not be used to remove non-significant predictors from the underlying model or to redesign the model solely on the basis of the present analysis. Such decisions would require formal model redevelopment or updating using appropriate predictive-modeling methodology.

This distinction is consistent with broader concerns about the implementation gap in healthcare ML. Seneviratne, Shah, and Chu (2020) emphasized that successful translation requires more than model development, while Preti et al. (2024) found that implementing ML applications in healthcare organizations involves technical, organizational, and contextual considerations [9, 10]. Mangold et al. (2021) similarly highlighted the need for external validation and calibration in neonatal prediction research [4].

Contextual Relevance for LMIC Neonatal Care

The findings are particularly relevant to neonatal prediction in LMIC settings because they show that routinely observable maternal and neonatal characteristics can be measurably related to early clinical outcomes, even within a relatively small, multisite implementation cohort [8, 11].

This contextual dimension matters because neonatal prediction models developed in one population may not behave the same way when transferred to another. Differences in case mix, referral pathways, clinical practice, timing of presentation, documentation, and access to interventions can change both the distribution of predictors and their relationship with outcomes.

The original Shukla model was deliberately developed in resource-limited settings, making it more contextually relevant than models developed exclusively in high-income tertiary-care environments. Nevertheless, the current findings demonstrate why local empirical assessment remains necessary even when the underlying model was developed for resource-limited settings.

This is consistent with the wider LMIC digital-health literature, which emphasizes contextual adaptation, infrastructure, workforce capacity, and integration with existing health systems [16–20]. The issue is, therefore, not simply whether a model can be transferred technically, but whether its predictors retain meaningful clinical relationships in the population and care environment in which the model is intended to operate.

Positioning the Study Within the Broader Evidence Pathway

The present study occupies a specific position between model development and clinical implementation.

The original ML model provides the predictive foundation [5]. The present analysis examines the empirical behavior of selected candidate predictors in the Kenyan implementation population. Further evidence is needed to establish whether the complete model demonstrates adequate discrimination and calibration in this population, whether predictions provide clinically useful information, and whether their use improves care or neonatal outcomes.

This distinction matters because implementing ML in healthcare is not a single-stage process. Seneviratne, Shah and Chu (2020) described the gap between model development and clinical implementation, while Preti et al. (2024) identified organizational and contextual determinants of ML implementation [9, 10]. McCool et al. (2022) and Husain et al. (2025) further emphasize the importance of contextual and health-system considerations in digital-health translation in LMICs [16, 17].

Strengths and Limitations

This study has several strengths. First, it examined three clinically distinct early neonatal outcomes, allowing the empirical behavior of candidate predictors to be considered across different dimensions of early neonatal risk. Second, the analysis included multiple candidate predictors and applied FDR adjustment, reducing the likelihood that the observed significant findings were attributable solely to multiple testing. Third, the data were generated across three Kenyan health facilities rather than a single institution, providing some contextual diversity.

The study also has important limitations. The most important is that the analyses were bivariate. Consequently, the findings do not establish independent associations and cannot distinguish predictor effects from confounding or correlation among clinical variables. This is particularly relevant for antenatal corticosteroid exposure, which is strongly linked to the clinical indication for treatment and, therefore, should not be interpreted causally.

Second, the binary outcomes had relatively few events. The relevant analysis included 42 resuscitation events and 35–38 hospitalization events, depending on the predictor-specific denominator. This limited statistical precision and may have contributed to failure to detect weaker associations.

Third, the analytical denominators varied across predictors because of differences in available observations. Fourth, the study examined early neonatal outcomes rather than the original model's primary mortality endpoint. Consequently, the present findings should not be interpreted as evidence of improved neonatal-mortality prediction.

Fifth, statistical significance does not establish clinical importance. Some observed associations were modest in magnitude, and the current analysis does not provide the information required to determine whether incorporating any individual predictor would materially improve prediction performance.

Finally, the study was conducted within three Kenyan facilities over a defined implementation period. External validation in other Kenyan facilities and populations will, therefore, be necessary before conclusions about broader transportability can be made.

Implications for Future Research

The findings provide a basis for several specific next steps. First, future studies should conduct multivariable analyses to determine whether the observed associations remain after adjustment for clinically relevant covariates. This would be particularly important for maternal sepsis, gestational age at

delivery, and antenatal corticosteroid exposure, where clinical relationships are likely to be interdependent.

Second, the complete ML predictor set should undergo external validation in the Kenyan population, including discrimination and calibration assessment. This is particularly important because the systematic review by Mangold et al. (2021) found that external validation and calibration were relatively uncommon among neonatal ML prediction studies [4].

Third, model updating may be warranted if systematic differences in predictor distributions or outcome relationships are identified. Such updating should be based on formal prediction-model methodology rather than simply deleting predictors that were not statistically significant in the present bivariate analysis.

Fourth, prospective studies should evaluate clinical utility and decision impact. A predictor can be statistically associated with an outcome without improving clinical decisions. Ultimately, the value of neonatal risk prediction depends on whether identification of elevated risk leads to timely and appropriate clinical action.

Finally, future studies should examine whether the observed cross-outcome signals, particularly maternal sepsis and gestational age at delivery, remain stable across facilities, populations, and longer observation periods.

Conclusion

This study demonstrates that selected candidate neonatal risk predictors show distinct, outcome-dependent empirical associations with early neonatal outcomes in a Kenyan multisite cohort. After FDR adjustment, seven statistically significant associations involving five candidate predictors remained. Maternal sepsis and gestational age at delivery demonstrated associations across two outcomes, whereas multiple birth, antenatal corticosteroid exposure, and antenatal hypertension/eclampsia demonstrated outcome-specific associations.

The findings provide contextual empirical evidence, rather than causal evidence or validation of the underlying ML model. Their principal contribution is to show that candidate predictors can behave differently depending on the neonatal outcome examined, reinforcing the importance of local, outcome-specific assessment when translating prediction models into new clinical environments.

Formal external validation, multivariable modeling, calibration assessment, and prospective evaluation of clinical utility are required before these associations can be translated into claims about predictive performance or improved neonatal care.

Abbreviations

FDR

false discovery rate

LMICs

low- and middle-income countries

ML

machine learning

WHO

World Health Organization

Declarations

Ethics approval and consent to participate

Ethical approval for the study was obtained from the Strathmore University Institutional Scientific and Ethical Review Committee (SU-ISERC2949/25) on July 28, 2025. The study was conducted in accordance with applicable Kenyan legislation and national ethical guidelines governing research involving human participants, including the Health Act, No. 21 of 2017, and the Science, Technology and Innovation Act, No. 28 of 2013. The study also complied with the approved ethical procedures and relevant regulatory requirements, including approved ethical amendments where applicable. Informed consent was obtained from participants before their involvement in the study, and privacy and confidentiality were maintained by omitting identifying information.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analyzed during this study are not publicly available because they contain potentially sensitive maternal and neonatal clinical information and are subject to institutional and ethical restrictions. Data may be available from the corresponding author or the relevant institutional authority upon reasonable request and subject to appropriate ethical, institutional, and data-governance requirements.

Competing interests

The authors declare no competing interests.

Authors' contributions

DN conceived and coordinated the study, contributed to study design, data collection, data analysis, and interpretation of the findings, and contributed to manuscript preparation. **EK, FS, BM,** and **PM** contributed to study implementation, data collection, data management, interpretation, and critical revision of the manuscript. **MSRJ** contributed to the study conceptualization and methodology, interpretation of the findings, manuscript development, and critical revision of the manuscript. All authors reviewed and approved the final manuscript and agree to be accountable for all aspects of the work.

Acknowledgements

The authors acknowledge the participating health facilities and healthcare professionals who supported the study and contributed to generating and collecting the clinical data. The authors also thank the study participants and their families for their contribution to the research.

Funding statements

A Frontiers Seed Funding (Grant No. FS-2425-23-127) from the Royal Academy of Engineering has been received to support this study.

AI disclosure

Generative artificial intelligence tools were used during manuscript preparation for language refinement, figure generation, and editorial support. The authors retained full responsibility for the study design, data analysis, interpretation of findings, scientific content, and final manuscript. No generative artificial intelligence tool was used to independently generate, analyze, or interpret the underlying study data.

References

1. Lawn JE et al (2023) Ending Preventable Neonatal Deaths: Multicountry Evidence to Inform Accelerated Progress to the Sustainable Development Goal by 2030. *Neonatology* 120(4):491–499
2. World Health Organization. Newborn mortality (2024) [cited 2026 21st February]; Available from: [https://www.who.int/news-room/fact-sheets/detail/newborn-mortality?](https://www.who.int/news-room/fact-sheets/detail/newborn-mortality?hl=en)
3. United Nations Inter-Agency Group for Child Mortality Estimation (UN IGME), Levels and trends in child mortality, U.N.C.s. Fund, Editor (2024) : New York
4. Mangold C et al (2021) Machine Learning Models for Predicting Neonatal Mortality: A Systematic Review. *Neonatology* 118(4):394–405
5. Shukla VV et al (2020) Predictive Modeling for Perinatal Mortality in Resource-Limited Settings. *JAMA Netw Open* 3(11):e2026750
6. Adlung L et al (2021) Machine learning in clinical decision making. *Med* 2(6):642–665

7. de Hond AAH et al (2022) Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *npj Digit Med* 5(1):2
8. Nyatuka RD, Siva F, Jabin MSR (2026) Governance pathways for scaling neonatal risk stratification tools in LMIC health systems: a multi-site qualitative study in Kenya. *Front Digit Health*, Volume 8–2026
9. Preti LM et al (2024) Implementation of Machine Learning Applications in Health Care Organizations: Systematic Review of Empirical Studies. *J Med Internet Res* 26:e55897
10. Seneviratne MG, Shah NH, Chu L (2020) Bridging the implementation gap of machine learning in healthcare. *BMJ Innovations* 6(2):45
11. Nyatuka RD et al (2026) Kenyan Neonatal Mortality Risk Predictor: Protocol for a User-Centered Design Evaluation. *JMIR Res Protoc* 15:e81996
12. Jabin MSR, Nyatuka RD (2026) Assessing the feasibility of solar-powered digital health infrastructure in rural and peri-urban Kenya: a mixed-methods study. *Front Digit Health*, Volume 8–2026
13. Lowry S et al (2014) *Integrating Electronic Health Records Into Clinical Workflow: An Application of Human Factors Modeling Methods to Two Specialty Care Areas*. Proceedings of the International Symposium of Human Factors and Ergonomics in Healthcare, 3: pp. 170–177
14. Hajesmaeel-Gohari S et al (2022) The most used questionnaires for evaluating satisfaction, usability, acceptance, and quality outcomes of mobile health. *BMC Med Inf Decis Mak* 22(1):22
15. Sauro J (2019) *10 Things to Know About the Post Study System Usability Questionnaire*. [cited 2026 22nd February]; Available from: <https://measuringu.com/pssuq/>
16. McCool J et al (2022) Mobile Health (mHealth) in Low- and Middle-Income Countries. *Annu Rev Public Health* 43:525–539
17. Husain L et al (2025) Bridging Implementation Gaps in Digital Health: A Translational Research Imperative for Equitable Healthcare Innovation. *Clin Transl Sci* 18(10):e70375
18. Jabin MSR, Nyatuka RD (2026) Bridging the data–power paradox: a conceptual–empirical analysis of solar-powered digital health systems in low-resource settings. *Front Digit Health*, Volume 8–2026
19. Nyatuka D, Jabin MSR, Dionne-Morris L (2026) Contextual Barriers to Health Information Systems Optimization in Underserved Communities in Kenya: Qualitative Study Informed by Frugal Innovation and Information and Communication Technologies for Development. *JMIR Form Res* 10:e78950
20. Nyatuka RD, Jabin MSR, Dionne-Morris L (2026) Frugal-Oriented Information and Communication Technology for Development Framework Toward Low-Cost Digital Maternal Health in Low- and Middle-Income Countries: Quantitative Descriptive Study. *JMIR Form Res* 10:e77330

Tables

Table 1
Analytical cohort and outcome-specific sample sizes

Analytical component	Overall analytical dataset	Predictor-specific analytical range
Newborn observations	394	—
Matuu	197	—
Nazareth	99	—
Kibra	98	—
Birth-weight analysis	—	375–391
Bag-and-mask resuscitation analysis	—	378–394
Hospitalization by Day 2 analysis	—	378–394
Bag-and-mask resuscitation events	42	36–42*
Hospitalization events	—	34–39*

NB: Event counts varied by predictor because of the number of available observations.

Table 2
Associations between candidate predictors and birth weight

Candidate predictor	n	Spearman ρ	Nominal P	FDR-adjusted P
Multiple births	391	-0.202	< 0.001	< 0.001
Gestational age at delivery	387	0.183	< 0.001	0.002
Antenatal corticosteroids	379	-0.163	0.001	0.006
Parity	388	0.089	0.080	0.239
Maternal age	389	0.070	0.171	0.401
Obstructed labor	389	-0.065	0.200	0.401
Teenage pregnancy	389	-0.050	0.328	0.495
Antenatal hypertension/eclampsia	388	-0.050	0.330	0.495
Antepartum hemorrhage	389	0.027	0.600	0.800
Delivery attendant	389	-0.017	0.740	0.829
Maternal sepsis	383	0.016	0.760	0.829
Gestational age at enrollment	375	0.011	0.829	0.829

NB: Bold indicates FDR-adjusted $P < 0.05$. Negative coefficients indicate an inverse association and positive coefficients indicate a positive association. The values are derived directly from the bivariate analysis reported in the project dataset.

Table 3
Associations between candidate predictors and bag-and-mask resuscitation

Candidate predictor	n	Events	Effect*	Nominal P	FDR-adjusted P
Maternal sepsis	386	39	0	0.004	0.030
Antenatal hypertension/eclampsia	391	42	0	0.005	0.030
Obstructed labor	392	42	0	0.042	0.168
Antepartum hemorrhage	392	41	0	0.173	0.520
Gestational age at delivery	390	41	1	0.247	0.592
Maternal age	392	42	-1	0.438	0.739
Antenatal corticosteroids	382	36	0	0.447	0.739
Teenage pregnancy	392	42	0	0.493	0.739
Parity	391	42	0	0.576	0.767
Delivery attendant	392	41	0	0.851	0.940
Multiple births	394	42	0	0.866	0.940
Gestational age at enrollment	378	42	0.5	0.940	0.940

NB: Bold indicates FDR-adjusted $P < 0.05$. Negative coefficients indicate an inverse association and positive coefficients indicate a positive association. The values are derived directly from the bivariate analysis reported in the project dataset.

Table 4
Associations between candidate predictors and hospitalization by Day 2

Candidate predictor	n	Events	Effect*	Nominal P	FDR-adjusted P
Maternal sepsis	386	35	0	< 0.001	< 0.001
Gestational age at delivery	390	38	-1	< 0.001	0.00018
Antenatal corticosteroids	382	34	0	0.0265	0.106
Antenatal hypertension/eclampsia	391	37	0	0.106	0.264
Delivery attendant	392	37	0	0.130	0.264
Antepartum hemorrhage	392	38	0	0.132	0.264
Obstructed labor	392	38	0	0.270	0.462
Multiple births	394	39	0	0.353	0.529
Parity	391	39	-1	0.402	0.537
Teenage pregnancy	392	39	0	0.682	0.748
Gestational age at enrollment	378	37	2	0.686	0.748
Maternal age	392	39	-1	0.887	0.887

NB: The source report labels this column "Effect"; because the underlying output does not provide a clinically interpretable standardized effect measure, these values are not interpreted as effect sizes. Bold indicates FDR-adjusted $P < 0.05$.

Figures

Analytical Framework for Evaluating Candidate Neonatal Risk Predictors and Early Neonatal Outcomes

Observational association analysis within a prospective implementation cohort in three Kenyan health facilities

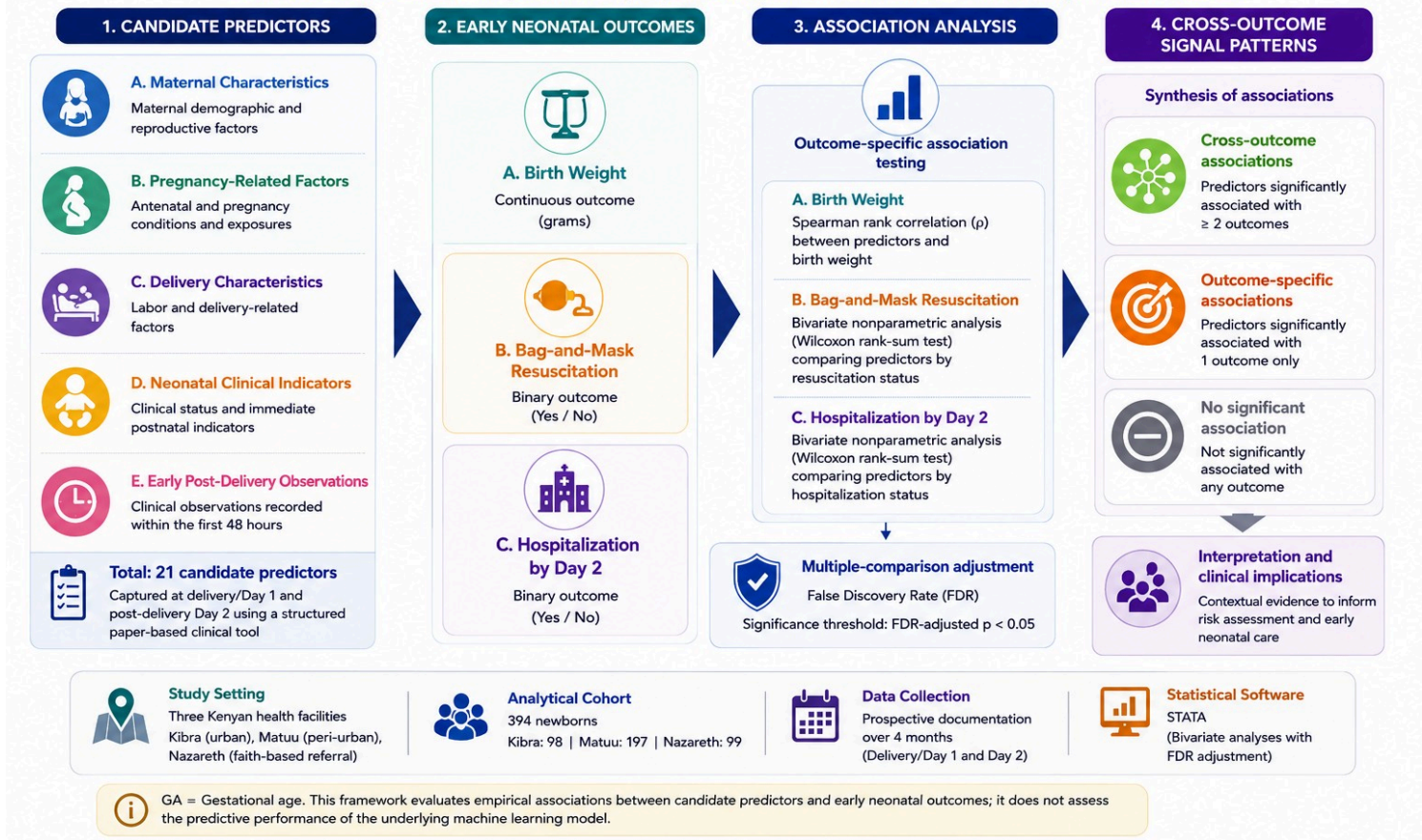


Figure 1

Analytical framework for examining candidate neonatal risk predictors and early neonatal outcomes

The framework outlines the analytical pathway from candidate predictors spanning maternal, pregnancy, delivery, neonatal, and early post-delivery domains to three selected early neonatal outcomes: birth weight, bag-and-mask resuscitation, and hospitalization by Day 2. The analysis evaluates outcome-specific and cross-outcome statistical associations without claiming validation of the underlying machine learning model.

Cross-outcome pattern of statistically significant candidate neonatal risk predictor associations

Associations that remained significant after false discovery rate (FDR) adjustment ($q < 0.05$)

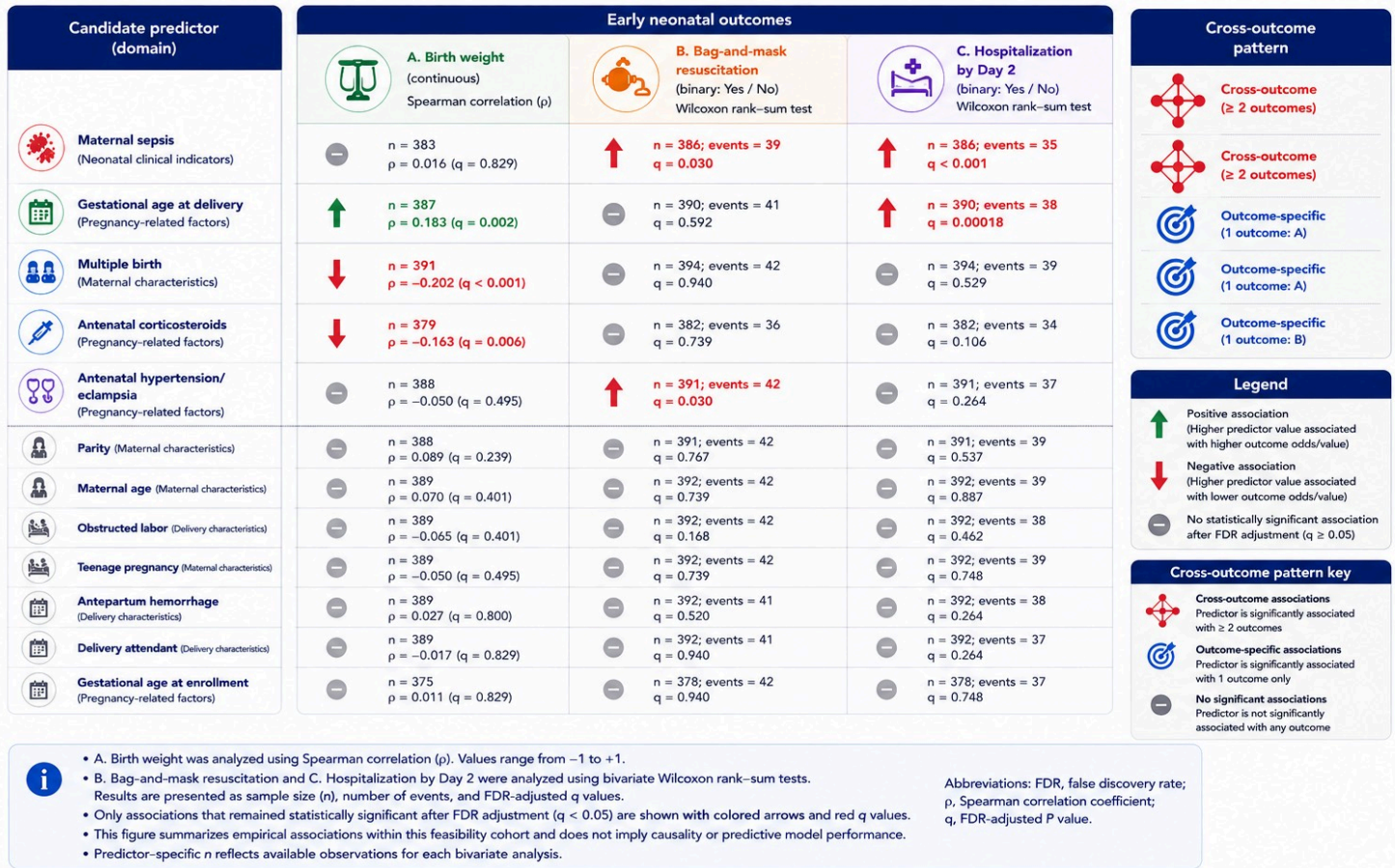


Figure 2

Cross-outcome pattern of statistically significant candidate predictor associations

The figure maps candidate predictors against the three early neonatal outcomes and identifies associations that remain statistically significant after false-discovery-rate adjustment. Predictors associated with two outcomes are visually distinguished from those associated with only one outcome. The figure summarizes the pattern of empirical signal across outcomes rather than comparing effect magnitudes across different statistical tests.