Kakaire *et al. Malaria Journal* (2025) 24:382
https://doi.org/10.1186/s12936-025-05627-9

Malaria Journal

REVIEW

Open Access

Structural equation modelling (SEM) for malaria prevalence and risk factors in Uganda

Grace Kakaire[1,2*], Edna Chepkemoi Chumoh[2], Mabula Salyungu[2,3], Gregory Kerich[2], Robert Kipchumba Too[2] and Mathew Kosgei[2]

Abstract

Background Malaria remains a leading cause of morbidity and mortality among children under five years in sub-Saharan Africa, despite extensive public health efforts. Its transmission is influenced by a complex interaction of socio-economic, environmental, maternal, and child health factors. Traditional analytic approaches often fail to capture these multifaceted relationships. This study employs Structural Equation Modelling (SEM) to explore the latent and observed predictors of child malaria prevalence, offering a comprehensive understanding of the underlying pathways.

Methods Utilizing secondary data from the 2018–2019 Uganda Malaria Indicator Survey (MIS), a SEM framework was constructed comprising four latent constructs: Socioeconomic Status (SES), Environment, Maternal Health, and Child Health. Each construct was defined by multiple observed variables, and the outcome of interest was the malaria status of the child. Factor loadings and regression coefficients were estimated using standardized model parameters. Model fit was assessed using indices including the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR).

Results The SEM analysis demonstrated significant pathways from latent variables to child malaria prevalence. Child Health had the strongest positive association with malaria status ($\beta = 0.22$, $p < 0.001$), followed by a marginal negative association with the Environment construct ($\beta = -0.36$, $p = 0.056$). Socioeconomic Status and Maternal Health were not statistically significant predictors. Model fit indices suggested a moderately acceptable fit (CFI = 0.786, TLI = 0.630, RMSEA = 0.064, SRMR = 0.053), indicating that the conceptual framework captured meaningful relationships.

Conclusion This study highlights the value of SEM in disentangling the intricate network of factors influencing child malaria. The findings underscore the importance of child-level health conditions and environmental factors, while pointing to the limited direct effect of socioeconomic and maternal health variables. These insights can inform integrated intervention strategies that simultaneously address child health and environmental exposures to reduce the burden of malaria.

Keywords Child malaria, Structural equation modeling, Child health, Sub-Saharan Africa

*Correspondence:
Grace Kakaire
kakairegrace2@gmail.com
Full list of author information is available at the end of the article

Background

Despite substantial global investment in malaria control, children under five in sub-Saharan Africa remain disproportionately affected, reflecting stalled progress and the unfinished agenda of malaria elimination. Despite global efforts and measurable progress in malaria control and elimination, the disease continues to disproportionately affect children under the age of five, contributing significantly to childhood morbidity and mortality. According to the World Health Organization (WHO), sub-Saharan Africa accounts for over 90% of global malaria cases and deaths, with children bearing the greatest burden due to their relatively underdeveloped immunity [1, 2]. While the biological mechanisms of malaria transmission are well-characterized, the relative contributions of socioeconomic disadvantage, environmental exposures (e.g., sanitation, housing quality), and maternal health factors remain contentious, with studies reporting inconsistent associations across settings [3, 4].

Malaria transmission is inherently multifactorial. Traditional studies have typically focused on single risk factors in isolation, such as insecticide-treated net (ITN) usage, indoor residual spraying (IRS), and proximity to mosquito breeding grounds [5, 6]. However, such an approach may obscure the broader, interconnected mechanisms that underpin malaria risk, particularly among vulnerable populations like children. Socioeconomic status (SES), for example, is not merely a background variable but a latent construct encompassing education, income, and living conditions, all of which influence both exposure to malaria vectors and the capacity to seek timely medical care [7, 8]. Similarly, maternal health factors, environmental conditions such as access to improved sanitation and water sources, and child-specific health variables (e.g., age, sex, recent illnesses) collectively shape the vulnerability landscape [9, 10].

Understanding the relative contributions and pathways through which these determinants affect child malaria outcomes requires more than just conventional statistical techniques. Structural Equation Modelling (SEM) offers a robust analytical framework to assess complex, latent constructs and their direct and indirect relationships with observed outcomes [5]. SEM enables researchers to model multi-layered relationships among variables, account for measurement error, and evaluate entire conceptual models in a statistically rigorous manner. This makes it especially well-suited to public health research, where latent constructs like SES and maternal health cannot be measured directly but can be inferred from observed indicators [1, 2].

Previous studies [11–16] have applied SEM in various domains of health research, including mental health, reproductive health, and health service utilization. SEM has been underutilized in malaria research. Although numerous studies have applied regression-based approaches to assess risk factors [7, 20, 25], very few, if any, have leveraged SEM to examine the latent pathways underlying child malaria risk. This represents an important methodological gap this study seeks to address. By leveraging SEM, it is possible to move beyond surface-level associations and toward a more integrated understanding of how various predictors interrelate and impact child malaria prevalence. For example, the effect of maternal education on child health might be mediated through antenatal care utilization and household wealth, while environmental exposures may be confounded by socioeconomic inequalities [5, 9, 14].

Moreover, as countries strive to achieve the United Nations Sustainable Development Goals (SDGs)-particularly Goal 3 (Good Health and Well-being) and Goal 6 (Clean Water and Sanitation)-there is a growing need to adopt data-driven, systems-level approaches to disease prevention [1, 2]. This necessitates a holistic understanding of social determinants of health, environmental factors, and access to care as mutually reinforcing contributors to health outcomes. Using SEM, researchers and policymakers can identify leverage points for targeted interventions that address root causes rather than symptoms.

This study, therefore, employs Structural Equation Modelling to investigate the latent and observed predictors of child malaria prevalence. Drawing upon nationally representative data, the model integrates four latent constructs Socioeconomic Status (SES), Environment, Maternal Health, and Child Health-each defined by a series of observed indicators. The aim is to elucidate the pathways through which these factors influence the likelihood of malaria infection in children. By doing so, the study seeks to contribute to a deeper understanding of the multifaceted etiology of malaria and to provide actionable insights for designing more effective, equity-oriented malaria control strategies.

Methods

Data source and study population

The study utilized data from the 2018–19 Uganda Malaria Indicator Survey (MIS). The analytic dataset was derived from two DHS/MIS standard files: The Household Members Recode (PR) file, which contains malaria rapid diagnostic test (RDT) and anaemia biomarker results, and the Children's Recode (KR) file, which contains demographic and maternal/child background information. The unit of analysis was children under 5 years of age who were tested for malaria. Women 15–49 years were part of the MIS sampling frame but were not analysed in this study.

The MIS was based on a stratified two-stage cluster design [15]. The MIS stratified the sample by region × urban/rural residence, yielding 34 sampling strata rather than the 20 that had previously reported. Within each stratum, clusters (enumeration areas) were selected, resulting in 210 clusters (78 urban and 132 rural).

Data linkage

The KR and PR files were merged using the household identifier (hhid), constructed from the KR file's case id, to match the household IDs in the PR file. Mismatches, duplicates, and inconsistencies were checked for, during the merging process. Records with conflicting or missing household identifiers were carefully reviewed and excluded if linkage could not be confidently established. The final merged dataset was assessed for representativeness by comparing key demographic variables before and after merging to ensure no systematic bias was introduced due to linkage errors.

Data collection and measurements

Data collection was conducted from December 2018 through February 2019, which corresponds to a period of relatively high malaria transmission in Uganda. Blood samples were taken from the fingers or heels of children aged less than 5 years and tested on the spot using Rapid Diagnostic Tests (RDTs). In addition, thick and thin blood smears were prepared and tested by microscopy. Results were recorded as either positive or negative if malaria parasites were found or not in the blood sample, respectively.

Capillary blood was obtained from children aged 0–59 months for malaria Rapid Diagnostic Testing (RDT) in the field, and thick/thin blood smears were prepared for centralized microscopy, which served as the reference standard. Smears were double-read by independent microscopists, with discrepancies adjudicated by a senior reader as part of the MIS quality-control protocol.

The primary outcome was microscopy-confirmed malaria (positive/negative). Sensitivity analyses were conducted using (i) RDT-only and (ii) a composite definition classifying a child as malaria-positive if either RDT or microscopy was positive. Under the composite definition, discordant results (RDT+/Mic– or RDT–/Mic+) were coded positive.

Sample selection

The analysis was restricted to de facto children under 5 years of age (0–59 months), i.e., those who were present in the household the night before the survey. Children without valid malaria RDT or microscopy results were excluded. Specifically, out of 7435 eligible children listed in the household roster and eligible for biomarker testing, 7214 provided valid malaria test results. After excluding children with missing outcome data, the final analytic sample included 6705 children (Fig. 1).

Study variables

Outcome variable

The primary outcome for the Structural Equation Modeling (SEM) was microscopy-confirmed malaria status (positive/negative), given its role as the reference standard. Sensitivity analyses using (i) RDT-only status and (ii) a composite malaria definition (positive if either RDT or microscopy was positive) were conducted separately to assess robustness but were not included in the main SEM. This approach ensures clarity and consistency in the primary model while acknowledging alternative malaria classification methods.

Latent constructs

Although malaria Rapid Diagnostic Test (RDT) results are described as a measure in this section, the primary malaria status indicator used throughout the analysis, including in latent variable modeling, was microscopy-confirmed malaria due to its higher specificity as a reference standard. RDT results were considered only in sensitivity analyses outside the main model. This

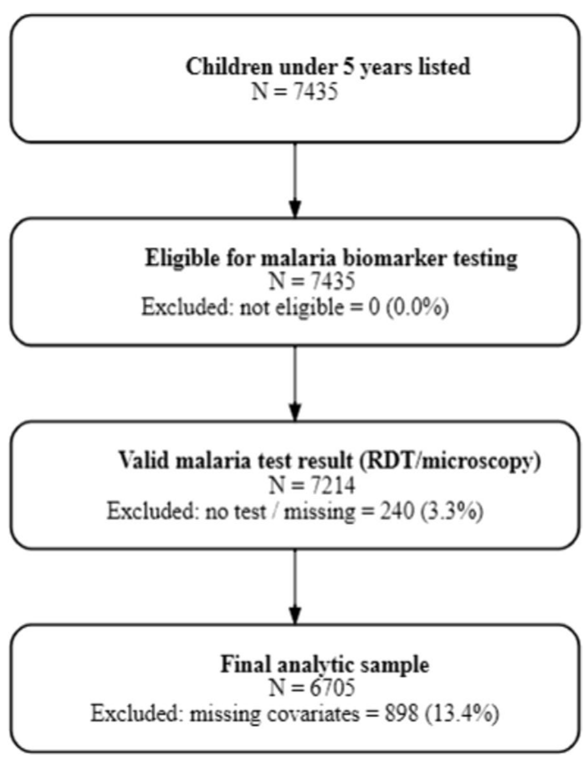

Fig. 1 Flow diagram of sample selection

consistent use of microscopy-confirmed status reduces measurement bias and maintains coherence across the study. All observed indicators were coded appropriately as binary or ordinal variables, with age treated categorically (0–11, 12–23 and 24+months).

Haemoglobin concentration was measured in the field using a HemoCue analyzer. The DHS program provides altitude-adjusted haemoglobin values in the recode files, and these pre-adjusted variables were used directly. Children were classified as anaemic if their adjusted haemoglobin concentration was < 11.0 g/dL, consistent with WHO guidelines for 0–59 months of age [21]. The child's malaria status was based on RDT results. A drop of blood was tested for the presence of *Plasmodium* parasites using the SD Bioline Pf/Pv RDT. This type of test is increasingly used as a diagnostic test when reliable microscopy is not available [15].

Maternal anaemia was defined using altitude-adjusted haemoglobin values measured similarly to child haemoglobin, with anaemia classified according to WHO guidelines (adjusted haemoglobin < 12.0 g/dL for non-pregnant women aged 15–49 years). Although maternal anaemia status was generated during data merging, it was not included as an outcome or latent variable in the Structural Equation Modelling (SEM) due to [reason, e.g., focus on child health outcomes]. Its potential role as a confounder or mediator could be explored in future analyses.

Statistical analysis

Descriptive statistics

Descriptive summaries were generated using R packages Table 1 and tableone, stratified by malaria status. To account for the complex survey design of the DHS/MIS, we applied sampling weights provided by DHS (PR weights for children), along with adjustment for clustering at the primary sampling unit and stratification by region–residence. All estimates therefore represent weighted, design-adjusted results. Weighted analyses were performed in R using the survey package. For categorical comparisons across malaria status, Rao–Scott corrected Chi-square tests were used (svychisq in the R survey package), which provide design-adjusted p-values under stratified cluster sampling with unequal weights.

Structural equation modelling

SEM was used to estimate direct and indirect effects. Missing data in the Structural Equation Modeling (SEM) were addressed using Full Information Maximum Likelihood (FIML) estimation, which allows for the inclusion of all available data under the assumption that data are missing at random (MAR). This approach maximizes statistical power and reduces bias compared to traditional list-wise deletion [5, 9]. Before proceeding with the full SEM, a Confirmatory Factor Analysis (CFA) was conducted to validate the adequacy of the latent constructs. Each construct (Socioeconomic Status, Environment, Maternal Health, and Child Health) was tested separately to ensure that the measurement model was acceptable. Factor loadings for the majority of observed indicators were above the recommended cutoff of 0.40, confirming that the items contributed meaningfully to their respective latent variables.

The structural paths specified were:

- SES → Environment.
- SES → Maternal health access.
- Environment → Child vulnerability.
- Maternal health access → Child vulnerability.
- Child vulnerability → Malaria status.

Table 2 shows fit indices used for model evaluation and their recommended cutoff thresholds.

Modification indices and residuals were reviewed to refine the model.

As shown in Fig. 2, it was hypothesized that each factor had a direct relationship with the outcome.

Table 2 Fit indices used for model evaluation and their recommended cutoff thresholds

Fit Index	Threshold for acceptable fit
RMSEA	< 0.08
CFI	> 0.90
TLI	> 0.90

Table 1 Latent constructs and their associated observed variables used in the structural equation model of malaria prevalence

Latent variable	Observed indicators
Socioeconomic status	Education attainment, Wealth index, Urban/rural residence
Household environment	Toilet facility, Water source
Maternal health access	Maternal anaemia status, Antenatal visits, Maternal age
Child vulnerability	Age of child, Sex of child, Recent fever status, Recent anaemia status

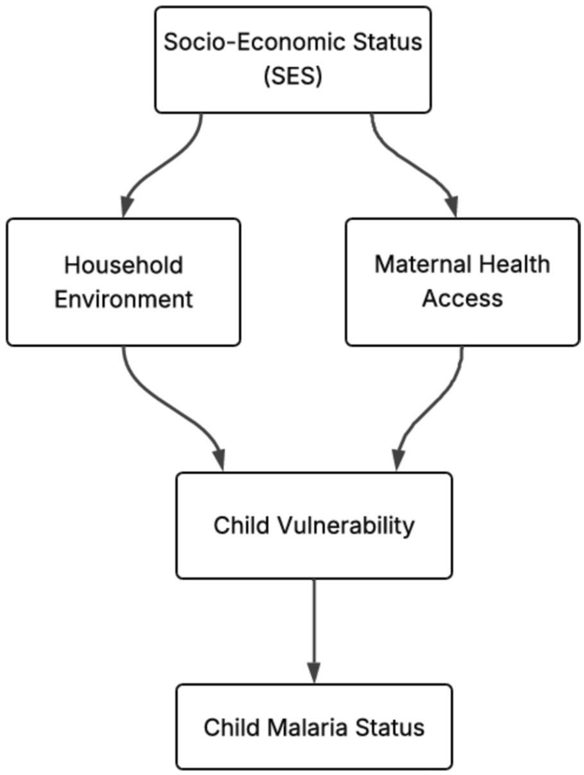

Fig. 2 Proposed model structure

Software

All analyses were performed in R (version 4.4.2). Structural Equation Modeling was conducted using the lavaan package (version 0.6–17.6), with additional visualization through semplot (version 1.1.6) and post-estimation summaries generated using parameters (version 0.21.7).

Ethical considerations

The 2018–2019 study was approved by two ethical review bodies that included the Makerere University School of Biomedical Sciences higher degrees research and ethics committee (sbs-HDREC) and the Uganda National Council for Science and Technology (UNCST). All participants interviewed gave their written informed consent to participate in the 2018–2019 MIS in addition to granting permission that information from survey could be published. Furthermore, written informed consent was sought from the mother or guardian of a child before any blood sample was taken. The data used in this analysis were anonymous with no individual names of participants captured.

Results

Descriptive characteristics of the study population by malaria status

Table 3 presents the weighted distribution of demographic, socioeconomic, and health-related characteristics of children and their mothers stratified by malaria status. All estimates were calculated using survey weights to account for the complex DHS sampling design. The age of the child was significantly associated with malaria status ($p = 0.005$). Malaria prevalence peaked among children aged 12–23 months (40.3%), followed by infants aged 0–11 months (37.5%), and was lowest among children aged 24 months and above (22.2%). Although the positivity rate appeared higher among infants, this likely reflects small subgroup sizes and sampling variability.

Sex of the child was not significantly associated with malaria positivity ($p = 0.604$). In contrast, residence status showed strong effects: children in rural areas had markedly higher malaria prevalence (94.5%) compared to urban areas (5.5%) ($p < 0.001$). Similarly, children from non-central regions exhibited higher prevalence (88.4%) than those from central Uganda (11.6%) ($p = 0.043$). Maternal education was inversely associated with malaria positivity ($p < 0.001$). Children of mothers with no education accounted for 30.8% of positive cases compared to 14.7% among negatives, while tertiary-educated mothers' children had the lowest positivity (0.9%). Household wealth index also showed a strong gradient: children from poor households had a prevalence of 87.6% compared to 12.4% in rich households ($p < 0.001$).

Sanitation and environment were also significant predictors. Only 3.3% of malaria-positive children had access to improved toilet facilities, compared to 16% among negatives ($p < 0.001$). However, water source was not significantly associated with malaria status ($p = 0.572$), possibly reflecting the grouping of piped and protected sources within the "improved" category.

Maternal anaemia was strongly associated with child malaria status ($p < 0.001$). Among children who tested positive, 73.9% had anaemic mothers. Recent fever in children also showed strong associations ($p < 0.001$), with 41.1% of malaria-positive children reported to have had fever, compared to 22.7% among negatives.

In summary, weighted analyses confirm that age, rural residence, region, maternal education, wealth, sanitation, maternal anaemia, and fever status are significantly associated with malaria positivity in Ugandan children under five, while sex of child and water source were not.

Measurement model: factor loadings

Table 4 presents the standardized factor loadings for the measurement model. Four latent constructs were specified: Socioeconomic Status (SES), Environment, Maternal Health, and Child Health. All loadings were statistically significant ($p < 0.05$), with the exception of fixed reference indicators, which by design do not have p-values.

Table 3 Characteristics of mothers/children by malaria status

Mother/child characteristic	Negative (N = 4767)	Positive (N = 1935)	Total (N = 6705)	p-value
Age of child (months)				0.005
0–11	2116 (45.1%)	760 (37.5%)	2876 (42.6%)	
12–23	1798 (36.9%)	786 (40.3%)	2584 (38.0%)	
24 +	853 (18.0%)	389 (22.2%)	1242 (19.4%)	
Sex of child				0.604
Female	2443 (49.9%)	947 (48.7%)	3390 (49.5%)	
Male	2425 (50.1%)	1022 (51.3%)	3447 (50.5%)	
Residence				0.000
Rural	3749 (74.2%)	1832 (94.5%)	5581 (80.6%)	
Urban	1119 (25.8%)	137 (5.5%)	1256 (19.4%)	
Region				0.043
Central	910 (26.6%)	122 (11.6%)	1032 (21.8%)	
Non-central	3464 (73.4%)	1712 (88.4%)	5176 (78.2%)	
Maternal education				0.000
No education	804 (14.7%)	648 (30.8%)	1452 (19.8%)	
Primary/secondary	3841 (80.8%)	1299 (68.2%)	5140 (76.8%)	
Tertiary	223 (4.5%)	22 (0.9%)	245 (3.4%)	
Wealth Index				0.000
Poor	3294 (62.4%)	1785 (87.6%)	5079 (70.4%)	
Rich	1574 (37.6%)	184 (12.4%)	1758 (29.6%)	
Toilet facility				0.000
Improved	605 (16.0%)	76 (3.3%)	681 (12.1%)	
Not improved	4218 (84.0%)	1882 (96.7%)	6100 (87.9%)	
Water source				0.572
Improved	3442 (75.6%)	1416 (79.0%)	4858 (76.7%)	
Not improved	1426 (24.4%)	553 (21.0%)	1979 (23.3%)	
Maternal anaemia				0.000
Anaemic	2122 (48.0%)	1265 (73.9%)	3387 (56.1%)	
Not anaemic	2469 (52.0%)	587 (26.1%)	3056 (43.9%)	
Fever status				0.000
No	3689 (77.3%)	1079 (58.9%)	4768 (71.4%)	
Yes	1078 (22.7%)	856 (41.1%)	1934 (28.6%)	

Table 4 Standardized factor loadings for latent constructs from SEM measurement model

Latent	Variable	Factor loading	Standard error	P-value	Standardized loading
SES	Education attainment	1.000	0.000		0.100
	Wealth index	7.978	0.846	0.000	0.477
Environment	Toilet improved	1.000	0.000		0.675
	Water improved	0.306	0.047	0.000	0.137
Maternal Health	Maternal anaemia	1.000	0.000		0.420
	Antenatal visits	0.374	0.035	0.000	0.174
Child Health	Fever recent	1.000	0.000		0.171
	Age of child	− 5.387	0.463	0.000	− 0.295
	Sex of child	0.199	0.085	0.019	0.031
CFI	0.786				
TLI	0.63				
RMSEA(CI)	0.064 (0.059, 0.068)				

For SES, wealth index demonstrated a moderate standardized loading (0.477), while education attainment loaded very weakly (0.100). This suggests that in the Ugandan MIS dataset, household wealth was a more salient measure of socioeconomic variation than maternal education. The weak contribution of education may reflect relatively homogeneous educational attainment across households, or a weaker association between education and health service access in this context. Education was retained in the model to remain consistent with epidemiological literature where maternal education has been strongly linked to malaria prevention knowledge and child health, but acknowledge that its measurement value here is limited.

The Environment construct was captured primarily through toilet facilities (loading = 0.675), whereas water source contributed only trivially (0.137). This imbalance suggests that sanitation was a stronger discriminator of household environment in relation to child malaria risk, while the broad categorization of water source (improved vs. not improved) may have obscured its potential effect. Accordingly, the Environment factor should be interpreted as reflecting sanitation more than water conditions.

Maternal Health was indicated by maternal anaemia (loading = 0.420) and antenatal visits (0.174). The latter, although statistically significant, was conceptually less plausible as a determinant of malaria in children aged 0–4 years, and its low loading confirms that it added little to the construct. Maternal anaemia, by contrast, provided a more consistent measure of maternal health status relevant to child malaria vulnerability. Future refinements of this construct should consider more appropriate indicators, such as maternal ITN use or treatment-seeking behaviours.

Taken together, the measurement model demonstrated uneven performance across constructs, with some indicators (education attainment, water source, antenatal visits) contributing weakly. This highlights both contextual limitations in the MIS variables and opportunities for future model refinement (Table 5).

Table 5 Standardized and unstandardized effects of predictors on child malaria

Factor	B (95% CI)	β (Std)	P-value
SES	0.99 (− 4.58, 6.57)	0.07	0.73
Environment	− 1.77 (− 3.57, 0.04)	− 0.36	0.06
Maternal health	0.18 (− 0.18, 0.54)	0.04	0.32
Child health	2.78 (2.06, 3.50)	0.22	0.00

Model fit indices

The overall fit of the SEM model was evaluated using multiple indices. The Comparative Fit Index (CFI = 0.786) and Tucker–Lewis Index (TLI = 0.630) were well below the conventional cutoff of 0.90, indicating poor model fit. The Root Mean Square Error of Approximation (RMSEA = 0.064, 90% CI: [0.059, 0.068]) and the Standardized Root Mean Square Residual (SRMR = 0.053) fell within generally accepted thresholds (≤ 0.08). While RMSEA and SRMR suggest a reasonable approximation of the data, the poor CFI and TLI scores demonstrate that the overall measurement and structural model did not achieve conventional SEM standards. This discrepancy highlights the limitations of the specified latent constructs.

Importantly, fit indices are known to be influenced by both sample size (N = 6,705 in this analysis) and the degrees of freedom of the model. Given the relatively large sample, even modest deviations between observed and model-implied covariance structures are amplified, making incremental fit indices such as CFI and TLI particularly stringent. Nonetheless, the results point to the need for cautious interpretation of the SEM findings.

Factor loadings

Although nearly all factor loadings were statistically significant ($p < 0.05$), several were extremely weak (e.g., sex of child = 0.031, water source = 0.137, antenatal visits = 0.174). Statistical significance should not be conflated with substantive importance. These small loadings indicate that the corresponding indicators contributed little to their latent constructs, raising concerns about construct validity. For example, education attainment (0.100) did not meaningfully capture SES in this dataset, and water source (0.137) did not meaningfully capture environmental risk. These limitations may have contributed to the poor overall model fit.

Path diagram results

The structural equation model (SEM) path diagram (Fig. 3) presents the estimated relationships among the four latent constructs—Socioeconomic Status (SES), Environment, Maternal Health, and Child Health—and their respective observed indicators, as well as the direct effects of these latent constructs on child malaria status.

Each latent variable was measured using two or more observed indicators, with acceptable standardized factor loadings. For instance, education attainment and wealth index loaded onto SES, with wealth index exhibiting a stronger standardized loading (0.477). The Environment factor included improved toilet and improved water source, with toilet facility being a particularly strong indicator (loading = 0.675). Maternal Health was defined by maternal anaemia and antenatal visits, while Child

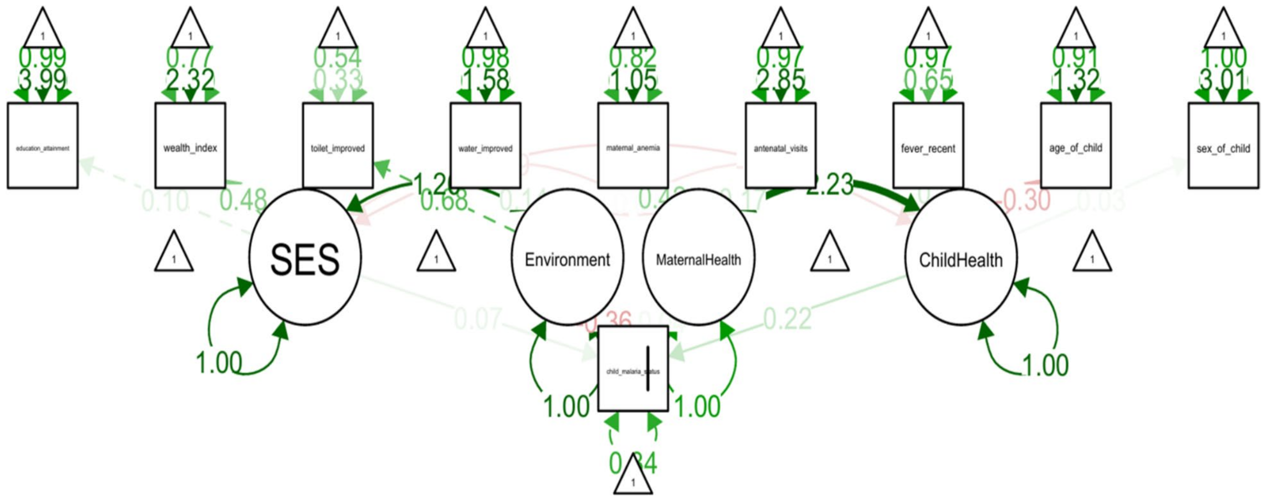

Fig. 3 SEM path diagram of latent and observed predictors of child malaria prevalence

Health comprised recent fever, age of the child, and sex of the child. Loadings were statistically significant for most observed variables, confirming their adequacy in representing latent constructs.

The path diagram indicates that Child Health was the only latent variable with a statistically significant direct effect on child malaria status, with a standardized path coefficient of 0.22 ($p < 0.001$). This suggests that the health status of the child, particularly recent fever and age, plays a central role in determining malaria risk. Environment showed a marginal negative association ($\beta = -0.36$, $p = 0.056$), implying a potential protective effect of improved water and sanitation infrastructure, although this did not reach conventional levels of significance.

Neither SES nor Maternal Health exhibited statistically significant direct paths to child malaria status ($\beta = 0.07$, $p = 0.728$ and $\beta = 0.04$, $p = 0.323$, respectively). This suggests that their effects may be more indirect or mediated through other latent variables rather than exerting a direct influence on malaria outcomes in children.

The diagram also displays covariances between latent variables, though these were not the primary focus of the current analysis. The model offers a holistic view of how individual, maternal, household, and environmental determinants collectively shape malaria risk in children.

Fit indices for the overall model were mixed. The Comparative Fit Index (CFI) was 0.786, and the Tucker-Lewis Index (TLI) was 0.630, both below the ideal threshold of 0.90. However, the Root Mean Square Error of Approximation (RMSEA) was 0.064, within acceptable limits (90% CI: 0.059–0.068), and the Standardized Root Mean Square Residual (SRMR) was 0.053, indicating an adequate but improvable model fit.

Discussion

This study employed structural equation modelling (SEM) to examine the latent pathways through which socioeconomic status (SES), environmental conditions, maternal health, and child health influence the risk of malaria infection among children. A notable and somewhat surprising finding of this study was that socioeconomic status (SES) and maternal health, factors widely recognized as structural determinants of malaria risk, were not significantly associated with child malaria infection in the structural equation model. This contrasts with prior evidence from malaria-endemic settings [4, 5, 17, 18] and represents the novel contribution of the present study. However, these non-significant effects should not be interpreted as evidence of absence, but rather as reflecting measurement and specification limitations in the modelling approach. While SEM enabled the exploration of multiple interrelated pathways, the results provide only a partial understanding of the multifactorial nature of malaria risk, as the model demonstrated poor overall fit (CFI = 0.786; TLI = 0.630) and specification weaknesses.

Among the latent constructs tested, Child Health exhibited the strongest and only statistically significant direct association with malaria status, whereas SES and maternal health showed weaker or non-significant direct associations in the structural model [1, 2, 19]. This construct was defined by age, sex, and recent fever. Younger children showed higher malaria positivity, consistent with the established epidemiological pattern that children under five, particularly those below 24 months, bear the greatest burden due to immature immunity [6, 7].

However, the inclusion of fever as an indicator introduces conceptual limitations. Fever is a diagnostic symptom of malaria rather than an independent risk factor, and its strong loading likely reflects tautological overlap with the malaria outcome [1, 2, 19]. This circularity undermines the construct validity of the Child Health factor and should be considered a limitation of the model [3, 20].

The Environment construct demonstrated a borderline, negative association with malaria status ($p = 0.056$). While this hints at a potential protective role of improved sanitation and water sources [8, 10, 18], the evidence is tentative and should be treated as hypothesis-generating rather than definitive. Furthermore, the trivial loading of water source (0.137) suggests that this indicator did not adequately capture environmental exposure pathways, which likely contributed to its nonsignificant role in the bivariate analysis. The argument that water collection practices may increase mosquito exposure is plausible [22], but cannot be directly substantiated with the MIS data, as no measures of collection frequency or distance were available. [18, 21].

Although the statistical significance was borderline, the practical implications are noteworthy. Environmental interventions remain a cornerstone of malaria prevention strategies. These findings lend support to the continued expansion of WASH (Water, Sanitation, and Hygiene) infrastructure, particularly in rural and peri-urban areas where malaria remains endemic [2, 23]. Programmes that combine health promotion with infrastructural improvements may be especially effective in reducing the risk of malaria and other vector-borne diseases [8, 10].

The SES and Maternal Health constructs did not show significant associations with malaria status. This finding may reflect weaknesses in construct measurement rather than true null effects. In the measurement model, maternal education (loading = 0.100) and antenatal visits (0.174) were extremely weak indicators. Such low factor loadings undermine construct validity and suggest that these proxies were insufficient to capture the broader influence of SES and maternal health on child malaria outcomes. Maternal anaemia, while moderately associated (0.420), may share common exposure pathways with child anaemia rather than independently influencing malaria risk [5, 14, 24]. Future research should consider alternative or additional indicators such as insecticide-treated net (ITN) use, maternal malaria history, healthcare-seeking behaviours, and household crowding to better capture these constructs [4, 8, 10, 25].

Model fit statistics further underscore the limitations of the current specification. While RMSEA (0.064) and SRMR (0.053) fell within conventional thresholds, both CFI (0.786) and TLI (0.630) were well below the accepted cutoff of 0.90 [9, 26]. This indicates that the model, as specified, demonstrated poor fit and did not adequately reproduce the observed covariance structure. The poor fit likely reflects both weak measurement indicators and omitted variables of known importance, including ITN use, malaria seasonality, and community-level prevalence, which were unavailable in the dataset but remain critical drivers of malaria risk [27, 28].

While this study provides important insights into the relative salience of child health and environmental conditions, it also demonstrates the limitations of SEM when applied to survey datasets with weak proxies for key constructs. These results should, therefore, be interpreted cautiously as hypothesis-generating, and they highlight the need for improved data collection and measurement strategies in future malaria research. [27, 28].

Taken together, these findings highlight the importance of proximal factors—particularly child health and environmental improvements—in shaping malaria risk, while also revealing serious limitations in the measurement of distal constructs such as SES and maternal health. The SEM path diagram further illustrated that only Child Health maintained a statistically significant pathway to malaria positivity, whereas other constructs either failed to reach significance or were weakly specified. These results underscore both the potential utility and the limitations of SEM for analysing complex health data in contexts where available indicators may be insufficiently sensitive.

The SEM path diagram provides a visual and statistical representation of the hypothesized relationships among latent variables and the outcome (child malaria status). The diagram reveals that child health was the only latent construct with a statistically significant direct path to malaria positivity, suggesting that variables such as recent fever, sex, and age of the child are most proximally related to infection status.

Future research should employ longitudinal designs and incorporate multilevel and geospatial modeling to capture temporal dynamics and contextual variations [29]. Expanding indicator sets to include ITN coverage, indoor residual spraying, maternal health-seeking practices, and seasonal malaria patterns may yield more accurate and theoretically coherent models. [29]. While the descriptive analyses accounted for the complex survey design using sampling weights, clustering, and stratification, the SEM was conducted using standard estimation methods in lavaan that assume independent observations. Given this limitation, survey weights were applied as probability weights in the SEM to partially adjust for unequal selection probabilities. However, clustering and stratification were not directly modelled, which may lead to underestimated standard errors. This limitation is acknowledged and recommend interpreting SEM results with caution. Future work could explore SEM approaches that fully integrate complex survey design, such as using pseudo-likelihood methods or bootstrap techniques.

Conclusion

This study underscores the critical role of child health and environmental conditions in shaping the risk of malaria infection among children. While socioeconomic status and maternal health were not directly associated with malaria outcomes in the final model, their influence may operate through indirect or context-specific pathways. These findings highlight the need for integrated malaria control strategies that prioritize child-centered health services and environmental improvements, while also addressing broader social determinants. Future research should build on these insights using longitudinal and multilevel approaches to better capture causal pathways and contextual variations.

Abbreviations

ANC	Antenatal care
CFI	Comparative Fit Index
CI	Confidence interval
DHS	Demographic and health surveys
FIML	Full information maximum likelihood
Hb	Hemoglobin
hhid	Household identifier
IRS	Indoor residual spraying
ITN	Insecticide-treated net
KR	Children's recode file (DHS/MIS dataset)
MIS	Malaria Indicator Survey
MAR	Missing at random
N	Sample size
PR	Household Members Recode file (DHS/MIS dataset)
RDT	Rapid diagnostic test
RMSEA	Root mean square error of approximation
SDG	Sustainable development goal
SEM	Structural equation modeling
SRMR	Standardized root mean square residual
TLI	Tucker–Lewis Index
UNCST	Uganda National Council for Science and Technology
WASH	Water, sanitation, and hygiene
WHO	World Health Organization

Acknowledgements

The authors wish to thank the MEASURE Demographic Health Survey (DHS) Program for granting them access to their dataset.

Author contributions

GK—Conceptualized the study and wrote the initial draft, GK, RKT, ECC, SM and MK-Contributed on methods and data analysis, RKT—Provided extensive edits and inputs. All authors read and approved the final draft.

Funding

None.

Data availability

The datasets generated and/or analysed during the current study are publicly available in the Demographic Health Survey repository, (https:/dhsprogram.com/data/dataset_admin/index.cfm).

Declarations

Ethics approval and consent to participate

The study obtained approval from MEASURE DHS to use their data, but consent to participate was not necessary as this was a secondary analysis of non-identifiable, publicly available data. The researchers treated the DHS data confidentially, and no attempt was made to identify individual women interviewed in the survey. All procedures were conducted in compliance with applicable guidelines and regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]School of Statistics and Planning, Makerere University, Kampala, Uganda. [2]School of Science and Aerospace Studies, Moi University, Eldoret, Kenya. [3]Department of Mathematics and Statistics, University of Dodoma (UDOM), Dodoma, Tanzania.

Received: 4 May 2025 Accepted: 15 October 2025
Published online: 07 November 2025

References

1. WHO. World Malaria Report 2023. Geneva: World Health Organization; 2023.
2. WHO and UNICEF. Progress on Household Drinking Water, Sanitation and Hygiene 2000–2020: Five Years into the SDGs. Geneva: World Health Organization and the United Nations Children's Fund; 2021.
3. White N, Pukrittayakamee S, Tinh Hien T, Faiz MA, Mokuolu O, Dondorp AM. Malaria. Lancet. 2014;383:723–35.
4. Eve W, Suprotik B, Kara H. Is malaria a disease of poverty? A review of the literature. Trop Med Int Health. 2005;10:1047–59.
5. Kline RB. Structural equation modeling. New York: Guilford Press; 1998.
6. Rogerson SJ, Hviid L, Duffy PE, Leke RFG, Taylor DW. Malaria in pregnancy: pathogenesis and immunity. Lancet Infect Dis. 2007;7:105–17.
7. Wnow SR, Sartorius B, Kyalo D, Maina J, Amratia P, Mundi CW, et al. The prevalence of *Plasmodium falciparum* in sub-Saharan Africa since 1900. Nature. 2017;550:515–8.
8. Tusting SK, Rek J, Arinaitwe E, Staedke SG, Kamya MR, Cano J, et al. Why is malaria associated with poverty? Findings from a cohort study in rural Uganda. Infect Dis Poverty. 2016;5:45–55.
9. Schreiber JB, Nora A, Stage FK, Barlow EA, King J. Reporting structural equation modeling and confirmatory factor analysis results: a review. J Educ Res. 2006;99:323–38.
10. Tusting LS, Bisanzio D, Alabaster G, Cameron E, Cibulskis R, Davies M, et al. Mapping changes in housing in sub-Saharan Africa from 2000 to 2015. Nature. 2019;18:391–4.
11. Andersen RM, Davidson PL, Baumeister SE. Improving access to care in America". In: Andersen RM, Rice TH, Kominski GF (editor) Changing the US Health Care System: key issues in health services policy and management. 3rd edn. San Francisco: Jossey-Bass. 2007; pp. 3–31.
12. Babitsch B, Gohl D, Von Lengerke T. Re-revisiting Andersen's behavioral model of health services use: a systematic review of studies from 1998–2011. GMS Psychosoc Med. 2012;25:1–15.
13. García-Pérez LE, Álvarez M, Dilla T, Gil-Guillén V, Orozco-Beltrán D. Adherence to therapies in patients with type 2 diabetes. Diabetes Ther. 2013;4:175–94.
14. Kline RB. Principles and practice of structural equation modeling. 5th ed. New York: Guilford Publications; 2023.
15. Muthén B, Muthen LK. Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. Alcohol Clin Exp Res. 2000;24:882–91.
16. Wang J, Wang X. Structural equation modeling: applications using Mplus. New Jersey: Wiley; 2019.

17. Grace JB, Anderson TM, Olff H, Scheiner SM. On the specification of structural equation models for ecological systems. Ecol Monogr. 2010;80:67–87.
18. Keiser J, Singer BH, Utzinger J. Reducing the burden of malaria in different eco-epidemiological settings with environmental management: a systematic review. Lancet Infect Dis. 2005;5:695–708.
19. D'Acremont V, Lengeler C, Genton B. Reduction in the proportion of fevers associated with *Plasmodium falciparum* parasitaemia in Africa: a systematic review. Malar J. 2010;9:240.
20. Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U, et al. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. Nature. 2015;526:207–11.
21. Prendergast AJ, Humphrey JH. The stunting syndrome in developing countries. Paediatr Int Child Health. 2014;34:250–65.
22. Yé Y, Hoshen M, Kyobutungi C, Louis VR, Sauerborn R. Local scale prediction of *Plasmodium falciparum* malaria transmission in an endemic region using temperature and rainfall. Glob Health Action. 2009;2:1923.
23. Cairncross S, Hunt C, Boisson S, Bostoen K, Curtis V, Fung IC, et al. Water, sanitation and hygiene for the prevention of diarrhoea. Int J Epidemiol. 2010;39:193–205.
24. Menéndez C, D'Alessandro U, ter Kuile FO. Reducing the burden of malaria in pregnancy by preventive strategies. Lancet Infect Dis. 2007;7:126–35.
25. Hill J, Hoyt J, van Eijk AM, D'Mello-Guyett L, ter Kuile FO, Steketee R, et al. Factors affecting the delivery, access, and use of interventions to prevent malaria in pregnancy in sub-Saharan Africa: a systematic review and meta-analysis. PLoS Med. 2013;10:e1001488.
26. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling. 1999;6:1–55.
27. Atieli HE, Zhou G, Afrane Y, Lee MC, Mwanzo I, Githeko AK, et al. Insecticide-treated net (ITN) ownership, usage, and malaria transmission in the highlands of western Kenya. Parasit Vectors. 2011;4:113.
28. Noor AM, Kinyoki DK, Mundia CW, Kabaria CW, Mutua JW, Alegana VA, et al. The changing risk of *Plasmodium falciparum* malaria infection in Africa: 2000–10: a spatial and temporal analysis of transmission intensity. Lancet. 2014;383:1739–47.
29. Gething PW, Casey DC, Weiss DJ, Bisanzio D, Bhatt S, Cameron E, et al. Mapping *Plasmodium falciparum* mortality in Africa between 1990 and 2015. N Engl J Med. 2016;375:2435–45.

Publisher's Note