JMIR Pediatrics and Parenting

Impact Factor (2023): 2.1
Volume 9 (2026) ISSN 2561-6722 Editor in Chief: Sherif Badawy, MS, MD, MBA

Contents

Original Papers

Review Paper

Modeling Zero-Dose Children in Ethiopia: A Machine Learning Perspective on Model Performance and Predictor Variables

Berhanu Fikadie Endehabtu[1,2], MPH; Kassahun Alemu[2,3], PhD; Shegaw Anagaw Mengiste[4], PhD; Meseret Zelalem[5], MD; Monika Knudsen Gullslett[6], PhD; Binyam Tilahun[1,2], PhD

[1]Department of Health Informatics, Institute of Public Health, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia

[2]Center for Digital Health and Implementation Science, University of Gondar, Gondar, Ethiopia

[3]Department of Epidemiology and Biostatistics, Institute of Public Health, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia

[4]School of Business, University of South-Eastern Norway, Drammen, Norway

[5]Department of Pediatrics and Child Health, University of Gondar, Gondar, Ethiopia

[6]Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway

Corresponding Author:
Berhanu Fikadie Endehabtu, MPH
Department of Health Informatics, Institute of Public Health, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia

Abstract

Background: Despite progress in childhood vaccination, many children in low- and middle-income countries, including Ethiopia, remain unvaccinated, presenting a significant public health challenge. The Immunization Agenda 2030 (IA2030) seeks to halve the number of unvaccinated children by identifying at-risk populations, but effective strategies are limited. This study leverages machine learning (ML) to identify Ethiopian children aged 12-35 months who are at higher risk of being zero dose (ZD). By analyzing demographic, socioeconomic, and health care access data, the study developed predictive models using different algorithms. The findings aim to inform targeted interventions, ultimately improving vaccination coverage and health outcomes.

Objective: This study aimed to develop an ML model to predict ZD children and to identify the most influential predictors of ZD in Ethiopia.

Methods: We examined how well the predictive algorithms can characterize a child at risk of being ZD based on predictor variables sourced from the recent National Immunization Evaluation Survey data. We applied supervised ML algorithms with the survey datasets, which included 13,666 children aged 12-35 months. Model performance was assessed using accuracy, area under the curve, precision, recall, and F_1-score. We applied Shapley Additive analysis to identify the most important predictors.

Results: The Light Gradient Boosting Machine (LGBM), Random Forest, Extreme Gradient Boosting (XGBoost), and AdaBoost classifiers effectively identified most ZD children as being at high risk. Among these, LGBM demonstrated the best performance, achieving an accuracy of 93%, an area under the curve of 97%, a precision of 94%, and a recall of 91%. The most significant features impacting the model included poor perception of vaccination benefits, lack of antenatal care utilization, distance from immunization services, and absence of maternal tetanus toxoid vaccinations.

Conclusions: The developed ML models effectively predict children at risk of being ZD, with the LGBM model showing the best performance. This model can guide targeted interventions to reduce ZD prevalence and address vaccination inequities. Key predictors include access to immunization sites, maternal health service utilization, and perceptions of immunization benefits. By focusing on these vulnerable groups, public health efforts can tackle disparities in vaccination coverage. Enhancing maternal care, raising caregiver awareness, and improving immunization access through outreach can significantly reduce the number of ZD children.

(JMIR Pediatr Parent 2026;9:e76712) doi:10.2196/76712

KEYWORDS

modeling; zero dose; children; machine learning; Ethiopia

Introduction

Child immunization is a cornerstone of public health, essential for safeguarding against life-threatening diseases and promoting the health of future generations [1]. Globally, significant advancements have been made in immunization programs, resulting in higher coverage rates [2] and a corresponding decline in vaccine-preventable disease [3]. However, as of 2023,

approximately 14.5 million children worldwide didn't receive the first dose of diphtheria, tetanus, and pertussis (DTP1) containing vaccines [4], a widely used indicator of access to immunization services [5]. This high number of zero-dose (ZD) children continues to be a pressing issue, intensifying health inequalities and heightening the likelihood of vaccine-preventable disease outbreaks [6]. These ZD children remain at high risk, creating considerable hurdles for public health efforts [7-9].

In a substantial portion of these ZD children, about 60% are concentrated in 10 low- and middle-income countries, including Ethiopia. Despite notable achievements in improving immunization coverage in Ethiopia, the country ranks third globally for ZD children, following Nigeria and India, accounting for 6% of the world's total [4].

Addressing the issue of children at risk of becoming ZD has emerged as a priority on both national and global agendas [10]. The Immunization Agenda 2030 (IA2030), endorsed by the World Health Assembly in November 2020, aims to reduce the number of ZD children by ensuring that every child is reached by 2030 [5]. However, effectively identifying and reaching these at-risk children poses significant operational challenges, and little is known about what strategies perform best.

Research in Ethiopia has identified various predictors of low immunization uptake, including low education levels and low wealth index [11-13], rural residence [12,14], limited access to health services [15,16], lack of antenatal care (ANC) and postnatal care (PNC) [13,15-18], home deliveries [13-16], absence of maternal tetanus toxoid (TT) vaccination [12,19], and poor caregiver knowledge [16]. However, there is a lack of evidence regarding how well these factors predict ZD status specifically and which factors are most relevant for optimal prediction.

Recent advancements in data science, coupled with available routine immunization data, present new opportunities to identify and reach at-risk children at both subnational and individual levels. Developing a robust algorithm to predict ZD children based on a set of variables could provide a valuable foundation for tailored interventions. Machine learning (ML) has emerged as a transformative tool in public health research particularly suited for this task which can capture complex relationships and interactions between variables [20-22]. Unlike traditional statistical methods that rely on predefined hypotheses, ML models can autonomously identify patterns and relationships within large datasets by learning from data rather than making prior assumptions [20,23,24]. This capability is particularly useful for multifactorial issues such as immunization uptake [25].

Using rule-based ML models can uncover hidden relationships among determinants of ZD children in large datasets, often represented through "if-then" statements that illustrate connections between variables [26]. This application of ML bridges the gap between theoretical research and practical applications, leading to advancements in the health care field [27].

This study aims to use ML algorithms to predict which Ethiopian children aged 12-35 months are at higher risk of being ZD and assess the predictive capabilities of the developed models. Findings from this study may provide actionable insights for policy makers and immunization program actors, informing the development of targeted strategies to effectively identify and reach those most at-risk children.

Methods

Study Design

The data for this study were sourced from the recent National Immunization Evaluation Survey in Ethiopia, which provides nationwide representation [28]. The survey included 11 regions and the 2 city administrations. A 2-stage stratified cluster sampling technique was used to select participants. The first stage is the enumeration areas (EAs), which served as clusters, randomly chosen with an urban-rural stratification approach, and the second stage is households within each EA. Sampling frames were prepared for each region and city administrations by the Ethiopian Statistical Services. The number of EAs required per region and city administration was determined based on the size within the stratum (study regions) and proportion of the Ethiopia population living in urban and rural areas (21.4% urban and 78.6% rural). A total of 468 EAs were randomly selected, comprising 100 from urban areas and 368 from rural regions, resulting in a total sample size of approximately 13,666 households with children aged 12 - 35 months.

We extracted information on immunization status for children aged 12 - 35 months. The vaccination status of children was assessed using 3 sources of information: caregiver reports, home-based vaccination cards, and facility-based records, following World Health Organization guidelines [29]. If a mother or a caregiver presented an immunization card, the child's vaccination status was assessed from that card. In cases where the card was unavailable, data collectors were instructed to verify the information at the nearest health facility if the caregiver reported that their child had been vaccinated. The mother's or caregiver's self-reports were considered only when neither the immunization card nor the facility records were available.

Using the operational definition set by Gavi, we defined a variable ZD status for each child, which is set to 1 if the child did not receive the first dose of the diphtheria, TTs, and pertussis-containing vaccine (DPT1), and set 0 otherwise [30,31].

We included a set of predictor variables or features to capture characteristics that have been associated with ZD status (Table 1). The factors influencing the outcome of interest are grouped into 3 groups: socioeconomic and demographic variables, health service utilization, and perceptions and attitudes. The first group of socioeconomic and demographic variables encompasses individual, household, and community-level characteristics that may affect the outcome of interest. The health service utilization represents the access to and use of various health care services, which can impact immunization status. The third category

focuses on the perceptions or attitudes that individuals or caregivers have toward the benefits of immunization. All the 3 categories of the variables gathered during the survey.

Table . The predictor variables used for analysis were extracted from the recent National Immunization Evaluation Survey in Ethiopia, 2023.

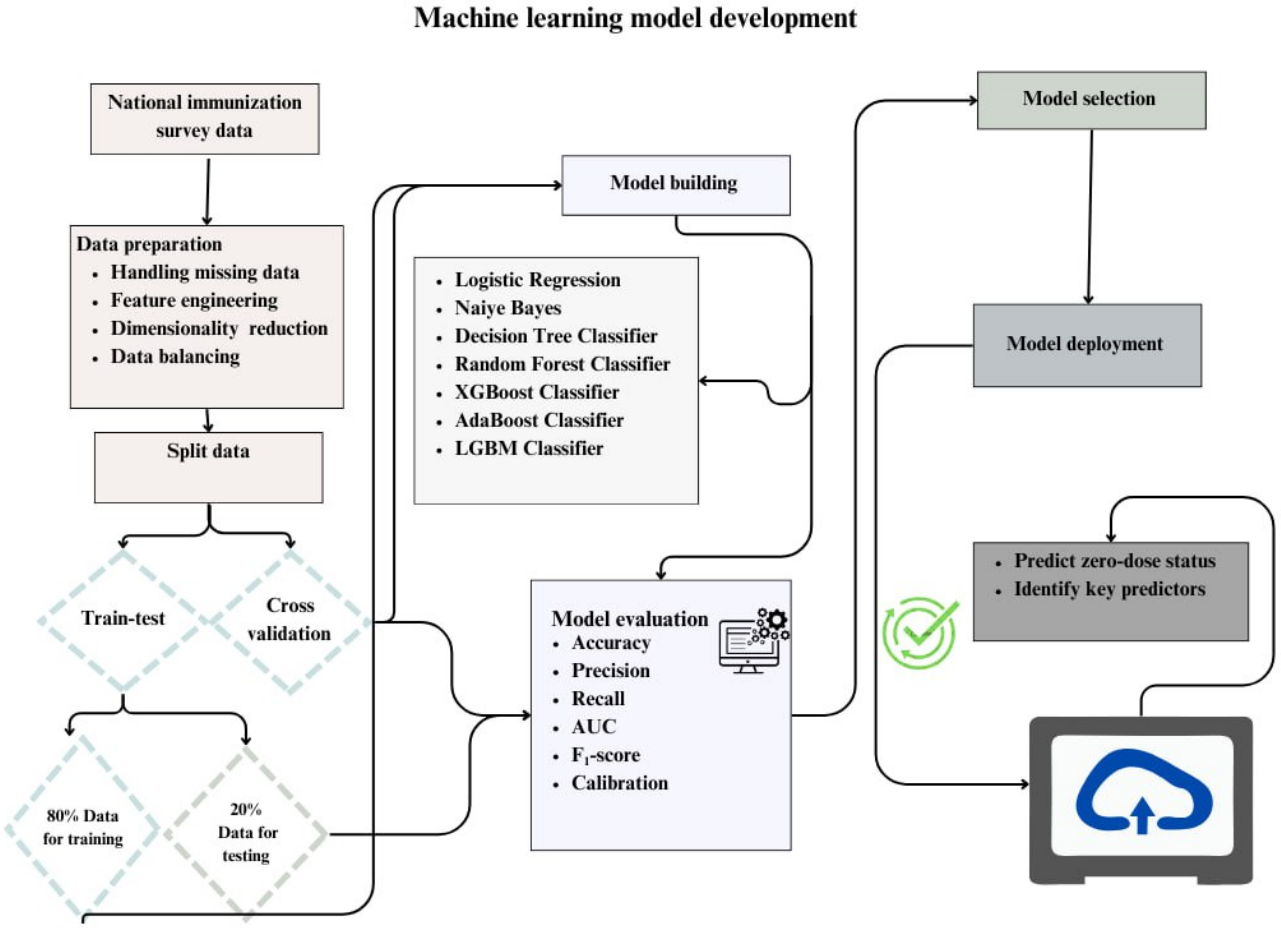
Category	Description	Response/type of data
Socioeconomic and demographic factor		
Residency	Type of living arrangement	Categorical (urban and rural)
Region	Geographic area of residence	Nominal (eg, Afar, Amhara...)
Religion	Cultural beliefs influencing health behaviors	Nominal or categorical (orthodox, Muslim, protestant, and others)
Marital status	Relationship status of the mother or the caregiver	Categorical (married and living together, married, married but not living together, and not in marital union)
Mother's or caregiver's educational status	Level of formal education attained	Categorical (no, primary, secondary, and higher education)
Occupation of mothers or caregivers	Employment status and type of work	Nominal categorical data
Birth order	Position of a child in relation to their siblings within a family	Categorical (first, second, third, and fourth and above)
Wealth index	Measures economic status	Categorical (poor, middle, and rich)
Health service utilization		
ANC[a] follow-up	History of ANC visits for the index child	Categorical (Yes/No)
History of maternal tetanus diphtheria vaccine	Previous vaccinations received	Categorical (Yes/No)
Distance to immunization site	Perceived impact of distance on immunization access	Categorical ("big problem," "not a problem")
Place of delivery	Location where the child was born	Categorical (Home/Facility)
Postnatal care	Follow-up care received after childbirth	Categorical (Yes/No)
Perceptions and attitudes		
Mother's or caregiver's perceived benefits on immunization	Beliefs regarding the advantages of vaccination	Was Likert (categorized into poor or good)
Trust in health care provider	The belief of mothers or caregivers on the services provided	Categorized into poor or good

[a]ANC: antenatal care.

Data Preprocessing and Transformation

We implemented several preprocessing steps to enhance model performance. First, we addressed missing values in the independent variables using the k-nearest neighbor approach. We then transformed categorical variables into numerical format through one-hot encoding, which is essential for preparing data for ML models. To standardize feature ranges, we applied minimum-maximum scaling and mean normalization, ensuring comparability among features (Figure 1).

Figure 1. Data preparation and analysis steps for zero-dose children prediction. AUC: area under the curve; LGBM: Light Gradient Boosting Machine.



We conducted sampling weight as instance weights during the training process for all algorithms. This was done by using the sample_weight parameter in the model's fitting functions, which adjusts the influence of each observation based on its probability of selection.

Next, we conducted a correlation analysis to identify and remove highly correlated features, thereby reducing multicollinearity and enhancing model robustness. Our correlation matrix showed a strong relation between parity and birth order (Figure 2), leading us to compute mutual information scores for each variable (Figure 3). This analysis highlighted ANC utilization and TT vaccination as significant predictors, while features such as marital status were excluded due to their minimal information value. Consequently, we retained birth order and omitted parity based on their scores.

Figure 2. Correlation analysis matrix for predictor variables for zero-dose children, Ethiopia, 2023. ANC: antenatal care; PNC: postnatal care; TT: tetanus toxoid.

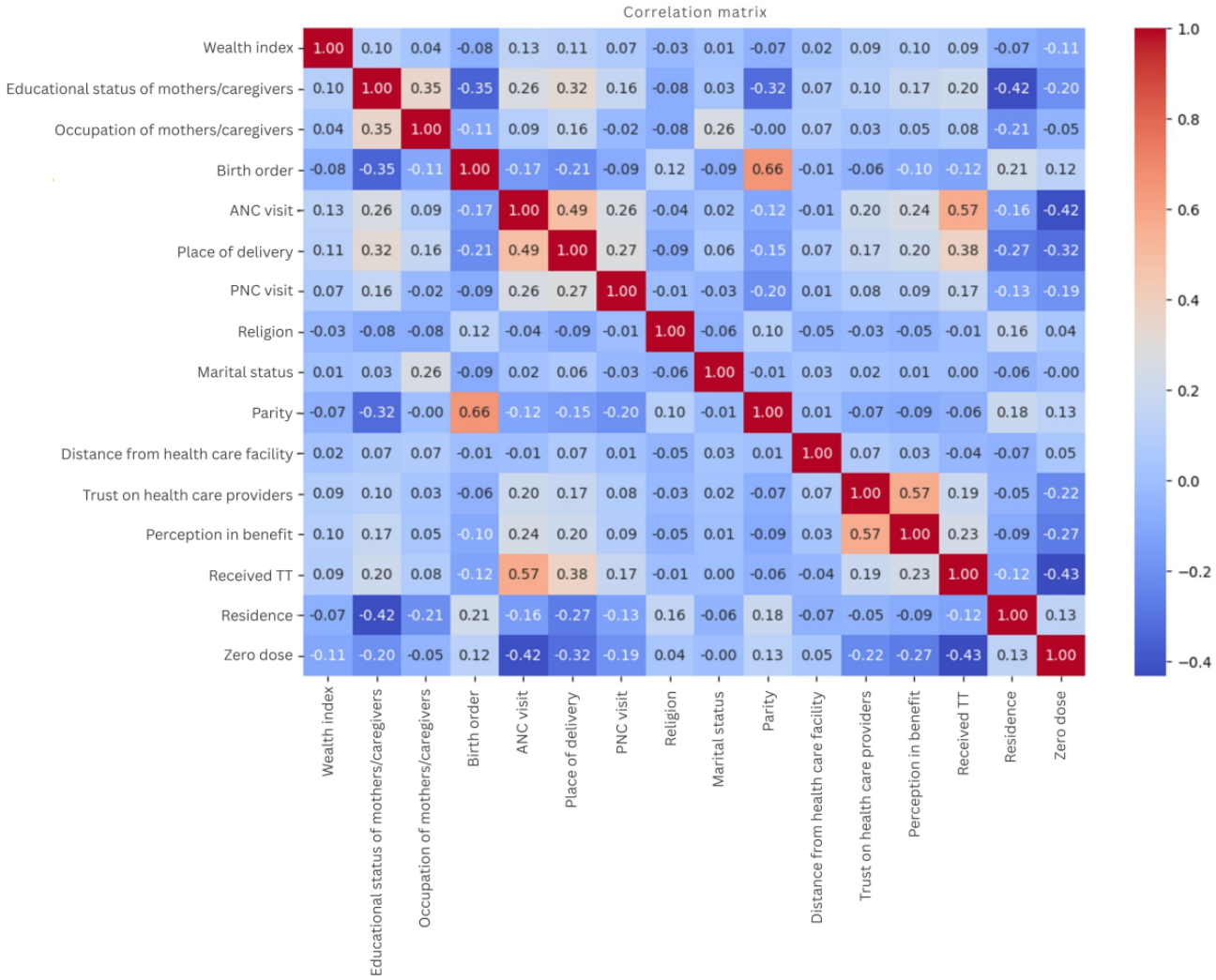
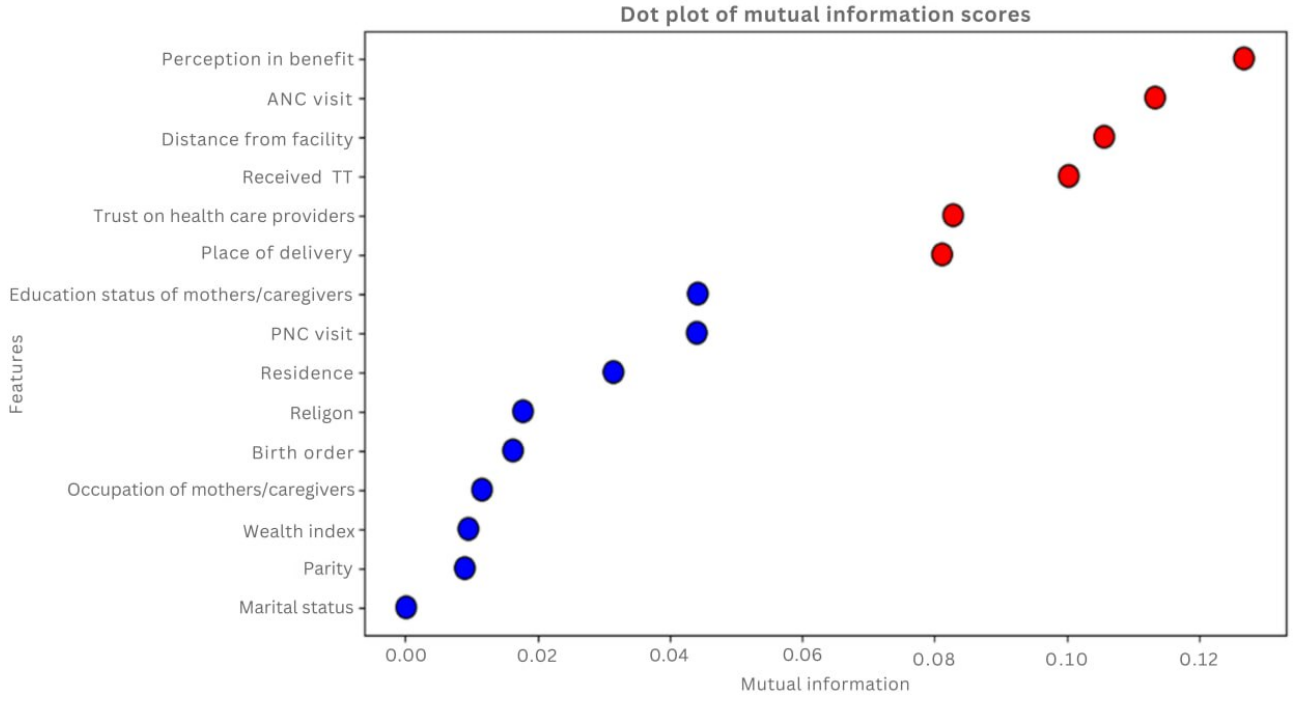


Figure 3. Mutual information score of predictor variables for zero-dose children, Ethiopia, 2023. ANC: antenatal care; PNC: postnatal care.



For dimensionality reduction, we used Forward Selection, Backward Elimination, and Recursive Feature Elimination methods. We opted for Recursive Feature Elimination due to its effectiveness in identifying the most significant predictors while simplifying the dataset. To address class imbalance, we applied the Synthetic Minority Oversampling Technique, which balanced the dataset from an initial skew of 82% majority and 18% minority to an equal distribution. This balancing supports the development of robust predictive models and mitigates bias toward the majority class (Multimedia Appendix 1).

Model Development

After the preprocessing, we split the dataset into 80% for training and 20% for testing (Figure 1). To avoid overfitting and underfitting, we applied 10-fold cross-validation, dividing the data into 10-folds and using one for validation while training on the others. The final performance is averaged across all folds.

The outcome variable, known as the class, is a binary variable indicating ZD status. A ZD status of 1 denotes a ZD child, while 0 indicates a non-ZD child. We applied supervised learning algorithms to develop a model from the training data to accurately predict this outcome in the test data.

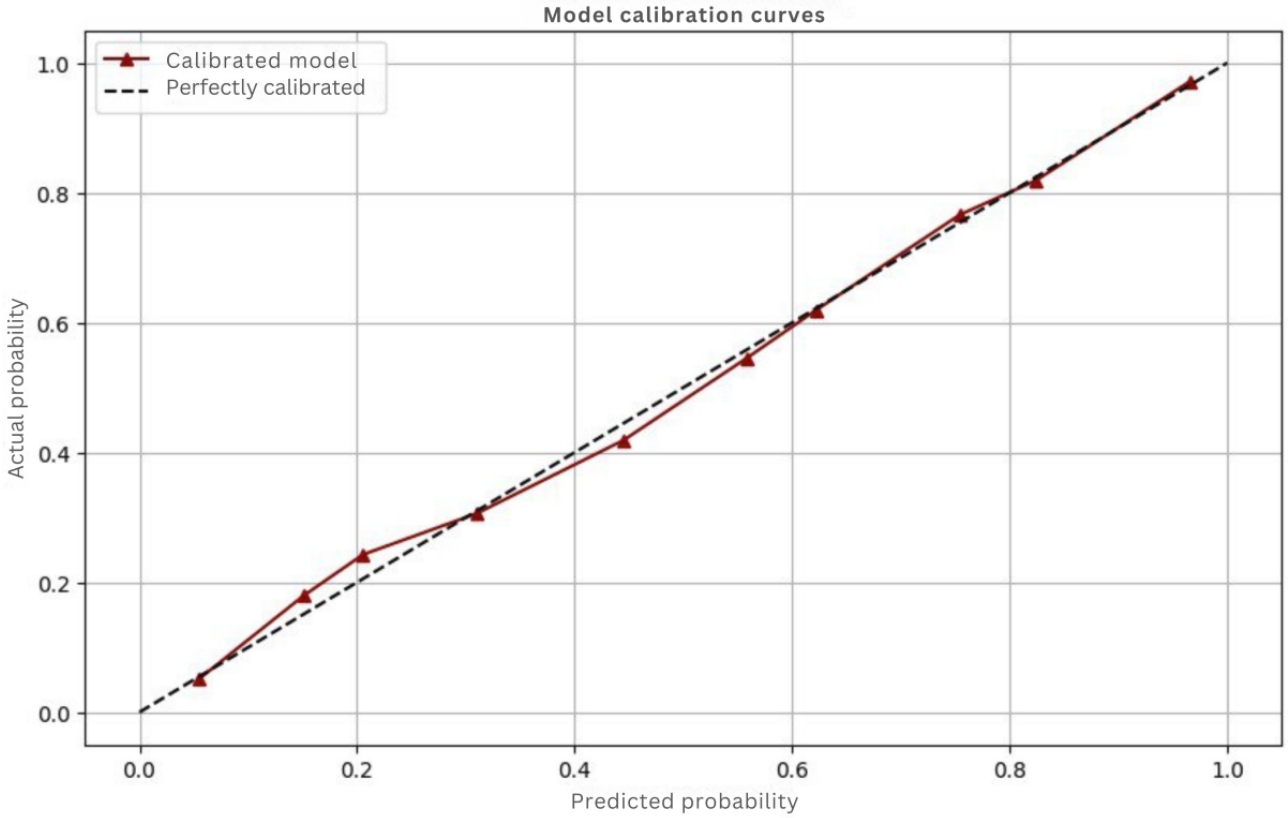
Given the categorical nature of the outcome variable, we used 7 classical classification algorithms: AdaBoost Classifier [32], Logistic Regression [33], Naive Bayes, Random Forest (RF) [34], Light Gradient Boosting Machine (LGBM) [35], Extreme Gradient Boosting (XGBoost) [35], and Decision Tree [36]. These models generate a predicted score between 0 and 1 for each child, which is then classified as ZD or non-ZD based on a defined threshold. Following the initial model comparison, hyperparametric tuning was conducted to further optimize the performance of the best performing algorithm using a RandomizedSearchCV with cross-validation. The search involved 100 iterations with each hyperparameter combination evaluated using 5-fold cross-validation. Finally, the performance of each model was tested before and after balancing the dataset to choose the best predictive model. The model comparison was carried out using the balanced dataset.

Model Evaluation

We evaluated model performance using both train-test split and cross-validation techniques, emphasizing both discrimination and calibration metrics to compare our classification of ZD status against the true ZD status of each child. Discrimination metrics included accuracy, precision, recall (sensitivity), F_1-score, and area under the curve and area under the receiver operating characteristic curve. Accuracy reflects the proportion of correctly classified instances among all tested cases [37], while precision indicates the ratio of true-positive predictions to all positive predictions [38]. Recall measures the proportion of actual positive cases that the model successfully identifies [39], and the F_1-score provides a balanced assessment of model performance, particularly useful in scenarios with class imbalances. In our application, as the objective is to assess the ability of a model to distinguish between positive and negative classes, area under the curve and area under the receiver operating characteristic curve emerged as the most critical measure, as it evaluates the model's ability to effectively distinguish between positive and negative classes by analyzing the trade-off between sensitivity and specificity [40].

In addition to discrimination metrics, we performed calibration to examine how well the predicted probabilities align with actual outcomes. While a model can demonstrate good discrimination, it may still exhibit biases in its risk predictions [41]. Calibration is essential to ensure that predicted probabilities accurately reflect the likelihood of outcomes. To visualize this alignment, we used calibration curves, which plot predicted probabilities against observed results (Figure 4). An ideally calibrated model would form a 45-degree diagonal line, signifying that predicted probabilities correspond closely to actual outcomes [42].

Figure 4. Calibration plot.



Important Feature Selection

Our second objective is to identify the most important predictors of ZD children. To achieve this, we used the best-performing ML model to determine the key features associated with identifying ZD cases. We used a unified framework developed by Lundberg and Lee [43], known as SHAP (SHapley Additive Explanations). This approach is based on Shapley values from cooperative game theory, which assign a value to each feature based on its contribution to the prediction, taking into account all possible combinations of features [44]. A waterfall plot is then created to visualize the cumulative effect of individual features on specific predictions, illustrating how each feature influences the final output. In addition, a beeswarm plot summarizes the distribution of SHAP values across multiple instances, revealing the variability and significance of feature contributions.

Rule Generation

We used rule mining techniques to uncover patterns and relationships within our dataset. We used association rule mining to identify correlations between features through Apriori algorithms [45]. In addition, we applied classification rule mining to generate rules that predict class labels, aiding in the identification of key predictors for ZD children, and explored sequential rule mining to capture temporal patterns where relevant. Following the mining process, we generated actionable insights by formulating human-readable rules that outline conditions (antecedents) and outcomes (consequents) [46]. We assessed the quality of these rules using metrics such as confidence and lift to ensure their reliability and relevance [47].

Ethical Considerations

The research was implemented in compliance with national and international ethical principles. The University of Gondar has provided ethical approval (CMHSSH-UOG IRERC/3/7/2024) to conduct this analysis. For this analyses we used the existing data with primary consent. We used deidentified data (summary data without individuals' identity) to ensure confidentiality. We followed the international standard of strengthening the reporting of cross-sectional studies in epidemiology.

Results

Children's and Mothers' or Caregivers' Characteristics

A total of 13,666 samples of children aged from 12 to 35 months were included for analysis. Nearly 57% (7727/13,666) of the children were younger than 24 - 35 months. The majority (10,204/13,666, 74.7%) of the children were from mothers or caregivers who live in rural areas. Half (6986/13,666, 51.1%) of the children were born from mothers who had not had formal education. More than half (6757/12,419, 54.4%) of the children were from mothers who had no PNC follow-up for the index children. The details are shown in Table 2.

Table . Sociodemographic and economic characteristics of mothers or caregivers of children aged 12 - 35 months in Ethiopia, 2023 (N=1366).

Variables	Frequency	Percentage
Age of the child		
12 - 23 months	5934	43.5
24 - 35 months	7727	56.5
Place of residency		
Rural	10,204	74.7
Urban	3462	25.3
Religion		
Orthodox	4430	32.4
Muslim	6158	45.2
Protestant	2944	21.5
Others[a]	134	1.0
Educational status		
No education	6986	51.1
Primary	3870	28.3
Secondary	1798	13.2
College and above	1012	7.4
Wealth status		
Poor	4558	33.4
Middle	4566	33.4
Richer	4542	33.2
Marital status		
Married and living together	12,765	93.4
Married but not living together	352	2.6
Not in marital union	549	4.0
Birth order		
First	4043	29.6
Second	4830	35.3
Third	2639	19.3
Fourth and above	2154	15.8
Parity		
Primipara	2499	20.1
Multipara (2-4)	6573	52.9
Grand multipara (5+)	3346	27.0
Perceived distance to health facility		
Big problem	5251	38.4
Not big problem	8415	61.6
Perceived benefit on immunization		
Poor	2502	19.3
Good	10,471	80.7
ANC[b] visit		
Yes	10,345	83.3

Variables	Frequency	Percentage
No	2074	16.7
Place of delivery		
Home	3807	30.6
Health facility	8612	69.4
PNC[c]		
Yes	5662	45.6
No	6757	54.4

[a]Others: Catholic, traditional, and others.

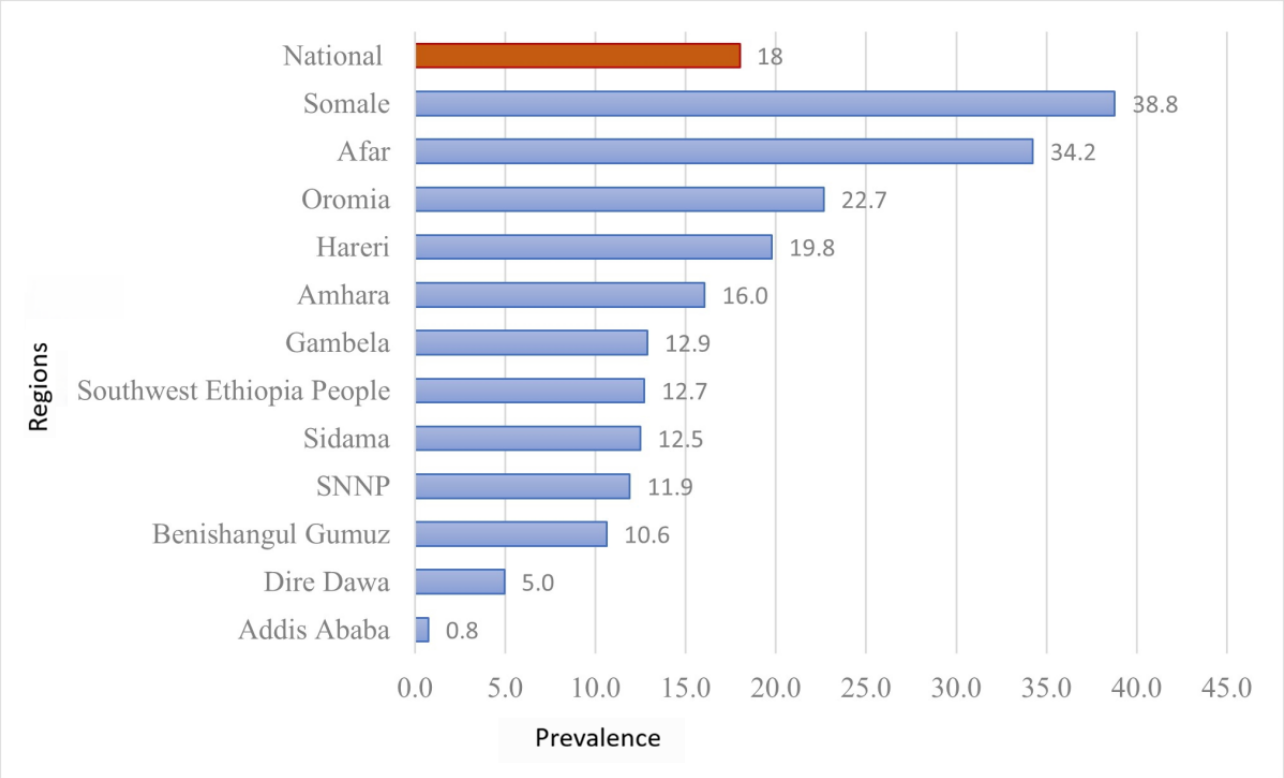
[b]ANC: antenatal care.

[c]PNC: postnatal care.

ZD Prevalence

The overall prevalence of ZD in Ethiopia was 18% (95% CI 17.4% - 18.7%). There were regional variations in the prevalence of ZD children. The higher prevalence was observed in Somali (38.8%), Afar (34.2%), and followed by Oromia (22.7%), and the lowest observed in Addis Ababa (0.8%) and Dire Dawa (5 %) (Figure 5).

Figure 5. Distribution of zero-dose children aged 12 - 35 months across regions in Ethiopia 2023. SNNP: Southern Nations, Nationalities, and Peoples'..



Performance of the Prediction Models

Seven ML algorithms were used to predict ZD status in Ethiopia, with the LGBM yielding the best performance for both unbalanced and balanced datasets (Table 3). It achieved accuracies of 89% and 93% for the unbalanced and balanced datasets, respectively. Most models showed improved accuracy when applied to the balanced dataset, except for Logistic Regression and Naive Bayes. After balancing the data, both XGBoost and LGBM reached an accuracy of 93%. Notably, the LGBM classifier excelled in terms of area under the curve (AUC) (98%) and sensitivity (92%).

Table . Model performance comparison before and after dataset balancing for predicting zero-dose children in Ethiopia, 2023.

Models and dataset	Accuracy (%)	AUC[a] (%)	Precision (%)	Sensitivity	F_1-score
Logistic Regression					
Unbalanced	88	88	77	48	59
Balanced	81	89	83	77	80
Naïve Bayes					
Unbalanced	85	87	59	63	61
Balanced	79	87	82	73	77
LGBM[b] Classifier					
Unbalanced	89	88	79	53	63
Balanced	93	97	94	91	92
DT[c] Classifier					
Unbalanced	86	75	64	51	57
Balanced	89	91	90	87	88
Random Forest Classifier					
Unbalanced	87	85	70	52	61
Balanced	91	96	91	90	91
XGBoost Classifier					
Unbalanced	88	87	75	52	61
Balanced	93	97	94	90	92
AdaBoost Classifier					
Unbalanced	88	88	77	46	58
Balanced	88	95	89	86	87

[a]AUC: area under the curve.

[b]LGBM: Light Gradient Boosting Machine.

[c]DT: Decision Tree.

Overall, while all ML models performed well on both datasets, those trained on balanced data especially XGBoost and LGBM proved to be more effective in identifying ZD children due to their higher recall and AUC. A comprehensive comparison of the ML algorithms used for ZD children is detailed in Table 3.

After the hyperparameter optimization conducted, the LGBM model achieved robust performance, with an accuracy of 92.4, an AUC of 97.4%, a precision of 93.2%, and a recall of 90.9%. The details are shown in Table 4.

Table . Model performance after hyperparameter tuning for predicting zero-dose children in Ethiopia, 2023.

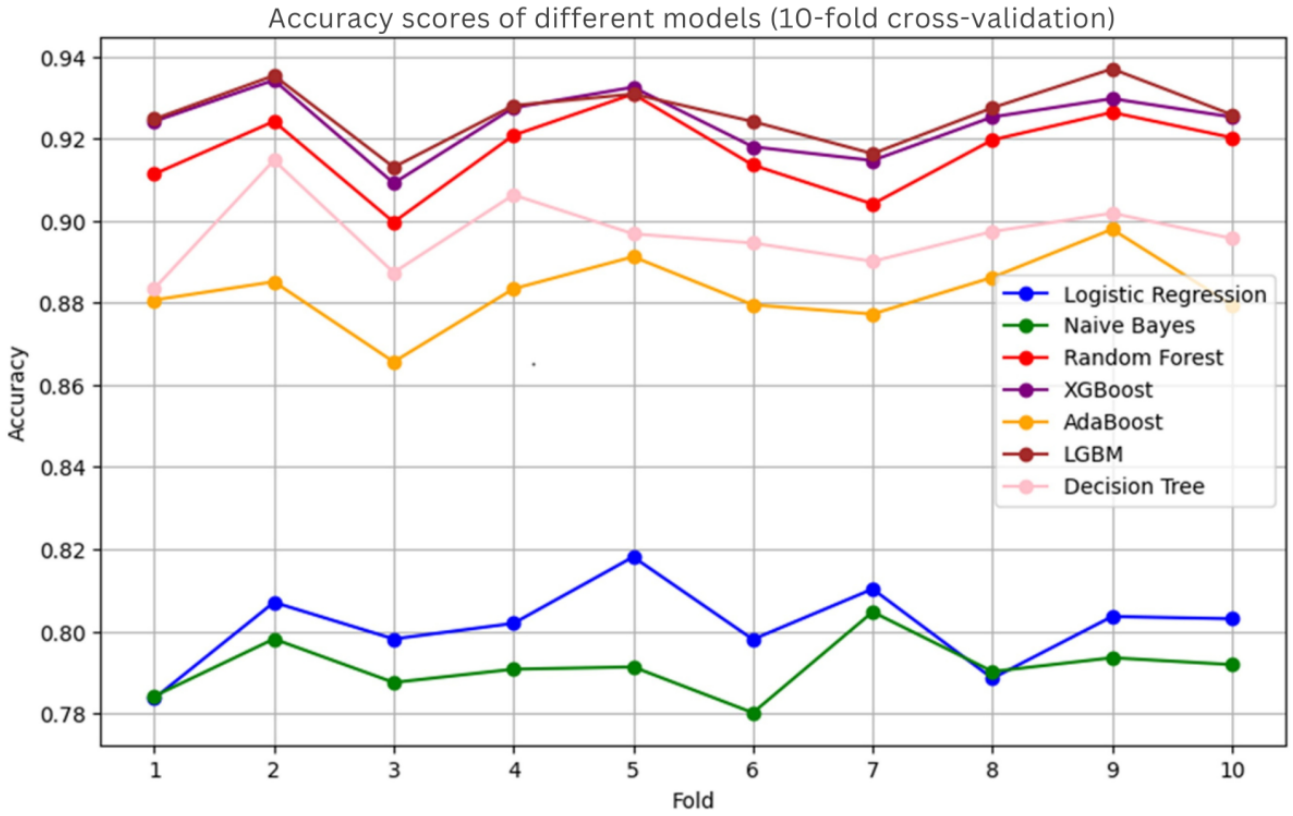
Model	Accuracy (%)	AUC[a] (%)	Precision (%)	Recall (%)	F_1-score (%)
Logistic Regression	80.8	89.1	82.5	76.9	79.6
Naive Bayes	79.1	87.3	82	73.1	77.3
Random Forest	91.6	96.7	91.9	90.6	91.3
XGBoost	92.2	97.3	93.9	89.8	91.8
AdaBoost	89.6	96.2	90.6	87.8	89.2
LGBM[b]	92.4	97.4	93.2	90.9	92.1
Decision Tree	89.9	94.5	89.5	89.7	89.6

[a]AUC: area under the curve.

[b]LGBM: Light Gradient Boosting Machine.

After parameter tuning, the models were further evaluated using 10-fold cross-validation, where XGBoost and LGBM demonstrated comparable accuracies of 93% (Figure 6).

Figure 6. Accuracy of models in 10-fold cross-validation after balancing the dataset for predicting zero-dose children in Ethiopia, 2023. LGBM: Light Gradient Boosting Machine.



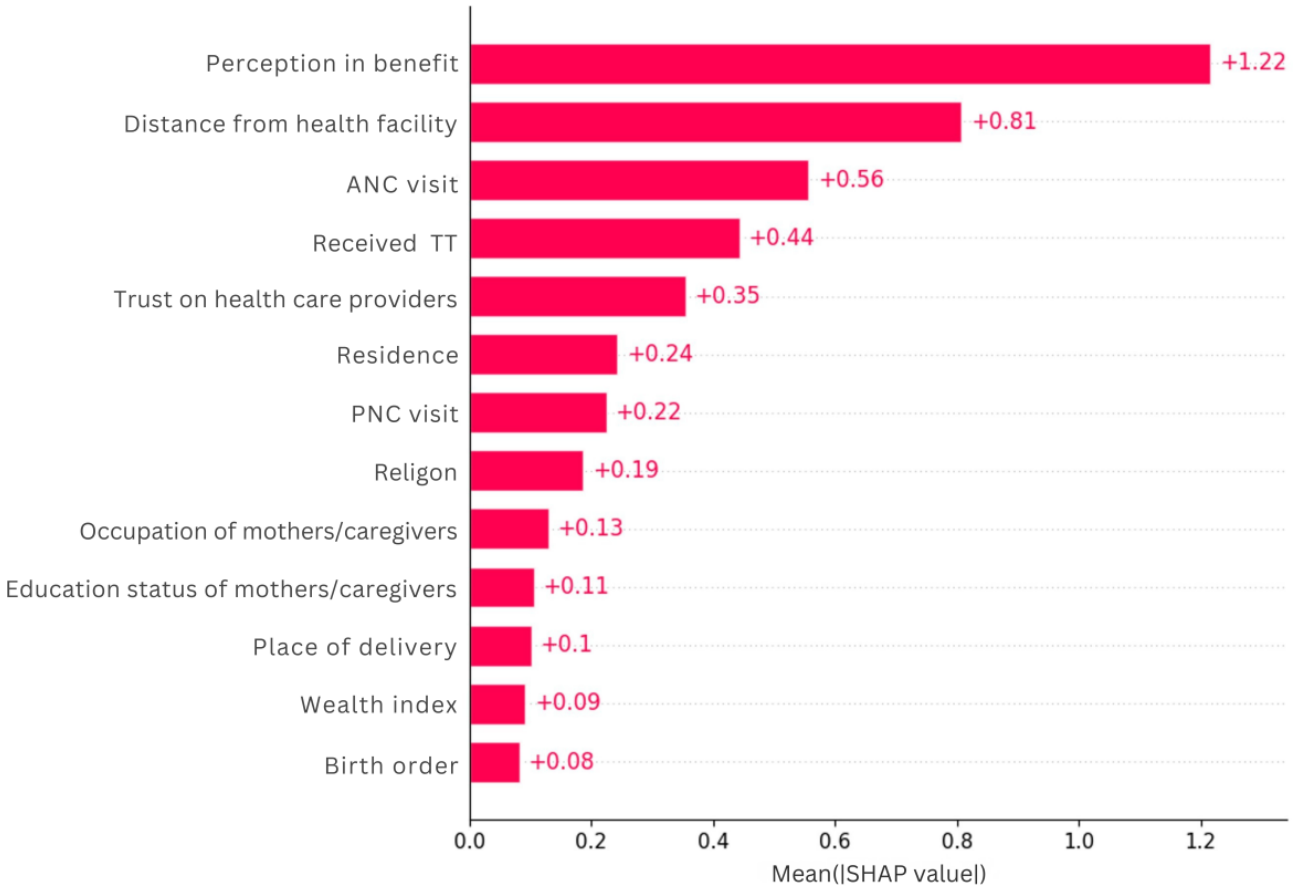
Predicting ZD

After building the model by using the training dataset, the performance of the LGBM model was evaluated by the testing dataset. From 2181 ZD children, the model predicted 1991 children correctly (true positive), and out of 2300 non-ZD children, the model predicted 2175 children correctly (true negative). However, the model incorrectly classified 190 ZD samples as non-ZD (false positive) and 125 non-ZD samples as ZD (false negative). The Matthews correlation coefficient was r=0.85 and Cohen κ=0.85. Overall, the model predicted with an accuracy of 93%, recall of 91%, F_1-score of 92%, and 94% precision on test data.

Feature Importance

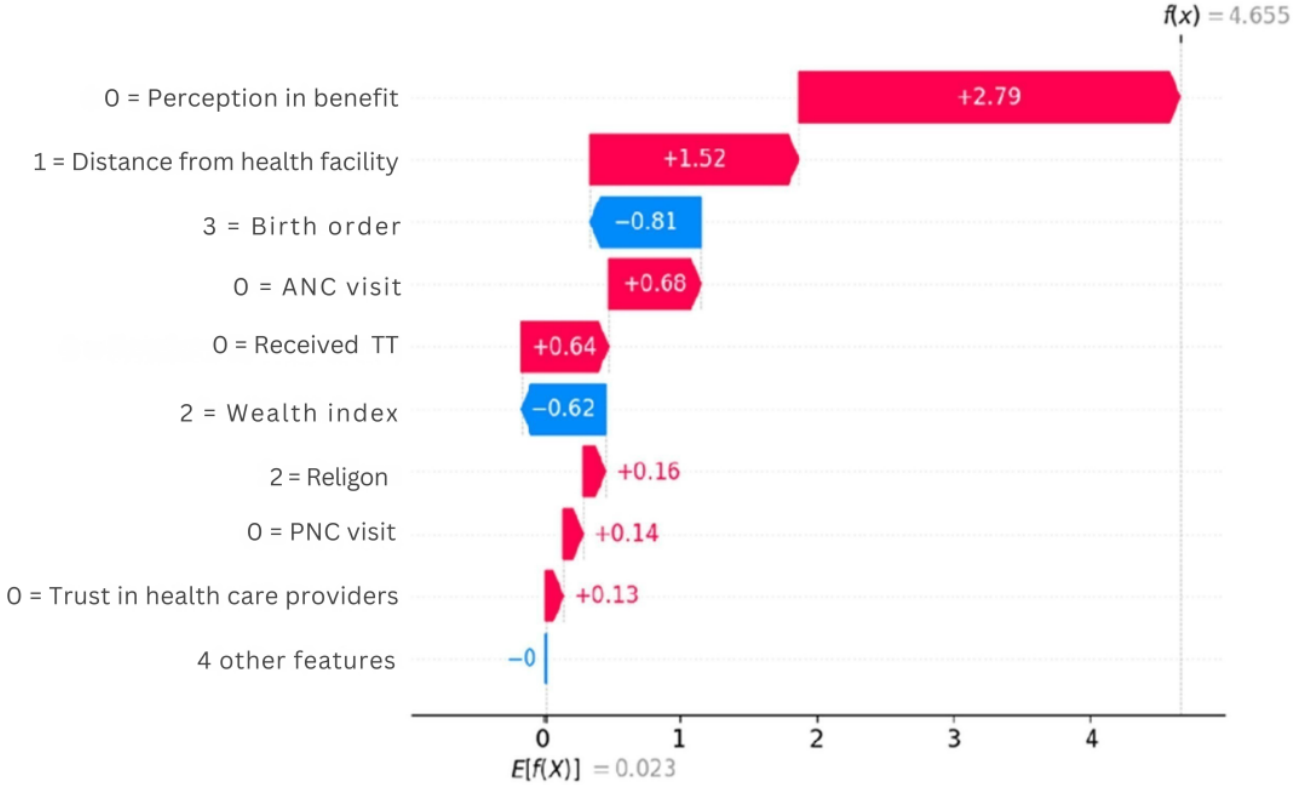
The Shapley Additive analysis identified that mother's or caregiver's perception of benefit of immunization (+1.13), with whether the distance to immunization site (+0.88), whether the mother received ANC (+0.55), whether the mother received TT (+0.42), and whether trust in health providers (0.41) were the most important features followed by place of residence (+0.35), and PNC visit (+0.25). Wealth index, birth order, and place of delivery were the features with low importance (Figure 7).

Figure 7. Important features for predicting zero-dose children in Ethiopia, 2023. ANC: antenatal care; PNC: postnatal care; SHAP: SHapley Additive Explanations.



The waterfall chart demonstrates how various factors influence the prediction of ZD vaccination status, starting from a baseline expected value of ($E[f(X)]$=0.023) and culminating in a final prediction of $f(x)$=4.655) indicating that the child is ZD. This indicated that poor perceptions of vaccination benefits, long distances to immunization sites, lack of antenatal and postnatal care visits, absence of TT vaccination, and low trust in health care providers are positively correlated with ZD. Conversely, being in a medium wealth index and having a third birth order is negatively correlated with ZD (Figure 8).

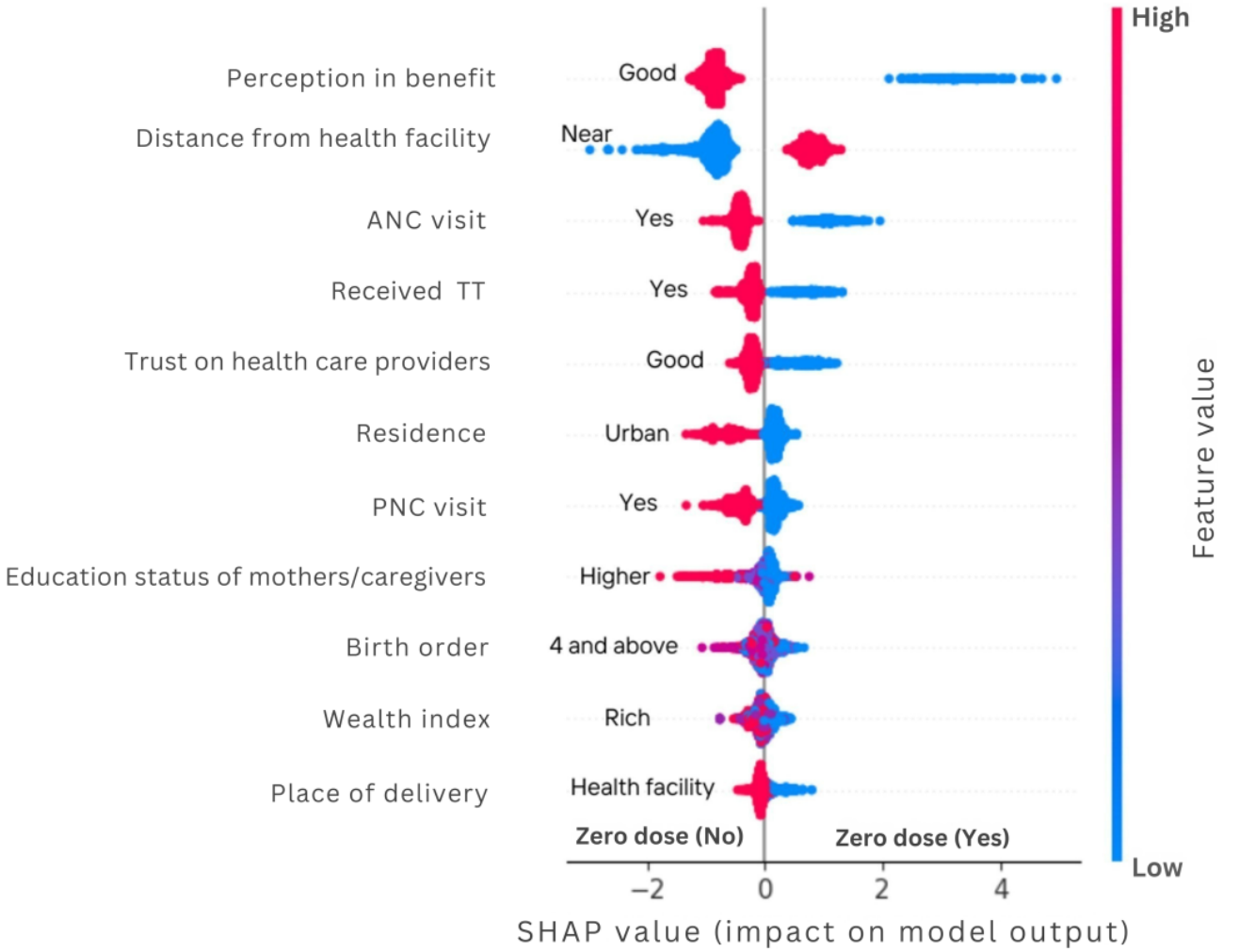
Figure 8. Waterfall plot of first observation value to predict zero-dose children in Ethiopia, 2023. ANC: antenatal care; PNC: postnatal care.



.

As shown in Figure 9, the beeswarm plot illustrates the impact of various predictor variables on ZD status, with distinct colors representing risk levels: red dots indicate high-risk values, while blue dots denote low-risk values for the predictor variables. The feature of perception exhibits a wide range of SHAP values, highlighting its significant influence on the model's predictions. A poor perception of the benefits of vaccination notably increases the likelihood of a child being classified as ZD. In addition, distance from health care facilities is strongly associated with ZD status, where far distances correlate with a higher likelihood of being unvaccinated. Other contributing factors include a lack of ANC visits, PNC visits, TT vaccination, low wealth index, low trust in health care providers, and home delivery, all of which contribute to the prediction of the positive class (ZD).

Figure 9. Zero-dose predictors for Light Gradient Boosting Machine model. SHAP summary plot of top predictors. ANC: antenatal care; PNC: postnatal care; SHAP: SHapley Additive Explanations.



Rule Generation

The rule generation process was done based on important attributes selected by the best performing ML model LGBM. Moving beyond individual feature importance, we used association rule mining to identify complex, multifactorial profiles of ZD children and to rigorously validate the interactions suggested by the SHAP analysis. This generated a set of human-interpretable "if then" rules, each validated by key metrics: support (prevalence of the rule in the data), confidence (conditional probability of the outcome), and lift (strength of the association above random chance). The rule generation process was done based on important attributes selected by the best performing ML model LGBM. The strongest rule (rule 1: lift = 2.17, confidence = 0.90) indicates that children whose caregivers live far from a health facility and have a poor perception of vaccination benefits have a 90% probability of being ZD, a risk 2.17 times higher than random. Rule 2 (confidence = 0.81) shows that combining distance with a lack of ANC and poor trust in providers creates another high-risk pathway, while rule 5 (confidence = 0.79) highlights the potent combination of no tetanus vaccination, no ANC, and distance. Crucially, these rules reveal critical synergies, while SHAP identified "distance" and "ANC" as top individual predictors, rule mining quantified how their combination with other factors (eg, rule 8: no ANC + Far distance, confidence = 0.72) creates a risk profile with a distinctly high probability of the outcome. This provides programmatically actionable insights, demonstrating that interventions must target these intersecting barriers simultaneously rather than in isolation to effectively reach ZD children. A total of 9 association rules were generated, and the details of the rules are shown in Textbox 1.

Textbox 1. Rule generation and knowledge extraction.

##Rule## 1: Distance from facility_far, Perception in benefit_Poor -> Zerodose_Yes

Support: 0.10897435897435898, Lift: 2.1738917080243128, Confidence: 0.9037974683544304

##Rule## 2: Distance from facility_far, Trust in healthcare provider_Poor, Anc visit_No -> Zerodose_Yes

Support: 0.10134310134310134, Lift: 1.947695283120232, Confidence: 0.8097560975609757

##Rule## 3: Perception in benefit_Poor -> Zerodose_Yes

Support: 0.15262515262515264, Lift: 1.9211552265274243, Confidence: 0.7987220447284347

##Rule## 4: Perception in benefit_Poor, Place Residence_rural -> Zerodose_Yes

Support: 0.10103785103785104, Lift: 1.9138215859030838, Confidence: 0.795673076923077

##Rule## 5: Received TT_No, Distance from facility_far, Anc visit_No -> Zerodose_Yes

Support: 0.13064713064713065, Lift: 1.9099490817552491, Confidence: 0.7940630797773656

##Rule## 6: Received TT_No, Distance from facility_far, Place delivery_Home -> Zerodose_Yes

Support: 0.12606837606837606, Lift: 1.7738986784140967, Confidence: 0.7374999999999999

##Rule## 7: PNC visit_Yes, Distance from facility_far, Anc visit_No -> Zerodose_Yes

Support: 0.10073260073260074, Lift: 1.7406677486668212, Confidence: 0.7236842105263158

##Rule## 8: Distance from facility_far, Anc visit_No -> Zerodose_Yes

Support: 0.17918192918192918, Lift: 1.736658159533137, Confidence: 0.7220172201722017

##Rule## 9: Distance from facility_far, Anc visit_No, Place Residence_rural -> Zerodose_Yes

Support: 0.1108058608058608, Lift: 1.7019862431408919, Confidence: 0.7076023391812866

Discussion

Principal Findings

Using the data from the most recent National Immunization Evaluation Survey in Ethiopia, we applied different supervised machine algorithms to assess how well the models predict whether a child is likely to be ZD and to identify the important predictor variables. We trained and compared 7 ML classifiers on both unbalanced and balanced datasets, using a train-test split, hyperparameter tuning, and 10-fold cross-validation for robust evaluation. A variety of socioeconomic, demographic, and health-related factors were included to enhance the model's predictions and facilitate important feature selection.

Our findings demonstrate that these ML algorithms are effective in identifying children at high risk of being ZD. Among the 7 models tested, LGBM emerged as the top performer, achieving an AUC of 97.4%, recall of 90.9%, accuracy of 92.4%, precision of 93.2%, and an F_1-score of 92.1%. These evaluation metrics underscore the model's strong capability in predicting ZD children. The high AUC indicates the model's effectiveness in distinguishing between children who receive immunization services and those who do not. Notably, a recall of 90.9% signifies that the model successfully identifies 90.9% of ZD children, who are often at greater risk for missing vaccines and vaccine-preventable diseases.

In addition to LGBM, both XGBoost and RF algorithms performed well, each achieving an accuracy of 92.2% and 91.6%, respectively. These results are consistent with previous studies that recognized XGBoost [48] and RF [49] as top performers in similar contexts. While these metrics indicate robust performance for critical health issues such as immunization, it is crucial to validate the model in real-world settings. Such testing will enhance its utility as a tool for guiding public health initiatives aimed at increasing vaccination rates and improving access to essential health care services for unvaccinated children.

Using an ML model, health care workers can pinpoint specific households and communities with ZD children, allowing them to shift from broad campaigns to targeted household visits. By leveraging the model's insights on local perceptions and socioeconomic barriers, they can tailor their communication and services, such as setting up mobile clinics, to overcome specific challenges and efficiently use scarce resources, ensuring that vaccines reach those most in need.

The second objective of the study was to identify important attributes that could predict ZD among children aged 12 - 35 months. Using SHAP analysis, the study found that perception of immunization benefit, ANC utilization, distance from vaccination site, maternal TT vaccination status, and trust in health providers were the most important features to identify at-risk children for ZD.

The top predictor was poor maternal perception with a SHAP value of 1.13 (Figure 7). This indicates that a negative perception of mothers or caregivers increases the likelihood of a child being ZD, likely because parental beliefs directly influence health care decisions regarding vaccination. This finding aligns with previous studies showing that parental beliefs and attitudes significantly affect a child's vaccination status [50,51].

ANC utilization was another important feature, with a SHAP value of 0.55, indicating that a lack of ANC is strongly linked to a child being ZD. This finding is in line with the previous

similar studies done [48,52,53]. This could be due to the fact that ANC visits enable mothers to access integrated health services, be more likely to receive information on immunization schedules, build trust in the health system, and improve adherence to health services [54-56].

In addition, the study found several important predictors. Maternal TT vaccination was a key factor; mothers who received the TT vaccine were more likely to have their children vaccinated, a finding consistent with studies from Sudan and Bangladesh [57,58]. Postnatal care visit was another important predictor. This service is likely gaining a better understanding of vaccination importance and feeding practices, thus reducing missed vaccinations [59,60]. In addition, maternal education was an important predictor, with uneducated mothers having a higher risk of ZD children than those with at least a primary education, firming up the known link between maternal literacy and vaccination rates and primary education. This finding is in line with previous research linking maternal literacy to vaccination completeness [61-65]. The other finding of this study is rule mining and generation. Using association rule mining with the Apriori algorithm, the study uncovered strong relationships between various socioeconomic, demographic, and health-related factors and ZD status. Key determinants, including distance from health facility, perception of vaccination benefits, trust in health care providers, ANC, place of delivery, place of residency, and TT vaccines were the most important features predicting ZD. Confidence levels for these findings ranged from 71% to 90%, indicating robust associations.

Findings from association rule 1 indicated that the probability of a child being ZD would be 90%, if and only if the mothers or caregivers were far from the health facility and had poor perception on immunization. This may be because mothers or caregivers who are far from the facility may not have access to health education directly or indirectly, affecting health-seeking behavior and health service utilization such as vaccination. The second rule also included poor trust in health care providers and lack of ANC visits as predictors for ZD. A child ZD would be 80% if mothers or caregivers have trust in providers and had no ANC follow-up for the index child.

Strengths and Limitations

This study had several strengths worth mentioning. We used national-level survey data from 463 EAs ensuring generalizability across the country and providing a current snapshot of the ZD situation. A key strength is that our analysis uses various ML algorithms from the field of data science, which significantly aids in identifying and targeting ZD children more effectively. These advanced analytical techniques allow us to process large datasets and uncover insights that may not be immediately apparent through traditional methods. At the same time, this study identified the risk factors of ZD that may help policy makers and planners to design tailored interventions to identify and reach the unvaccinated children.

This study was subject to some limitations. First, although we used national-level data, we did not include data from the Tigray region, which is one of the administrative regions of the country, due to security issues. Second, the study did not include health system side predictors such as availability of vaccination supplies and vaccines. Finally, we could not do external validation for the modes due to the lack of real-world data.

Conclusions

The developed ML models effectively predict children at risk of being ZD and identify associated risk factors. Among these models, the LGBM model demonstrated the best performance in predicting ZD children. Key features linked to ZD status include access to immunization sites, maternal health service utilization (such as antenatal and postnatal care, place of delivery, and TT vaccination), and perceptions regarding immunization.

By implementing ML models, public health interventions can be more precisely targeted at the most vulnerable groups. This approach may address inequities in vaccination coverage by identifying specific sociodemographic, economic, and health-related factors associated with ZD children. Consequently, it aids in the formulation and implementation of effective policies and strategies to improve vaccination rates. Strengthening the continuum of care for mothers, raising awareness among caregivers, and improving immunization access through outreach strategies may help in reducing the high burden of ZD children.

Acknowledgments

The authors would like to thank the University of Gondar for the unwavering support. In addition, the authors extend their sincere thanks to the Ministry of Health of Ethiopia for granting them access to the data that were critical for their study.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

BFE, KA, SAM, MZ, MKG, and BT conceptualized and designed the study. All authors contributed significantly to the analysis and interpretation of the results. The original draft was written by BFE and reviewed by all authors, who also provided critical revisions for important intellectual content. Each author has read and approved the final manuscript. They have agreed to take personal responsibility for their contributions and to ensure that any questions regarding the accuracy or integrity of any part of the work, even if they were not directly involved, are properly investigated and documented in the literature.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Zero-dose status among children aged 12 - 35 months in Ethiopia, before and after data balancing, using the 2023 survey dataset. SMOTE: Synthetic Minority Oversampling Technique.
[PNG File, 162 KB - pediatrics_v9i1e76712_app1.png]

References

1. Clark H, Coll-Seck AM, Banerjee A, et al. A future for the world's children? A WHO–UNICEF–Lancet Commission. Lancet 2020 Feb 22;395(10224):605-658. [doi: 10.1016/S0140-6736(19)32540-1] [Medline: 32085821]
2. Kaur G, Danovaro-Holliday MC, Mwinnyaa G, et al. Routine vaccination coverage—worldwide, 2022. MMWR Morb Mortal Wkly Rep 2023 Oct 27;72(43):1155-1161. [doi: 10.15585/mmwr.mm7243a1] [Medline: 37883326]
3. Sharrow D, Hug L, You D, et al. Global, regional, and national trends in under-5 mortality between 1990 and 2019 with scenario-based projections until 2030: a systematic analysis by the UN Inter-agency Group for Child Mortality Estimation. Lancet Glob Health 2022 Feb;10(2):e195-e206. [doi: 10.1016/S2214-109X(21)00515-5] [Medline: 35063111]
4. WHO/UNICEF estimates of national immunization coverage. World Health Organization. 2023 Sep. URL: https://www.who.int/news-room/questions-and-answers/item/who-unicef-estimates-of-national-immunization-coverage [accessed 2026-01-19]
5. Immunization Agenda 2030: a global strategy to leave no one behind. : World Health Organization; 2020 URL: https://www.who.int/publications/m/item/immunization-agenda-2030-a-global-strategy-to-leave-no-one-behind [accessed 2024-11-22]
6. Restrepo-Méndez MC, Barros AJ, Wong KL, et al. Inequalities in full immunization coverage: trends in low- and middle-income countries. Bull World Health Organ 2016 Nov 1;94(11):794-805B. [doi: 10.2471/BLT.15.162172] [Medline: 27821882]
7. Ozawa S, Yemeke TT, Evans DR, Pallas SE, Wallace AS, Lee BY. Defining hard-to-reach populations for vaccination. Vaccine (Auckl) 2019 Sep;37(37):5525-5534. [doi: 10.1016/j.vaccine.2019.06.081]
8. Johri M, Rajpal S, Subramanian SV. Progress in reaching unvaccinated (zero-dose) children in India, 1992-2016: a multilevel, geospatial analysis of repeated cross-sectional surveys. Lancet Glob Health 2021 Dec;9(12):e1697-e1706. [doi: 10.1016/S2214-109X(21)00349-1] [Medline: 34798029]
9. Johri M, Ng ES, Sharkey A, Bosson-Rieutort D, Kone GK, Subramanian SV. Effects of zero-dose vaccination status in early childhood and level of community socioeconomic development on learning attainment in preadolescence in India: a population-based cohort study. bmjph 2023 Nov;1(1):e000022. [doi: 10.1136/bmjph-2023-000022]
10. Chopra M, Bhutta Z, Chang Blanc D, et al. Addressing the persistent inequities in immunization coverage. Bull World Health Organ 2020 Feb 1;98(2):146-148. [doi: 10.2471/BLT.19.241620] [Medline: 32015586]
11. Teferi E. Factors influencing coverage and key challenges to achieving targets of routine immunization in Africa: a systematic review. Ethiop J Pediatr Child Health 2016;12(2):34-45 [FREE Full text]
12. Nour TY, Farah AM, Ali OM, Osman MO, Aden MA, Abate KH. Predictors of immunization coverage among 12–23 month old children in Ethiopia: systematic review and meta-analysis. BMC Public Health 2020 Dec;20(1):1-19. [doi: 10.1186/s12889-020-09890-0]
13. Tamir TT, Zegeye AF, Mekonen EG, et al. Prevalence, spatial variation and determinants of zero-dose children in Ethiopia: spatial and multilevel analyses. Public Health (Fairfax) 2024 Nov;236:365-372. [doi: 10.1016/j.puhe.2024.09.011]
14. Yadita ZS, Ayehubizu LM. Full immunization coverage and associated factors among children aged 12-23 months in Somali Region, Eastern Ethiopia. PLoS One 2021;16(12):e0260258. [doi: 10.1371/journal.pone.0260258] [Medline: 34874949]
15. Biset G, Woday A, Mihret S, Tsihay M. Full immunization coverage and associated factors among children age 12-23 months in Ethiopia: systematic review and meta-analysis of observational studies. Hum Vaccin Immunother 2021 Jul 3;17(7):2326-2335. [doi: 10.1080/21645515.2020.1870392]
16. Girmay A, Dadi AF. Full immunization coverage and associated factors among children aged 12-23 months in a hard-to-reach areas of Ethiopia. Int J Pediatr 2019;2019:1924941. [doi: 10.1155/2019/1924941] [Medline: 31263502]

17. Gurmu E, Etana D. Factors influencing children's full immunization in Ethiopia. Afr Popul Stud 2016;30:2. [doi: 10.11564/30-2-813]
18. Asresie MB, Dagnew GW, Bekele YA. Changes in immunization coverage and contributing factors among children aged 12-23 months from 2000 to 2019, Ethiopia: multivariate decomposition analysis. PLoS One 2023;18(9):e0291499. [doi: 10.1371/journal.pone.0291499] [Medline: 37703252]
19. Nozaki I, Hachiya M, Kitamura T. Factors influencing basic vaccination coverage in Myanmar: secondary analysis of 2015 Myanmar demographic and health survey data. BMC Public Health 2019 Feb 28;19(1):242. [doi: 10.1186/s12889-019-6548-0] [Medline: 30819127]
20. Song X, Mitnitski A, Cox J, Rockwood K. Comparison of machine learning techniques with classical statistical models in predicting health outcomes. Stud Health Technol Inform 2004;107(Pt 1):736-740. [Medline: 15360910]
21. Bonaccorso G. Machine Learning Algorithms: Popular Algorithms for Data Science and Machine Learning: Packt Publishing Ltd; 2018. URL: https://cir.nii.ac.jp/crid/1130000794529220864 [accessed 2024-11-22]
22. Sarker IH. Machine learning: algorithms, real-world applications and research directions. SN Comput Sci 2021;2(3):160. [doi: 10.1007/s42979-021-00592-x] [Medline: 33778771]
23. Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. Presented at: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data; Jun 1, 1993; Washington, DC p. 207-216. [doi: 10.1145/170035.170072]
24. Cheong Q, Au-Yeung M, Quon S, Concepcion K, Kong JD. Predictive modeling of vaccination uptake in US counties: a machine learning-based approach. J Med Internet Res 2021 Nov 25;23(11):e33231. [doi: 10.2196/33231] [Medline: 34751650]
25. Carrieri V, Lagravinese R, Resce G. Predicting vaccine hesitancy from area-level indicators: a machine learning approach. Health Econ 2021 Dec;30(12):3248-3256. [doi: 10.1002/hec.4430] [Medline: 34523180]
26. Avirappattu G, Pach Iii A, Locklear CE, Briggs AQ. An optimized machine learning model for identifying socio-economic, demographic and health-related variables associated with low vaccination levels that vary across ZIP codes in California. Prev Med Rep 2022 Aug;28:101858. [doi: 10.1016/j.pmedr.2022.101858] [Medline: 35706686]
27. Zhang A, Xing L, Zou J, Wu JC. Shifting machine learning for healthcare from development to deployment and from models to data. Nat Biomed Eng 2022 Dec;6(12):1330-1345. [doi: 10.1038/s41551-022-00898-y] [Medline: 35788685]
28. National immunization program evaluation research in Ethiopia. : Consortium of CBMP Universities; 2024 URL: https://www.scribd.com/document/875138308 [accessed 2025-06-29]
29. World Health Organization vaccination coverage cluster surveys: reference manual. World Health Organization. 2018. URL: https://www.who.int/publications/i/item/WHO-IVB-18.09 [accessed 2024-11-23]
30. Wonodi C, Farrenkopf BA. Defining the zero dose child: a comparative analysis of two approaches and their impact on assessing the zero dose burden and vulnerability profiles across 82 low- and middle-income countries. Vaccines (Basel) 2023 Sep 28;11(10):1543. [doi: 10.3390/vaccines11101543] [Medline: 37896946]
31. Zero dose children and missed communities. Gavi. URL: https://www.gavi.org/our-alliance/strategy/phase-5-2021-2025/equity-goal/zero-dose-children-missed-communities [accessed 2026-01-19]
32. Schapire RE, Freund Y. Boosting: foundations and algorithms. Kybernetes 2013 Jan 4;42(1):164-166. [doi: 10.1108/03684921311295547]
33. Cessie SL, Houwelingen JCV. Ridge estimators in logistic regression. Appl Stat 1992;41(1):191. [doi: 10.2307/2347628]
34. Pal M. Random forest classifier for remote sensing classification. Int J Remote Sens 2005 Jan;26(1):217-222. [doi: 10.1080/01431160412331269698]
35. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 13-17, 2016; San Francisco, CA p. 785-794. [doi: 10.1145/2939672.2939785]
36. Breiman L, Friedman J, Olshen RA, Stone CJ. Classification and Regression Trees: Routledge; 2017. [doi: 10.1201/9781315139470]
37. Šimundić AM. Measures of Diagnostic Accuracy: Basic Definitions. EJIFCC 2009 Jan;19(4):203-211. [Medline: 27683318]
38. Flach P, Kull M. Precision-recall-gain curves: PR analysis done right. Adv Neural Inf Process Syst 2015;28 [FREE Full text]
39. Santini A, Man A, Voidăzan S. Accuracy of diagnostic tests. J Crit Care Med 2021 Jul 1;7(3):241-248. [doi: 10.2478/jccm-2021-0022]
40. Kumar R, Indrayan A. Receiver operating characteristic (ROC) curve for medical researchers. Indian Pediatr 2011 Apr;48(4):277-287. [doi: 10.1007/s13312-011-0055-4] [Medline: 21532099]
41. Collins GS, Reitsma JB, Altman DG, Moons KGM, TRIPOD Group. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. The TRIPOD Group. Circulation 2015 Jan 13;131(2):211-219. [doi: 10.1161/CIRCULATIONAHA.114.014508] [Medline: 25561516]
42. Huang Y, Li W, Macheret F, Gabriel RA, Ohno-Machado L. A tutorial on calibration measurements and calibration models for clinical prediction models. J Am Med Inform Assoc 2020 Apr 1;27(4):621-633. [doi: 10.1093/jamia/ocz228]

43. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. : Curran Associates Inc; 2017 Presented at: Proceedings of the 31st International Conference on Neural Information Processing Systems; Dec 4-9, 2017; Long Beach, CA p. 4768-4777 URL: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html [accessed 2026-01-25]
44. Bifarin OO. Interpretable machine learning with tree-based SHapley Additive Explanations: application to metabolomics datasets for binary classification. PLoS One 2023;18(5):e0284315. [doi: 10.1371/journal.pone.0284315] [Medline: 37141218]
45. Altaf W, Shahbaz M, Guergachi A. Applications of association rule mining in health informatics: a survey. Artif Intell Rev 2017 Mar;47(3):313-340. [doi: 10.1007/s10462-016-9483-9]
46. Khare S, Gupta D. Association rule analysis in cardiovascular disease. Presented at: 2016 Second International Conference on Cognitive Computing and Information Processing (CCIP); Aug 12-13, 2016; Mysuru, India p. 1-6. [doi: 10.1109/CCIP.2016.7802881]
47. Agrawal R, Srikant R. Fast algorithms for mining association rules. 1994 Presented at: Proceedings of the 20th International Conference on Very Large Data Bases; Sep 12-15, 1994; San Francisco, CA, USA p. 487-499 URL: https://www.columbia.edu/~rd2537/docu/apriori(abstract).pdf [accessed 2024-11-24]
48. Tadese ZB, Nigatu AM, Yehuala TZ, Sebastian Y. Prediction of incomplete immunization among under-five children in East Africa from recent demographic and health surveys: a machine learning approach. Sci Rep 2024 May 21;14(1):11529. [doi: 10.1038/s41598-024-62641-8] [Medline: 38773175]
49. Chandir S, Siddiqi DA, Hussain OA, et al. Using predictive analytics to identify children at high risk of defaulting from a routine immunization program: feasibility study. JMIR Public Health Surveill 2018 Sep 4;4(3):e63. [doi: 10.2196/publichealth.9681] [Medline: 30181112]
50. Chambongo PE, Nguku P, Wasswa P, Semali I. Community vaccine perceptions and its role on vaccination uptake among children aged 12-23 months in the Ileje District, Tanzania: a cross section study. Pan Afr Med J 2016;23:162. [doi: 10.11604/pamj.2016.23.162.8925] [Medline: 27303578]
51. Puri YE, Murti B, Demartoto A. Analysis of the effect of maternal perception on completeness of child immunization status with health belief model. J Health Promot Behav 2016;01(3):211-222. [doi: 10.26911/thejhpb.2016.01.03.08]
52. Demsash AW, Chereka AA, Walle AD, Kassie SY, Bekele F, Bekana T. Machine learning algorithms' application to predict childhood vaccination among children aged 12–23 months in Ethiopia: evidence 2016 Ethiopian Demographic and Health Survey dataset. PLoS One 2023;18(10):e0288867. [doi: 10.1371/journal.pone.0288867]
53. Biswas A, Tucker J, Bauhoff S. Performance of predictive algorithms in estimating the risk of being a zero-dose child in India, Mali and Nigeria. BMJ Glob Health 2023 Oct;8(10):e012836. [doi: 10.1136/bmjgh-2023-012836] [Medline: 37821114]
54. Dixit P, Dwivedi LK, Ram F. Strategies to improve child immunization via antenatal care visits in India: a propensity score matching analysis. PLoS One 2013;8(6):e66175. [doi: 10.1371/journal.pone.0066175] [Medline: 23824555]
55. Jiao B, Iversen I, Sato R, et al. Association between achieving adequate antenatal care and health-seeking behaviors: a study of demographic and health surveys in 47 low- and middle-income countries. PLoS Med 2024 Jul;21(7):e1004421. [doi: 10.1371/journal.pmed.1004421] [Medline: 38968303]
56. Krishnamoorthy Y, Rehman T. Impact of antenatal care visits on childhood immunization: a propensity score-matched analysis using nationally representative survey. Fam Pract 2022 Jul 19;39(4):603-609. [doi: 10.1093/fampra/cmab124] [Medline: 34564727]
57. Ibrahim ZA, Sabahelzain MM, Elhadi YAM, Malande OO, Babiker S. Predictors of tetanus vaccine uptake among pregnant women in Khartoum State, Sudan: a hospital-based cross-sectional study. Vaccines (Basel) 2023 Jul 21;11(7):7. [doi: 10.3390/vaccines11071268] [Medline: 37515083]
58. Amin MB, Roy N, Meem AE, Hossain E, Aktarujjaman M. Trends and determinants of taking tetanus toxoid vaccine among women during last pregnancy in Bangladesh: country representative survey from 2006 to 2019. PLoS One 2022;17(10):e0276417. [doi: 10.1371/journal.pone.0276417]
59. Laryea DO, Abbeyquaye Parbie E, Frimpong E. Timeliness of childhood vaccine uptake among children attending a tertiary health service facility-based immunisation clinic in Ghana. BMC Public Health 2014 Jan 29;14:1-5. [doi: 10.1186/1471-2458-14-90] [Medline: 24476339]
60. Sarker AR, Akram R, Ali N, Chowdhury ZI, Sultana M. Coverage and determinants of full immunization: vaccination coverage among Senegalese children. Medicina 2019;55(8):480. [doi: 10.3390/medicina55080480]
61. Acharya P, Kismul H, Mapatano MA, Hatløy A. Individual- and community-level determinants of child immunization in the Democratic Republic of Congo: a multilevel analysis. PLoS One 2018;13(8):e0202742. [doi: 10.1371/journal.pone.0202742] [Medline: 30138459]
62. Rauniyar SK, Munkhbat E, Ueda P, Yoneoka D, Shibuya K, Nomura S. Timeliness of routine vaccination among children and determinants associated with age-appropriate vaccination in Mongolia. Heliyon 2020 Sep;6(9):e04898. [doi: 10.1016/j.heliyon.2020.e04898] [Medline: 32995607]
63. Budu E, Darteh EKM, Ahinkorah BO, Seidu AA, Dickson KS. Trend and determinants of complete vaccination coverage among children aged 12-23 months in Ghana: analysis of data from the 1998 to 2014 Ghana Demographic and Health Surveys. PLoS One 2020;15(10):e0239754. [doi: 10.1371/journal.pone.0239754]

64. Noh JW, Kim YM, Akram N, et al. Factors affecting complete and timely childhood immunization coverage in Sindh, Pakistan; a secondary analysis of cross-sectional survey data. PLoS One 2018;13(10):e0206766. [doi: 10.1371/journal.pone.0206766]
65. Sarker AR, Akram R, Ali N, Sultana M. Coverage and factors associated with full immunisation among children aged 12–59 months in Bangladesh: insights from the nationwide cross-sectional demographic and health survey. BMJ Open 2019 Jul;9(7):e028020. [doi: 10.1136/bmjopen-2018-028020]

Abbreviations

ANC: antenatal care
AUC: area under the curve
DPT: diphtheria, tetanus, and pertussis
EA: enumeration area
LGBM: Light Gradient Boosting Machine
ML: machine learning
RF: Random Forest
SHAP: SHapley Additive Explanations
TT: tetanus toxoid
XGBoost: Extreme Gradient Boosting

Edited by S Badawy; submitted 29.Apr.2025; peer-reviewed by F Pabon-Rodriguez, M Adisu, N Meng; revised version received 18.Sep.2025; accepted 14.Oct.2025; published 02.Feb.2026.

Please cite as:
Endehabtu BF, Alemu K, Mengiste SA, Zelalem M, Gullslett MK, Tilahun B
Modeling Zero-Dose Children in Ethiopia: A Machine Learning Perspective on Model Performance and Predictor Variables
JMIR Pediatr Parent 2026;9:e76712
URL: https://pediatrics.jmir.org/2026/1/e76712
doi:10.2196/76712

Increasing Use of a Postpartum and Newborn Chatbot among Birthing Individuals and Caregivers: Development and Implementation Study

Jessica N Rivera Rivera[1], PhD, MPH; Marjanna Smith[1], MPH; Shrey Mathur[2], MS; Katarina E AuBuchon[3], PhD; Angela D Thomas[1], DrPH, MPH, MBA; Hannah Arem[1,4], PhD, MHS

[1]Healthcare Delivery Research, MedStar Health Research Institute, 100 Irving St. NW, Washington, DC, United States

[2]Center for Biostatistics, Informatics and Data Science, MedStar Health Research Institute, Washington, DC, United States

[3]Palliative Care Research Group, MedStar Health Research Institute, Washington, DC, United States

[4]Department of Oncology, Georgetown University, Washington, DC, United States

Corresponding Author:
Jessica N Rivera Rivera, PhD, MPH
Healthcare Delivery Research, MedStar Health Research Institute, 100 Irving St. NW, Washington, DC, United States

Abstract

Background: The 42 days following childbirth are a high-risk period for birthing individuals and newborns. We created 2 rule-based chatbots, 1 for birthing individuals and 1 for newborn caregivers, to deliver information on postpartum and newborn warning signs, follow-up care, and other relevant resources during this high-risk period.

Objective: This study aims to examine strategies for implementing the chatbot following discharge from a large hospital center, initial chatbot reach, and subsequent reach after chatbot refinement based on end-user feedback.

Methods: Reach was defined as the number of users opening the chatbot out of those who received it. Birthing individuals' demographic (age, ethnicity, race, language, and insurance type) and clinical characteristics (delivery method and prenatal care location) and newborns' time in the hospital were obtained from the medical record. Descriptive statistics, chi-square tests, and multiple logistic regression models were used to analyze the association between demographic and clinical characteristics and chatbot reach.

Results: Both chatbots were developed and revised based on clinician, community, and patient feedback. Overall, 65.9% (4933/7489) of newborn caregivers discharged between October 2, 2022, and January 15, 2025, opened the newborn chatbot, and 63.6% (4140/6505) of birthing individuals discharged between November 21, 2022, and January 15, 2025, opened the postpartum chatbot. Older age (odds ratio [OR] 1.02, 95% CI 1.01-1.03), Black race (OR 0.73, 95% CI 0.61- 0.88; reference: White), languages other than English or Spanish (OR 1.90, 95% CI 1.21-2.98; reference: English), receipt of prenatal care external to the hospital system (federally qualified health center: OR 0.52, 95% CI 0.45-0.60; Kaiser: OR 0.34, 95% CI 0.29-0.39; reference: within the hospital system), and public insurance (OR 0.72, 95% CI 0.64-0.82; reference: private insurance) were significant predictors of postpartum chatbot reach. Older age (OR 1.02, 95% CI 1.01-1.03), Black race (OR 0.61, 95% CI 0.50-0.74; reference: White), receipt of prenatal care external to the hospital system (federally qualified health center: OR 0.50, 95% CI 0.44-0.57; Kaiser: OR 0.30, 95% CI 0.26-0.35; reference: within the hospital system), public insurance (OR 0.63, 95% CI 0.55-0.71) and self-pay (OR 0.56, 95% CI 0.38-0.83; reference: private insurance), and newborn time in the hospital of 2 - 4 days (OR 1.21, 95% CI 1.09-1.35; reference: less than 2 d) were significant predictors of newborn chatbot reach. Including a Spanish-language version in the newborn chatbot improved reach among Spanish-preferring caregivers (from 58% to 66.2%), but additional chatbot content revision and the addition of chatbot information to discharge paperwork did not change chatbot reach.

Conclusions: While there were differences in chatbot reach by patient demographics, the chatbot showed delivery of time-sensitive information and support to >60% of individuals. This intervention demonstrated that chatbots can be used to supplement patient care and help bridge the gaps in connecting patients to care and support after hospital discharge. Future work should address additional ways to improve chatbot reach and explore the impact on targeted health outcomes.

(JMIR Pediatr Parent 2026;9:e81844) doi:10.2196/81844

KEYWORDS

postpartum period; newborn care; health information; chatbot; implementation

Introduction

The 42-day period after childbirth is widely viewed as high-risk both for the birthing individual and for the newborn [1-3]. Access to timely information about potential warning signs for when to seek emergency care, along with proper follow-up care and support, has the potential to improve maternal and infant health outcomes after discharge from the hospital. Top causes of death for birthing individuals in the 42-day period after delivery are mental health conditions, cardiovascular conditions, and infections [4], while top causes of mortality among infants are birth defects, preterm birth and low birth weight, and sudden infant death syndrome [5]. Previous work has cited the importance of monitoring women post-delivery, with an estimated 15% of severe maternal morbidity cases occurring after discharge [6]. Furthermore, a prior randomized control trial found that enhanced caregiver education via SMS text messaging, timed to the infant's age and most common reasons for emergency department visits, reduced emergency department visits in infants' first year of life [7]. In recent years, SMS text messaging services such as Text4Baby (National Healthy Mothers, Healthy Babies Coalition and Voxiva) [8,9] and apps such as BabyScripts (1EQ, Inc) [10,11] have been developed to provide information to patients during the prenatal and postpartum periods on maternal and infant health.

Chatbots are another mobile health (mHealth) strategy used to deliver timely information. Maternal chatbots are an acceptable and feasible strategy for postpartum and caregiving information and support [12,13]. For example, Rosie is a chatbot that leverages artificial intelligence (AI) to deliver personalized assistance related to pregnancy, labor, postpartum care, and newborn care [13,14]. Another chatbot, Dr. Joy, is an obstetric and mental health-related question-and-answer knowledge-based chatbot that also leverages AI for prenatal and postpartum care [15]. While these chatbots are effective in disseminating relevant health information among select patients, they have been deployed in small trials, as opposed to part of standard of care.

We developed and piloted 2 rule-based chatbots (one for birthing individuals and one for newborn caregivers) to provide timely health information and resources, including connection to care. This study aims to describe the process of developing and piloting an mHealth program aimed at improving education on postpartum and newborn warning signs and appropriate connection to care and resources after hospital discharge. We also aim to describe the chatbot reach (number of patients who opened the chatbot out of those who received it) among patients discharged from a large mid-Atlantic hospital that serves a socioeconomically and racially diverse population overall. We assessed reach after initial launch and after refining the chatbot in response to patient feedback. We hypothesized that with each chatbot refinement, there would be an improvement in reach. Finally, we explored how individuals' demographic and clinical factors were associated with reach for each chatbot.

Methods

Overview

This study was prepared in accordance with the iCHECK-DH (Guidelines and Checklist for the Reporting on Digital Health Implementations) [16].

Chatbot Development, Initial Implementation Strategies, and Setting

The postpartum and newborn chatbots were developed as part of a larger initiative, Safe Babies Safe Moms (SBSM), which was aimed at reducing infant and maternal disparities in the District of Columbia (DC) [17]. These chatbots were designed with key goals of helping patients connect with care teams (eg, information about recommended pediatric and postpartum appointments, list of pediatricians in the area), educating them on warning signs, and providing additional postpartum and newborn information and resources (eg, breastfeeding and wound healing after C-section) [18]. The chatbots operated on a rule-based system with fixed logic for interaction (ie, patients could not ask open-ended questions). Content was developed by a multidisciplinary team of experts in obstetrics, pediatrics, social work, psychiatry, mHealth, and health equity, designed to meet the needs of diverse patients.

We created 2 separate chatbots to account for instances where a birthing individual and newborn did not go home together (eg, adoption, surrogacy, neonatal intensive care unit, or postpartum complication) and provide appropriate information at the appropriate time. Thus, within a nonprofit health care system, caregivers of newborn patients discharged at a large, mid-Atlantic hospital between August 29, 2022, and January 15, 2025, received the newborn chatbot, and birthing individuals discharged between November 21, 2022, and January 15, 2025, received the postpartum chatbot within 24 hours of hospital discharge as standard of care. Individuals discharged after January 15, 2025, did not receive the chatbots due to the funding period ending.

The postpartum chatbot messages were sent in the morning, delivering messages weekly for the first 42 days post-discharge (Figure 1 and Multimedia Appendix 1). The newborn messages started 24 hours after discharge, asked about whether a newborn visit was scheduled, offered support and resources when the caregiver answered no, and followed up with weekly informational outreach messages. A total of 7 messages with unique content were sent for the postpartum chatbot, and 5 messages were sent for the newborn chatbot. Each unique set of chatbot content is referred to as an experience. The topics included in each chatbot experience are listed in Table 1. In short, topics included appointment reminders, warning signs, nutrition recommendations, and developmental milestones for newborns and postpartum care. Both postpartum and newborn messages were sent as standard of care to all individuals who delivered at the hospital, except for one community clinic that opted to only implement the newborn chatbot because they already had a postpartum follow-up process and did not want to confuse patients. There was some content tailoring in the

postpartum chatbot; for instance, birthing individuals who had a C-section received information about wound care.

Table . Newborn and postpartum chatbot reach for caregivers and birthing individuals, and topics covered by each chatbot experience.

Experiences and topics covered in the chatbot[a]	General reach by experience, n/N (%)
Newborn chatbot experience[b]	
Experience 1 (day 1 post discharge)	
Original: pediatric appointment reminder; resources to address common challenges for scheduling an appointment (only offered to patients without an appointment); parental leave	1757/4438 (39.6)
Revised: pediatric appointment reminder; resources to address common challenges for scheduling an appointment (offered to all patients); parental leave	1285/2935 (43.8)
Experience 2 (day 7)	
Original: pediatric warning signs; newborn sleep recommendations	930/4384 (21.2)
Revised: pediatric warning signs; newborn sleep recommendations (added recommendations for premature newborns)	652/2989 (21.8)
Experience 3 (day 14)	
Original: newborn nutrition recommendations; resources for breastfeeding and formula; food assistance program	973/4320 (22.5)
Revised: newborn nutrition recommendations; resources for breastfeeding and formula; colic education; newborn bowel movement education; food assistance program	691/3053 (22.6)
Experience 4 (day 28)	
Original: newborn developmental milestones	1104/4192 (26.3)
Revised: pediatric 1-month visit; tips for communicating with providers; newborn developmental milestones; recommended activities with newborns	886/3181 (27.9)
Experience 5 (day 38)	
Original: recommended pediatric visits and vaccines	1608/4104 (39.2)
Revised: 2-month checkup and recommended vaccines; support groups; resources for diapers and baby essentials	1126/3269 (34.4)
Postpartum chatbot experience[c]	
Experience 1 (day 1 post discharge)	
Original: postbirth warning signs; baby blues	1336/3895 (34.3)
Revised: postbirth warning signs; baby blues and recommendations	907/2610 (34.8)
Experience 2 (day 3)	
Original: postbirth warning signs; postpartum check-up; recovering after a C-section; resources to address common challenges for scheduling a postpartum check-up	1317/3874 (34)
Revised: postbirth warning signs; postpartum recovery; recovering and management after a C-section or vaginal birth; breastfeeding tips and resources; pain management; paid family leave	846/2631 (32.2)
Experience 3 (day 7)	
Original: nutrition	903/3846 (23.5)
Revised: nutrition and resources; stress management tips, postpartum check-up and resources to address common challenges for scheduling or attending a postpartum check-up	544/2659 (20.5)
Experience 4 (day 14)	
Original: postpartum depression; sleep recommendations	691/3794 (18.2)
Revised: postpartum depression; sleep recommendations and tips	390/2711 (14.4)
Experience 5 (day 21)	
Original: family planning and sex after birth	1002/3728 (26.9)

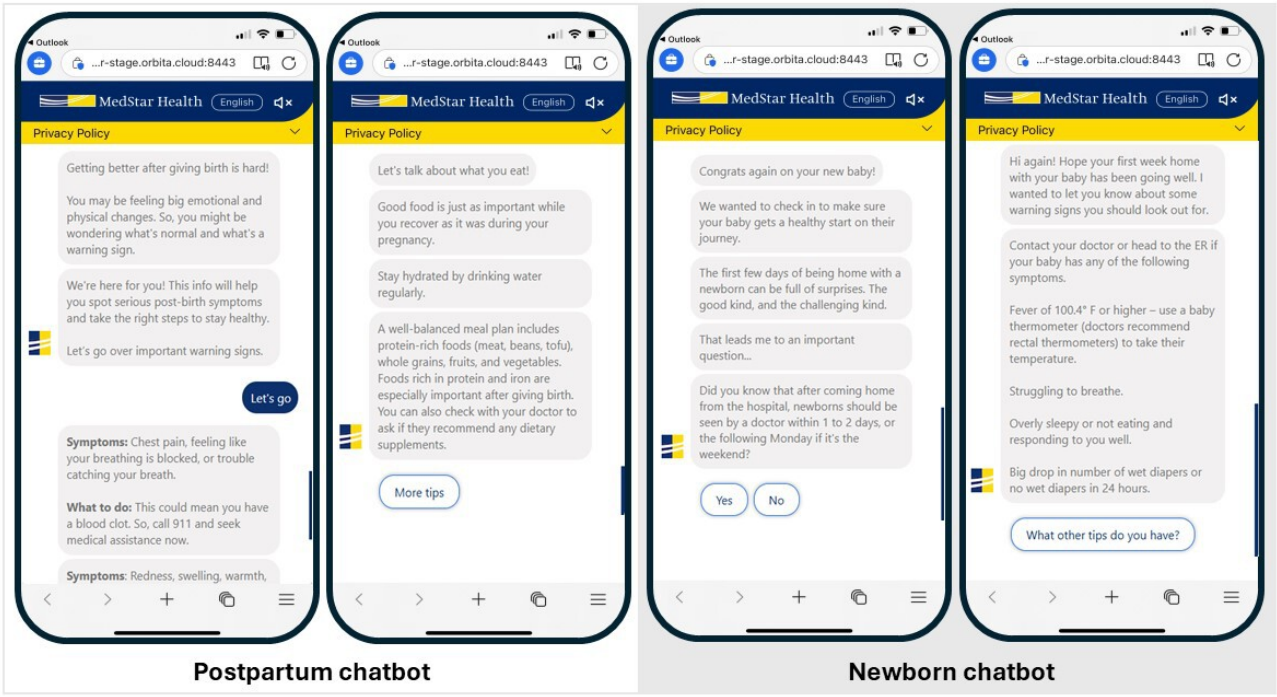
Experiences and topics covered in the chatbot[a]	General reach by experience, n/N (%)
Revised: family planning and sex after birth	706/2777 (25.4)
Experience 6 (day 28)	
Original: pelvic floor - Kegel exercises	656/3660 (17.9)
Revised: physical activity; Pelvic floor - Kegel exercises	511/2845 (18)
Experience 7 (day 42)	
Original: social support; postpartum depression	413/3560 (11.6)
Revised: social support and resources; resources for postpartum depression; health care after birth	250/2945 (8.5)

[a]The topics covered in each experience were revised, and the new content for all experiences was launched on February 21, 2024, for both chatbots. Thus, the total number of patients who received the new content will differ by experience.

[b]Caregivers of newborns discharged from the hospital between October 2, 2022, and January 15, 2025, received the newborn chatbot. A total of 116 caregivers who received the newborn chatbot were excluded due to missing data by experience.

[c]Only birthing individuals discharged from the hospital between November 21, 2022, and January 15, 2025 received the postpartum chatbot.

Figure 1. Chatbot screenshots of the postpartum and newborn chatbots.



Electronic Health Record Data Extraction and Integration

The chatbot system relied on structured electronic health record (EHR) data to identify discharge dates for birthing individuals and newborns. A native EHR data extraction program using Cerner Command Language was designed to retrieve relevant patient information including contact information, preferred language, and prenatal care provider. The data extraction was fully automated and ran daily. These data were securely transmitted to the chatbot vendor system via the enterprise Interface Engine using secured File Transfer Protocol to ensure end-to-end transport layer security, consistent with industry-standard security practices. Upon receiving the contact list, the chatbot vendor delivered the chatbot content through SMS text messaging and email. The vendor generated daily analytics files that included user reach metrics (eg, overall chatbot use) that were securely transferred back to the health care system.

Study Design and Participants

This is a pragmatic implementation study where we described the strategies used to develop the postpartum and newborn chatbots and performed a cross-sectional analysis to evaluate the postpartum and newborn reach. Because patients received the chatbots as standard of care, we included all the birthing individuals and caregivers of newborn patients discharged from a large, mid-Atlantic hospital in the data analyses. Newborn chatbot reach was evaluated for newborn patients discharged from October 2, 2022, to January 15, 2025, who received the newborn chatbot. We were unable to report chatbot delivery and reach during the first month of chatbot launch due to inconsistent documentation by the vendor; these issues were resolved during team meetings. For postpartum reach, all

birthing individuals discharged from November 21, 2022, to January 15, 2025, and who received the postpartum chatbot were included in the analysis.

Measures

Implementation Strategies and Chatbot Reach

The chatbots were designed to fill a gap in follow-up after patient discharge, particularly among those patients who did not seek care within the health care system before or after birth. Implementation strategies specific to increasing reach and engagement for the chatbot focused on developing stakeholder relationships, reexamining implementation, and engaging consumers. The primary outcome of this analysis is chatbot reach and is defined as the number of people who opened the chatbot link out of those who received outreach messaging. Chatbot reach data were obtained from the chatbot vendor, and the research team linked the data back to the patient's medical record.

Demographic and Clinical Data Collection

Birthing individuals' and newborns' demographic and clinical characteristics were obtained from the EHR. These independent variables were categorized as follows: birthing individual age (<20, 20 - 29, 30 - 39, 40+ years), ethnicity (Hispanic or non-Hispanic), race (Black or African American, White, other, or unknown), preferred language (English, Spanish, other [eg, Amharic, Arabic, French, etc], or unknown), and insurance type (private or commercial, public, other, self-pay, or unknown). Regarding medical history, we included delivery method (vaginal birth or C-section) for the birthing individuals, newborns' weight at birth (very low birthweight: <1500g, low birthweight: 1500g to <2500g, or normal birthweight: ≥2500g), gestational age (<37 wk or ≥37 wk), time in the hospital between birth and discharge (<2 d, 2 - 4 d, or >4 d), and prenatal care location (care within the MedStar Health system where they delivered, Kaiser [where both care and insurance are provided within the same system], other external clinics (largely federally qualified health centers [FQHCs] or unknown sources of care). Birthweight and gestational age data were only available for newborns discharged from October 2, 2022, to September 30, 2024, as they were obtained from the EHR for a prior SBSM project [17].

Data Analysis

We used descriptive statistics to describe reach and chi-square tests to evaluate demographic and clinical factors related to the postpartum and newborn chatbot reach. We also assessed changes in reach after modifying chatbot outreach and content. We used backward selection of the independent variables (birthing individual age, ethnicity, race, preferred language, insurance type, and prenatal care location) to finalize the multivariate logistic regression models. For the postpartum chatbot reach, we also included birthing individual delivery method (vaginal vs cesarian), and for the newborn chatbot, we included time in the hospital. In addition, we tested the impact of adding birthweight and gestational age as independent variables to the newborn chatbot logistic regression limiting the population to the date range October 2, 2022, to September 30, 2024, (as these 2 variables were only available between these dates).

Ethical Considerations

The study protocol was approved by the Institutional Review Board (IRB) from MedStar Health Research Institute (IRB #5741). This project was completed in accordance with the ethical standards of the MedStar Health Research Institute IRB and the Helsinki Declaration of 1975 and the 2000 revision. We did not ask for consent or compensate patients in this study as the chatbots were automatically sent to patients as standard of care. Our IRB granted a Health Insurance Portability and Accountability Act waiver to link chatbot engagement to the patients' demographic and clinical data (eg, age, race, preferred language, type of insurance, and prenatal care location). Identifiers were needed to link patients to their relevant EHR information. All patients' identifiers were deleted from all files after data analysis was completed.

Results

Implementation Strategies

Developing Stakeholder Relationships

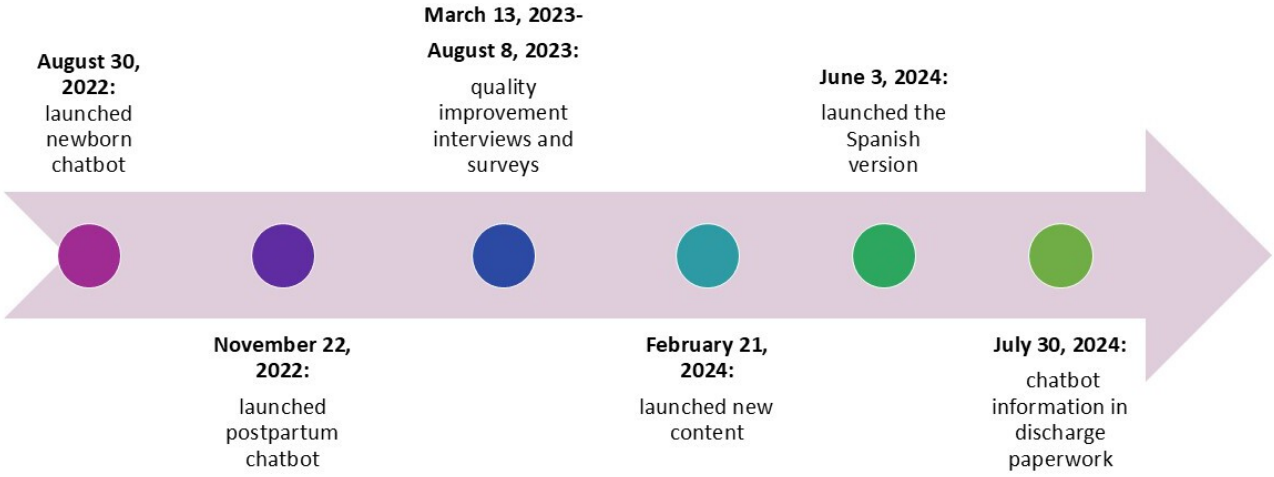
Before developing the chatbot, we met with the clinical leaders in labor and delivery, including physicians, nurses, and administrators. We also consulted with our community partners on key content and outreach strategies, as two-thirds of those delivering at the target hospital seek prenatal and postpartum care outside of the hospital's health care system. We then identified a lead in each of the clinical areas involved (pediatrics, labor and delivery, and behavioral health) to help with the development of the chatbot content.

We iteratively developed the chatbot content, incorporating feedback from birthing individuals and caregivers (n=9) recruited from two pediatric clinics, who reviewed the content and participated in an individual semistructured interview. We notified providers and health care system leaders once the program was ready to launch and invited feedback on integration with existing workflows. Throughout this project, clinical leads were invited to participate in scientific abstracts to increase a sense of buy-in and were given opportunities to reflect on progress. Additionally, we presented ongoing successes and challenges to the parent project's strategic advisory board, which includes clinical and community maternal and infant health equity experts, to solicit feedback. In response to provider feedback about the busy pace and competing priorities of the labor and delivery unit, we automated outreach to eliminate any need to ask providers to enroll patients. Despite the automation, building stakeholder relationships was vital to ensure that clinical teams were comfortable with their patients receiving the outreach, optimize the outreach content, and increase the likelihood that they would support introducing it into discharge paperwork.

Iterative Processes, Including Purposefully Re-Examining the Implementation and Reaching Consumers

The timeline for chatbot launch and iterative refinement is shown in Figure 2. The research team met with the chatbot vendor weekly to review data. With input from the research team, the vendor created a dashboard to monitor reach and certain engagement metrics (eg, user response to the question about scheduling their first newborn pediatric appointment). After assessing baseline rates of reach and preliminary engagement, between March and August of 2023, we collected qualitative and quantitative feedback from diverse users (46% identified as Black, 25% as Hispanic, and 54% had public insurance) to solicit input on increasing reach and use of chatbot content [18].

Figure 2. Postpartum and newborn chatbot implementation and refinement timeline.



Based on initial reach data and patient feedback (reported elsewhere [18]), we created the following changes: (1) we modified the outreach messages to include more information about each week's content, (2) we added relevant topics and resources within each chatbot experience (revised topic content listed in Table 1), (3) we translated the chatbot content into Spanish, (4) we added information about the chatbot to the discharge paperwork so that patients who reviewed the paperwork might have a frame of reference for the outreach.

The outreach messages were changed from a generic script to specify the topics covered in the corresponding experience. Overall, language was revised to improve readability and clarity (eg, "postpartum" was replaced with "after giving birth"). The chatbot content was changed in various ways (Table 1). First, we added additional tailored information for recovery after a C-section or vaginal birth, and options to read more about breastfeeding, with sensitivity to patient experiences around whether they chose to or were able to breastfeed. Second, additional resources such as Maternal Mental Health Hotline, Postpartum Support International, Breastfeeding Center of Greater Washington, and Safe Sleep Program from DC Department of Health were added. Third, we included a voice-over feature to facilitate reviewing the chatbot information using audio, as well as an option for patients to download the content as a PDF so that they could refer back to the information and share it with their partners or family members. We reviewed the revised content with 5 patients at an obstetrics clinic and 5 caregivers at a pediatric clinic to solicit in-person feedback via brief interviews. The interviewer took structured notes on patients' and caregivers' recommendations to share with the study team. Recommendations included: slowing down the speed that chatbot messages added text, shortening the messages, bolding the most important information, and adding a summary of the information included in the shared links. The speed at which new messages appeared in the chatbot was slowed to 2500 milliseconds, and we divided the information into more segments where patients could choose response options. This allowed us to shorten the messages and increase the potential for interaction within the chatbot. The updated outreach messages and chatbot content were launched on February 21, 2024.

A Spanish version of each chatbot was implemented on June 3, 2024, after professional translation, given that Spanish is the second most common primary language among our patient population. We also revised the outreach message to include a Spanish explanation for changing the chatbot language to Spanish as preferred language in the EHR may not capture all appropriate patients. Finally, on July 30, 2024, we added screenshots about the chatbot outreach and content in the discharge paperwork for both birthing individuals and newborns to inform patients about the chatbot and address concerns that the chatbot outreach messages were potentially a scam.

Chatbot Reach Overall

A total of 6505 individuals out of 6684 birthing individuals (97.3%) discharged from the hospital successfully received the postpartum chatbot outreach message; those who did not receive it either had a landline phone number or did not have a working phone number or email in the EHR. Overall, 7489 out of 7525 caregivers (99.5%) successfully received the chatbot outreach message for each newborn. Less than 1% of recipients opted out of the chatbots. Approximately 63.3% of patients had a valid email (postpartum: 4107/6505 and newborn: 4757/7489), and 99.9% (postpartum: 6505/6505 and newborn: 7488/7489) had a valid phone number. A total of 6107 birthing individuals/caregivers received the postpartum and newborn chatbot messages, 398 birthing individuals only received the postpartum chatbot, 1076 unique caregivers only received the

newborn chatbot, and 298 caregivers received additional newborn chatbot messages for each newborn (due to multiparous, or births at different times; eg, 8 individuals had 3 newborns each during the project period).

Approximately two-thirds of recipients opened messages from either chatbot (Table 2). Newborn chatbot reach by experience ranged from 41.3% (3042/7373, experience 1) to 21.5% (1582/7373, experience 2), and the postpartum chatbot reach ranged from 34.5% (2243/6505, experience 1) to 10.2% (663/6505, experience 7; Table 1). Birthing individuals and caregivers who had a valid phone and email had a significantly higher chatbot reach (postpartum chatbot: 2980/4107, 72.6% and newborn chatbot: 3561/4756, 74.9%) than patients with only a valid phone number (postpartum chatbot: 1160/2398, 48.4% and newborn chatbot: 1372/2732, 50.2%). Only one caregiver had a valid email only and was not reached. For both chatbots, approximately 57% of users opened the messages by SMS text messaging only (postpartum: 2388/4140 and newborn: 2831/4933), 14% by email only (postpartum: 601/4140 and newborn: 678/4933), and 28% by both SMS text messaging and email (postpartum: 1151/4140 and newborn: 1424/4933).

Table . Patients' demographic and clinical characteristics by postpartum and newborn chatbot reach.

Patients' demographic	Postpartum chatbot[a]				Newborn chatbot[b]			
	Total received	Total reach	Reach[c], %	*P* value[d]	Total received	Total reach	Reach[c], %	*P* value[d]
Total	6505	4140	63.6		7489	4933	65.9	
Age (years)				<.001				<.001
<20	285	153	53.7		332	189	56.9	
20 - 29	2310	1349	58.4		2746	1669	60.8	
30 - 39	3466	2337	67.4		3945	2720	69.0	
40+	444	301	67.8		466	355	76.2	
Race				<.001				<.001
Black	2926	1749	59.8		3527	2217	62.9	
White	940	713	75.9		1001	823	82.2	
Other	1294	860	66.5		1468	953	64.9	
Unknown	1345	818	60.8		1493	940	63.0	
Ethnicity				.01				<.001
Hispanic	470	318	67.7		517	355	68.7	
Non-Hispanic	3937	2539	64.5		4589	3108	67.7	
Other	101	67	66.3		113	69	61.1	
Unknown	1997	1216	60.9		2270	1401	61.7	
Language				.003				<.001
English	5633	3589	63.7		6498	4321	66.5	
Spanish	699	436	62.4		793	480	60.5	
Other	113	86	76.1		123	93	75.6	
Unknown	60	29	48.3		75	39	52.0	
Insurance				<.001				<.001
Private	2670	1866	69.9		2955	2197	74.4	
Public	2941	1749	59.5		3498	2103	60.1	
Self-pay	106	67	63.2		120	72	60.0	
Other	79	59	74.7		68	52	76.5	
Unknown	709	399	56.3		848	509	60.0	
Delivery method				.13				N/A[e]
Vaginal	3987	2509	62.9		N/A	N/A	N/A	
Cesarean	2518	1631	64.8		N/A	N/A	N/A	
Baby weight[fg]				N/A				.07
Very low birthweight (<1500 g)	N/A	N/A	N/A		64	47	73.4	
Low birthweight (1500 to <2500 g)	N/A	N/A	N/A		633	396	62.6	
Normal birthweight (≥2500 g)	N/A	N/A	N/A		5863	3890	66.4	
Gestational age[fg]				N/A				.003
<37 weeks	N/A	N/A	N/A		738	452	61.3	

Patients' demographic	Postpartum chatbot[a]				Newborn chatbot[b]			
	Total received	Total reach	Reach[c], %	*P* value[d]	Total received	Total reach	Reach[c], %	*P* value[d]
≥37 weeks	N/A	N/A	N/A		5822	3881	66.7	
Time in hospital after birth [g]				N/A				<.001
<2 days	N/A	N/A	N/A		3070	1972	64.2	
2 - 4 days	N/A	N/A	N/A		3304	2257	68.3	
>4 days	N/A	N/A	N/A		1115	704	63.1	
Prenatal care location				<.001				<.001
Within the hospital integrated health system	2895	2148	74.2		3042	2380	78.2	
Kaiser clinics[h]	1234	630	51.1		1342	721	53.7	
Other external clinics (eg, FQHCs)[i]	1864	1072	57.5		2609	1546	59.3	
Unknown clinics	512	290	56.6		496	286	57.7	

[a]Among birthing individuals discharged from the hospital between November 21, 2022, and January 15, 2025, who received the postpartum chatbot.

[b]Among caregivers of newborns discharged from the hospital between October 2, 2022, and January 15, 2025, who received the newborn chatbot.

[c]Reach is defined as the number of people who opened the chatbot link out of those who received outreach messaging.

[d]Chi-square statistics.

[e]N/A: not applicable.

[f]Data are only available through September 30, 2024.

[g]Data specific to newborn patients obtained from the electronic health record.

[h]Kaiser clinics are external clinics where both care and insurance are provided within the same system.

[i]FQHC: federally qualified health center.

Significant differences in reach were identified by age, ethnicity, race, preferred language, and insurance type across both chatbots (Table 2; all significant *P* values <.001). Birthing individuals who opened messages were more likely to be 30 years and older (2638/3910, 67.5%) compared to 29 years and younger (1502/2595, 57.9%), Hispanic (318/470, 67.7%) compared to non-Hispanic (2539/3937, 64.5%); White (713/940, 75.9%) compared to Black (1749/2926, 59.8%), had private insurance (1866/2670, 69.9%) compared to public insurance (1749/2941, 59.5%), and had prenatal care within the hospital's integrated health system (2148/2895, 74.2%) compared to external prenatal clinics (Kaiser patients 630/1234, 51.1%), other largely FQHC clinics (1072/1864, 57.5%), unknown prenatal care location (290/512, 56.6%). No significant differences in reach were found by the delivery method (ie, vaginal vs C-section).

Similarly, newborn chatbot caregiver reach was higher among birthing individuals who were 30 years and older (3075/4411, 69.7%) compared to 29 years and younger (1858/3078, 60.4%), White (823/1001, 82.2%) compared to Black (2217/3527, 62.9%), and had private insurance (2197/2955, 74.4%) compared to public insurance (2103/3498, 60.1%). Newborn chatbot reach was also higher when the newborn had a gestational age of 37 weeks or more (3881/5822, 66.7%) compared to a gestational age of less than 37 weeks (452/738, 61.3%), stayed in the hospital 2 - 4 days (2257/3304, 68.3%) compared to less than 2 days (1972/3070, 64.2%) and more than 4 days (704/1115, 63.1%), and had prenatal care within the hospital integrated health system (2380/3042, 78.2%) compared to external prenatal clinics Kaiser (721/1342, 53.7%), other largely FQHC clinics (1546/2609, 59.3%), and unknown prenatal care location (286/496, 57.7%). No significant differences were found for birth weight.

Postpartum Chatbot Analyses

In the final postpartum multivariate logistic regression model, we included age, race, preferred language, type of insurance, and prenatal location (Table 3). The odds of opening the postpartum chatbot were significantly lower for individuals identified in the EHR as Black (odds ratio [OR] 0.73, 95% CI 0.61- 0.88) compared to White individuals. Patients with a preferred language of "other" (OR 1.90, 95% CI 1.21-2.98) had greater odds of postpartum chatbot reach compared to English-preferring patients. Patients with public insurance (OR 0.72, 95% CI 0.64-0.82) and unknown insurance (OR 0.57, 95% CI 0.47-0.69) had lower odds of postpartum chatbot reach compared to individuals with private insurance. Patients who received prenatal care at clinics external to the hospital, including Kaiser clinics (OR 0.34, 95% CI 0.29-0.39), other

clinics (OR 0.52, 95% CI 0.45-0.60), or unknown clinics (OR 0.45, 95% CI 0.37-0.55), had lower odds of postpartum chatbot reach compared to patients who received prenatal care within the hospital integrated health system. Age was also a significant predictor, with each one-year increase in age associated with a 2% higher likelihood of chatbot reach (OR 1.02, 95% CI 1.01-1.03).

Table . Multivariate logistic regression for postpartum chatbot reach among birthing individuals discharged from the hospital between November 21, 2022, and January 15, 2025. Reach is defined as the number of people who opened the chatbot link out of those who received outreach messaging.

Characteristics	OR[a] (95% CI)	*P* value
Age (continuous)	1.02 (1.01-1.03)	<.001
Race		
White (reference)	N/A[b]	N/A
Black	0.73 (0.61-0.88)	<.001
Other	1.10 (0.88-1.37)	.40
Unknown	0.92 (0.75-1.13)	.40
Language		
English (reference)	N/A	N/A
Spanish	1.04 (0.86-1.27)	.66
Other	1.90 (1.21-2.98)	.005
Unknown	0.69 (0.41-1.16)	.16
Insurance		
Private (reference)	N/A	N/A
Public	0.72 (0.64-0.82)	<.001
Self-pay	0.75 (0.49-1.14)	.18
Other	1.07 (0.63-1.82)	.80
Unknown	0.57 (0.47-0.69)	<.001
Prenatal location		
Within the hospital system where delivery occurred (reference)	N/A	N/A
Kaiser clinics[c]	0.34 (0.29-0.39)	<.001
Other external clinics (eg, FQHCs)[d]	0.52 (0.45-0.60)	<.001
Unknown clinics	0.45 (0.37-0.55)	<.001

[a]OR: odds ratio.

[b]N/A: not applicable.

[c]Kaiser clinics are external clinics where both care and insurance are provided within the same system.

[d]FQHC: federally qualified health center.

Newborn Chatbot Analyses

For the final newborn multivariate logistic regression model, we included the birthing individual's age, race, insurance type, prenatal care location, and their newborn's time in the hospital (Table 4). Age was a significant predictor, with each one-year increase in age associated with a 2% higher likelihood of chatbot reach (OR 1.02, 95% CI 1.01-1.03). Birthing individuals who were listed in the EHR as Black (OR 0.61, 95% CI 0.50-0.74), other race (OR 0.73, 95% CI 0.59-0.91), and unknown race (OR 0.74, 95% CI 0.60-0.91) had lower odds of newborn chatbot reach than when the newborn's birthing individual was White. Patients with public insurance (OR 0.63, 95% CI 0.55-0.71), self-pay (OR 0.56, 95% CI 0.38-0.83), and unknown insurance (OR 0.58, 95% CI 0.48-0.69) had lower odds of reach than patients with private insurance. Patients who received prenatal care at clinics external to the hospital, including Kaiser clinics (OR 0.30, 95% CI 0.26-0.35), other clinics (OR 0.50, 95% CI 0.44-0.57), unknown clinics (OR 0.40, 95% CI 0.32-0.48), had lower odds of postpartum chatbot reach compared to patients who received prenatal care within the hospital health system. Finally, caregivers of newborns who stayed in the hospital for 2 - 4 days after birth (OR 1.21, 95% CI 1.09-1.35) had greater odds for chatbot reach compared to newborns who stayed in the hospital less than 2 days, but no significant differences were found for the newborns who stayed more than 4 days (OR 0.99, 95% CI 0.86-1.15).

Table . Multivariate logistic regression for newborn chatbot reach among caregivers of newborns discharged from the hospital between October 2, 2022, and January 15, 2025. Reach is defined as the number of people who opened the chatbot link out of those who received outreach messaging.

Characteristics	OR[a] (95% CI)	*P* values
Age (continuous)	1.02 (1.01-1.03)	<.001
Race		
White (reference)	N/A[b]	N/A
Black	0.61 (0.50-0.74)	<.001
Other	0.73 (0.59-0.91)	.004
Unknown	0.74 (0.60-0.91)	.005
Insurance		
Private (reference)	N/A	N/A
Public	0.63 (0.55-0.71)	<.001
Self-pay	0.56 (0.38-0.83)	.004
Other	0.85 (0.48-1.53)	.60
Unknown	0.58 (0.48-0.69)	<.001
Time in hospital		
<2 days (reference)	N/A	N/A
2 - 4 days	1.21 (1.09-1.35)	.001
>4 days	0.99 (0.86-1.15)	.92
Prenatal location		
Within the hospital integrated health system (reference)	N/A	N/A
Kaiser clinics[c]	0.30 (0.26-0.35)	<.001
Other external clinics (eg, FQHCs)[d]	0.50 (0.44-0.57)	<.001
Unknown clinics	0.40 (0.32-0.48)	<.001

[a]OR: odds ratio.

[b]N/A: not applicable.

[c]Kaiser clinics are external clinics where both care and insurance are provided within the same system.

[d]FQHC: federally qualified health center.

We conducted additional analyses limited to those where we had information on birthweight and gestational age (newborns discharged between October 2, 2022, and September 30, 2024). Birthing individual age, race, insurance, newborn time in the hospital, and prenatal care location remained in the model when including and excluding newborn birthweight and gestational age; ethnicity and preferred language were eliminated for both final models to fit the best model. Caregivers of newborns with very low birthweight (OR 2.44, 95% CI 1.32-4.50) had greater odds of opening the chatbot compared to newborns with normal birthweight, although no significant differences were found for the newborns with low birthweight (OR 1.11, 95% CI 0.88-1.39). In relation to gestational age, for newborns born at <37 weeks (OR 0.77, 95% CI 0.61-0.97), the caregivers were less likely to open the chatbot than caregivers with newborns born at term (37+ weeks). The ORs for other variables in the model were similar to those in the model with the full cohort that did not include birthweight and gestational age (Multimedia Appendix 2).

Chatbots Reach Changes Over Time

No significant changes in reach were found after launching the updated version of the newborn and postpartum outreach and chatbot content, nor after including the chatbot information in patients' discharge paperwork (Table 5); thus, we focused on results in the combined analyses for the full time period. We found significant improvements in reach for Spanish-speaking patients after deploying the Spanish version of the newborn chatbot (OR 1.42, 95% CI 1.03-1.96), but no significant differences in reach were found for the postpartum chatbot (OR 1.09, 95% CI 0.78-1.53). For a graphic description of the evolution of postpartum and newborn chatbots' reach over time (Multimedia Appendices 3 and 4).

Table . Change in chatbot reach before and after changes in outreach, content, and awareness efforts.

Implementation strategies	Postpartum chatbot				Newborn chatbot			
	Total received	Total reach[a]	Reach, %	OR[b] (95% CI)	Total received	Total reach[a]	Reach, %	OR (95% CI)
Outreach and content revisions[c]								
Original content (reference)	3554	2278	64.1	N/A[d]	4212	2775	65.9	N/A
New content	2610	1639	62.8	0.95 (0.85-1.05)	2935	1914	65.2	0.97 (0.88-1.07)
Content in Spanish[e]								
Pre-Spanish language (reference)	448	277	61.8	N/A	519	301	58	N/A
Spanish language	213	136	63.9	1.09 (0.78-1.53)	237	157	66.2	1.42 (1.03-1.96)
Discharge paperwork[f]								
Prior (reference)	5150	3274	63.6	N/A	5967	3950	66.2	N/A
Included in discharge paperwork	1355	866	63.9	1.02 (0.90-1.15)	1522	983	64.6	0.93 (0.83-1.05)

[a]Reach is defined as the number of people who opened the chatbot link out of those who received outreach messaging.

[b]OR: odds ratio.

[c]Individuals discharged before January 9, 2024, for the postpartum chatbot and January 13, 2024, for the newborn chatbot were included in the analysis as the original content group, and individuals discharged between February 20, 2024, and January 15, 2025, for both chatbots were included in the new content group.

[d]N/A: not applicable.

[e]Only Spanish-speaking individuals identified in the medical record were included in the analysis. Individuals discharged before April 21, 2024, for the postpartum chatbot and April 25, 2024, for the newborn chatbot were included in the analysis as the pre-Spanish language group, and patients discharged from June 2, 2024, through January 15, 2025, were included in the Spanish-language group.

[f]Individuals discharged before July 29, 2024, were included in the analysis as before discharge paperwork, and individuals discharged from July 29, 2024, through January 15, 2025, were included in the discharge paperwork group.

Lessons Learned

This project highlighted the importance of involving stakeholders, including clinical staff, providers, patients, and community before and during the chatbot implementation to ensure acceptability, usefulness, and reach of the chatbot. The clinical teams were supportive of the rollout throughout the project. However, there were significant challenges in data collection and reporting from the third-party vendor, which may be better addressed in future work by working through specific needs, data accuracy, and various use cases for data before deploying the chatbot. Also, relying on a third-party vendor with ongoing expenses for hosting the chatbot limited the sustainability of this project after the project funding ended. Other health care systems or groups may consider options that are integrated into existing licenses so that ongoing fees are not prohibitive. Finally, given the design of the project and the vendor's limited capability to track individual level engagement with the program over time, we were unable to assess the impact of the chatbot on health outcomes.

Discussion

Principal Findings

This study describes strategies to develop a postpartum and newborn chatbot that was offered as standard of care for birthing individuals and caregivers of newborns discharged at a large hospital that serves a socioeconomically and racially diverse population. The chatbots' reach was close to two-thirds of the eligible birthing individuals and caregivers. Differences in reach were identified by age, race, prenatal care location, and insurance status for both chatbots, while newborn weight at birth, gestational age, and newborn time in the hospital were also significant predictors of the newborn chatbot reach. While disparities were present, this study demonstrated that chatbots have the potential to reach a significant proportion of this at-risk

population and elucidates opportunities to conduct additional targeted supports.

Previous commercial and research programs have shown satisfaction among those who choose to receive messages about maternal and infant care. In other large-scale programs such as Text4Baby, a national SMS text messaging strategy in the US, about 150,000 individuals have chosen to enroll annually to receive SMS text messaging about prenatal, postpartum, and infant health recommendations, with users reporting high satisfaction [8]. Other interventions have also shown that patients and caregivers report high satisfaction when receiving SMS text messaging with health recommendations about postpartum depression [19] and infant care [7]. Health chatbots are also positively viewed [20,21], and parents tend to appreciate the informational and emotional support provided by chatbots [21]. In addition to high acceptability, mHealth strategies are also effective in improving health behaviors. For example, mHealth strategies have been shown to be effective in infant emergency room visit reduction [7], blood pressure monitoring [22], prenatal and postpartum weight management [23], and smoking cessation [24], among others [25].

In this project, about 65% of the individuals opened the chatbots, suggesting interest in reviewing information about postpartum and newborn care. In a randomized control trial, patients who received Rosie, an AI chatbot for new mothers, found that 87% (13/15) of the users reported using the chatbot at least weekly [13]. Another study of a postpartum AI chatbot found that of the 290 enrolled patients in the study, 99% of the patients responded to the platform at least once, and 52% asked a question to the chatbot [26]. However, to our knowledge, there are no comparable standard of care for newborn and postpartum mHealth educational outreach strategies that have evaluated reach in real-world settings, underscoring the need for this work. Prior real-world behavioral interventions on various health topics have documented SMS text messaging reach ranging from 14% to 60% [27-29]; our findings thus surpassed expectations. For example, a study among smokers and at-risk drinkers found that a total of 14% of the patients clicked on the embedded link to the apps in the SMS text messaging [27]. Another project using EHR data to identify smokers found that 15% responded to the smoking cessation SMS text messaging program after a single recruitment text [28]. Another project for patients aged 65 years and older found that 60% of the patients responded to the COVID-19 vaccine SMS text messaging outreach strategy [29].

We found notable differences in chatbot reach by age, race, insurance type, and prenatal care location for both chatbots. Lower chatbot reach among younger patients, Black patients, patients with public insurance, and those who received prenatal care at clinics external to the hospital health care system may be attributed to differences in desire for information, disparities in health/technological literacy, concerns about data plans, or trust in the health care system or chatbots [29,30]. We were also unable to deliver the chatbot to patients without a valid email or cell phone number in the EHR. To improve equity, it is crucial to develop interventions that address these issues, whether through additional in-clinic outreach to inform patients and caregivers about the chatbot, ensuring a valid email and phone number is documented in the EHR before discharge, or improving patient access to health information via different channels (eg, phone calls, SMS text messaging, and community outreach). Costs to the health care system and potential models for covering such low-cost services should also be explored. Future research could also consider intentional partnership with community groups focused on Black birthing individuals and underserved populations to avoid disparities in chatbot implementation.

While significant changes to the chatbot content and strategies were made to improve the chatbot's reach, these strategies did not result in significant improvements. Similar findings were found in a prior study where no significant improvements in acceptability were found after intervention refinement based on participants' feedback [30]. Our study suggests that further research is needed to understand the drivers of chatbot reach post-hospital discharge, such as user experience and how outreach and content can be optimized to capture and retain user interest over time. Additional strategies, such as identifying the best outreach time of day, further tailoring the outreach messages/content, and potentially including AI for personalization [13] may play a role in how users access and engage with similar chatbots. Community-based approaches may also be useful for addressing concerns about chatbots and identifying ways to increase the attractiveness of chatbots to patients who experience disparities. In addition, these changes could potentially impact other areas of user experience that we did not measure, like usability, satisfaction, and engagement. Ongoing audit and feedback may also prevent gaps in study metrics and provide timely information on how patients are using these types of mHealth interventions.

Limitations and Strengths

In this study, patients did not have the opportunity to request that the chatbots be sent to their preferred contact information, which could have impacted reach. Furthermore, because our chatbot followed a rule-based system with fixed logic for user interaction, it did not allow patients to ask questions or review the information in their preferred order. While efforts were made to match the content to when patients might need this information, for some patients, the information presented in the chatbots might not have been relevant. Future integration of conversation AI is recommended to provide personalized responses; however, it needs to be done with caution to guarantee that patients are receiving accurate information.

Two-thirds of the birthing individuals and most newborns seek postpartum or newborn care, respectively, outside the system. Thus, we could not evaluate the impact on patient health outcomes. Even within our system, there was limited ability to link engagement with specific outcomes due to the variability in the ways that individuals could engage with the chatbot and view tailored information. Future longitudinal studies within an integrated health care system, through partnership across facilities or designed in a manner where relevant outcome data can be linked to specific aspects of chatbot engagement, are also needed to assess the impact of chatbot reach and engagement on maternal and infant health outcomes, such as postpartum depression, follow-up care, and visits to the

emergency room. Further research is also needed to understand patterns of reach related to gestational age, birthweight, or duration of hospital stay.

Despite these limitations, this study has many strengths. Delivering this program at scale provides unique information on who might engage with this type of digital outreach. This chatbot was intended to reach all patients and especially those who might not have a regular care provider to help them establish appropriate care. Finally, to our knowledge, our study presents the first example about chatbot reach to birthing individuals and newborn caregivers as standard of care within an urban hospital.

Conclusions

Our intervention demonstrated that chatbots can deliver important health information to many individuals efficiently, offering timely guidance on postpartum recovery, infant care, and follow-up appointments. These tools may serve as a supplement to patient care, helping bridge gaps in communication and support between health care providers and patients, especially for those who do not seek regular care. Understanding how constant interaction with digital health tools influences clinical outcomes, including a focus on racially and socioeconomically diverse populations, will help refine usability and effectiveness of these technologies.

Acknowledgments

We want to acknowledge the support of the DC Safe Babies Safe Moms Team, Doriun Brown, Christine Laccay, Dr Amalia Londono Tobon for reviewing the Spanish version of the chatbots, and Orbita, the company that ran the chatbots.

Funding

This project was funded by the A James and Alice B Clark Foundation. The funder had no involvement in the study design, data collection, analysis, interpretation, or the writing of the article.

Data Availability

The datasets generated or analyzed during this study are not publicly available because the data analyzed in this study could be used to identify patients. However, deidentified data obtained for this study will be made available (as allowable according to institutional review board standards) by emailing the corresponding author.

Authors' Contributions

JNRR wrote the first draft with input from HA. JNRR, MS, SM, and HA engaged in data collection, data curation, and analysis. JNRR and HA were responsible for study methodology, project administration, and supervision of the study team. KEA contributed to the development of the chatbot, and ADT contributed to the project conceptualization. All authors reviewed and edited the final article. There was no generative AI used in article writing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Postpartum and newborn chatbot outreach message schedules.
[DOCX File, 94 KB - pediatrics_v9i1e81844_app1.docx]

Multimedia Appendix 2
Multivariate logistic regression for newborn chatbot reach including birthweight and gestational age variables for patients discharged from October 2, 2022, to September 30, 2024.
[DOCX File, 17 KB - pediatrics_v9i1e81844_app2.docx]

Multimedia Appendix 3
Postpartum chatbot reach by week of initial outreach.
[DOCX File, 69 KB - pediatrics_v9i1e81844_app3.docx]

Multimedia Appendix 4
Caregivers' newborn chatbot reach by week of initial outreach.
[DOCX File, 74 KB - pediatrics_v9i1e81844_app4.docx]

References

1. Tully KP, Stuebe AM, Verbiest SB. The fourth trimester: a critical transition period with unmet maternal health needs. Am J Obstet Gynecol 2017 Jul;217(1):37-41. [doi: 10.1016/j.ajog.2017.03.032] [Medline: 28390671]
2. WHO recommendations on maternal and newborn care for a positive postnatal experience. : World Health Organization; 2022 URL: https://iris.who.int/server/api/core/bitstreams/73dec697-c033-449c-8323-1cd04a8d8f20/content [accessed 2025-12-23]
3. Dol J, Hughes B, Bonet M, et al. Timing of maternal mortality and severe morbidity during the postpartum period: a systematic review. JBI Evid Synth 2022 Sep 1;20(9):2119-2194. [doi: 10.11124/JBIES-20-00578] [Medline: 35916004]
4. Pregnancy-related deaths: data from maternal mortality review committees. CDC. URL: https://www.cdc.gov/maternal-mortality/php/data-research/mmrc/?CDC_AAref_Val=https%3A%2F%2Fwww.cdc.gov%2Fmaternal-mortality%2Fphp%2Fdata-research%2Findex.html&cove-tab=3 [accessed 2025-12-23]
5. Infant mortality. CDC. URL: https://www.cdc.gov/maternal-infant-health/infant-mortality/index.html [accessed 2025-12-11]
6. Chen J, Cox S, Kuklina EV, Ferre C, Barfield W, Li R. Assessment of incidence and factors associated with severe maternal morbidity after delivery discharge among women in the US. JAMA Netw Open 2021 Feb 1;4(2):e2036148. [doi: 10.1001/jamanetworkopen.2020.36148] [Medline: 33528553]
7. Ladley A, Hieger AW, Arthur J, Broom M. Educational text messages decreased emergency department utilization among infant caregivers: a randomized trial. Acad Pediatr 2018 Aug;18(6):636-641. [doi: 10.1016/j.acap.2018.02.003] [Medline: 29432907]
8. Whittaker R, Matoff-Stepp S, Meehan J, et al. Text4baby: development and implementation of a national text messaging health information service. Am J Public Health 2012 Dec;102(12):2207-2213. [doi: 10.2105/AJPH.2012.300736] [Medline: 23078509]
9. Text4baby. Alabama Public Health. 2023. URL: https://www.alabamapublichealth.gov/perinatal/text4baby.html [accessed 2025-12-30]
10. Marko KI, Krapf JM, Meltzer AC, et al. Testing the feasibility of remote patient monitoring in prenatal care using a mobile app and connected devices: a prospective observational trial. JMIR Res Protoc 2016 Nov 18;5(4):e200. [doi: 10.2196/resprot.6167] [Medline: 27864167]
11. About us. BabyScripts. 2025. URL: https://babyscripts.com/about-us [accessed 2025-12-11]
12. Suharwardy S, Ramachandran M, Leonard SA, et al. Feasibility and impact of a mental health chatbot on postpartum mental health: a randomized controlled trial. AJOG Glob Rep 2023 Aug;3(3):100165. [doi: 10.1016/j.xagr.2023.100165] [Medline: 37560011]
13. Nguyen QC, Aparicio EM, Jasczynski M, et al. Rosie, a health education question-and-answer chatbot for new mothers: randomized pilot study. JMIR Form Res 2024 Jan 12;8:e51361. [doi: 10.2196/51361] [Medline: 38214963]
14. Mane HY, Channell Doig A, Marin Gutierrez FX, et al. Practical guidance for the development of rosie, a health education question-and-answer chatbot for new mothers. J Public Health Manag Pract 2023;29(5):663-670. [doi: 10.1097/PHH.0000000000001781] [Medline: 37478093]
15. Chung K, Cho HY, Park JY. A chatbot for perinatal women's and partners' obstetric and mental health care: development and usability evaluation study. JMIR Med Inform 2021 Mar 3;9(3):e18607. [doi: 10.2196/18607] [Medline: 33656442]
16. Perrin Franck C, Babington-Ashaye A, Dietrich D, et al. iCHECK-DH: guidelines and checklist for the reporting on digital health implementations. J Med Internet Res 2023 May 10;25:e46694. [doi: 10.2196/46694] [Medline: 37163336]
17. Thomas AD, Biel M, Rethy J, et al. D.C. safe babies safe moms: a novel, multigenerational model to reduce maternal and infant health disparities. NEJM Catalyst 2025 Jan 15;6(2). [doi: 10.1056/CAT.24.0161]
18. Rivera Rivera JN, AuBuchon KE, Smith M, et al. Development and refinement of a chatbot for birthing individuals and newborn caregivers: mixed methods study. JMIR Pediatr Parent 2024 Nov 14;7:e56807. [doi: 10.2196/56807] [Medline: 39541147]
19. Lawson A, Dalfen A, Murphy KE, Milligan N, Lancee W. Use of text messaging for postpartum depression screening and information provision. Psychiatr Serv 2019 May 1;70(5):389-395. [doi: 10.1176/appi.ps.201800269] [Medline: 30717643]
20. Abd-Alrazaq AA, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M. Perceptions and opinions of patients about mental health chatbots: scoping review. J Med Internet Res 2021 Jan 13;23(1):e17828. [doi: 10.2196/17828] [Medline: 33439133]
21. Chua JYX, Choolani M, Chee CYI, et al. Insights of parents and parents‐to‐be in using chatbots to improve their preconception, pregnancy, and postpartum health: a mixed studies review. J Midwifery Womens Health 2023;68(4):480-489. [doi: 10.1111/jmwh.13472] [Medline: 36734375]
22. Arkerson BJ, Finneran MM, Harris SR, et al. Remote monitoring compared with in-office surveillance of blood pressure in patients with pregnancy-related hypertension. Obstet Gynecol 2023 Oct 1;142(4):855-861. [doi: 10.1097/AOG.0000000000005327] [Medline: 37734091]
23. Litman EA, Kavathekar T, Amdur R, Sebastian A, Marko K. Remote gestational weight gain monitoring in a large low-risk US population. Obes Sci Pract 2022 Apr;8(2):147-152. [doi: 10.1002/osp4.554] [Medline: 35388344]
24. Abroms LC, Boal AL, Simmens SJ, Mendel JA, Windsor RA. A randomized trial of Text2Quit: a text messaging program for smoking cessation. Am J Prev Med 2014 Sep;47(3):242-250. [doi: 10.1016/j.amepre.2014.04.010] [Medline: 24913220]

25. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence-based chatbots for promoting health behavioral changes: systematic review. J Med Internet Res 2023 Feb 24;25:e40789. [doi: 10.2196/40789] [Medline: 36826990]
26. Leitner K, Cutri-French C, Mandel A, et al. A conversational agent using natural language processing for postpartum care for new mothers: development and engagement analysis. JMIR AI 2025 Apr 22;4:e58454. [doi: 10.2196/58454] [Medline: 40605798]
27. Khadjesari Z, Brown TJ, Ramsey AT, et al. Novel implementation strategy to electronically screen and signpost patients to health behavior apps: mixed methods implementation study (optimine study). JMIR Form Res 2022 Jul 11;6(7):e34271. [doi: 10.2196/34271] [Medline: 35816374]
28. Kruse GR, Kelley JHK, Chase K, Rigotti NA. Feasibility of a proactive text messaging intervention for smokers in community health centers: pilot study. JMIR Form Res 2018;2(1):e11. [doi: 10.2196/formative.9608] [Medline: 30506038]
29. Ahmed N, Boxley C, Dixit R, et al. Evaluation of a text message-based COVID-19 vaccine outreach program among older patients: cross-sectional study. JMIR Form Res 2022 Jul 18;6(7):e33260. [doi: 10.2196/33260] [Medline: 35724339]
30. Keyworth C, Quinlivan L, Leather JZ, O'Connor RC, Armitage CJ. Does refining an intervention based on participant feedback increase acceptability? An experimental approach. BMC Public Health 2023 Aug 22;23(1):1598. [doi: 10.1186/s12889-023-16344-w] [Medline: 37608317]

Abbreviations

EHR: electronic medical record
FQHC: federally qualified health center
iCHECK-DH: Guidelines and Checklist for the Reporting on Digital Health Implementations
IRB: Institutional Review Board
mHealth: mobile health

Edited by S Badawy; submitted 04.Aug.2025; peer-reviewed by J Xie, R Garcia-Fernandez; accepted 27.Nov.2025; published 09.Jan.2026.

Please cite as:
Rivera Rivera JN, Smith M, Mathur S, AuBuchon KE, Thomas AD, Arem H
Increasing Use of a Postpartum and Newborn Chatbot among Birthing Individuals and Caregivers: Development and Implementation Study
JMIR Pediatr Parent 2026;9:e81844
URL: https://pediatrics.jmir.org/2026/1/e81844
doi:10.2196/81844

Original Paper

Facilitating Communication With Children and Young Adults With Special Health Care Needs Through a Web-Based Application: Qualitative Descriptive Study

Jessica R Hanks[1], MD; Ashley M Hughes[2], MS, PhD; Safura Sultana[3], MBA; Ryan Klute[4], MS; Kyle Formella[4], BSED, MS; Connor Flynn[5], BS; Allison Wallenfang[5], BS; Divya Krishnakumar[6], BS; Masah Mourad[7], BS; Yoonje Cho Morse[5], BS; Matthew J Mischler[8,9], MD

[1]Department of Health Sciences Education and Pathology, College of Medicine Peoria, University of Illinois Chicago, Peoria, IL, United States
[2]Department of Medicine at Metrohealth, School of Medicine, Case Western Reserve University, Cleveland, OH, United States
[3]Healthcare Analytics, OSF, Peoria, IL, United States
[4]Innovation, OSF, Peoria, IL, United States
[5]College of Medicine Peoria, University of Illinois Chicago, Peoria, IL, United States
[6]College of Medicine, University of Illinois Chicago, Chicago, IL, United States
[7]School of Public Health, University of Illinois Chicago, Chicago, IL, United States
[8]Department of Medicine, College of Medicine Peoria, University of Illinois Chicago, Peoria, IL, United States
[9]OSF Saint Francis Medical Center, Peoria, IL, United States

Corresponding Author:
Jessica R Hanks, MD
Department of Health Sciences Education and Pathology
College of Medicine Peoria
University of Illinois Chicago
1 Illini Drive
Peoria, IL, 61603
United States
Phone: 1 309 671 8407
Fax: 1 3096808480
Email: jhanks1@uic.edu

Abstract

Background: Children and young adults with special health care needs comprise a significant portion of the pediatric population in the United States, where 1 in every 5 children has a complex health care need. These patients are more likely to receive unsafe care and have their needs unmet in part due to lack of accessible information and limited training support. Barriers in communication may contribute to detrimental outcomes for this vulnerable, high-risk population.

Objective: This project aims to identify barriers to communication in children and young adults with special health care needs in the health care setting. These barriers will inform prototype development using human-centered design approaches to create a web-based application. Feedback from patients, caregivers, and health care providers (HCPs) was obtained on the usability and usefulness of the tool within the health care setting.

Methods: A needs assessment was conducted in which participants shared their experiences in providing or receiving health care services via a semistructured interview that was recorded and transcribed. Transcripts were analyzed inductively for themes, coded, and used to categorize the data. On the basis of these themes, iterative development of a web-based application for social stories took place. Focus groups were held to provide relevant feedback on the prototype.

Results: There were 15 participants (n=10, 67% HCPs and n=5, 33% patients and caregivers) interviewed for the needs assessment that informed prototype development. A web-based application for social stories depicting different aspects of health care interactions was created. Focus group feedback from 19 participants (n=12, 63% HCPs and n=7, 37% patients and caregivers) on usability through the System Usability Scale, along with narrative feedback, was obtained. Overall, the usability of the application was supported by caregivers and HCPs.

Conclusions: Children and young adults with special health care needs require medical services that their peers generally do not, thereby compounding potential barriers in communication surrounding health care delivery. Using social stories geared toward health care interactions may help reduce anxiety and difficulty.

(JMIR Pediatr Parent 2026;9:e76512) doi:10.2196/76512

KEYWORDS

pediatrics; complex care; human-centered design; qualitative; communication

Introduction

Complex care is defined as "a person-centered approach to address the needs of people whose combinations of medical, behavioral health, and social challenges result in extreme patterns of healthcare utilization and cost" [1]. Approximately 20% of adolescents (aged 12-17 years) in North America live with at least one chronic condition or special health care need, >90% of whom will require ongoing care into adulthood [1-3]. Unfortunately, 67% to 75% of individuals living with special health care needs experience frequent visits to the emergency room, forego recommended care (including lack of annual checkups), and frequently have multiple comorbid health issues [2-4]. As a result, their health care expenditures are estimated to be 5 times greater than those of the general population [5,6]. In addition, patients with complex, special health care needs often require individualized treatment plans to overcome the unique barriers they face in obtaining care [7]. These barriers often result in inequities in patient safety and health care outcomes in this population, particularly in individuals with intellectual disabilities [8]. The complexity of navigating health care systems may be lessened with care coordination in a medical home model; there have been reports of improved family satisfaction with overall care and improved health outcomes with dedicated care coordination [9]. As a result, high-quality and timely access to care services [10] and care coordination is cited as a top priority for individuals with disabilities living with specialized health care needs [11].

Complex care approaches focus on the patient, treating them as an individual embedded within a social context. Complex care programs benefit greatly from strong patient–health care provider (HCP) relationships, excellent communication practices, time, and use of interdisciplinary teams who work with specialized care providers to coordinate and provide patient-centered care [12].

Social stories are personalized narratives used to help teach children and young adults with autism spectrum disorder how to navigate social situations [13]. To do this, social stories use a combination of visual aids and text to teach social skills and increase understanding of social context and cues. The structure and predictability of a social story can decrease anxiety in new or unexpected environments. In addition, they can increase a patient's independence [14] and communication skills, and the skills taught via social stories may then generalize to other social contexts [15].

Patients with complex, special health care needs face several barriers in the health care system when seeking care. These barriers include difficulty communicating with HCPs, a lack of processes to accommodate individual needs, and difficulty accessing recommended care [16]. Specifically for individuals with autism spectrum disorder, patient behavior in combination with deficits in expressive and receptive communication may contribute to challenging medical encounters. A combination of environmental challenges in the setting where medical care is provided, demands placed on the patient (physical examinations or procedure based), and challenges with HCP communication and interaction may invoke challenging behavior in a patient population with reduced communication ability [17].

We sought to address some of these barriers within the health care system for this complex population of patients using the interface between technology and communication. We engaged in a multiphasic study through which we developed an application to address the information and design needs of patients; caregivers; and HCPs who engage in the direct care of patients living with complex, specialized health care needs. The widespread use of technology provides a digital space to create social stories describing health care interactions through visual and narrative means. This intersection of social context in the health care setting may help reduce communication barriers with children and young adults with special health care needs and their caregivers.

We hypothesize that (1) a web-based application can be developed to facilitate communication in the context of health care–specific interactions and flexibility to customize the technology for the patient and (2) a customized web-based application can help close the communication gap that exists for patients with complex, special health care needs, which will lead to fewer poor experiences for patients and caregivers.

Methods

Study Design and Setting

We conducted a human-centered design study in three phases consisting of (1) identifying end user needs using qualitative interviews; (2) conducting rapid, iterative prototyping of a web-based application; and (3) holding focus group sessions with end users for prototype feedback. This study relied on a convenience sample recruited from a large academic-affiliated health system (from both outpatient clinic and inpatient units) located in the Midwest region of the United States. The study population included patients with special health care needs, their caregivers, and HCPs who provide health care to patients with special health care needs.

Ethical Considerations

This project was approved by the University of Illinois College of Medicine Peoria Institutional Review Board under expedited

review (2053251). Assent for pediatric patients aged <18 years or adults without decision-making capacity was obtained along with parent or guardian consent. In addition, adults with decision-making capacity provided consent. Transcripts from interviews and focus groups were deidentified before analysis. Participants in the study were not provided with compensation for taking part.

Human-Centered Design Approach

The first phase of this study comprised semistructured interviews. The interviews were designed to elicit participants' perceptions of end user communication needs to inform the development of a prototype of a web-based application to facilitate communication among patients, caregivers, and health care personnel. We conducted qualitative interviews to inform the development of a prototype, which was subsequently tested via focus groups with anticipated end users.

Inclusion Criteria and Participant Eligibility

We gathered a convenience sample from a tertiary care hospital and outpatient care clinic that provides care to patients with special health care needs. To be eligible to take part in all phases of the study, participants needed to meet one of the following criteria: (1) being employed as health care personnel participating in the care of patients with special health care needs in the inpatient or outpatient setting, (2) being a pediatric or adult patient with special health care needs with decision-making capacity who presented for inpatient or outpatient care, or (3) being a caregiver of children or adults with special health care needs who presented for outpatient or inpatient visits.

Recruitment

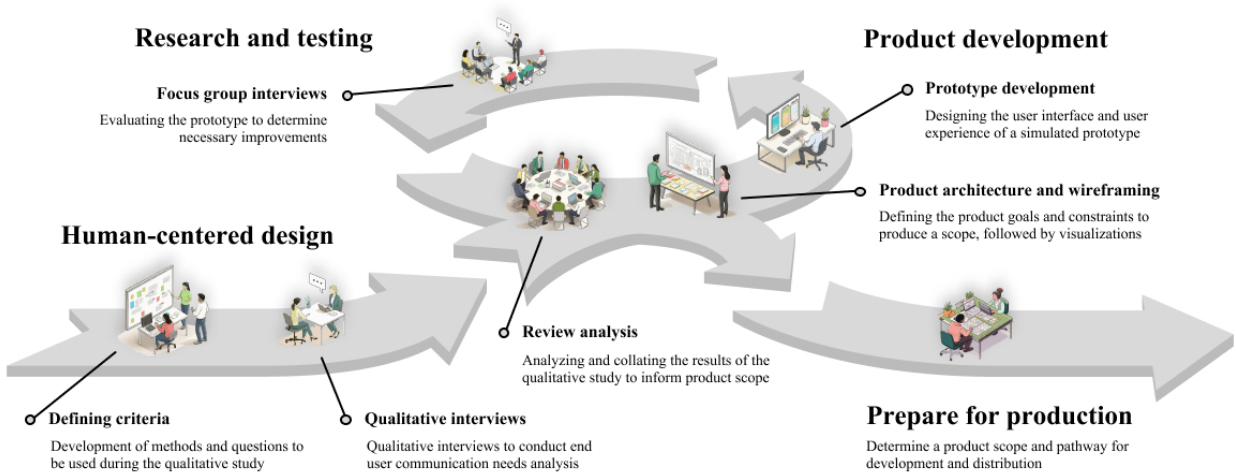
Recruitment for participation in the interviews to aid in the development of the web-based application and for the focus groups to provide feedback on the prototype took place via informational study fliers that were distributed electronically via email and physically within inpatient and outpatient clinical settings. The flier included a QR code to a participation questionnaire, as well as QR codes to the various types of age-appropriate consent information, giving prospective participants an opportunity to take part in the study. The participants provided their email address to be contacted for the needs assessment interview via phone call or audio-only Zoom videoconference (Zoom Video Communications).

HCPs providing inpatient or outpatient care were identified by their respective departments and emailed the study details and invitation to participate, in addition to the posted fliers.

Caregivers and patients were primarily recruited using fliers posted in clinical settings. To maximize the likelihood of reaching saturation, patients and caregivers were further identified through convenience sampling. They were contacted via phone or email directly for recruitment. For the purposes of saturation in qualitative methods, we targeted a minimum of 5 participants per category (HCPs, caregivers, and patients) [18].

After completing the initial interview, participants were asked whether they would like to take part in future focus groups providing feedback on the web-based application prototype (Figure 1).

Figure 1. Process map outlining the 3 phases of the study protocol. Phase 1 provided an overview of the human-centered design approach through qualitative interviews, phase 2 incorporated thematic feedback for prototype development, and phase 3 included focus group interviews on the prototype.



Interview Guide Development

The semistructured interview guide underwent several iterations. First, a draft interview guide was developed based on the study aims and clinical context. This was done via discussion among the investigative team; the initial guide was then pilot-tested with 1 caregiver and 1 HCP and subsequently revised. After 4 interviews, transcripts were reviewed in relation to the study's overarching research question. It was determined that the interview questions were too broad in relation to the study's research questions, and the guide underwent subsequent revision. Upon approval of the revised questions, follow-up interviews using the new guide were conducted with 2 of the previous 4 participants.

Before conducting interviews, all interviewers underwent interview training. The training materials were developed based on content made available by the University of Illinois Chicago's

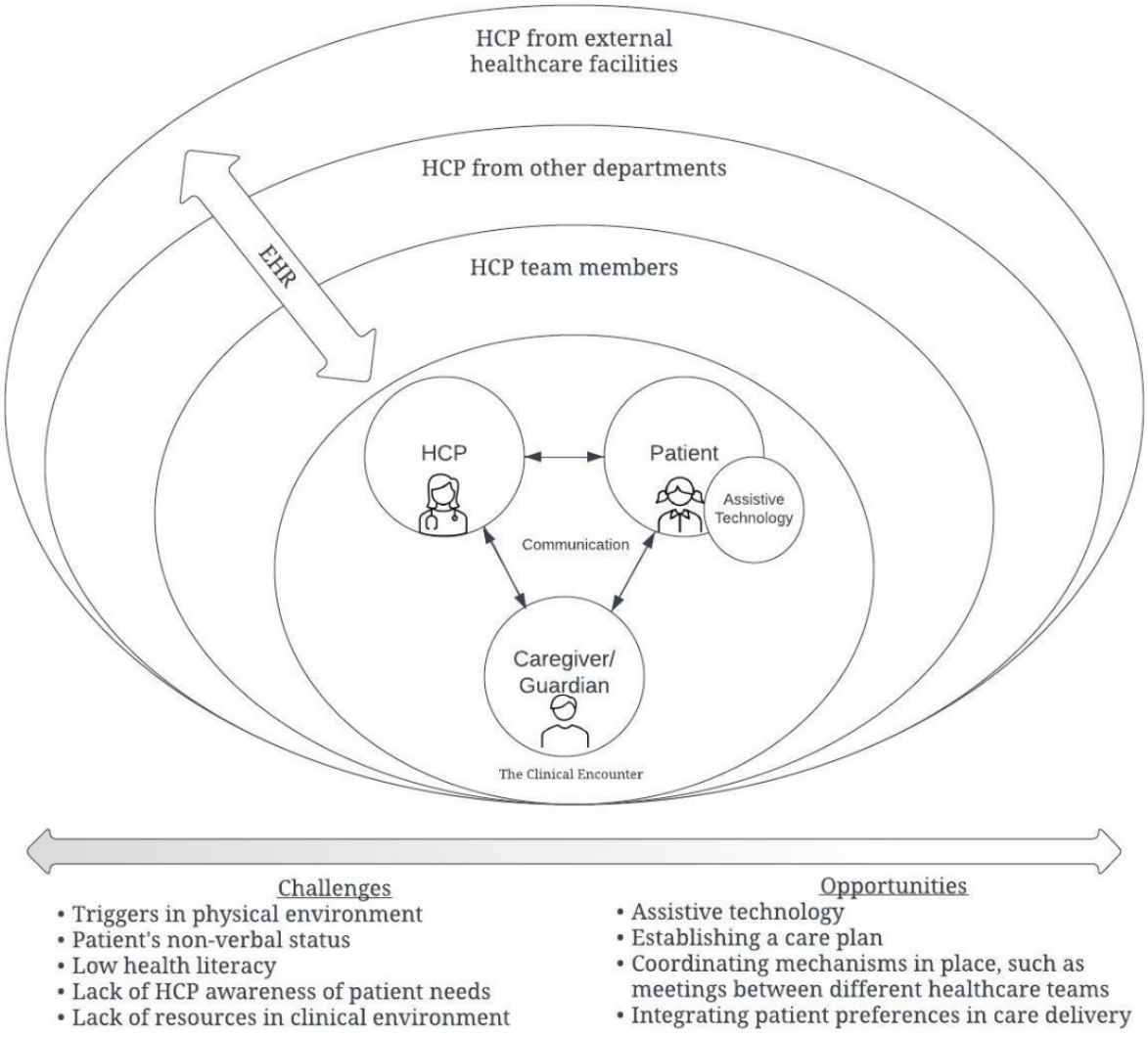
School of Public Health Collaboratory for Health Justice [19]. Training materials were presented, followed by opportunities to practice using the guide. The first 2 to 4 interviews were conducted under the supervision of a faculty member with expertise in qualitative human-centered design methods (AMH). Interviews were recorded and transcribed verbatim using Zoom (audio only). A member of the research team reviewed the auto-generated Zoom transcriptions for accuracy and to strip transcripts of identifiable information before analysis.

Qualitative Analysis

Deidentified transcripts were uploaded to ATLAS.ti (Scientific Software Development GmbH) [20] for subsequent analysis. Approaches to coding were inductive. Coders with backgrounds in medicine (DK), premedicine (MM), and human factors (AMH) reviewed the first 5 transcripts to identify codes and develop a codebook [21]. The remaining transcripts were coded independently by at least 2 trained coders; all discrepancies were identified and resolved for 100% consensus. Emergent themes were identified, sorted into challenges and opportunities, and interpreted for design recommendations. Distributed cognition and communication theories that consider complex interdependencies among multiple individuals, coordinating mechanisms, communication, and information sharing patterns were used for interpretation [22].

The results identified layers of the health system reliant on communication and coordination practices (Figure 2); challenges and opportunities arise within each layer of the complex system. However, at the core of these interactions is communication, which occurs within the clinical encounter. We thereby focused on codes and emergent themes that attested to challenges and opportunities regarding communication as identified in the clinical encounter.

Figure 2. Phase 1 interviews identified opportunities and challenges in communication with health care providers (HCPs) during health care interactions. There are many layers of communication that occur in the coordination of the care of complex patients and their families.



Product Development

Prototyping

In phase 2, qualitative themes from the interviews were provided to the design team responsible for the development of a prototype for a web-based application through a series of meetings with members of the investigative team (AMH, JRH, and MJM) and designer (KF) to develop multiple iterations of the application. Iterative feedback was received from the investigative team and qualitative lead at that point for clinical expertise (JRH and MJM) and participant voices (AMH) and was incorporated. Each iteration of the prototype was shared with the research team to meet the specifications of the prospectively identified communication and end user needs. Approximately 3 iterative cycles were completed to achieve a final prototype.

Clinical Scenario Development

Interviews revealed a common need in both inpatient and outpatient settings: to convey routine tests and procedures in an understandable and nonthreatening manner. To this end, members of the study team developed 2 clinical scenarios through which interaction with the application would guide and convey the steps of a visit or procedure. Our principal investigators, subject matter experts, and internal medicine/pediatric physicians (JRH and MJM) worked closely with content creators to develop a realistic clinical scenario that met the identified areas of need and preference relevant to both inpatient and outpatient care settings. This was developed and refined over several iterations and was ultimately used to inform the flow of the application.

Product Design Integration

A user flow was built to represent high-level user experience. An information flow diagram helped identify how this application might connect to other utilities in the health system. A technical assessment began to explore the feasibility of certain features. On the basis of the feedback received, the user flow and technical assessment were refined to address any identified issues and further focus the scope of the application. In conjunction with clinical scenario development, wireframes were built to visualize the prototype layout and structure and develop a preliminary art style for the content. As a result of project team feedback, the wireframes were revised to represent a final scope for the prototype to be used in focus group user testing. Art styles were explored considering the original research data, market evaluation, and potential full-scale production technical constraints. The final phase involved iterating on the art style and developing all necessary design assets to produce a prototype for user testing.

Research and Prototype Testing

The third phase of our human-centered design study tested the resulting prototype with end user groups. These user groups consisted of patients, caregivers, and HCPs. For participants aged <18 years, caregivers provided verbal consent, and the participating children provided assent. Decisions to participate were audio recorded. Caregivers of adult patients without decision-making capacity provided consent for the patient to participate.

HCPs, patients, and caregivers who consented to participate convened in person and met in separate focus groups (ie, HCPs met separately from patients and caregivers). Focus group participants were welcomed, provided with the project aims and objectives, provided with a tablet and QR code, and briefed on how to access the prototype application from the tablet provided to them. To interact with the prototype, patients and caregivers were given a medical social story to simulate receiving the social story before an upcoming appointment. They then chose whether they wanted to be represented by an animal or person avatar throughout the medical social story.

HCPs were given the same medical social story under the pretense of having to prepare and send the social stories before an upcoming medical encounter with a patient with special health care needs. All participants were presented with the System Usability Scale (SUS) questionnaire upon completion of the scenarios and then took part in a facilitated focus group interview session to provide more granular feedback. Focus group interviews were recorded and transcribed verbatim with participant consent.

Focus Group Data Collection Forms

To evaluate prototype success in the focus group, we adapted survey questions from the SUS [23] and from existing educational immersion scales [24]. Furthermore, we used a semistructured guide to elicit open-ended feedback on the application experience from end users. The SUS is a scale designed to measure participants' perceived usability of a product. Scores range from 0 to 100, with the average score (50th percentile) being 68 [25]. A score above 70 is generally considered acceptable [26-28]. The Learning Immersion Scale in Simulation is a psychometrically validated and reliable survey consisting of 4 factors: cognitive assimilation, emotional buy-in, focused attention, and autotelic experience [24]. We focused on the cognitive assimilation subscale, which measures to what extent an individual differentiates between interaction with the simulated environment and reality [29]. The questionnaire uses a 7-question, 5-point scale adapted from Ko et al [24]. It ranges from "strongly disagree" (1) to "strongly agree" (5). This scale was only administered to caregivers and patients.

Focus Group Data Analysis

SUS surveys were pooled and analyzed descriptively, examining overall usability and usability differences by role (caregivers and patients vs HCPs). The caregiver or parent survey asked 3 additional questions about how they thought their child or the individual under their care would feel about the prototype, which caused the SUS scale score range to change to 0 to 130. Scores are reported as both out of 130 and normalized to fit the original scale from 0 to 100. The resulting focus group transcripts were reviewed iteratively for trends, including areas of feedback (positive and negative) on the prototype interaction experience.

Results

Overview

Results are presented and discussed in the order in which the phase of research took place. There were 15 participants (n=10,

67% HCPs and n=5, 33% patients and caregivers) in the phase 1 individual interviews (Table 1).

Table 1. Participant demographics.

	Patients and caregivers, n (%)	Health care providers, n (%)
Phase 1 interviews (n=15)	5 (33)	10 (67)
Phase 3 focus groups recruited in phase 1 (n=9)	4 (44)	5 (56)
Phase 3 focus groups recruited in phase 3 (n=10)	3 (30)[a]	7 (70)
Phase 3 total (n=19)	7 (37)	12 (63)

[a]One pediatric participant.

Human-Centered Design

Qualitative analysis of the interviews completed in phase 1 identified both challenges and opportunities of care, illustrating design needs for our web-based application. Themes were identified to help prioritize the creation of a web-based application as a potential solution to reduce challenges and facilitate interaction (Multimedia Appendix 1).

Challenges in Communication Based on Cognitive Status

Equating verbal status with cognitive ability emerged as a barrier to effective communication. This challenge was mentioned by HCPs, with 1 patient and 1 caregiver noting this barrier from their perspective. Mainly, HCPs indicated difficulty assessing patient capabilities and capacity to communicate and understand interactions autonomously due to time constraints or unfamiliarity with the patient:

> *Because a lot of times they are so afraid of what's going on, and they're not understanding. And sometimes...providers in the room aren't understanding them. And that can cause a lot of problems.* [HCP]

> *That they're special needs and they don't have the capacity when so many of them have the capacity we [HCPs] just don't have the time to spend with them to understand what they do have.* [HCP]

Meanwhile, patients and caregivers described how this challenge manifested on their end in that HCPs may not provide enough information on what they do with a patient, expressing a desire for more explanation that warranted more communication. One caregiver described their intensive care unit experience as follows:

> *[A] nurse will come in and start something [a routine medical procedure]. You're like, wait a minute, what are you doing? So I don't know if like just a little more communication as to the doctors thought processes of: This is what we're doing. This is what we're thinking and what we're going to try.* [Caregiver]

Several caregivers noted that the barrier was the verbal status (eg, the patient's ability to talk) rather than their cognitive ability or capacity for understanding. One caregiver noted the following:

> *Just because he can't talk doesn't mean he doesn't have feelings and doesn't understand everything.* [Caregiver]

> *Now his mom never left she was able to communicate a lot with us [HCP team].... I didn't know him [the patient] very well. And so if mom worked or couldn't be here, there would have been a lot of gap(s) in communication and understanding what he needed.* [HCP]

Some HCPs corroborated this element of patient understanding despite verbal status:

> *[Patient name] is very, very smart and understands a lot of what we are saying or doing. So even though a child may have special needs and are non-verbal and not able to communicate what they want, how we do, I feel like it's important to know that they still sometimes are aware of their surroundings and are smarter than we really realize.* [HCP]

Opportunities to Improve Communication

The challenge of ascertaining a patient's cognitive status is prevalent, often requiring intervention and the constant presence of a caregiver. Overall, there was a preference for more streamlined and direct communication between HCPs and patients. Ideas for supporting direct communication during an encounter included the use of simplified messaging. For instance, one HCP highlighted the need to use simple language to explain routine procedures in terms readily understood by patients and caregivers:

> *Yes, and it fits people that are aware of medical jargon...[but] our special needs are not. And so we have to be able to. Adjust [using jargon] as according to our patients you know.* [HCPs]

Assistive devices were noted in their ability to promote more direct and patient-centered communication during clinical encounters, potentially bypassing the otherwise constant need for caregiver presence. These devices were often noted to be available through specialty hospital services or otherwise belong to the patient privately for at-home use, and it was noted that they were useful in the health care environment to aid in direct communication between HCPs and their patients. In other cases, involving the caregivers in the health care delivery provides insights into the patient's mannerisms and needs in a way that only those who know them best can decipher. One HCP noted the following:

> *[D]epending on what's going on with them. Absolutely. Yeah, so more individuality in a system, you know. Ability to change it to a specific patient would be helpful.* [HCP]
>
> *So if they are like not actually having access to smart tablets or for understanding and navigating the My Chart system. So that technology.... To feel comfortable with doing that is one of the barriers.* [HCP]

When describing aspects of an assistive device used for communication, several facilitating device features were noted to enhance communication during the clinical encounter:

> *So from, provider to patient communication, just having some, a bunch of preset sort of procedures, the very, you know, illustrations that sort of simplify it and make it easy to understand for the patient and then on the opposite, you know, things that allowed them to quickly say and communicate with you without.... That doesn't mean they're the same as everyone else. And so I think a goal with the app is to sort of a breakdown communication barriers so that there's an easier time getting to know them and then also you know those individual preferences as individual traits and things like that can be quickly communicated and just sort of be embedded and quickly understood by providers.* [HCP]

Prototype Development

In the process of prototype development in phase 2, we focused on actionable themes that met the following criteria: (1) opportunities to improve communication in clinical encounters (the reason for this being that broader coordination issues worsened in part because of communication problems, ie, communication is often necessary for coordination) and (2) common or routine tests and procedures experienced in both inpatient and outpatient encounters.

A tappable prototype was built and prepared for testing on tablet devices. Story development and key features targeted emergent qualitative themes, with a focus on routine procedures performed in both inpatient and outpatient visits and on a standardized outpatient visit.

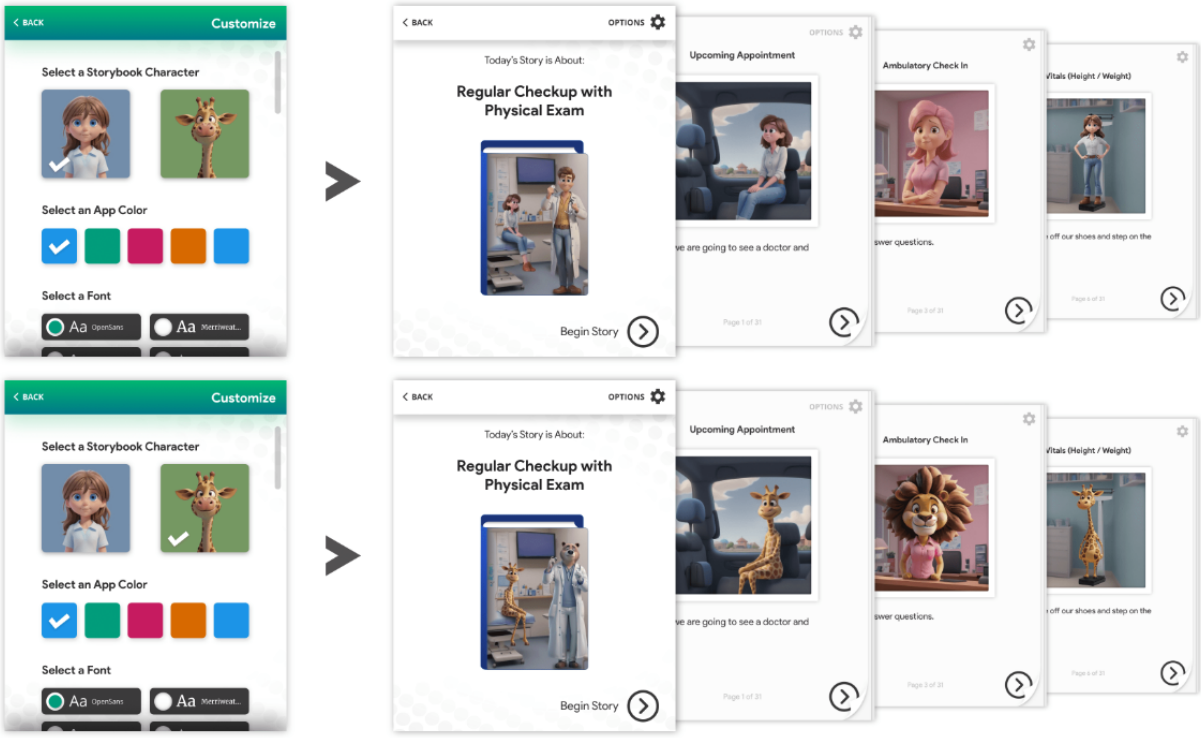
Common procedures patients undergo in both settings are phlebotomy and measurement of vital signs, and a common visit focus in the outpatient setting is a preventative examination or annual physical examination (including obtaining vital signs). An outline of narrative details was created that would help shape the final visual representation of each step within the procedure or visit. A landing page of potential scenarios for social stories was created (Figures S1 and S2 in Multimedia Appendix 2), with the standardized outpatient visit fully developed with 2 different representations, an animal or a person avatar. This allowed the end user to choose the avatar that best reflected their desire to represent themselves in an upcoming encounter.

Research and Prototype Testing

The scope of the themes was then identified by the research team to provide the framework for the iterative prototyping of the web-based application. The web-based application focused on providing a visual and narrative aid to explain treatments or procedures to the patients and caregivers. It also provided the opportunity for the patient to react to the content (Table S1 and Figure S1 in Multimedia Appendix 3).

In the phase 3 focus groups, there were a total of 19 participants (Table 1) who provided feedback on the visual and narrative examples (Figure 3) for the web-based application.

Figure 3. Social story outlining the steps of a routine outpatient preventative examination. The patient or caregiver can select between an animal and a person avatar that represents them throughout the social story.



SUS Results

HCP Results

On the basis of the HCP SUS scores (Table 2), the prototype ranked in the "best imaginable" range with an average score of 93.125 (SD 6.67; 96th-100th percentile). However, due to the small sample size, more participants would be needed to accurately judge the usability and generalizability of the prototype.

Table 2. System Usability Scale scores from health care providers.

Respondent ID	Score (range 0-100)[a]
HCP1	97.5
HCP2	100
HCP3	90
HCP4	92.5
HCP5	100
HCP6	90
HCP7	97.5
HCP8	100
HCP9	97.5
HCP10	87.5
HCP11	80
HCP12	85

[a]Average score: 93.125 (SD 6.67).

Caregiver Results

The caregiver SUS scores (Table 3) also ranked the prototype in the "best imaginable" range, with an average normalized score of 88.08 (96th-100th percentile).

Table 3. System Usability Scale scores from caregivers.

Respondent ID	Original score (0-130)[a]	Normalized score (range 0-100)[b]
CGP1	127.5	98.08
CGP2	112.5	86.54
CGP3	97.5	75
CGP4	125	96.15
CGP5	110	84.62

[a]Average score: 114.5 (SD 12.17).

[b]Average score: 88.08 (SD 9.36).

Patient Results

The 2 patients surveyed rated the prototype just below the average SUS score with an average score of 66.25 (SD19.45; 41st-59th percentile). This would place it in the "OK" range.

Cognitive Assimilation

Caregivers (n=5) consistently provided high scores on the cognitive assimilation scale, with the lowest score being 4.2 for "I was able to see it I was doing it right" and "The situation seemed to flow smoothly" (Table S1 in Multimedia Appendix 4). However, patients (n=2) provided low scores on cognitive assimilation, with an average of 2 for most of the questions (Table S2 in Multimedia Appendix 4).

Focus Group Themes

Common themes from both HCPs and patients and caregivers on the prototype centered on the user-friendly application design, including the avatar choice between an animal and a person. The design was overall appreciated as a general medical social story. However, broad use among patients with special health care needs of all age ranges may be limited by the spectrum of ability. The current narrative explanations and visual aids are most applicable to patients of a certain cognitive ability regardless of chronological age. When used outside of that scope, they may not be as effective for end user interaction. However, both groups generalized applicability to neurotypical children of a similar cognitive level:

> *Yeah, I thought that it was really well laid out. I thought that as a mom and then also as like I'm a pediatric nurse. So from both of those aspects, I feel like this is something that we're currently missing. We're not able to make those connections with our pediatric patients because you know we don't have a lot of child friendly resources to kind of help them prepare for different things that they may go through in their health care process.* [Caregiver]

Both patients and caregivers and HCPs provided feedback on preferences regarding the ability to modify the scope of the medical social story for various end users, including narrative explanations, length, and increased avatar options, which may increase general applicability and end user satisfaction:

> *You know some more of an adult or approaching adults. A tween? Yeah, a tween. I would expect most children would pick the giraffe. Yeah, but I would think more teens and tweens would choose a teenager type avatar. That would kind of bridge the gap between the very childlike and juvenile appearance and the more capturing the ages in between.* [HCP]

Caregivers and patients with specific health care needs identified potential areas of improvement to meet broader patient care accessibility needs. The addition of an audio version with sign language should be considered so that individuals that are hard of hearing or cannot read can still use the application fully.

In the HCP feedback specifically, there was some concern about being able to capture the many different permutations through which an HCP could approach a case to meet the needs of each individual patient or account for variances within a health care system in terms of what information is included. In addition, as the nature of patients with special health care needs may result in frequent encounters with the health care system, the application needs to have the ability to be personalized for repeated similar health care interactions for patients:

> *We use freezy spray so I can speak to that, but you know the hospital will use numbing cream, you know, and it is, you know, you use these words, but they have different things that we can use. Some people use a little light to see your vein. Sometimes you know.* [HCP]

> *Going back to this surgery example, kids have to be under a certain weight to get a mask induction versus the IV induction and so you can't even break it down by 8 years if you have a 7-year-old that weighs the same as in 12-year-old. The seven-year-old will get the IV right? Yeah, it it's very specific.* [HCP]

Discussion

Principal Findings

The initial objective of this project was to create a web-based application that would facilitate communication for patients with complex, special health care needs during interactions with their health care team. Through a 3-phase human-centered design process, we identified communication needs for patients

with complex, special health care needs; their caregivers; and their HCPs. Key themes informed prototype development. The prototype was then reviewed by a group of stakeholders; the results of focus groups with end users support overall usability and utility for certain patient populations.

The results from the phase 1 interviews identified areas of opportunities and challenges in communication for patients with special health care needs in health care encounters. Specifically, HCPs' level of comfort and inclusive communication with the patients and their caregivers during clinical encounters were two areas identified as needing improvement. Patients notably display a broad range of communication abilities, particularly when using alternative and augmentative communication methods. Interestingly, capacity for direct verbal communication may not directly reflect an individual's cognitive ability. This finding highlights how assumptions on patient function and ability and possible ableism may create additional communication barriers during a patient encounter. Additionally, a high reliance on caregivers for verbal communication of the patient status may further complicate the patient's relationship with the health care team, whereas using communication devices available allows for direct communication with the patient. This was consistent with our initial hypothesis and experience within the health care system. Although our data suggest additional multiple areas of opportunity to enhance communication, we focused on the findings most relevant to the interactions among HCPs, patients, and caregivers during an encounter for prototype development. In phase 2, construction of the prototype elements was important to consider for scalability and future digital interfaces. The design of the medical social stories was chosen to depict routine experiences for all patients and their families but potentially significant barriers for patients with complex, special health care needs due to the physical environment and impaired communication and rapport with the health care team. Generated images of an animal or person avatar were incorporated to provide options for individual preferences, but creating avatars using personalized variables was not possible during prototype creation to remain within the scope. Striking an appropriate balance among the development of a web-based application, functionality, and scalability was at the forefront of our decisions when finalizing prototype details. In phase 3, the qualitative feedback from stakeholders highlighted consistencies and potential applicability to a focused patient population with communication barriers and intellectual disabilities. The qualitative feedback from HCPs and caregivers identified that this prototype may be applicable to patients with typical development at the same cognitive level. Overall, caregiver and HCP testing was supportive of the prototype meeting the standards for usability [30]. The combination of the qualitative and quantitative data was positively congruent with HCPs and caregivers, but the small sample size prevents solid conclusions on usability and generalizability to both other institutions and other patient groups with special health care needs.

Limitations of this study include the representativeness of the sample in both size and diversity of disease processes that result in specialized health care needs. Therefore, the full scope of opportunities and challenges faced by patients across the continuum may not be identified. Thus, the broad applicability of a prototype designed for all patients with special health care needs would need further validation across a larger sample size and across multiple areas of care delivery within the health care system. In addition, there was a broad range of levels of ability among patients in the phase 1 interviews, which created variable outcomes in the design of the prototype. Building a prototype that would meet the needs of an innately unique population of patients with varying levels of ability within the same diagnoses exceeded the scope and timeline of the first cycle of this project but is a rich area for further refinement and implementation. Finally, just as the patients are unique, so are the individual approaches that health care team members bring to their patient care. For example, the approach to anesthesia for a surgical procedure may vary depending on the HCP and the patient's unique needs, which, therefore, could limit the generalization of the medical social story.

Conclusions

Our study identified an area of further exploration for increasing successful health care interactions with patients with special health care needs through using social stories in a web-based application. Providing a mechanism to prepare patients and their caregivers for health care interactions by introducing a standardized process allows for structure, visibility, and appropriate anticipation. This also allows for the engagement of HCPs to communicate with patients and families before busy encounters to help set expectations and also potentially reduce anxiety about the unknown. As noted in the stakeholder feedback, these features are also applicable for health care interactions involving neurotypical patients. Future areas of research will be to further the development of the web-based application, including expansion of the medical social stories for increased applicability and evaluating implementation in the workplace for feasibility. In future ambulatory workflow states, distributing the relevant social stories through the electronic medical record before an upcoming health care encounter would mimic the use of social stories in other settings. During inpatient admissions, the social stories may be deployed by caregivers or bedside nursing staff for individualized preparation for the health care process and procedures. Once the workflows are established, an evaluation of patient, caregiver, and HCP satisfaction is warranted to ensure usability and impact. In addition, consideration of a built-in functionality for a caregiver or patient to select an estimated cognitive ability, which may increase applicability to broader neurotypical and neurodivergent patient populations, would require flexibility in terms of social story content and avatar selection. Finally, other common themes that were identified as areas of need were the ability to provide an updated interface with the health care team highlighting important, individualized care elements; patient advocacy; and preferred interfaces with electronic health records.

Acknowledgments

The views expressed in this manuscript are those of the authors and do not necessarily reflect the positions or policies of the US Department of Veterans Affairs or the US government. The authors thank Lauren Stufflebeam and Jakob Plotts as part of the design team, and Lidia Barraza for contributions with coding.

Funding

This project was funded by the Community Health Advocacy (CHA) program through the University of Illinois Chicago Office of the Vice Chancellor for Innovation.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Codes used in the qualitative analysis of phase 1 interviews with patients, caregivers, and health care providers.
[DOCX File , 16 KB - pediatrics_v9i1e76512_app1.docx]

Multimedia Appendix 2
Screenshot of the web-based application home page and of the landing page depicting the table of contents of potential social stories for health care interactions.
[DOCX File , 144 KB - pediatrics_v9i1e76512_app2.docx]

Multimedia Appendix 3
Outline of the thematic takeaway from phase 1 for the creation of the prototype and potential directions for future iterations and an accompanying visualization of the function of the prototype.
[DOCX File , 420 KB - pediatrics_v9i1e76512_app3.docx]

Multimedia Appendix 4
Cognitive assimilation scores for caregivers and patients. Scores measure to what extent an individual differentiates between interaction with the simulated environment and reality.
[DOCX File , 15 KB - pediatrics_v9i1e76512_app4.docx]

References

1. Stevens AC, Carroll DD, Courtney-Long EA, Zhang QC, Sloan ML, Griffin-Blake S, et al. Adults with one or more functional disabilities - United States, 2011-2014. MMWR Morb Mortal Wkly Rep 2016 Sep 30;65(38):1021-1025 [FREE Full text] [doi: 10.15585/mmwr.mm6538a1] [Medline: 27684532]
2. Disability impacts all of us infographic. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/disability-and-health/articles-documents/disability-impacts-all-of-us-infographic.html [accessed 2025-02-25]
3. Gulley SP, Rasch EK, Chan L. The complex web of health: relationships among chronic conditions, disability, and health services. Public Health Rep 2011 Jul 01;126(4):495-507 [FREE Full text] [doi: 10.1177/003335491112600406] [Medline: 21800744]
4. Kennedy J, Wood EG, Frieden L. Disparities in insurance coverage, health services use, and access following implementation of the Affordable Care Act: a comparison of disabled and nondisabled working-age adults. Inquiry 2017;54:46958017734031 [FREE Full text] [doi: 10.1177/0046958017734031] [Medline: 29166812]
5. Kraus L, Lauer E, Coleman R, Houtenville A. 2017 Disability statistics annual report. University of New Hampshire. 2018. URL: https://files.eric.ed.gov/fulltext/ED583258.pdf [accessed 2025-12-09]
6. Crabb C, Owen R, Stober K, Heller T. Longitudinal appraisals of family caregiving for people with disabilities enrolled in Medicaid managed care. Disabil Rehabil 2020 Aug 29;42(16):2287-2294. [doi: 10.1080/09638288.2018.1557266] [Medline: 30696289]
7. Ghandour RM, Hirai AH, Kenney MK. Children and youth with special health care needs: a profile. Pediatrics 2022 Jun 01;149(Suppl 7):e2021056150D. [doi: 10.1542/peds.2021-056150D] [Medline: 35642877]

8. Mimmo L, Harrison R, Travaglia J, Hu N, Woolfenden S. Inequities in quality and safety outcomes for hospitalized children with intellectual disability. Dev Med Child Neurol 2022 Mar 25;64(3):314-322 [FREE Full text] [doi: 10.1111/dmcn.15066] [Medline: 34562021]
9. Boudreau AA, Perrin JM, Goodman E, Kurowski D, Cooley WC, Kuhlthau K. Care coordination and unmet specialty care among children with special health care needs. Pediatrics 2014 Jun 26;133(6):1046-1053. [doi: 10.1542/peds.2013-2174] [Medline: 24864172]
10. McLellan SE, Mann MY, Scott JA, Brown TW. A blueprint for change: guiding principles for a system of services for children and youth with special health care needs and their families. Pediatrics 2022 Jun 01;149(Suppl 7):e2021056150C. [doi: 10.1542/peds.2021-056150C] [Medline: 35642876]
11. Association of Maternal and Child Health Programs. Models of care for children and youth with special health care needs: promising models for transforming California's system of care. Lucile Packard Foundation Children's Health. 2009 Nov 1. URL: https://tinyurl.com/2f35n6eh [accessed 2024-07-16]
12. Karam M, Chouinard MC, Poitras ME, Couturier Y, Vedel I, Grgurevic N, et al. Nursing care coordination for patients with complex needs in primary healthcare: a scoping review. Int J Integr Care 2021 Mar 19;21(1):16 [FREE Full text] [doi: 10.5334/ijic.5518] [Medline: 33776605]
13. Wahman CL, Pustejovsky JE, Ostrosky MM, Santos RM. Examining the effects of social stories™ on challenging behavior and prosocial skills in young children: a systematic review and meta-analysis. Top Early Child Spec Educ 2019 Aug 02;41(4):267-279. [doi: 10.1177/0271121419855692]
14. Ivey ML, Heflin LJ, Alberto P. The use of social stories to promote independent behaviors in novel events for children with PDD-NOS. Focus Autism Other Dev Disabl 2004 Aug 01;19(3):164-176. [doi: 10.1177/10883576040190030401]
15. Wright LA, McCathren RB. Utilizing social stories to increase prosocial behavior and reduce problem behavior in young children with autism. Child Dev Res 2012 May 08;2012:1-13. [doi: 10.1155/2012/357291]
16. Cassidy L, Quirke MB, Alexander D, Greene J, Hill K, Connolly M, et al. Integrated care for children living with complex care needs: an evolutionary concept analysis. Eur J Pediatr 2023 Apr 13;182(4):1517-1532 [FREE Full text] [doi: 10.1007/s00431-023-04851-2] [Medline: 36780041]
17. Wilson SA, Peterson CC. Medical care experiences of children with autism and their parents: a scoping review. Child Care Health Dev 2018 Nov 22;44(6):807-817. [doi: 10.1111/cch.12611] [Medline: 30136407]
18. Yates KA, Feldon DF. Advancing the practice of cognitive task analysis: a call for taxonomic research. Theor Issues Ergonomics Sci 2011 Nov;12(6):472-495. [doi: 10.1080/1463922X.2010.505269]
19. Collaboratory for health justice. University of Illinois Chicago. URL: https://publichealth.uic.edu/community-engagement/collaboratory-for-health-justice/ [accessed 2024-06-04]
20. Hwang S. Utilizing qualitative data analysis software: a review of Atlas.ti. Soc Sci Comput Rev 2007;26(4):519-527. [doi: 10.1177/0894439307312485]
21. Hennink MM, Kaiser BN, Marconi VC. Code saturation versus meaning saturation: how many interviews are enough? Qual Health Res 2017 Mar;27(4):591-608 [FREE Full text] [doi: 10.1177/1049732316665344] [Medline: 27670770]
22. Hutchins E. Cognition, distributed. In: Smelser NJ, Baltes PB, editors. International Encyclopedia of the Social & Behavioral Sciences. Oxford, UK: Pergamon Press; 2001:2068-2072.
23. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation in Industry. London, UK: CRC Press; 1996.
24. Ko EJ, Nam KA, Kim EJ. Development and psychometric testing of learning immersion scale in clinical simulation: a methodological study. Nurse Educ Today 2022 Jun;113:105363. [doi: 10.1016/j.nedt.2022.105363] [Medline: 35439632]
25. Sauro J, Lewis JR. Quantifying the User Experience: Practical Statistics for User Research. Burlington, MA: Morgan Kaufmann Publishers; 2016.
26. Melnick ER, Dyrbye LN, Sinsky CA, Trockel M, West CP, Nedelec L, et al. The association between perceived electronic health record usability and professional burnout among US physicians. Mayo Clin Proc 2020 Mar;95(3):476-487 [FREE Full text] [doi: 10.1016/j.mayocp.2019.09.024] [Medline: 31735343]
27. Sauro J. Measuring Usability with the System Usability Scale (SUS). MeasuringU. 2011 Feb 3. URL: https://measuringu.com/sus/ [accessed 2024-04-25]
28. Sauro J. 5 ways to interpret a SUS score. MeasuringU. 2018 Sep 19. URL: https://measuringu.com/interpret-sus-score/ [accessed 2024-04-23]
29. Tian Z, Zhang K, Zhang T, Dai X, Lin J. Application of Ausubel cognitive assimilation theory in teaching/learning medical biochemistry and molecular biology. Biochem Mol Biol Educ 2020 May 20;48(3):202-219 [FREE Full text] [doi: 10.1002/bmb.21327] [Medline: 31859445]
30. Hyzy M, Bond R, Mulvenna M, Bai L, Dix A, Leigh S, et al. System usability scale benchmarking for digital health apps: meta-analysis. JMIR Mhealth Uhealth 2022 Aug 18;10(8):e37290 [FREE Full text] [doi: 10.2196/37290] [Medline: 35980732]

Abbreviations

HCP: health care provider
SUS: System Usability Scale

Edited by S Badawy; submitted 03.Jun.2025; peer-reviewed by A Roundtree; comments to author 02.Jul.2025; revised version received 23.Nov.2025; accepted 27.Nov.2025; published 06.Jan.2026.

Please cite as:
Hanks JR, Hughes AM, Sultana S, Klute R, Formella K, Flynn C, Wallenfang A, Krishnakumar D, Mourad M, Morse YC, Mischler MJ
Facilitating Communication With Children and Young Adults With Special Health Care Needs Through a Web-Based Application: Qualitative Descriptive Study
JMIR Pediatr Parent 2026;9:e76512
URL: https://pediatrics.jmir.org/2026/1/e76512
doi:10.2196/76512
PMID:

Feasibility, Diagnostic Accuracy, and Satisfaction of an Acute Pediatric Video Interconsultation Model in Rural Primary Care in Catalonia: Prospective Observational Study

Marta Castillo-Rodenas[1,2], MD; Núria Solanas Bacardit[1], RNC; Clotilde Farràs Company[1], MD; Queralt Miró Catalina[2,3], MSc, PhD; Laia Solà Reguant[3,4], MSc; Aïna Fuster-Casanovas[5,6,7], RPh, PhD; Francesc López Seguí[8,9], PhD; Josep Vidal-Alaball[2,3,10], MD, MPH, PhD

[1]Centre d'Atenció Primària Cardona, Gerència d'Atenció Primària i a la Comunitat de la Catalunya Central, Institut Català de la Salut (ICS), Cardona, Barcelona, Spain

[10]Department of Medicine, Faculty of Medicine, University of Vic - Central University of Catalonia (UVIC-UCC), Vic, Catalonia, Spain

[2]Intelligence for Primary Care Research Group (I4PC), Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina, Manresa, Catalonia, Spain

[3]Unitat de Recerca i Innovació Gerència d'Atenció Primària i a la Comunitat de la Catalunya Central, Institut Català de la Salut (ICS), Carrer de Soler i March, 6, CAP Bages, Manresa, Catalonia, Spain

[4]Health Promotion in Rural Areas Research Group, Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina, Manresa, Catalonia, Spain

[5]Unitat d'Innovació, Direcció de Qualitat, Processos i Innovació, Hospital Universitari Vall d'Hebron, Barcelona, Catalonia, Spain

[6]Grup de recerca en Serveis Sanitaris, Vall d'Hebron Research Institute, Barcelona, Catalonia, Spain

[7]eHealth Lab Research Group School of Health Sciences and eHealth Centre, Universitat Oberta de Catalunya (UOC), Barcelona, Catalonia, Spain

[8]Chair in ICT and Health Centre for Health and Social Care Research, University of Vic - Central University of Catalonia (UVIC-UCC), Vic, Catalonia, Spain

[9]Centre de Recerca en Economia i Salut (CRES), Universitat Pompeu Fabra (UPF), Barcelona, Catalonia, Spain

Corresponding Author:
Josep Vidal-Alaball, MD, MPH, PhD
Intelligence for Primary Care Research Group (I4PC), Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina, Manresa, Catalonia, Spain

Abstract

Background: In Catalonia, Spain, pediatric primary care is undergoing restructuring, including greater integration of information and communication technologies. The adoption of digital health solutions has increased notably since the COVID-19 pandemic. In areas with limited availability of health care professionals, digital tools are a key strategy for facilitating access to services and ensuring continuity of care.

Objective: This study aimed to evaluate the feasibility, diagnostic agreement, and satisfaction of users and professionals of an acute pediatric video consultation model, referred to as video interconsultation, that includes a synchronous remote physical examination and takes place between health care professionals.

Methods: This was a 20-month prospective within-patient diagnostic accuracy study including 200 children (aged 0 - 14 y) with acute conditions in rural primary care in Catalonia. A secure, closed, real-time, web-based, clinician-assisted video consultation platform enabled remote pediatric assessment—visual examination, audio auscultation via a digital stethoscope, and caregiver-reported symptoms—with a pediatrician remotely guiding a nurse physically present with the child. The intervention was compared, in all cases, with a standard in-person pediatric assessment as the reference standard. Outcomes were feasibility, diagnostic accuracy, and user and professional satisfaction. The platform was developed based on telemedicine usability and clinical safety principles.

Results: Of the 200 children enrolled, remote video consultations were successfully completed in 64.5% (129/200) of cases. Diagnostic agreement with in-person assessment was 78.2% (129/165). Overall mean diagnostic accuracy across all diagnoses was 0.99 (95% CI 0.98 - 1.00), with a mean specificity of 0.99 (95% CI 0.98 - 1.00) and a mean sensitivity of 0.90 (95% CI 0.84 - 0.95), varying by condition, with lower performance for pathologies requiring detailed physical examination. Overall, 95% (190/200) of users and 74% (148/200) of professionals reported a positive experience.

Conclusions: The proposed pediatric video consultation model was feasible, accurate, and well accepted for managing a substantial proportion of acute pediatric conditions in primary care. Its implementation could improve access to medical care in

rural areas and help reduce health care disparities. Further research is needed to support scalability and implementation in routine clinical practice.

International Registered Report Identifier (IRRID): RR2-10.2196/52946

(JMIR Pediatr Parent 2026;9:e82133) doi:10.2196/82133

KEYWORDS

interconsultation; pediatrics; primary health care; remote consultation; rural health services; telemedicine; video consultation

Introduction

Background

In Catalonia, children and adolescents represent nearly 20% of the total population. These life stages are critical for development and have specific health needs and challenges. Health interventions during childhood and adolescence have both short- and long-term effects on adult health. Therefore, prevention, health promotion, and equitable access to high-quality pediatric care are essential priorities for the health system [1]. Ensuring appropriate care requires access to a pediatric referral team composed of pediatricians and specialized nurses [2].

When primary care teams include professionals with formal pediatric training, clinical practice becomes more efficient and better aligned with children's needs. Appropriate prescribing improves (particularly of antibiotics), vaccination coverage increases, and unnecessary diagnostic tests and specialist referrals decrease [3]. Similarly, pediatric-trained nurses play a key role in primary care by promoting child health within the community and schools, supporting the management of pediatric demand, and contributing to improved overall quality of care [4].

However, many regions, particularly rural and underserved areas, face a shortage of pediatric specialists, which challenges the continuity and quality of pediatric health care by limiting timely access to diagnosis and treatment. According to the Catalan Pediatrics Society, all primary care pediatric positions in Catalonia are currently filled; however, more than one-third are occupied by general practitioners who, although not pediatric specialists, provide pediatric care in primary care settings [5]. This proportion has increased in recent years, and the geographic distribution of pediatric providers remains uneven, making recruitment especially difficult in rural areas [6]. Similar patterns have been reported in other European countries and in the United States [7,8].

In this context, digital health tools have emerged as promising strategies to support primary care teams, improve access to pediatric expertise, and reduce disparities between urban and rural populations.

In Catalonia, the primary care system already incorporates information and communication technologies to enhance communication between patients and health care professionals, including *eConsulta*—an asynchronous teleconsultation platform—complementing telephone consultations [9]. The use of video consultations, however, remains limited despite a temporary increase during the COVID-19 pandemic. Although still rarely used in routine primary care, this experience has prompted renewed interest in exploring their potential applications in daily clinical practice [10].

Simple digital devices can now be integrated into video consultations to enable remote physical examinations using a digital camera, video otoscope, and digital stethoscope [11]. Initially designed for home use by caregivers, these devices allow pediatricians to receive real-time clinical information through a virtual connection. Similar telemedicine solutions are already implemented in several European countries and the United States, mainly in private health care settings [12,13].

When combined with these digital tools, video consultations can facilitate real-time collaboration between health care professionals. In this model, a nurse physically present with the patient performs the remote examination under the pediatrician's guidance, allowing both history taking and a basic remote physical examination to be conducted synchronously. As this interaction occurs between health care professionals, it is referred to as a video interconsultation.

This approach is particularly suitable for acute pediatric cases—commonly referred to as same-day or urgent visits—that require pediatric assessment within 48 hours. Implementing this model could enhance access to pediatric care in remote areas, promote territorial equity, and reduce unnecessary emergency department referrals.

However, the limited evidence on the use of digital tools such as video interconsultation in pediatric primary care highlights the need to develop and evaluate new technology-integrated models of care, especially in rural areas [14].

Objectives

This study aimed to evaluate the feasibility of a synchronous acute pediatric video interconsultation model that integrates a remote physical examination and is conducted between health care professionals, one of whom is physically present with the patient, in the rural primary care setting of Catalonia.

The study also sought to assess diagnostic adequacy compared with in-person visits and satisfaction among users and health care professionals, considering quality of care, patient safety, and key influencing factors such as reason for consultation, patient age, and visit duration.

This pediatric video interconsultation model is hypothesized to be a feasible, diagnostically adequate, and well-accepted approach in rural primary care settings.

Methods

Study Protocol

The study protocol has been published in a separate publication [15].

Study Design

This was a prospective observational diagnostic accuracy study conducted in a real-world primary care setting, without modification of routine clinical practice. Each participant underwent both the index test (video interconsultation) and the reference standard (in-person visit) during the same clinical encounter, enabling a within-participant comparison of diagnostic performance.

Blinding of participants, clinicians, or outcome assessors was not feasible due to the design: the same pediatrician performed both evaluations sequentially in the same visit, and therefore, all parties were aware of the modality used. As the intervention was limited to a single synchronous video consultation per episode, the study design did not involve repeated use or longitudinal tracking of digital engagement.

Standard in-person pediatric assessment was selected as the reference comparator because it represents current clinical practice and the diagnostic gold standard in primary care.

The study follows the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines, incorporates relevant STARD (Standards for Reporting Diagnostic Accuracy Studies) principles, and adheres to CONSORT-EHEALTH recommendations, an extension of the CONSORT (Consolidated Standards of Reporting Trials) statement for reporting digital health interventions [16-18].

Setting and Period

The study was conducted within the primary care network of the Central Catalonia Health Region (Institut Català de la Salut) at the Cardona Primary Care Center by the pediatric care team. This rural area provides services to approximately 5000 residents, including approximately 800 children aged 0 to 14 years, with a population density of 68 inhabitants per square kilometer.

Data collection took place from June 7, 2023, to January 22, 2025. Diagnostic confirmation occurred immediately after the video consultation through the in-person reassessment; therefore, no additional follow-up was required.

No technological or operational changes occurred during the study that could have affected the feasibility or diagnostic performance of video interconsultations.

Participants

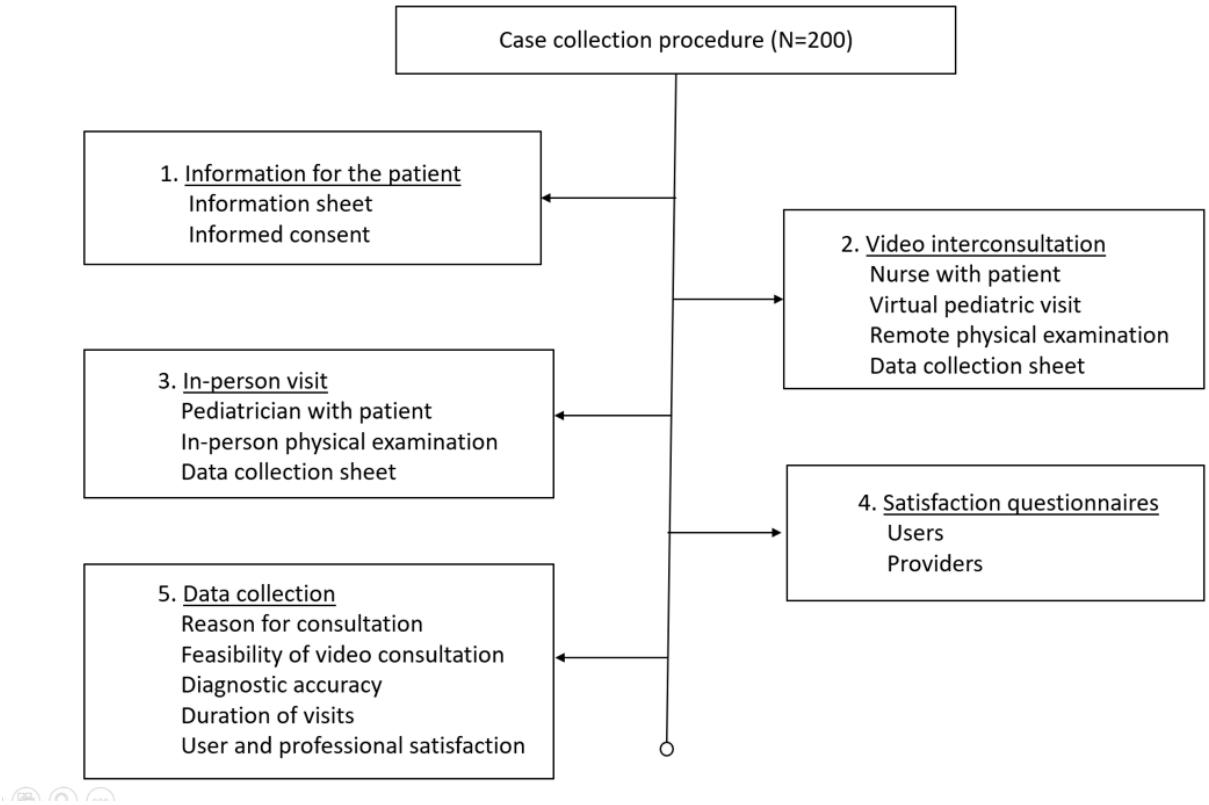
Eligibility criteria included children aged 0 to 14 years presenting with acute conditions requiring care within 48 hours. Acute illness was defined as a condition requiring medical attention within 48 hours. Parental informed consent was mandatory.

The exclusion criteria were routine checkups, chronic condition follow-ups, emergencies requiring immediate in-person care, cases that could be managed autonomously by nursing staff, and an absence of informed consent.

Sample Size and Sampling

A convenience sample of all eligible cases was included. To ensure adequate power for the main objective, the minimum required sample size was estimated at 170 cases to assess the feasibility and diagnostic concordance of video interconsultations compared with in-person visits. The calculation was conducted using the GRANMO-DATARUS online tool, with a 95% CI, an 8% margin of error, and an anticipated 10% dropout rate [19]. A total of 200 cases were recruited. Participant flow is shown in Figure 1.

Figure 1. Sequential case collection process from recruitment and video interconsultation to in-person assessment, satisfaction evaluation, and data recording.



Digital Health Intervention and Procedures

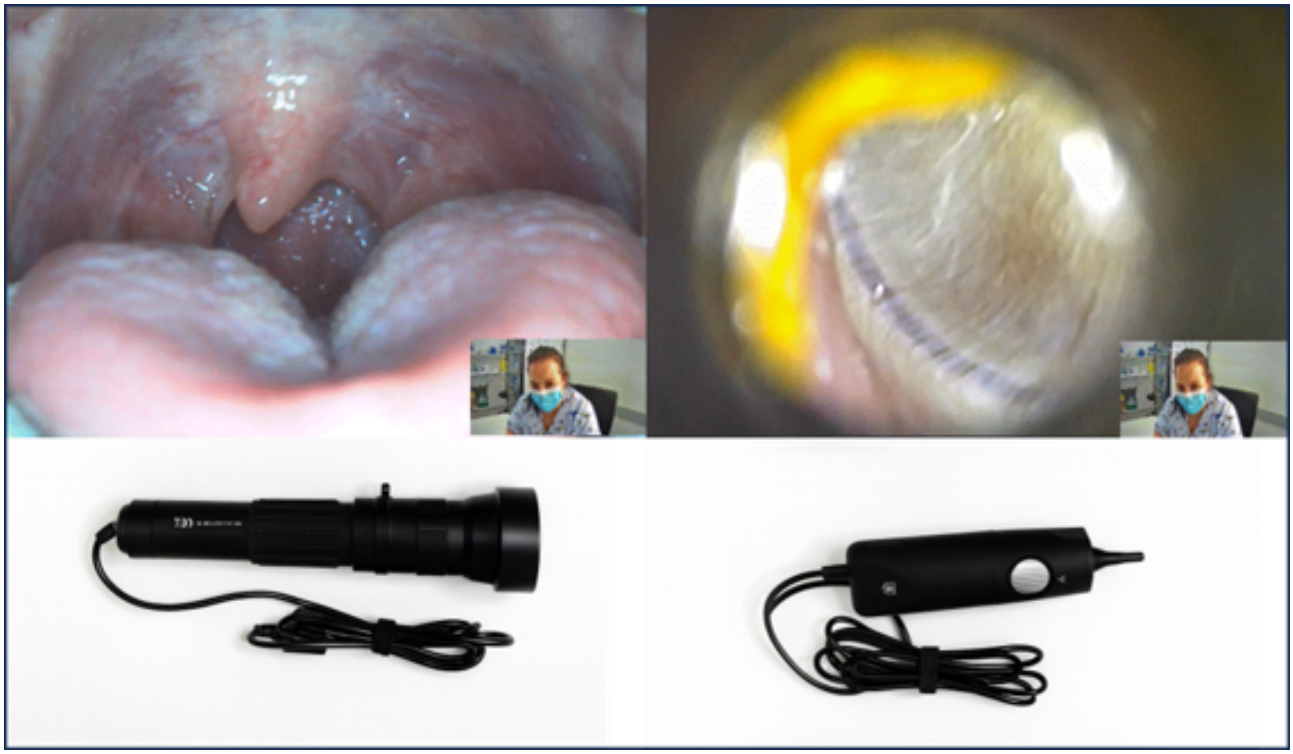
Information and Consent

Families were informed in person by the pediatric nurse during the visit, before enrollment.

Video Interconsultation (Index Test)

A secure, encrypted Microsoft Teams video call connected the pediatrician remotely with the onsite nurse and patient. With the pediatrician's guidance, the nurse performed a physical examination using a digital camera, a video otoscope, and a digital stethoscope. A remote diagnostic impression was recorded. Example images are shown in Figure 2.

Figure 2. Images of the pharynx and tympanic membrane captured during the remote physical examination and the certified digital camera and digital video otoscope used in the study.



In-Person Reassessment (Reference Standard)

Immediately afterward, the same pediatrician performed a face-to-face evaluation and confirmed the final diagnosis.

Satisfaction Assessment

At the end of the visit, two brief questionnaires were administered: one for users (patients and families) and one for health care professionals (categorized as receiving pediatricians, assistant nurses, or observers).

Data Collection

Data were collected using Microsoft 365 Forms via encrypted institutional accounts and stored on secure health services servers. All participants were exposed; no comparison group was included. As this trial was conducted in a rural primary care environment, all video consultations were mediated by trained health care staff.

A total of three data collection tools were used: (1) a clinical form, (2) a professional satisfaction questionnaire (validated Catalan version of the Health Optimum Telemedicine Acceptance Questionnaire) [20], and (3) a user satisfaction questionnaire (adapted from the Northern Saskatchewan Telehealth Network) [21].

No privacy breaches or adverse events occurred. Technical issues were managed by redirecting participants to an in-person assessment when needed.

Variables

The following variables were collected to describe the characteristics of the sample and to assess the feasibility, diagnostic adequacy, and satisfaction associated with pediatric video consultations:

- Sociodemographic variables
  - Age: grouped into five 3-year intervals
  - Sex: female, male, or nonbinary
- Clinical variables
  - Reason for consultation: recorded individually and categorized by affected system or area (respiratory; otorhinolaryngology or ear, nose, and throat; gastrointestinal; infectious diseases; dermatology; musculoskeletal; ocular; and other)
  - Diagnosis: recorded individually and categorized using the same classification
- Feasibility variables
  - Feasibility: feasible or infeasible based on the ability to establish a safe and appropriate diagnosis. Patient safety indicators are included in feasibility, defined as the ability to reach a correct diagnosis without causing harm.
  - Duration of both consultations (in minutes)
  - Limiting factors of nonfeasibility: classified into 6 groups (need for in-person physical examination, telematic auscultation difficulty, camera visualization difficulty, video otoscope visualization difficulty, urgent demand, and lack of patient cooperation)
- Diagnostic adequacy variables
  - Diagnostic concordance: correct or incorrect, using the in-person visit as the gold standard
  - Sensitivity, specificity, and accuracy of the video interconsultation

- Satisfaction variables
- - Professional satisfaction score
  - User satisfaction score

Outcomes Measures

The primary outcomes were the feasibility and diagnostic accuracy of the pediatric video interconsultation model compared with in-person visits, as well as satisfaction levels among health care professionals and users.

The secondary outcomes included factors limiting feasibility (reason for consultation, patient age, and visit duration) and potential barriers identified during the implementation process.

Statistical Analysis

Categorical variables were summarized using absolute frequencies and percentages, whereas continuous variables were described using means and SDs. Associations between categorical variables were analyzed using the Pearson chi-square test or Fisher exact test when expected cell counts were less than 5. For continuous variables, comparisons were made using the *t* test or, when normality assumptions were not met, the Mann-Whitney *U* test.

Diagnostic performance of video consultations was evaluated using sensitivity, specificity, and overall diagnostic accuracy, considering the in-person assessment as the reference standard. CIs for accuracy estimates were calculated using the Wilson method. Diagnostic concordance was assessed using the Cohen κ coefficient, and Gwet's first-order agreement coefficient (AC1) was additionally computed to account for potential prevalence and bias effects. The binomial test was used to analyze the type and direction of diagnostic disagreements. When cell counts were small, estimates were interpreted with caution due to limited statistical power.

All estimates were reported with 95% CIs, and statistical significance was set at $P<.05$. Statistical analyses were performed using R software (version 4.0.3; R Foundation for Statistical Computing).

Ethical Considerations

This study was approved by the Ethical Committee for Research in Medicines of the Jordi Gol i Gurina Primary Care Research Institute (Barcelona, Spain; registration number 22/236-P; March 8, 2023).

As participants were minors, written informed consent from their parents or legal guardians was mandatory. Families received both oral and written information about the study at the time of the visit, before providing consent. The information included the study purpose, procedures (first, a video consultation performed with the pediatric nurse onsite while the pediatrician participated remotely, followed by a conventional in-person visit with the same pediatrician), eligibility criteria, potential risks, confidentiality, and lawful data protection. Participation was voluntary, and withdrawal was possible at any time without consequences for clinical care. Contact details of the principal investigator were provided to address any present or future questions. No images or videos were recorded during the video consultations, except in isolated cases where photographs of clinical findings were taken with explicit informed consent and used exclusively for research or training purposes, ensuring that patients could not be identified. Participants did not receive any financial or other form of compensation for participation in the study.

The telemedicine assessment was supplementary to standard in-person clinical care, ensuring that no diagnostic or therapeutic decisions relied solely on the digital evaluation. Video calls were conducted using secure and encrypted systems to ensure the confidentiality of clinical information.

All researchers signed a confidentiality agreement concerning the treatment and use of study data. No direct personal identifiers were collected, and all data were pseudonymized and processed confidentially. Access to the data was restricted exclusively to the research team. The project database is hosted on secure servers of the Primary Care Management and Community of the Catalan Health Institute (Institut Català de la Salut), which acts as the data processor. Data retention is planned for 10 years, and no international data transfers are anticipated.

The research team will only use the coded database for scientific purposes (eg, journal articles, scientific reports, and book chapters). The study was conducted in full compliance with the ethical principles of the Declaration of Helsinki (1964) and its latest amendment (Fortaleza, 2013), as well as with the European General Data Protection Regulation (GDPR EU 2016/679) and Spanish Organic Law 3/2018 on the Protection of Personal Data and Guarantee of Digital Rights.

Results

Participant Characteristics

A total of 200 pediatric video interconsultations for acute conditions were conducted in rural primary care. All diagnosis-related data were complete for all included cases; no missing diagnostic data were present. The sample distribution by age and sex, along with the main reasons for consultation, is summarized in Table 1. These reasons are provided both individually and grouped by system to facilitate analysis: respiratory, otorhinolaryngology or ENT, gastrointestinal, infectious diseases, dermatology, trauma, ocular, and other. The most prevalent reasons for consultation were respiratory, otorhinolaryngologic, and dermatologic conditions, with cough, earache, and skin lesions being the most frequent symptoms, followed in frequency by fever and odynophagia. Specific reasons for consultation for each organ system are provided in Multimedia Appendix 1.

Table . Characteristics of the sample (N=200).

Characteristics	Values
Patient sex, n (%)	
Female	98 (49)
Male	102 (51)
Patient age (y), n (%)	
0 - 2	38 (19)
3 - 5	44 (22)
6 - 8	48 (24)
9 - 11	37 (18.5)
12 - 14	33 (16.5)
Grouped consultation reasons, n (%)	
ENT[a]	57 (28.5)
Respiratory	55 (27.5)
Dermatology	23 (11.5)
Infectious	21 (10.5)
Trauma	13 (6.5)
Gastrointestinal	11 (5.5)
Ocular	7 (3.5)
Other	13 (6.5)
Duration (min[b]), mean (SD)	
Video interconsultation	7.13 (3.85)
In-person visit	3.96 (1.57)
Feasibility video interconsultation, n (%)	
Feasible	129 (64.5)
Infeasible	71 (35.5)

[a]ENT: ear, nose, and throat.

[b]*P* value <.001 based on an independent samples Student *t* test.

Regarding consultation duration, video interconsultations had a significantly longer mean duration (*P*<.001) of 7.13 (SD 3.85) minutes compared with in-person visits, which had a mean duration of 3.96 (SD 1.57) minutes.

Feasibility

Video interconsultation was feasible in 129 (64.5%) of the 200 cases. In these visits, the video interconsultation could be completed appropriately, providing the necessary data to issue a reliable diagnosis while maintaining quality of care and patient safety.

In 71 (35.5%) cases, video interconsultation was not feasible. The causes were analyzed, and the most frequent cause of infeasibility was the inability to perform a complete physical examination electronically, requiring redirection to an in-person visit (27/71, 38%). Other limitations included difficulties in interpreting online auscultation (18/71, 25.4%), problems viewing images obtained with the digital camera (12/71, 16.9%), and problems with the video otoscope (10/71, 14.1%). In 5 (7%) cases, the consultation was urgent and could not be completed via video, and in 2 (2.8%) cases, the patient's lack of cooperation prevented completion. In 3 (4.2%) cases, there were combined technical difficulties, with simultaneous problems in interpreting images from both the digital camera and the video otoscope or in the quality of auscultation through the electronic stethoscope (Table 2).

Table . Reasons for infeasibility of video interconsultations and corresponding frequencies (N=71).

Reasons for video interconsultation infeasibility[a]	Frequency, n (%)
Need for an in-person physical examination	27 (38)
Difficulty with remote auscultation	18 (25)
Limited visibility through digital camera	12 (17)
Limited visibility through video otoscope	10 (14)
Urgent consultation required	5 (7)
Lack of patient cooperation	2 (3)

[a]Multiple reasons were reported in 3 cases.

A bivariate analysis was conducted to identify demographic and clinical variables potentially associated with the feasibility of video interconsultations.

Categorical variables, including patient sex, age group, diagnostic adequacy, and the clinical category of the consultation reason, were compared between feasible and nonfeasible cases using the chi-square test. A *P* value <.05 was considered statistically significant. No significant associations were found between feasibility and the patient's sex or age group.

However, a statistically significant relationship was observed between feasibility and the clinical category of the consultation reason. Consultations related to gastrointestinal, musculoskeletal, and other conditions showed a higher proportion of nonfeasible cases, whereas those for dermatologic, ocular, and otorhinolaryngologic conditions demonstrated a higher proportion of feasible cases.

Diagnostic adequacy was also significantly associated with feasibility, as all cases classified as feasible presented correct diagnostic agreement (*P*<.001; Table 3).

Table . Bivariate analysis of main variables by feasibility of telemedicine visits.

Variables	Infeasible (n=71)	Feasible (n=129)	*P* value[a]
Patient sex, n (%)			.21
Male	30 (42.3)	68 (52.7)	
Female	41 (57.7)	61 (47.3)	
Patient age (y), n (%)			.19
0 - 2	15 (21.1)	23 (17.8)	
3 - 5	9 (12.7)	35 (27.1)	
6 - 8	20 (28.2)	28 (21.7)	
9 - 11	13 (18.3)	24 (18.6)	
12 - 14	14 (19.7)	19 (14.7)	
Diagnostic agreement, n (%)			<.001
Correct	35 (49.3)	129 (100)	
Incorrect	36 (50.7)	0 (0)	
Reason for consultation, n (%)			<.001
Respiratory	20 (28.2)	35 (27.1)	
ENT[b]	12 (16.9)	45 (34.9)	
Gastrointestinal	10 (14.1)	1 (0.8)	
Trauma	9 (12.7)	4 (3.1)	
Other	9 (12.7)	4 (3.1)	
Infectious	8 (11.3)	13 (10.1)	
Dermatology	3 (4.2)	20 (15.5)	
Ocular	0 (0)	7 (5.4)	

[a]*P* values calculated using the χ^2 test.

[b]ENT: ear, nose, and throat.

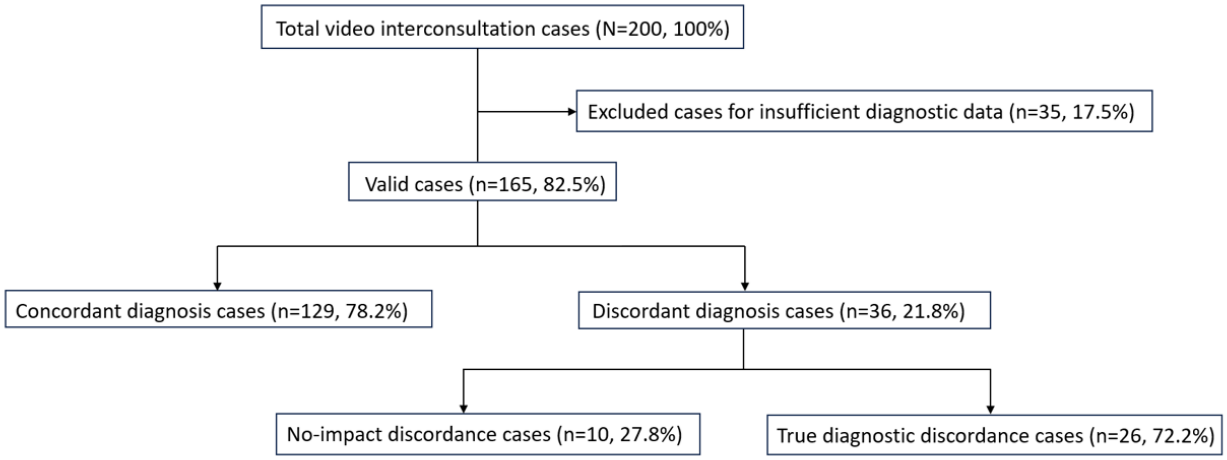
Accuracy

Regarding diagnostic accuracy, diagnoses were grouped by system, using the same eight categories as the reasons for consultation: respiratory, otorhinolaryngology, gastrointestinal, infectious diseases, dermatology, trauma, ocular, and other. In all 129 cases in which video interconsultation was feasible, diagnostic concordance with the in-person visit was observed, as this was considered inherent to the concept of feasibility.

In contrast, among the 71 nonfeasible cases, diagnostic discrepancies between the video interconsultation and the subsequent in-person visit were identified in 36 instances. To calculate diagnostic agreement, 165 valid cases were included, as 35 cases were excluded because of insufficient data for diagnostic evaluation. Diagnostic concordance between the video interconsultation and the in-person assessment was observed in 129 (78.2%) of 165 feasible cases. In the remaining 36 (21.8%) cases, the diagnoses differed between the 2 assessment modalities. Figure 3 shows the flow of participants through the study and the diagnostic concordance analysis.

Cohen κ coefficient for telematic–in-person diagnostic concordance was 0.36, indicating fair agreement. However, given the unbalanced distribution of diagnostic categories, Gwet's AC1 coefficient was also computed, yielding a value of 0.67, which indicates substantial agreement, according to the Landis and Koch scale.

Figure 3. Flow of participants throughout the study according to diagnostic agreement analysis. The different percentages for all subgroups are indicated in relation to the corresponding group.



Among the 36 (21.8%; total 165) cases of showing a diagnostic discrepancy between the telematic and in-person assessments, 10 (27.8%) cases were attributed to the need for a complete physical examination to establish an accurate diagnosis. The distribution of these cases was as follows: 5 involved the musculoskeletal system, 3 were related to the gastrointestinal system, 1 was related to infectious diseases, and 1 was related to other conditions.

Of the remaining 26 (72.2%) discordant cases, the type of diagnostic discrepancy and its distribution by organ system and specific diagnosis were analyzed (Table 4). Most discrepancies were underdiagnoses (23/26, 88.5%), while 5 (19.2%) cases represented overdiagnoses. The exact binomial test showed that the proportion of underdiagnoses was significantly greater than 50% (95% CI 69.8% - 97.6; P<.001).

Table . Types and frequencies of diagnostic discrepancies between video interconsultations and in-person visits (n=26).

Discrepancy type (video interconsultation vs in person)	Cases[a] n (%)
Underdiagnosis of respiratory conditions	
Bronchospasm	6 (23.1)
Respiratory superinfection	4 (15.4)
Bronchiolitis	1 (3.8)
Overdiagnosis of respiratory conditions	
Bronchospasm	2 (7.7)
Underdiagnosis of ENT[b] conditions	
Acute otitis media	3 (11.5)
Herpangina	3 (11.5)
Otitis externa	2 (7.7)
Streptococcal pharyngitis	1 (3.8)
Dental abscess	1 (3.8)
Overdiagnosis of ENT conditions	
Acute otitis media	3 (11.5)
Underdiagnosis of dermatology conditions	
Atopic dermatitis	1 (3.8)
Scarlet fever	1 (3.8)

[a]One case involved both underdiagnosis in the respiratory system and overdiagnosis in ENT. Another case showed both underdiagnosis and overdiagnosis within ENT.

[b]ENT: ear, nose, and throat.

Of the 23 (88.5%; total 26) underdiagnosed cases in the video interconsultation assessment, 11 (%) corresponded to the respiratory system, 10 (%) corresponded to otorhinolaryngology, and 2 (%) corresponded to dermatology. The distribution of underdiagnoses by organ system did not differ significantly from a uniform distribution (Fisher exact test, P=.08).

Regarding overdiagnosis, of the 5 (19.2%) detected cases, 2 (%) involved the respiratory system, and 3 (%) involved the otorhinolaryngology system. Given the small sample size, no specific pattern could be confirmed or ruled out, and the Fisher exact test also showed no significant deviation from a uniform distribution (P≈.6).

In 2 cases, both underdiagnosis and overdiagnosis occurred simultaneously and were therefore classified in both categories. In 1 case, overdiagnosis involved an otorhinolaryngologic condition and underdiagnosis a respiratory one; in the other, both diagnoses were within otorhinolaryngology, where acute otitis media was incorrectly diagnosed instead of otitis externa during the video interconsultation.

The accuracy, sensitivity, and specificity of each diagnosis obtained via video interconsultation were estimated using the in-person diagnosis as the gold standard. Importantly, the mean diagnostic accuracy of video interconsultations across all conditions was 0.99 (95% CI 0.98 - 1.00). The mean overall specificity was 0.99 (95% CI 0.98 - 1.00), and the mean overall sensitivity was 0.90 (95% CI 0.84 - 0.95).

The diagnostic performance metrics for each clinical category, organized by organ system, are summarized in Table 5. A high level of diagnostic agreement was observed between video and in-person consultations across all categories, with accuracy values exceeding 0.92 and ranging from 0.93 to 1.00. Video interconsultations demonstrated the best performance for otorhinolaryngologic, dermatologic, and trauma-related conditions, which also showed high sensitivity and specificity. In contrast, respiratory diagnoses had lower sensitivity (0.68, 95% CI 0.51 - 0.82), suggesting a potential risk of diagnostic underestimation during video interconsultations. Detailed diagnostic performance metrics by specific condition, along with the corresponding frequency distribution, are provided in Multimedia Appendix 2. CIs should be interpreted with caution in categories with small sample sizes.

Table . Diagnostic frequency by system in video and in-person consultations, with corresponding accuracy, sensitivity, and specificity (N=200).

Organ system	Video visit, n (%)	In-person visit, n (%)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
ENT[a]	106 (53)	96 (48)	0.92 (0.87 - 0.95)	0.97 (0.91 - 0.99)	0.88 (0.80 - 0.93)
Respiratory	28 (14)	38 (19)	0.93 (0.89 - 0.96)	0.68 (0.51 - 0.82)	0.99 (0.96 - 1.00)
Dermatology	21 (10.5)	22 (11)	0.99 (0.97 - 1.00)	0.95 (0.77 - 1.00)	1.00 (0.98 - 1.00)
Trauma	12 (6)	13 (6.5)	0.99 (0.97 - 1.00)	0.92 (0.64 - 1.00)	1.00 (0.98 - 1.00)
Gastrointestinal	10 (5)	9 (4.5)	0.99 (0.97 - 1.00)	1.00 (0.66 - 1.00)	0.99 (0.97 - 1.00)
Other	10 (5)	9 (4.5)	0.98 (0.96 - 1.00)	0.89 (0.52 - 1.00)	0.99 (0.96 - 1.00)
Infectious	7 (3.5)	7 (3.5)	0.99 (0.96 - 1.00)	0.86 (0.42 - 1.00)	0.99 (0.97 - 1.00)
Ocular	7 (3.5)	7 (3.5)	1.00 (0.98 - 1.00)	1.00 (0.59 - 1.00)	1.00 (0.98 - 1.00)

[a]ENT: ear, nose, and throat.

Professional Satisfaction

Professional satisfaction, assessed through the validated Catalan version of the Health Optimum Telemedicine Acceptance Questionnaire, was analyzed by comparing responses across three professional groups—receivers (pediatricians), assistants (nurses), and observers (residents and students)—using the chi-square tests (Table 6).

In terms of perceived quality, 74% (148/200) of professionals rated the video interconsultation as equal to or higher than in-person care, whereas 26% (52/200) perceived it as lower. This assessment differed significantly among professional groups (P<.001). Nursing assistants expressed the most favorable opinions (42/47, 89.4%, positive ratings), similar to observers (41/46, 89.1%), whereas receivers (pediatricians) were less favorable, with 50.5% (54/107) positive and 43% (46/107) negative ratings.

Overall, 94.5% (189/200) of participants believed that telemedicine could have a positive impact on patient health, 4.5% (9/200) thought it had no influence, and 1% (2/200) considered that it might worsen patient outcomes. When analyzed by professional group, receivers reported in 99.1% (106/107) of visits that telemedicine could improve patient health, compared with 78.7% (37/47) among assistants and 100% (200/200) among observers. Among assistants, 4.3% (2/47) indicated that telemedicine could negatively affect patients' health.

With respect to the continuity of telemedicine use, 83.4% (166/199) of professionals indicated that improvements were needed in infrastructure or organization. Receivers were particularly likely to request such improvements (96/107, 89.7%), compared with assistants (15/47, 31.9%) and observers (7/45, 15.6%; one observer did not provide a response to this item).

Differences among professional groups were statistically significant for both perceived quality (P<.001) and willingness to continue using telemedicine (P=.004).

Table . Health care professionals' satisfaction with telemedicine and video interconsultation.

Opinion professionals	Total (N=200), n (%)	Assistant (nurse; n=47), n (%)	Receiver (pediatrician; n=107), n (%)	Observer[c] (other professionals[a]; n=46), n (%)	*P* value[b]
Perceived quality					<.001
Very good	63 (31.5)	25 (53.2)	12 (11.2)	26 (56.5)	
Good	74 (37)	17 (36.2)	42 (39.3)	15 (32.6)	
Fair	11 (5.5)	1 (2.1)	7 (6.5)	3 (6.5)	
Poor	35 (17.5)	3 (6.4)	31 (29)	1 (2.2)	
Very poor	17 (8.5)	1 (2.1)	15 (14)	1 (2.2)	
Perceived impact					<.001
No	9 (4.5)	8 (17)	1 (0.9)	0 (0)	
Yes, positive	189 (94.5)	37 (78.7)	106 (99.1)	46 (100)	
Yes, negative	2 (1)	2 (4.3)	0 (0)	0 (0)	
Intent to continue[c]					.004
Unchanged	33 (16.6)	15 (31.9)	11 (10.3)	7 (15.6)	
Improved	166 (83.4)	32 (68.1)	96 (89.7)	38 (84.4)	

[a]Other professionals included residents and students.

[b]*P* values were calculated using the χ^2 test.

[c]One observer did not provide a response to this item.

User Satisfaction

User experience, measured using a survey adapted from the Telehealth Network Questionnaire (Northern Saskatchewan Telehealth Network) and completed by the accompanying adults responsible for the child, was rated as very good in 74.5% (149/200) of the 200 cases and good in 20.5% (41/200). A total of 4.5% (9/200) of participants described the experience as poor, and 0.5% (1/200) of participants described the experience as very poor (Table 7).

In addition, 92% (184/200) of families reported that they would be willing to repeat the video interconsultation in the future (142/200, 71%, very likely, and 42/200, 21%, likely), whereas only 8% (16/200) expressed reluctance (12/200, 6%, unlikely and 4/200, 2%, very unlikely).

To explore whether these perceptions were shared by health care professionals, a correlation analysis between user and professional satisfaction for each visit revealed a weak but statistically significant positive correlation (r=0.182, 95% CI 0.044 - 0.312; P=.009).

Table . Users' satisfaction with telemedicine and video interconsultation (N=200).

Opinion users	Total, n (%)
Overall quality	
Very good	149 (74.5)
Good	41 (20.5)
Poor	9 (4.5)
Very poor	1 (0.5)
Willingness to repeat	
Very likely	142 (71)
Likely	42 (21)
Unlikely	12 (6)
Very unlikely	4 (2)

Discussion

Principal Findings

This study evaluated whether a synchronous video consultation model—including remote physical examination guided by a pediatrician and supported by an onsite pediatric nurse—is a feasible, clinically appropriate, and well-accepted approach to managing acute pediatric conditions in rural primary care in Catalonia.

Feasibility was achieved in nearly two-thirds of cases, allowing completion of a safe and adequate remote assessment. Among feasible cases, diagnostic performance was high, with substantial agreement between telemedicine and in-person evaluations. Video consultations also demonstrated high diagnostic accuracy and sensitivity, along with near-optimal specificity. However, performance varied by clinical condition, suggesting that limitations inherent to remote assessment (eg, reduced ability to detect subtle clinical signs) may reduce sensitivity in certain scenarios.

Process outcomes further supported implementation: satisfaction was very high among families and positive among health care professionals, and consultation duration remained acceptable despite the inclusion of a remote physical examination.

Together, these findings indicate that—when appropriate feasibility and safety criteria are met—this telemedicine model can effectively complement, rather than replace, in-person pediatric care in underserved regions, thereby contributing to improved health equity and accessibility.

Furthermore, these findings could be applicable to other pediatric primary care settings with similar digital infrastructure and staffing resources. However, as this was an observational, single-center study without randomization, these findings should be interpreted with caution regarding their external validity.

Comparison With Prior Work

These results are consistent with the current literature suggesting that telemedicine in pediatrics can achieve high diagnostic validity when applied under well-defined conditions. That said, most previous studies were not conducted in primary care settings, did not address acute conditions, and primarily examined video consultations between professionals and patients, rather than between professionals themselves [22-24]. There is limited evidence on video consultations that incorporate physical examinations. In pediatrics, Wagner et al [25] found that remote physical examination using medical devices similar to those used in this study was comparable to in-person assessment. Their findings align with ours, showing high diagnostic accuracy for otoscopy, oropharyngeal evaluation, and dermatological examination, but lower accuracy for assessing abdominal pathology.

Several studies have found that the effectiveness of video consultations is particularly high for diagnoses that rely primarily on medical history and visual assessment, such as upper respiratory tract infections (eg, pharyngitis and otitis) or dermatologic lesions [26,27]. In contrast, conditions that require a full physical examination, such as abdominal pain, which necessitates palpation, or headache, which may call for a neurological evaluation, may show reduced sensitivity, as also observed in this study [25,28].

This limitation may increase the risk of incomplete or inaccurate diagnoses, as reflected in the 21.8% (36/165) of cases that showed diagnostic discordance between virtual and in-person visits. While most discrepancies were classified as underdiagnoses or overdiagnoses, some involved mixed errors. Several cases required referral to in-person care to complete the physical examination and establish a reliable diagnosis, underscoring the complexity of remote clinical assessment. Furthermore, in some cases classified as infeasible, the online and in-person diagnoses were consistent, and the infeasibility was attributed to other limitations. In pediatric care, particularly for younger children, a comprehensive physical examination is often performed regardless of the presenting complaint. These findings underscore the need for triage and feasibility criteria tailored to the specific characteristics of each patient and condition to optimize the safety and effectiveness of video interconsultations. They also highlight the importance of establishing clear clinical guidelines to ensure the safe, high-quality use of video interconsultations in primary care pediatrics.

Pediatric patients with specialty conditions, such as rheumatologic, cardiologic, or endocrinologic disorders, were not included, as no such cases were seen during the study period. However, a few neurology and gynecology cases were included in the “other” category due to their small number. It is worth noting that telemedicine has also proven useful for the follow-up of pediatric rheumatic diseases, particularly in rural areas and during the COVID-19 pandemic [29].

Regarding consultation duration, a 2022 review by the Catalan Agency for Health Quality and Evaluation reported that video or telephone consultations are typically 1.5 to 4 minutes shorter than in-person visits [30]. In contrast, in this study, video interconsultations lasted nearly twice as long as in-person visits, likely due to the inclusion of a physical examination, which extended the consultation time. Additionally, this model involved 2 health care professionals. While it may enhance accessibility, it also requires greater resource allocation in terms of time and staffing.

Regarding users’ perceptions of telemedicine, the findings of this study indicate a high level of satisfaction, consistent with previous research [31]. The main reasons families expressed concerns about remote interconsultations were technical issues related to sound, connectivity, and image quality, as well as fears that in-person care might be replaced by virtual services. Other reasons, such as a perceived lack of safety, were infrequent. In any case, it is essential to provide families with clear and transparent information so they can make informed decisions about using this technology. Ultimately, they remain at the center of care [32].

With regard to health care professionals’ perspectives on telemedicine, the results indicate that professionals acknowledge the value of pediatric video interconsultations, and most consider them beneficial for patient health. However, differences among

professional groups suggest that perceptions vary depending on each participant's clinical role and expectations.

These findings are consistent with those of Martín-Masot et al [33], who analyzed the views of Spanish pediatricians following the rapid digitalization of health care delivery during the COVID-19 pandemic. In their study, most pediatricians regarded digital consultations as time-efficient and a valuable resource, aligning with the present results. Among nurses, the majority reported that the quality of video consultations was equal to or better than that of in-person visits. These results align with the findings of Navarro-Martínez et al [34], who reported that telenursing is positively perceived in routine clinical practice. However, most health care professionals in this study identified the need to improve the application of telemedicine in clinical settings, particularly in terms of technology, organizational processes, and bioethical considerations. This observation is echoed in a study conducted in Catalonia by Vidal-Alaball et al [35]. Similarly, other studies, such as that by Inoue et al [36] in Japan, have reached the same conclusions. In this context, it is essential to train professionals not only in the use of digital tools, but also in what Finkelstein et al [37] refer to as "webside manner," a more effective approach to online clinical communication.

Furthermore, although telemedicine can improve communication between doctors and patients and help reduce health care costs, it may compromise the quality of care, therapeutic effectiveness, and patient safety if not implemented properly [38]. Therefore, legislation and bioethical frameworks must evolve to accommodate these emerging models of care [39].

Limitations

This study has several limitations. First, there is a potential risk of diagnostic inaccuracy in video consultations compared with in-person visits, especially for conditions that require a direct physical examination, such as abdominal pain, trauma, or headache. Technical issues affecting image or audio quality may also hinder adequate remote assessment. Additionally, some diagnostic subgroups were small, which reduced the precision of accuracy estimates and widened the CIs.

Confirmation bias may have occurred because the same pediatrician conducted both the telemedicine and in-person evaluations. A role bias may also have occurred, as those receiving telemedicine may have different experiences from those delivering or observing it. If recipients represent the majority, as in this case, overall satisfaction may primarily reflect their perspective. Likewise, observers may have overemphasized methodological aspects while underestimating the actual user experience.

Another limitation is a possible social desirability bias, as the project was conducted by the patients' regular and trusted pediatrician and nurse. This relationship of trust may have influenced participants' responses, although no intentional intervention or influence was exerted during data collection.

Operational limitations should also be considered. Video consultations took longer than in-person visits and required the presence of both a pediatrician and a nurse, which may not be feasible in all settings and could increase the workload. Economic, organizational, and technological implications were not evaluated in this study, and the environmental impact of digital health equipment warrants further attention.

Finally, data were collected in a single rural primary care center, which may limit the generalizability to other health care settings. Despite strict data protection measures, concerns about confidentiality and the potential for digital care to reduce the human component of clinical interaction remain relevant considerations.

Future Directions

Future research should focus on defining evidence-based triage and feasibility criteria to select patients who can be safely and effectively managed through video interconsultation. Larger multicenter studies are needed to confirm diagnostic performance and satisfaction outcomes across broader pediatric populations and health care contexts. Evaluating economic, organizational, and environmental sustainability will be essential to inform real-world implementation. In addition, exploring strategies to streamline the workflow, optimize technical reliability, and maintain the human aspects of care will help ensure successful and responsible integration into routine pediatric practice.

Conclusions

The proposed model of synchronous video consultation between health care professionals, including physical examination, has proven to be a feasible option. It shows good diagnostic agreement with in-person visits and has been positively evaluated by both users and health care professionals.

This approach may serve as a valuable tool for managing acute pediatric conditions in rural primary care settings in Catalonia, provided it is implemented appropriately and maintains patient safety and quality of care. Although it cannot replace in-person visits, it can complement them within the ongoing reorganization of pediatric primary care, contributing to improved accessibility, territorial equity, and system efficiency.

The implementation of this model involves several challenges, including longer consultation times, training requirements, the development of standardized protocols, economic and environmental costs, and the management of data confidentiality.

Acknowledgments

This study was conducted within the PhD Program in Medicine and Biomedical Sciences at the University of Vic – Central University of Catalonia. The study also received support through a research release fellowship granted by Institut Català de la

Salut - IDIAP Jordi Gol. The authors would like to express their sincere gratitude to the patients and their families for their altruistic participation.

Funding

This study was supported by a financial grant from Consultori Bayés (Vic), which was used to fund the acquisition of the digital devices used in the project. The authors would also like to express their gratitude to Institut Català de la Salut, who kindly covered the publication fees for this article.

Data Availability

Data will be made available by the corresponding author upon reasonable request.

Authors' Contributions

MCR served as the principal investigator. CFC and JVA contributed the original idea. NSB assisted in case collection. AFC, QMC, and LSR contributed to the statistical analysis and provided support throughout the research process. JVA and FLS acted as supervisors and provided continuous guidance throughout the project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Specific reasons for consultation for each organ system.
[DOCX File, 8 KB - pediatrics_v9i1e82133_app1.docx]

Multimedia Appendix 2
Frequencies and diagnostic performance metrics of video interconsultations by condition, with in-person diagnoses used as the gold standard.
[DOCX File, 12 KB - pediatrics_v9i1e82133_app2.docx]

References

1. Proposta sobre el model d'atenció pediàtrica a l'atenció primària a Catalunya. Societat Catalana de Pediatria Secció d'Atenció Primària. 2022. URL: http://www.scpediatria.cat/adjunts_mails/prop-model-At-pediatrica-AP_01-06-2022.pdf [accessed 2025-05-01]
2. Rodrigo MA, Sanz AC, Pina CS, et al. The role of pediatricians in providing greater-quality care for children: an ongoing debate. J Pediatr 2021;231:303-304. [doi: 10.1016/j.jpeds.2020.12.072] [Medline: 33373669]
3. Aparicio Rodrigo M, Ruiz Canela J, Buñuel Álvarez JC, et al. Paediatricians provide higher quality care to children and adolescents in primary care: a systematic review. Acta Paediatr 2020 Oct;109(10):1989-2007. [doi: 10.1111/apa.15324] [Medline: 32311805]
4. Dall'Oglio I, Rosati GV, Biagioli V, et al. Pediatric nurses in pediatricians' offices: a survey for primary care pediatricians. BMC Fam Pract 2021 Jun 29;22(1):136. [doi: 10.1186/s12875-021-01457-1] [Medline: 34187392]
5. Societat Catalana de Pediatria. Secció d'Atenció Primària. Pediatria d'Atenció primària en fallida. PediatresAP.cat. 2020. URL: https://pediatresap.cat/wp-content/uploads/bsk-pdf-manager/2020/10/Pediatria-dAtencio-Primaria-en-fallida.pdf [accessed 2025-05-01]
6. van Esso D, del Torso S, Hadjipanayis A, Primary-Secondary Working Group (PSWG) of European Academy of Paediatrics (EAP), et al. Paediatric primary care in Europe: variation between countries. Arch Dis Child 2010 Oct;95(10):791-795. [doi: 10.1136/adc.2009.178459] [Medline: 20403821]
7. Neumeier S. Pediatric care proximity in Germany: a comparative study of regional accessibility. KN J Cartogr Geogr Inf 2024 Dec;74(3-4):185-204. [doi: 10.1007/s42489-024-00175-9]
8. Bettenhausen JL, Winterer CM, Colvin JD. Health and poverty of rural children: an under-researched and under-resourced vulnerable population. Acad Pediatr 2021;21(8S):S126-S133. [doi: 10.1016/j.acap.2021.08.001] [Medline: 34740419]
9. Saigí-Rubió F, Vidal-Alaball J, Torrent-Sellens J, et al. Determinants of Catalan public primary care professionals' intention to use digital clinical consultations (eConsulta) in the post-COVID-19 context: optical illusion or permanent transformation? J Med Internet Res 2021 May 31;23(6):e28944. [doi: 10.2196/28944] [Medline: 34097638]
10. Jiménez-Rodríguez D, Santillán García A, Montoro Robles J, Rodríguez Salvador MDM, Muñoz Ronda FJ, Arrogante O. Increase in video consultations during the COVID-19 pandemic: healthcare professionals' perceptions about their implementation and adequate management. Int J Environ Res Public Health 2020 Jul 15;17(14):5112. [doi: 10.3390/ijerph17145112] [Medline: 32679848]
11. Firefly telemedicine. Firefly Global. 2025. URL: https://fireflyglobal.com/ [accessed 2025-12-24]

12. TytoCare for professionals: advanced telemedicine solutions. TytoCare. 2025. URL: https://www.tytocare.com/ [accessed 2025-12-24]
13. Higo: a small medical device and telemedical solution. Higo sense. 2025. URL: https://higosense.com/ [accessed 2025-12-24]
14. Vidal-Alaball J, Descals Singla E. Abordaje de la telemedicina entre proveedores: ejemplos de uso. Atenc Prim Práct 2021 Dec;3(1):100112. [doi: 10.1016/j.appr.2021.100112]
15. Castillo-Rodenas M, Vidal-Alaball J, Solanas-Bacardit N, et al. Feasibility of a pediatric acute video consultation process among health care professionals in primary care in a rural setting: protocol for a prospective validation study. JMIR Res Protoc 2024 Feb 1;13:e52946. [doi: 10.2196/52946] [Medline: 38300693]
16. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008 Apr;61(4):344-349. [doi: 10.1016/j.jclinepi.2007.11.008] [Medline: 18313558]
17. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 2015 Oct 28;351:h5527. [doi: 10.1136/bmj.h5527] [Medline: 26511519]
18. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
19. Marrugat J. Calculadora de grandària mostral GRANMO. DATARUS – Applications for Biomedical Research. 2022. URL: https://www.datarus.eu/aplicaciones/granmo/ [accessed 2025-07-29]
20. Vidal-Alaball J, Flores Mateo G, Garcia Domingo JL, et al. Validation of a short questionnaire to assess healthcare professionals' perceptions of asynchronous telemedicine services: the Catalan version of the health optimum telemedicine acceptance qestionnaire. Int J Environ Res Public Health 2020 Mar 25;17(7):2202. [doi: 10.3390/ijerph17072202] [Medline: 32218310]
21. Telehealth satisfaction questionnaire (TSQ). Agency for Healthcare Research and Quality. URL: https://digital.ahrq.gov/sites/default/files/docs/survey/telehealthsatisfactionquestionnaire_comp.pdf [accessed 2025-07-28]
22. Casey SD, Huang J, Parry DD, Lieu TA, Reed ME. Health care utilization with telemedicine and in-person visits in pediatric primary care. JAMA Health Forum 2024 Nov 1;5(11):e244156. [doi: 10.1001/jamahealthforum.2024.4156] [Medline: 39576615]
23. Southgate G, Yassaee AA, Harmer MJ, Livesey H, Pryde K, Roland D. Use of telemedicine in pediatric services for 4 representative clinical conditions: scoping review. J Med Internet Res 2022 Oct 26;24(10):e38267. [doi: 10.2196/38267] [Medline: 36287609]
24. Pathak PR, Stockwell MS, Lane MM, et al. Access to primary care telemedicine and visit characterization in a pediatric, low-income, primarily Latino population: retrospective study. JMIR Pediatr Parent 2024 Dec 17;7:e57702. [doi: 10.2196/57702] [Medline: 39689903]
25. Wagner R, Lima TC, Silva MD, et al. Assessment of pediatric telemedicine using remote physical examinations with a mobile medical device: a nonrandomized controlled trial. JAMA Netw Open 2023 Feb 1;6(2):e2252570. [doi: 10.1001/jamanetworkopen.2022.52570] [Medline: 36729459]
26. Gao Y, Magin P, Tapley A, et al. Prevalence of antibiotic prescribing for aute respiratory tract infection in telehealth versus face-to-face consultations: cross-sectional analysis of general practice registrars' clinical practice. J Med Internet Res 2025 Mar 13;27:e60831. [doi: 10.2196/60831] [Medline: 40080812]
27. Bittmann S, Moschüring-Alieva E, Bittmann L, Luchter E, Villalon G. Preliminary results of a telemedicine questionnaire in pediatrics as an innovative new tool to diagnose and treat children virtually in an ambulatory setting: analysis of 400 pediatric consultations. J Regen Biol Med 2021;3(6):1-3. [doi: 10.37191/Mapsci-2582-385X-3(6)-098]
28. García Ron A, Arias Vivas E, Martínez del Río C, et al. Usefulness of telemedicine in urgent pediatric care during the COVID-19 pandemic [article in Spanish]. Rev Pediatr Aten Primaria 2022;24(93):23-29 [FREE Full text]
29. Mukusheva Z, Assylbekova M, Poddighe D. Management of pediatric rheumatic patients in Kazakhstan during the coronavirus disease 2019 (COVID-19) pandemic. Rheumatol Int 2020 Aug;40(8):1351-1352. [doi: 10.1007/s00296-020-04613-5] [Medline: 32514608]
30. Moltó-Puigmartí C, Segur-Ferrer J, Berdún Peñato J, et al. Evaluación de la seguridad, eficacia/efectividad y eficiencia de la teleconsulta en atención primaria, y de los aspectos organizativos, éticos, sociales y legales ligados a su uso. Scientia. 2022. URL: https://scientiasalut.gencat.cat/handle/11351/7977?show=full [accessed 2025-02-12]
31. Kodjebacheva GD, Culinski T, Kawser B, Coffer K. Satisfaction with telehealth services compared with nontelehealth services among pediatric patients and their caregivers: systematic review of the literature. JMIR Pediatr Parent 2023 Apr 27;6:e41554 [FREE Full text] [doi: 10.2196/41554] [Medline: 37000504]
32. Carta de drets i deures de la ciutadania en relació amb la salut i l'atenció sanitària. Departament de salut. 2015. URL: https://scientiasalut.gencat.cat/handle/11351/1840 [accessed 2025-05-28]
33. Martín-Masot R, Diaz-Martin JJ, Santamaría-Orleans A, Navas-López VM. Impact of the COVID-19 pandemic on the digitization of routine pediatric practice in Spain: a nationwide survey study. Front Pediatr 2023;11:1106488. [doi: 10.3389/fped.2023.1106488] [Medline: 36865698]

34. Navarro-Martínez O, Martinez-Millana A, Traver V. Use of tele-nursing in primary care: a qualitative study on its negative and positive aspects. Aten Primaria 2024 May;56(5):102843. [doi: 10.1016/j.aprim.2023.102843] [Medline: 38215687]
35. Vidal-Alaball J, López Seguí F, Garcia Domingo JL, et al. Primary care professionals' acceptance of medical record-based, store and forward provider-to-provider telemedicine in Catalonia: results of a web-based survey. Int J Environ Res Public Health 2020 Jun 8;17(11):4092. [doi: 10.3390/ijerph17114092] [Medline: 32521740]
36. Inoue Y, Kishi T, Sato T, et al. Telemedicine for paediatric rheumatic diseases in Japan: a national survey of physicians' perspectives. Mod Rheumatol 2025 Dec 22;36(1):83-89. [doi: 10.1093/mr/roaf075] [Medline: 40794858]
37. Finkelstein JB, Tremblay ES, Van Cain M, et al. Pediatric clinicians' use of telemedicine: qualitative interview study. JMIR Hum Factors 2021 Dec 2;8(4):e29941. [doi: 10.2196/29941] [Medline: 34860669]
38. Zafra RP, Parramon NA, Albiol-Perarnau M, Torres OY. Análisis de retos y dilemas que deberá afrontar la bioética del siglo XXI, en la era de la salud digital. Atención Primaria 2024 Jul;56(7):102901. [doi: 10.1016/j.aprim.2024.102901]
39. Telemedicina: com i quan utilitzar-la en la pràctica assistencial quaderns de la bona praxi. Issuu.: Col·legi de Metges de Barcelona URL: https://issuu.com/comb/docs/bona_praxi_37_telemedicina [accessed 2025-05-14]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by S Badawy; submitted 09.Aug.2025; peer-reviewed by D Poddighe, M Lemonde; revised version received 22.Nov.2025; accepted 27.Nov.2025; published 26.Jan.2026.

Please cite as:
Castillo-Rodenas M, Solanas Bacardit N, Farràs Company C, Miró Catalina Q, Solà Reguant L, Fuster-Casanovas A, López Seguí F, Vidal-Alaball J
Feasibility, Diagnostic Accuracy, and Satisfaction of an Acute Pediatric Video Interconsultation Model in Rural Primary Care in Catalonia: Prospective Observational Study
JMIR Pediatr Parent 2026;9:e82133
URL: https://pediatrics.jmir.org/2026/1/e82133
doi:10.2196/82133

Predicting Infant Sleep Patterns From Postpartum Maternal Mental Health Measures: Machine Learning Approach

Rawan AlSaad[1], PhD; Raghad Burjaq[2], PhD; Majid AlAbdulla[3,4], MD; Alaa Abd-alrazaq[1], PhD; Javaid Sheikh[1], MD; Rajat Thomas[1], PhD

[1]Weill Cornell Medical College in Qatar, 2700 Education City, Doha, Qatar

[2]Women's Wellness and Research Center, Hamad Medical Corporation, Doha, Qatar

[3]Mental Health Services, Hamad Medical Corporation, Doha, Qatar

[4]College of Medicine, Qatar University, Doha, Qatar

Corresponding Author:
Rawan AlSaad, PhD
Weill Cornell Medical College in Qatar, 2700 Education City, Doha, Qatar

Abstract

Background: Postpartum maternal mental health (MMH) symptoms, including depression, anxiety, and childbirth-related post-traumatic stress disorder, are known to influence infant sleep trajectories. While previous research has examined their individual and combined associations, the predictive utility of these MMH symptoms for the early identification of infant sleep problems through machine learning (ML) remains understudied.

Objective: This study aimed to examine whether postpartum MMH measures can predict infant sleep outcomes during the first year of life. The analysis focused on 2 clinically relevant sleep indicators: (1) nocturnal sleep duration and (2) night awakening frequency.

Methods: A total of 409 mother-infant dyads were included in the study. Predictor variables comprised postpartum MMH symptoms assessed between 3 and 12 months postpartum, along with sociodemographic characteristics of mothers and infants. MMH symptoms were measured using 3 validated instruments: the Edinburgh Postnatal Depression Scale, the Hospital Anxiety and Depression Scale, and the City Birth Trauma Scale. Infant sleep outcomes were assessed using the Brief Infant Sleep Questionnaire. Six supervised ML algorithms were evaluated: logistic regression, random forest, support vector classifier, extreme gradient boosting, Light Gradient Boosting Machine, and multilayer perceptron. Post hoc feature importance analyses were conducted to identify the most influential predictors associated with each infant sleep outcome.

Results: All models demonstrated high predictive performance. The best model achieved a precision-recall area under the curve of 0.92, F_1-score of 0.84, and accuracy of 0.88 for predicting short nocturnal sleep duration. For frequent night awakenings, the top precision-recall area under the curve was 0.91, with an F_1-score of 0.78 and accuracy of 0.85. Key predictors included maternal age and total scores from the Edinburgh Postnatal Depression Scale, Hospital Anxiety and Depression–Anxiety subscale, and City Birth Trauma Scale, with individual symptom items offering additional discriminative value.

Conclusions: ML models can accurately predict which infants are at risk for suboptimal sleep based on MMH measures, enabling personalized, responsive, and developmentally informed postpartum care that promotes long-term maternal and infant well-being.

(JMIR Pediatr Parent 2026;9:e78937) doi:10.2196/78937

KEYWORDS

artificial intelligence; sleep; postpartum; mental health; depression; women's health.

Introduction

Infant sleep plays a foundational role in early neurodevelopment, with significant implications for cognitive functioning, emotional regulation, physical growth, and long-term health outcomes [1-3]. During the first year postpartum, infant sleep patterns are highly dynamic and marked by individual variability in both nocturnal sleep duration and the frequency of night awakenings. Insufficient or fragmented sleep during this critical period has been associated with impaired memory consolidation, behavioral dysregulation, and suboptimal emotional development [4,5].

A growing body of research [6-12] has demonstrated that maternal mental health (MMH) symptoms during the postpartum period, including depression [13,14], anxiety [8,15], and childbirth-related post-traumatic stress disorder (CB-PTSD) [16,17], are associated with alterations in infant sleep architecture. However, the underlying mechanisms through which these maternal conditions influence infant sleep remain

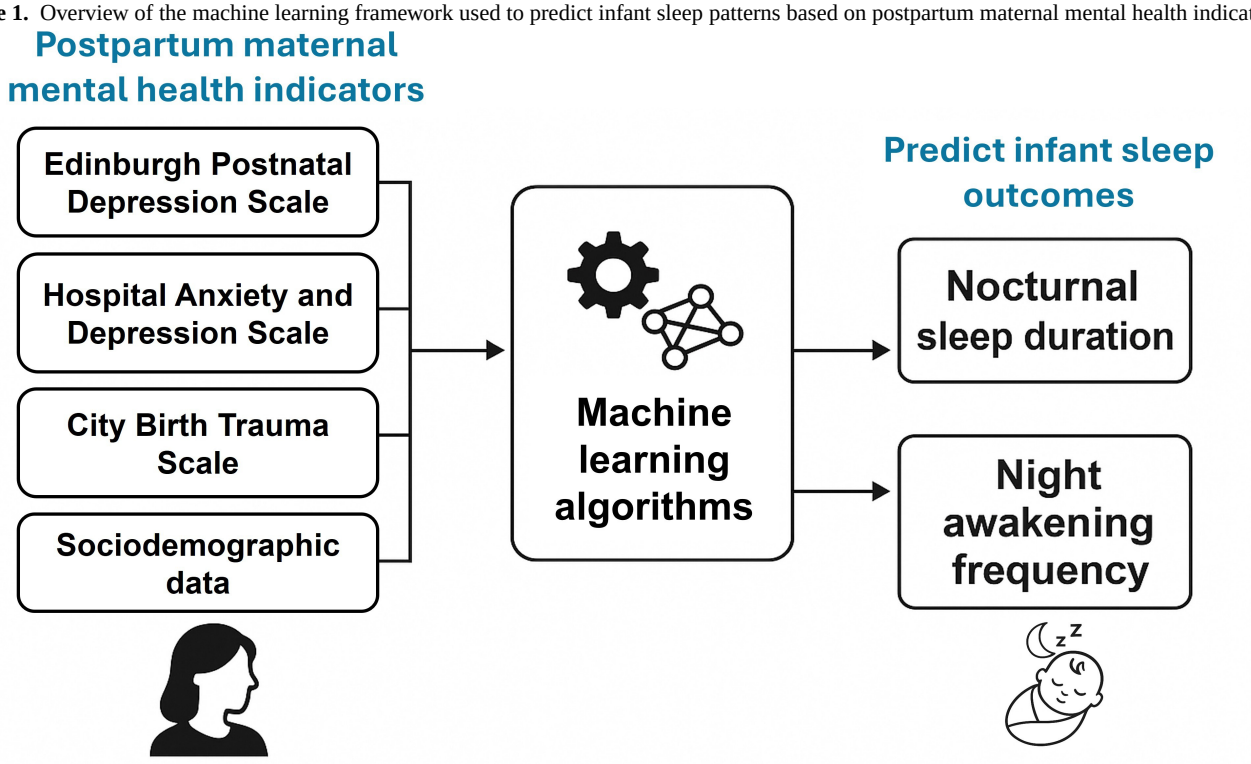
poorly understood [18]. Furthermore, most prior studies have examined these symptoms in isolation or as covariates, without evaluating their collective predictive value using integrative modeling approaches.

In parallel, machine learning (ML) approaches offer a powerful alternative to traditional regression techniques. Beyond simply reproducing associations already established using regression or structural equation models, ML offers added value by flexibly capturing nonlinear relationships and higher-order interactions among MMH symptoms and covariates. In this context, ML models can benchmark a range of algorithms on predictive performance, support individual-level risk stratification, and highlight symptom patterns that are most informative for early identification of infants at risk of sleep disturbance. In turn, this can refine existing theoretical models and guide more targeted, data-driven clinical decision support.

Recent studies have demonstrated the utility of ML in various infant sleep and postpartum mental health applications [19]. For example, Wang et al [20] developed an automated sleep-stage classifier using heart rate and respiratory rate data to predict white matter development in preterm infants. Similarly, Werth et al [21] designed a deep learning–based system for sleep-stage classification in preterm infants using electrocardiogram (ECG) signals. Additionally, the Sleep Well Baby project introduced a real-time sleep-wake state prediction algorithm based on physiological signals, facilitating improved monitoring in neonatal intensive care units [22]. In another study, Chang et al [23] utilized a multimodal wearable device to collect audio, ECG, and motion data, employing transformer-based neural networks to classify infant sleep/wake states with high accuracy. Furthermore, Huang et al [24] applied ML models to classify and identify infant sleep positions. However, existing ML-based studies have mostly focused on characterizing infant sleep problems using demographic, behavioral, or sensor-derived features, without explicitly leveraging MMH symptoms as primary predictors. To date, no study has examined whether postpartum MMH symptoms can be used, in conjunction with ML methods, to predict infant sleep patterns during the first year of life.

The present study addresses this gap by leveraging ML methods to predict infant sleep trajectories across the first year postpartum based on MMH symptoms and sociodemographic characteristics of mothers and infants. Specifically, we aimed to evaluate the performance of six supervised ML models in predicting two clinically relevant sleep outcomes: nocturnal sleep duration and the frequency of night awakenings (Figure 1). In addition, feature importance analyses were conducted to identify key MMH predictors associated with each outcome.

Figure 1. Overview of the machine learning framework used to predict infant sleep patterns based on postpartum maternal mental health indicators.



We hypothesized that postpartum MMH symptom measures, in combination with basic maternal-infant characteristics, would enable supervised ML models to accurately predict infant sleep outcomes. To operationalize this hypothesis, we addressed the following research questions: (1) Can MMH indicators and sociodemographic characteristics accurately predict infant sleep outcomes (ie, nocturnal sleep duration and night awakening frequency) during the first year postpartum using ML models? and (2) Which MMH features are most predictive of infant sleep outcomes across the first year postpartum? By characterizing the predictive utility of MMH symptoms and elucidating the most influential features, this study seeks to inform early screening and intervention strategies to optimize both MMH and infant developmental well-being.

Methods

Study Population and Data Sources

This study utilized a publicly available dataset [11] comprising 410 mother-infant dyads, collected via an online cross-sectional survey conducted between June and September 2020 at a university hospital in Switzerland. Eligible participants were birth mothers aged 18 years or older with infants between 3 and 12 months of age at the time of data collection and with no reported major neonatal complications.

The dataset included measures of MMH symptoms, infant sleep outcomes, and sociodemographic characteristics of both mothers and infants. A detailed description of the input features used in the analysis is provided in Multimedia Appendix 1. One mother-infant dyad was excluded due to missing information on nocturnal sleep duration, resulting in a final sample of 409 dyads.

Data Elements

MMH Measures

MMH symptoms were assessed using 3 validated self-report instruments: the Edinburgh Postnatal Depression Scale (EPDS), the Hospital Anxiety and Depression Scale-Anxiety subscale (HADS-A), and the City Birth Trauma Scale (CBTS). These measures were selected to comprehensively capture postpartum symptoms of depression, anxiety, and CB-PTSD, respectively.

The EPDS is a 10-item screening tool designed to detect symptoms of postnatal depression in women [25]. It focuses on emotional and cognitive symptoms experienced during the preceding week, excluding somatic complaints that may overlap with normal postpartum changes. Each item is scored on a 4-point Likert scale, and total scores ranging from 0 to 40, with higher scores indicating greater symptom severity.

The HADS-A is the anxiety subscale of the Hospital Anxiety and Depression Scale [26]. It consists of 7 items that assess the frequency and severity of anxiety symptoms experienced during the preceding week. Responses are rated on a 4-point scale, yielding a total score ranging from 0 to 21, where higher scores reflect more severe anxiety symptomatology.

The CBTS is a 29-item instrument [27] specifically developed to assess CB-PTSD symptoms based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (*DSM-5*). The scale is divided into 2 subscales: the birth-related symptoms subscale, which assesses intrusion and avoidance symptoms as well as a subset of negative mood items, and the general symptoms subscale, which captures remaining negative cognition and hyperarousal symptoms. The total score for the *DSM-5*–based items ranges from 0 to 60, with higher scores indicating greater severity of CB-PTSD symptoms. Together, these 3 instruments provided a multidimensional assessment of postpartum MMH, enabling the identification of symptom patterns relevant to infant sleep outcomes.

Infant Sleep Measures

Infant sleep was assessed using the Brief Infant Sleep Questionnaire (BISQ), a widely used and validated parent-report instrument designed to evaluate sleep behavior in infants and toddlers [28]. Mothers were asked to report on their infant's sleep patterns over the preceding week, including total nocturnal sleep duration (between 7:00 PM and 7:00 AM), frequency of night awakenings, and method of falling asleep. For the purposes of this study, 2 primary sleep outcomes were derived and categorized as binary variables: nocturnal sleep duration and night awakenings, both of which serve as indicators of infant sleep quality.

Nocturnal sleep duration was classified as either normal (coded as 0) or insufficient (coded as 1). Infants were categorized as having normal nocturnal sleep if their reported sleep duration was ≥9 hours, for all infant age groups. Infants who slept for less than 9 hours per night were classified as having insufficient nocturnal sleep duration. This threshold aligns with prior research and pediatric sleep guidelines that recommend a minimum of 9 hours of nighttime sleep for infants aged 3 to 12 months [29] .

Night awakenings were categorized based on age-specific thresholds. For infants aged 3 to 6 months, normal was defined as ≤3 awakenings per night. For infants aged 6 to 9 and 9 to 12 months, normal was defined as ≤2 awakenings per night. Infants exceeding these thresholds were classified as having frequent night awakenings, consistent with existing sleep research indicating that night waking typically decreases with age as self-regulation improves [30].

Nocturnal sleep duration and night-awakening frequency were modeled as separate primary outcomes because they capture distinct dimensions of infant sleep—quantity (duration) versus continuity (awakenings)—that may have partially different determinants (eg, circadian scheduling/feeding patterns versus arousal regulation) and lead to different clinical actions. Their measurement properties also differ (duration: continuous; awakenings: count/ordinal), warranting distinct modeling approaches and metrics. Analyzing them separately preserves interpretability of feature effects and supports symptom-targeted guidance. Although the 2 domains can co-occur, our predictive focus is outcome-specific.

Data Preprocessing

Preprocessing steps included computing total scores for the maternal mental health instruments (EPDS, HADS-A, and CBTS) according to their manuals and recoding response options for consistency across instruments. For each instrument, both the individual item responses and the derived total scores were retained as candidate predictors, allowing the models to leverage overall symptom burden as well as more fine-grained symptom patterns (eg, specific anxiety, depression, or trauma-related items). Missing values were imputed (numerical features: mean; categorical features: mode). Numerical features were then standardized (z score), and categorical features were one-hot encoded to ensure consistent transformations during model training and evaluation. This procedure yielded a clean, model-ready feature matrix for all classifiers. Because the study's primary aim was to evaluate predictive performance rather than coefficient-level inference, we did not perform formal multicollinearity diagnostics (eg, VIF). Including both total scores and item-level responses intentionally introduces

some correlation among predictors; however, many of the algorithms employed (eg, tree-based learners and regularized linear models) are designed to handle correlated and partially redundant features by down-weighting or shrinking less informative variables. Potential overfitting from the expanded feature space was further mitigated through cross-validated hyperparameter tuning and evaluation on a held-out test set.

ML Models

Six ML algorithms were employed to predict infant sleep outcomes based on MMH measures and demographic features. These models were selected to represent a diverse range of linear and nonlinear classifiers, including both ensemble and neural network-based approaches.

- Logistic regression: A linear classification algorithm that estimates the probability of a binary outcome based on a weighted combination of input features.
- Random forest: An ensemble learning method that constructs multiple decision trees during training and outputs the class that is the mode of the predictions of the individual trees.
- Support vector classifier: A kernel-based method that identifies the optimal hyperplane separating classes in a high-dimensional feature space.
- extreme gradient boosting (XGBoost): A gradient-boosted decision tree algorithm known for its scalability and performance. It builds an ensemble of weak learners sequentially, optimizing residual errors from prior iterations.
- Light gradient boosting machine (LightGBM): A gradient boosting framework that uses histogram-based learning and leaf-wise tree growth.
- Multilayer perceptron (MLP): A feedforward artificial neural network composed of fully connected layers. It captures complex, nonlinear interactions among features and is trained using backpropagation.

Each questionnaire item and each derived total score was treated as a separate candidate predictor. Modern supervised ML algorithms (eg, tree-based ensembles and regularized models) are generally robust to moderately correlated predictors and can down-weight or ignore redundant features during training, so including both item-level and total-score features does not compromise model learning or model behavior; instead, it allows the algorithm to "decide" whether predictive signal is better captured at the composite-score or item level.

Models Training and Evaluation Strategy

Both outcome variables exhibited class imbalance. For nocturnal sleep duration, 359/409 infants (87.8%) were classified as normal (class 0) and 50/409 (12.2%) as insufficient (class 1). For night awakenings, 346/409 infants (84.6%) were classified as normal (class 0) and 63/409 (15.4%) as elevated (class 1). To address this, we applied 2 strategies. First, we evaluated each model using 4 sampling methods: no sampling, random upsampling, random downsampling, and synthetic minority oversampling technique (SMOTE). This allowed us to assess the impact of different data distributions on model performance. Second, we used evaluation metrics suited for imbalanced data. In addition to accuracy, we computed the precision-recall area under the curve (PR-AUC), which focuses on the minority class and is not influenced by the number of true negatives. We also reported the F_1-score, the harmonic mean of precision and recall, which balances false positives and false negatives. Together, these strategies ensured reliable evaluation of model performance in the context of class imbalance.

All analyses were conducted at the level of the mother-infant dyad. The dataset (N=409 dyads) was randomly split into training (327/409, 80%) and test (82/409, 20%) sets, using stratified sampling to preserve the proportion of infants with nocturnal sleep disturbance and frequent night awakenings in both partitions. All model development (including hyperparameter tuning and internal validation) was performed exclusively on the training set. All analyses were implemented in Python and performed on a high-performance computing node equipped with an NVIDIA A100 GPU (80 GB memory).

To quantify the variability and robustness of model performance, we additionally performed stratified 5-fold cross-validation within the training set for each model-sampling combination. For every fold, we computed PR-AUC, accuracy, and F_1-score and summarized their distribution across folds. Final performance for each model was then evaluated on the held-out test set.

Models Explainability

To characterize which MMH and covariate features contributed most to predictions, we first computed model-based feature importance for the best-performing model for each outcome (as determined by PR-AUC on the held-out test set), using the model's native importance measure. To further enhance interpretability, we then performed a post hoc explainability analysis using Shapley additive explanations (SHAP). For each outcome, we computed SHAP values for all input features. SHAP values quantify the marginal contribution of each feature to the predicted probability of the positive (sleep disturbance) class for individual mother-infant dyads. We summarized global importance by the mean absolute SHAP value across participants and visualized the distribution of feature effects using SHAP summary (beeswarm) plots, as complementary views to the main feature-importance analyses.

Ethical Considerations

This study did not involve the collection or generation of original human subject data. Instead, it utilized publicly available, deidentified data from a licensed source. As such, institutional review board approval and informed consent were not required.

Results

Participant Characteristics

A total of 409 mother-infant dyads were included. Participant characteristics and summary measures are shown in Table 1. The mean maternal age was 30.20 (SD 4.36) years. Nearly half held a university degree (192/409, 46.9%); 388 out of 409 (94.9%) were in a couple relationship. Overall, 51.6% (211/409) of the infants were female and 48.4% (198/409) were male. The mean gestational age at birth was 39.11 (SD 1.90) weeks. At

assessment, infants were distributed as follows: 147/409 (35.9%) were aged 3 to <6 months, 133/409 (32.5%) aged 6 to <9 months, and 129/409 (31.5%) aged 9 to <12 months. MMH means were 9.06 (SD 6.76) on the EPDS, 7.85 (SD 4.26) on the HADS-A, and 13.15 (SD 10.81) on the CBTS.

Table . Sample characteristics and key measures (N=409).

Domain and variable	Value
Maternal	
Age (y), mean (SD)	30.20 (4.36)
Education, n (%)	
University degree	192 (46.9)
Applied Science/Tech diploma	88 (21.5)
Postsecondary/apprenticeship	103 (25.2)
Completed compulsory school	24 (5.9)
No formal education	2 (0.5)
Marital status, n (%)	
Couple relationship	388 (94.9)
Single	14 (3.4)
Separated/divorced/widowed	7 (1.7)
Pregnancy/birth	
Gestational age at birth (wk), mean (SD)	39.11 (1.90)
Infant	
Sex, n (%)	
Female	211 (51.6)
Male	198 (48.4)
Age group, n (%)	
3 to <6 mo	147 (35.9)
6 to <9 mo	133 (32.5)
9 to <12 mo	129 (31.5)
Maternal mental health, mean (SD)	
EPDS[a] total	9.06 (6.76)
HADS-A[b] total	7.85 (4.26)
CBTS[c] total	13.15 (10.81)

[a]EPDS: Edinburgh Postnatal Depression Scale.

[b]HADS-A: Hospital Anxiety and Depression Scale-Anxiety Subscale.

[c]CBTS: City Birth Trauma Scale.

Prediction of Nocturnal Sleep Duration

Models Performance

Figure 2 presents PR-AUC values for each model across 4 sampling strategies. All configurations achieved PR-AUC values above 0.88, with XGBoost with SMOTE highest (0.931), followed by logistic regression with SMOTE (0.924). Accuracy (Figure 3) showed greater variability, with XGBoost without sampling highest (0.886), followed by random forest without sampling or with upsampling (0.878). F_1-scores (Figure 4) mirrored these trends: XGBoost without sampling achieved the highest F_1 (0.840), followed by random forest (0.821) and support vector classifier/MLP (0.820) with either no sampling or SMOTE.

Figure 2. Comparison of the precision-recall area under the curve (PR-AUC) across models and sampling methods for outcome nocturnal sleep duration. PR-AUC quantifies how well a model can distinguish positive cases (infants with insufficient nocturnal sleep duration) from negative ones across various thresholds, especially under class imbalance conditions. LightGBM: light gradient boosting machine; MLP: multilayer perceptron; SMOTE: synthetic minority oversampling technique; SVC: support vector classifier; XGBoost: extreme gradient boosting.

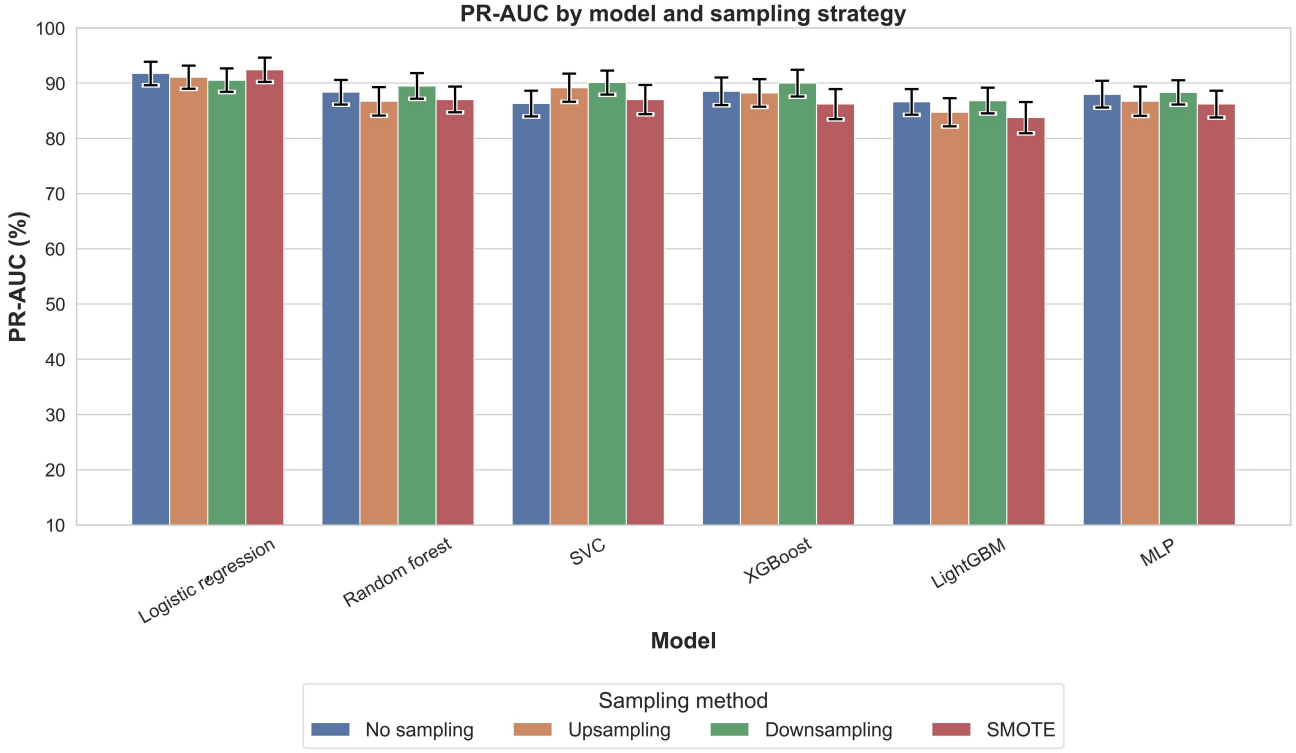


Figure 3. Comparison of accuracy across models and sampling methods for outcome nocturnal sleep duration. Accuracy represents the overall proportion of correct predictions, combining both positive and negative cases, and provides a broad measure of model correctness. LightGBM: light gradient boosting machine; MLP: multilayer perceptron; SMOTE: synthetic minority oversampling technique; SVC: support vector classifier; XGBoost: extreme gradient boosting.

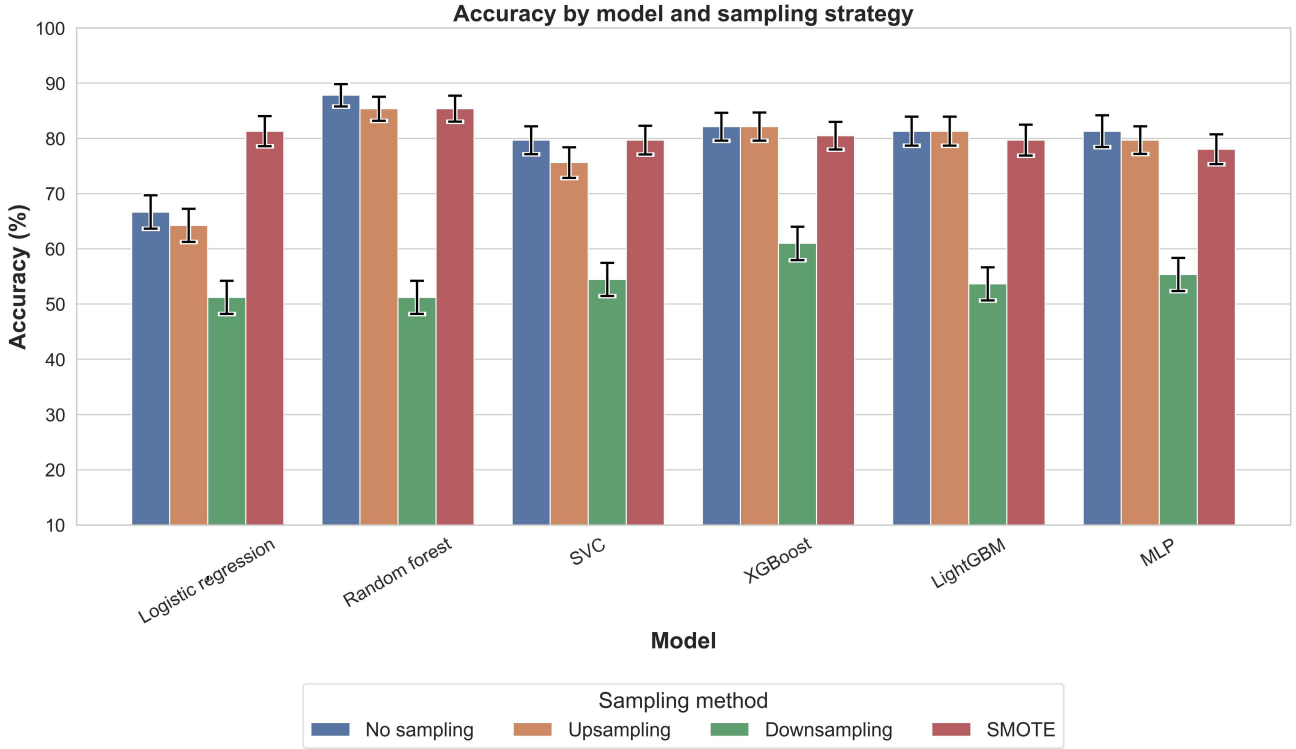
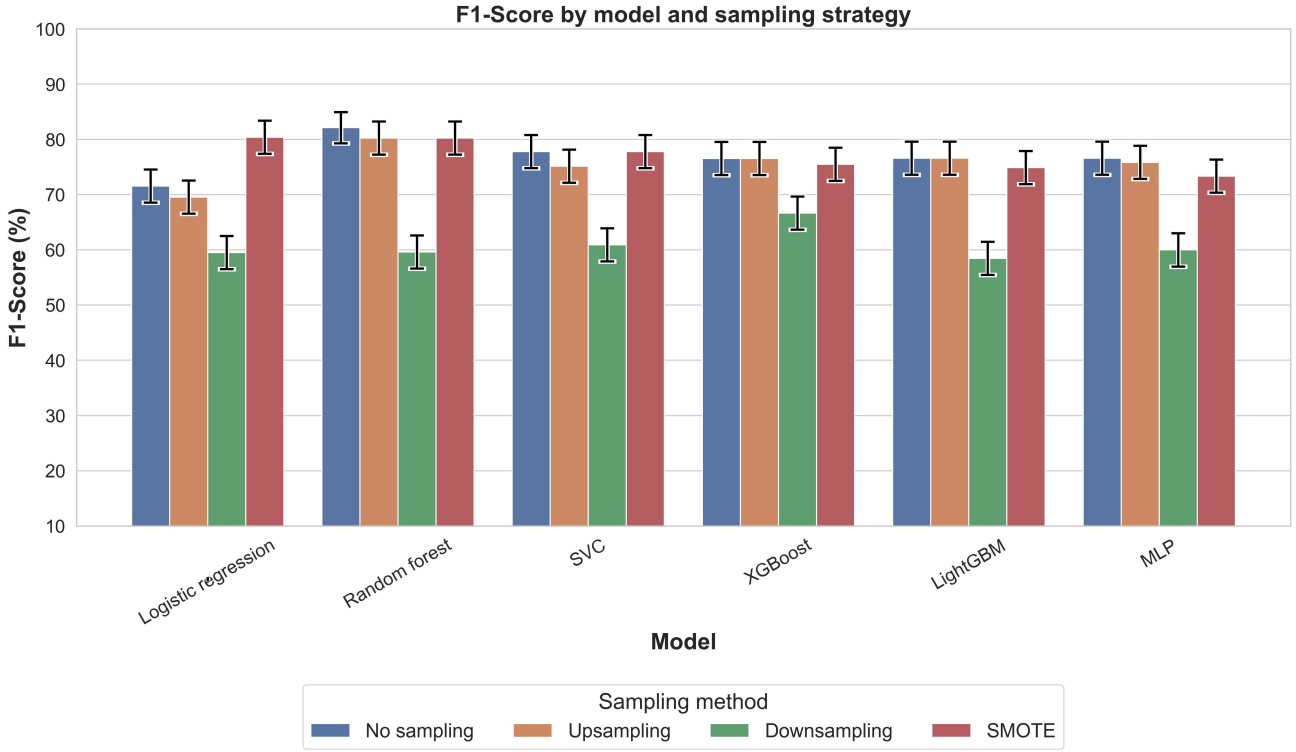


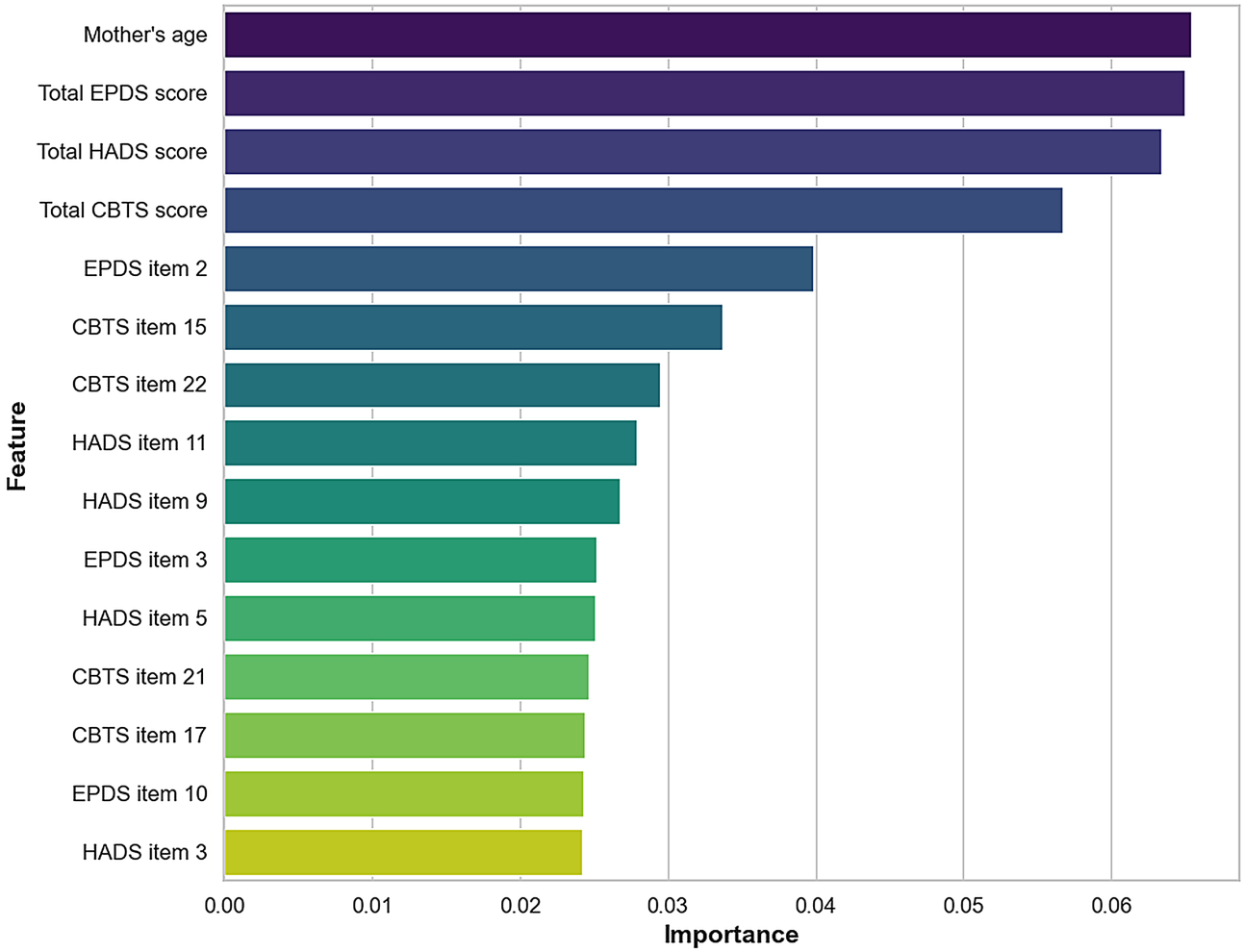
Figure 4. Comparison of F_1-score across models and sampling methods for outcome nocturnal sleep duration. The F_1-score balances precision and recall, making it a valuable metric for assessing model performance in the context of imbalanced datasets. LightGBM: light gradient boosting machine; MLP: multilayer perceptron; SMOTE: synthetic minority oversampling technique; SVC: support vector classifier; XGBoost: extreme gradient boosting.



Feature Importance Analysis

Figure 5 shows the most influential predictors of short nocturnal sleep duration: maternal age and total scores on the EPDS, HADS-A, and CBTS. Individual items also contributed meaningfully, particularly EPDS Item 2 (I have looked forward with enjoyment to things) and CBTS Item 15 (Feeling detached from others). To further probe how individual feature values contributed to predictions for the best-performing model, we examined SHAP summary plots for the nocturnal sleep outcome (Multimedia Appendix 1). These global SHAP patterns were broadly consistent with the main feature-importance rankings and illustrate how higher maternal symptom scores tend to shift predictions toward increased risk of nocturnal sleep disturbance.

Figure 5. Feature importance analysis for outcome nocturnal sleep duration. EPDS: Edinburgh Postnatal Depression Scale; HADS: Hospital Anxiety and Depression Scale; CBTS: City Birth Trauma Scale.



Prediction of Frequent Night Awakenings

Models Performance

Figure 6 reports PR-AUC for predicting night awakenings frequency across models and sampling strategies. All models performed well, typically exceeding 0.83, with logistic regression highest (0.91) and MLP close behind (0.89). Figure 7 shows accuracy, with random forest without sampling highest (0.85) and MLP and XGBoost without sampling at 0.81; downsampling reduced accuracy for all models. F_1-scores (Figure 8) mirrored accuracy, with MLP and XGBoost without sampling at 0.76 and random forest with SMOTE at 0.78. Models trained on downsampled data had the lowest F_1-scores, underscoring the performance cost of sample reduction despite improved class balance.

Figure 6. Comparison of the precision-recall area under the curve (PR-AUC) across models and sampling methods for outcome night awakenings frequency. LightGBM: light gradient boosting machine; MLP: multilayer perceptron; SMOTE: synthetic minority oversampling technique; SVC: support vector classifier; XGBoost: extreme gradient boosting.

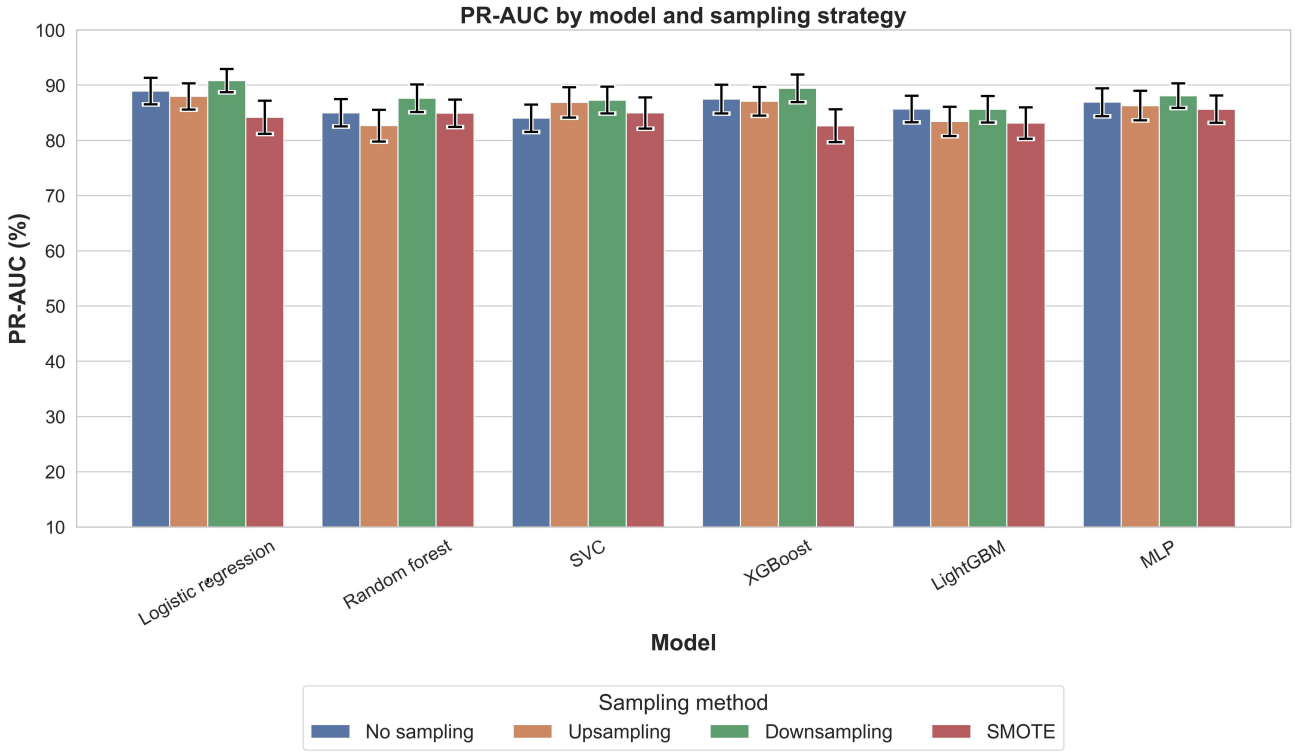


Figure 7. Comparison of accuracy across models and sampling methods for outcome night awakenings frequency. LightGBM: light gradient boosting machine; MLP: multilayer perceptron; SMOTE: synthetic minority oversampling technique; SVC: support vector classifier; XGBoost: extreme gradient boosting.

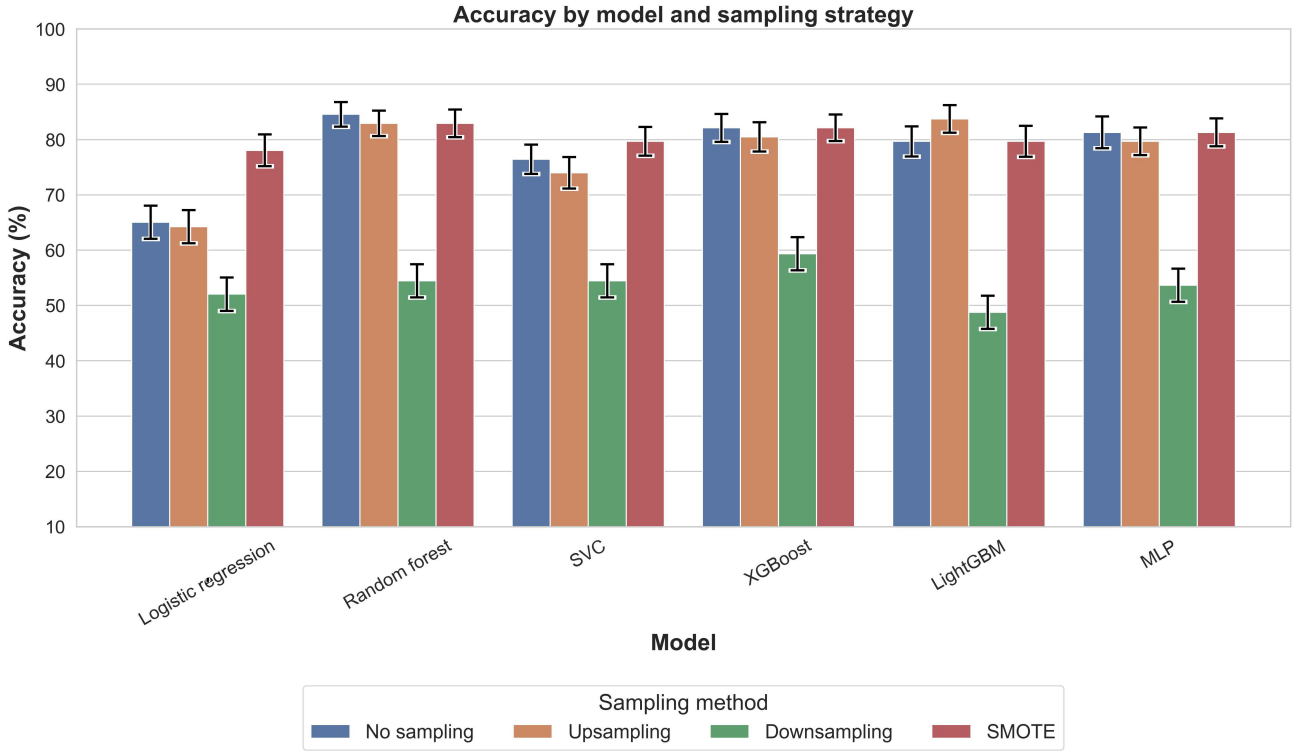
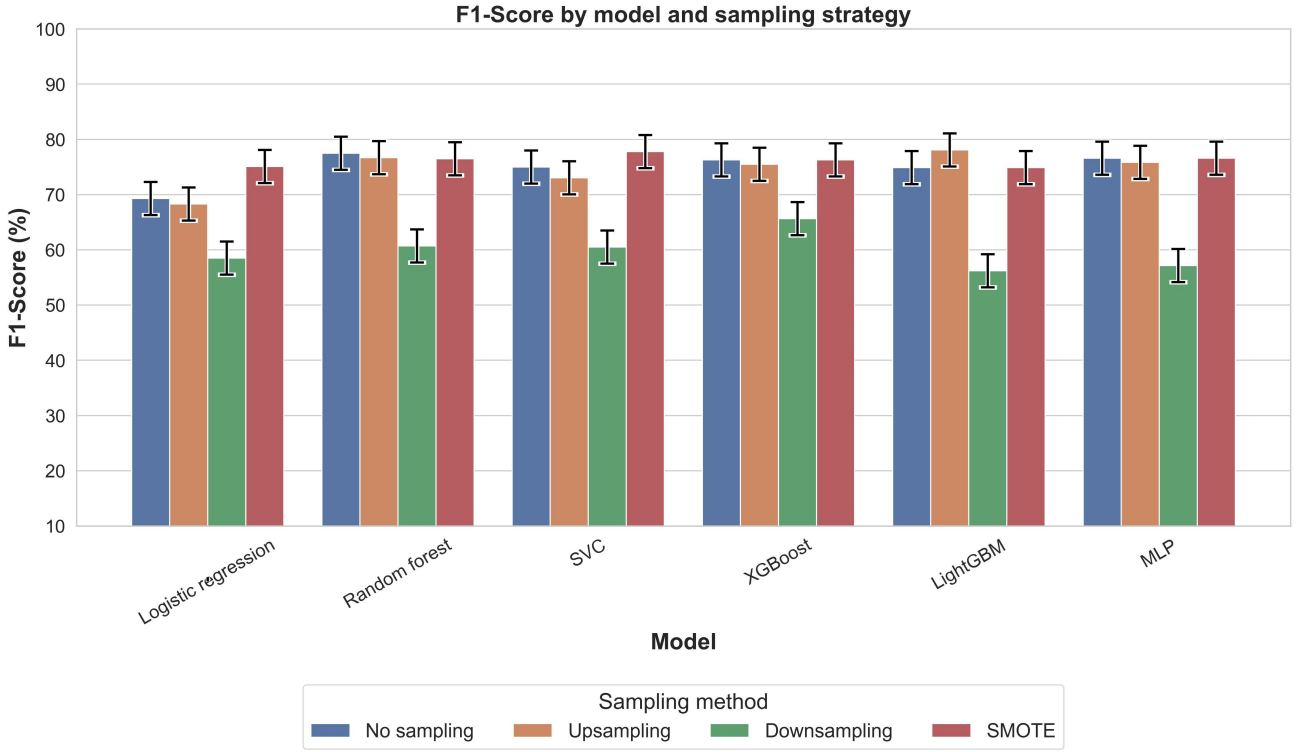


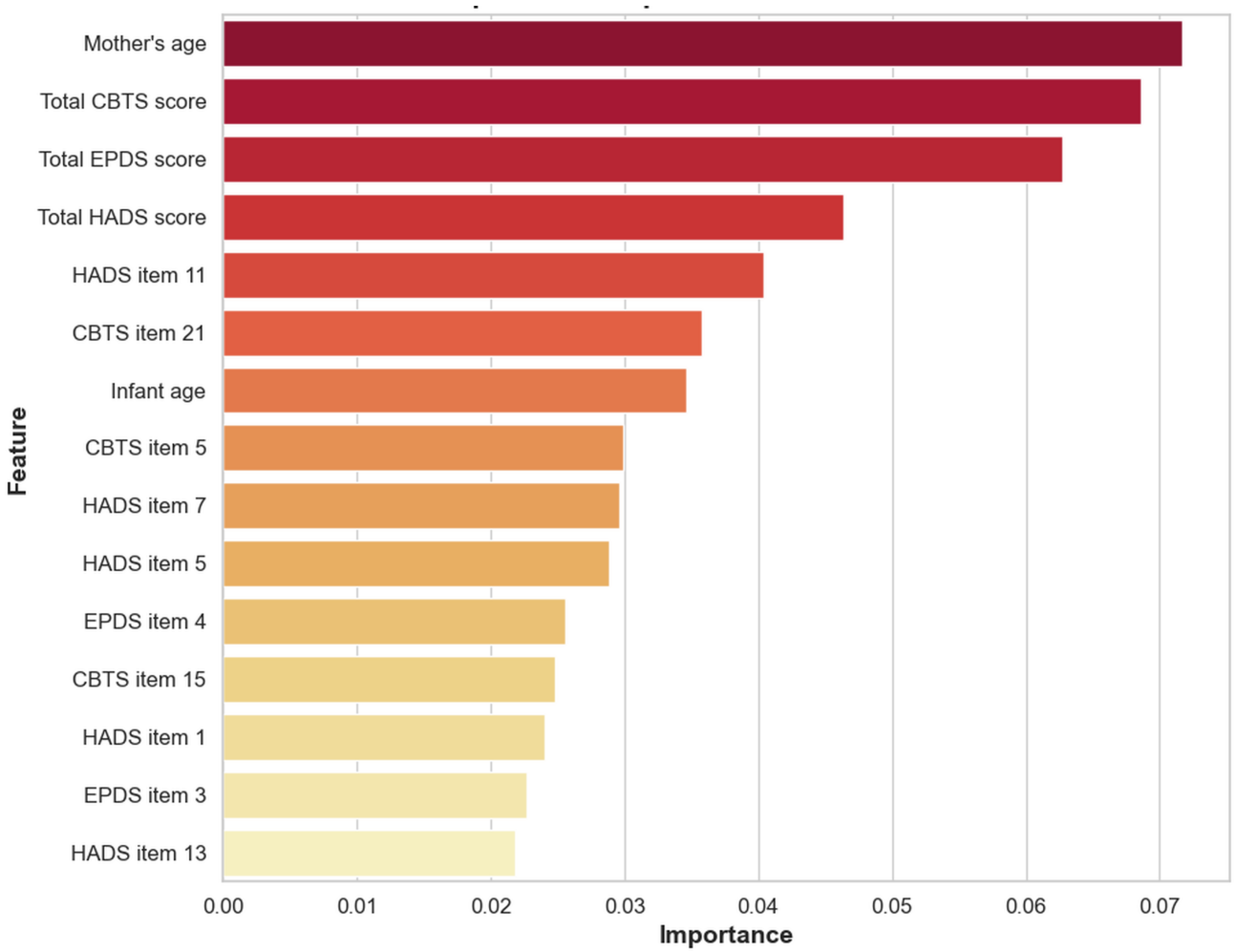
Figure 8. Comparison of F_1-score across models and sampling methods for outcome night awakenings frequency. LightGBM: light gradient boosting machine; MLP: multilayer perceptron; SMOTE: synthetic minority oversampling technique; SVC: support vector classifier; XGBoost: extreme gradient boosting.



Feature Importance Analysis

Feature importance analysis identified the most influential predictors of elevated night awakening frequency (Figure 9). As with nocturnal sleep duration, maternal age and total EPDS, HADS-A, and CBTS scores were top predictors. Individual items also contributed, notably HADS-A Item 11 (I feel restless and cannot seem to stay still) and CBTS Item 21 (Having difficulty concentrating). Infant age also emerged as a relevant predictor. An analogous SHAP summary plot for night awakenings (Multimedia Appendices 1-3) confirms the prominence of MMH features and illustrates how variations in these scores and sociodemographic factors shift individual predictions toward higher or lower night-awakening risk.

Figure 9. Feature importance analysis for outcome night awakenings frequency. EPDS: Edinburgh Postnatal Depression Scale; HADS: Hospital Anxiety and Depression Scale.



Discussion

Key Findings

This study evaluated the utility of postpartum MMH measures in predicting infant sleep patterns using an ML approach. Notably, supervised ML models trained on standardized psychological screening instruments (EPDS, HADS-A, and CBTS), combined with basic demographic and maternal variables, demonstrated high predictive accuracy for both outcomes: insufficient nocturnal sleep duration and frequent night awakenings. These findings indicate the feasibility of using MMH symptoms to identify infants at risk for suboptimal sleep patterns. It also confirms the feasibility of integrating ML tools into postpartum care pathways to facilitate early risk identification.

Predictors of Infant Sleep Patterns

Maternal age emerged as a top predictor for both outcomes, potentially reflecting links with parenting experience, physiological resilience, and contextual factors such as social support and caregiving efficacy. Prior work associating younger age with higher postpartum depression and poorer infant sleep [7,31] is consistent with its predictive strength in our models.

Total scores from the EPDS, HADS-A, and CBTS were among the most influential features in both models. Higher total EPDS scores indicate more severe postpartum depressive symptoms, which can impact maternal responsiveness and infant sleep regulation. Elevated anxiety levels (HADS-A) in mothers may lead to increased nighttime interactions, potentially disrupting infant sleep. Additionally, higher CBTS scores reflect greater CB-PTSD symptoms, which can affect maternal-infant bonding and sleep routines. These aggregate scores likely reflect the cumulative burden of postpartum psychological distress, which has been linked in prior research to disruptions in maternal caregiving behavior, nighttime responsiveness, and the emotional climate surrounding infant sleep routines [13,32,33].

Beyond aggregate symptom scores, several individual questionnaire items provided fine-grained insights. For example, EPDS Item 2 (“I have looked forward with enjoyment to things”)—a measure of anhedonia—was highly predictive of short nocturnal sleep duration. Lower scores on this item suggest anhedonia, a core symptom of depression, which may influence maternal engagement in establishing infant sleep routines. Similarly, CBTS Item 15 (“I felt distant or cut off from other people”) was ranked highly predictive of short nocturnal sleep duration, suggesting that maternal emotional withdrawal and social detachment, characteristic of childbirth-related trauma, may negatively impact the ability to establish secure and consistent nighttime routines.

In the models predicting frequent night awakenings, several additional features emerged as specific to this outcome. These included HADS-A Item 11 (“I feel restless and can’t seem to sit still”) and CBTS Item 21 (“I had difficulty concentrating”), both of which reflect maternal hyperarousal and cognitive dysregulation. These symptoms may manifest in heightened maternal vigilance or difficulty in promoting infant self-soothing, thereby contributing to fragmented infant sleep. Infant age also appeared as a differentiating predictor for this outcome, likely reflecting developmental maturation of sleep consolidation and age-dependent thresholds used in classifying night awakening frequency.

Research and Clinical Implications

From a research perspective, this study illustrates the value of combining symptom-level data with advanced modeling approaches to move beyond correlational frameworks and toward predictive analytics in maternal–infant health. The identification of both composite scores and individual symptom items as key predictors offers a granular understanding of how distinct psychological dimensions—such as anhedonia, emotional detachment, and hyperarousal—may differentially impact infant sleep regulation. These findings advocate for future investigations that examine not only the additive burden of MMH symptoms, but also the specific affective and cognitive pathways through which maternal distress shapes caregiving practices and infant behavioral development. Because MMH and infant sleep outcomes were assessed at the same time point, our models characterize concurrent statistical associations rather than temporal or causal effects. In this context, we use the term “prediction” to denote out-of-sample statistical prediction within the cross-sectional dataset, not longitudinal forecasting. The original analysis of this dataset by Sandoz et al [12] examined the cross-sectional associations between MMH symptom profiles and infant sleep outcomes using traditional statistical methods. In contrast, the present study focuses exclusively on evaluating the predictive performance of supervised ML models that use these MMH measures to classify infant sleep outcomes. Longitudinal studies are particularly needed to clarify the temporal sequence between MMH symptom fluctuations and changes in infant sleep architecture. Moreover, item-level granularity opens avenues for psychometric refinement of postpartum screening instruments, enabling the development of targeted subscales that better predict specific infant outcomes.

Integrating wearable technologies (eg, smartwatches, sleep trackers, biosensors) could passively capture continuous physiological and behavioral data from mothers and infants, reducing reliance on retrospective self-report. When combined with symptom-level psychological data, these rich data streams may improve ML predictive accuracy, enable earlier detection of risk patterns, and support more responsive, personalized interventions.

From a clinical perspective, our findings are best viewed as proof of concept for generating individualized risk scores rather than as a ready-to-deploy screening tool. In practice, such risk scores could be integrated into routine postpartum or well-baby contacts to flag mother-infant dyads who may benefit from closer follow-up (eg, additional monitoring visits or phone check-ins), brief psychoeducation on infant sleep and maternal self-care, targeted support around bedtime routines and soothing strategies, or referral to perinatal mental health services for more structured interventions (such as brief CBT-based programs, parenting support groups, or trauma-focused care where indicated). The exact decision thresholds would need to be codesigned with clinicians and policymakers, balancing sensitivity (minimizing missed high-risk dyads) against specificity and available resources. Our analyses therefore focus on overall discrimination metrics (eg, PR-AUC, F_1) rather than on a single “optimal” cut-off; future work should calibrate and validate context-specific thresholds and decision rules in real-world postpartum care pathways. In addition, findings from this study may inform the design of preventive intervention trials. For instance, trials could test whether tailoring interventions to specific symptom clusters (eg, anhedonia-focused therapies for mothers at risk of short infant sleep duration) yields superior outcomes. Finally, these results highlight the importance of interdisciplinary collaboration—integrating mental health, pediatrics, and data science—to advance personalized, responsive, and developmentally informed postpartum care that promotes long-term maternal and infant well-being.

Limitations and Future Directions

Several limitations should be carefully considered when interpreting the findings of this study.

First, the data relied entirely on maternal self-report questionnaires, which introduces potential response and recall biases. Mothers experiencing psychological distress may perceive or report their infant’s sleep differently, potentially inflating associations between MMH symptoms and infant sleep disturbances due to shared method variance. Furthermore, infant sleep during the first year is influenced by a complex interplay of biological, environmental, and caregiving factors. The exclusive focus on MMH, without integrating other relevant variables such as infant temperament, feeding methods, family routines, or the home sleep environment, limits the comprehensiveness of the predictive models. Future studies should incorporate multimodal, multi-informant data sources, including reports from partners or caregivers and objective sleep measures such as actigraphy or polysomnography, alongside contextual and behavioral variables to more accurately capture the multifactorial nature of infant sleep regulation.

Second, the analysis was limited to 409 mother-infant dyads, all recruited from a single university hospital in Switzerland. This relatively modest sample size and geographically restricted setting may limit the generalizability of the findings to broader, more diverse populations. Sociocultural factors, health care systems, parental practices, and support structures can vary significantly across regions and may influence both MMH and infant sleep patterns. Future studies should validate these predictive models using larger, more heterogeneous samples across multiple countries and health care settings to ensure greater external validity and applicability of the results.

Third, the cross-sectional design limits causal inference. Although we examine associations between MMH symptoms and infant sleep, we cannot determine directionality or

temporality. MMH may influence infant sleep, but the reverse is also plausible, with persistent sleep disturbances worsening maternal distress. Longitudinal studies are needed to disentangle these bidirectional effects and to capture trajectories of MMH and infant sleep over time.

Conclusions

This study demonstrates the feasibility and utility of applying supervised ML models to postpartum MMH symptom measures, together with basic maternal–infant characteristics, to predict infant sleep outcomes—specifically nocturnal sleep duration and night awakening frequency—during the first year of life. The combination of high-performing models and consistent variable importance patterns suggests that both maternal psychological well-being (eg, depressive, anxiety, and CB-PTSD symptoms) and non–mental-health factors such as maternal and infant age are associated with infant sleep patterns in this sample. By integrating scalable mental health screening tools with predictive analytics, this approach holds promise for early identification of at-risk dyads and for informing targeted, preventive interventions that support both maternal and infant health outcomes.

Data Availability

The data used in this study are publicly available and can be accessed through the Zenodo repository [11]. The code used is available at GitHub [34].

Authors' Contributions

Conceptualization: RA, RT, JS
Data Curation: RA
Methodology: RA, RT
Visualization: RA
Writing – Original Draft: RA
Writing – Review & Editing: RT, RB, MA, AA, JS
All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data Dictionary.
[DOCX File, 23 KB - pediatrics_v9i1e78937_app1.docx]

Multimedia Appendix 2
Shapley additive explanations (SHAP) summary plot for nocturnal sleep disturbance.
[DOCX File, 280 KB - pediatrics_v9i1e78937_app2.docx]

Multimedia Appendix 3
Shapley additive explanations (SHAP) summary plot for night awakening.
[DOCX File, 282 KB - pediatrics_v9i1e78937_app3.docx]

References

1. Finkel MA, Duong N, Hernandez A, et al. Associations of infant sleep characteristics with childhood cognitive outcomes. J Dev Behav Pediatr 2024;45(6):e560-e568. [doi: 10.1097/DBP.0000000000001311] [Medline: 39140879]
2. Butler B, Burdayron R, Mazor Goder G, et al. The association between infant sleep, cognitive, and psychomotor development: a systematic review. Sleep 2024 Nov 8;47(11):zsae174. [doi: 10.1093/sleep/zsae174] [Medline: 39158050]
3. Andrade Correia LT, Coimbra DG, Gitaí DLG, Gitaí LLG, de Andrade TG. Associations between chronotype, sleep quality, maternal mental health, and child development in mother-infant dyads. Sleep Med 2023 Jun;106:90-96. [doi: 10.1016/j.sleep.2023.03.023] [Medline: 37075531]
4. Liang X, Zhang X, Wang Y, van IJzendoorn MH, Wang Z. Sleep problems and infant motor and cognitive development across the first two years of life: the Beijing longitudinal study. Infant Behav Dev 2022 Feb;66:101686. [doi: 10.1016/j.infbeh.2021.101686] [Medline: 35007866]
5. Hernandez-Reif M, Gungordu N. Infant sleep behaviors relate to their later cognitive and language abilities and morning cortisol stress hormone levels. Infant Behav Dev 2022 May;67:101700. [doi: 10.1016/j.infbeh.2022.101700] [Medline: 35158217]
6. Cai S, Phua DY, Tham EKH, et al. Mid-pregnancy and postpartum maternal mental health and infant sleep in the first year of life. J Sleep Res 2023 Jun;32(3):e13804. [doi: 10.1111/jsr.13804] [Medline: 36511597]

7. Mazandarani AA, Zare Bahramadbadi M. Maternal self-efficacy as a moderator in the relationship between infant/toddler sleep and maternal mental health. Behav Sleep Med 2025;23(4):515-525. [doi: 10.1080/15402002.2025.2483961] [Medline: 40125868]
8. Mathiasen S, Parsons CE, Fusaroli R, Paavonen EJ, Karlsson H, Karlsson L. Maternal depression and anxiety symptoms across pregnancy and the postnatal period: modest associations between depression symptoms and infant sleep outcomes. Sleep Med 2023 Dec;112:291-300. [doi: 10.1016/j.sleep.2023.10.006] [Medline: 37950940]
9. AlSaad R, Alabdulla M, Tabassum A, Sheikh J, Thomas R. From mother to infant: predicting infant temperament using maternal mental health measures and tabular machine learning models. Front Public Health 2025;13:1659987. [doi: 10.3389/fpubh.2025.1659987] [Medline: 41048269]
10. Škodová Z, Kelčíková S, Maskálová E, Mazúchová L. Infant sleep and temperament characteristics in association with maternal postpartum depression. Midwifery 2022 Feb;105:103232. [doi: 10.1016/j.midw.2021.103232] [Medline: 34971869]
11. Sandoz V, Horsch A. Dataset of the distinct influence of different maternal mental health symptom profiles on infant sleep during the first year postpartum: a cross-sectional survey. Zenodo. 2021. URL: https://zenodo.org/records/5070945 [accessed 2026-01-14]
12. Sandoz V, Lacroix A, Stuijfzand S, Bickle Graz M, Horsch A. Maternal mental health symptom profiles and infant sleep: a cross-sectional survey. Diagnostics (Basel) 2022 Jul 4;12(7):1625. [doi: 10.3390/diagnostics12071625] [Medline: 35885530]
13. Dias CC, Figueiredo B. Mother's prenatal and postpartum depression symptoms and infant's sleep problems at 6 months. Infant Ment Health J 2020 Sep;41(5):614-627. [doi: 10.1002/imhj.21869] [Medline: 32589320]
14. Rudzik AEF, Robinson-Smith L, Tugwell F, Ball HL. Relationships between postpartum depression, sleep, and infant feeding in the early postpartum: an exploratory analysis. Front Psychiatry 2023;14:1133386. [doi: 10.3389/fpsyt.2023.1133386] [Medline: 37032920]
15. Wang S, Cai Z, Gao G, Yan S. Correlation between maternal anxiety during mid-pregnancy and subsequent infant sleep issues: a cross-sectional study from 2015 to 2016. Int J Gen Med 2024;17:623-633. [doi: 10.2147/IJGM.S445604] [Medline: 38410242]
16. Sorondo BM, Reeb-Sutherland BC. Associations between infant temperament, maternal stress, and infants' sleep across the first year of life. Infant Behav Dev 2015 May;39:131-135. [doi: 10.1016/j.infbeh.2015.02.010] [Medline: 25837288]
17. Dagla M, Dagla C, Mrvoljak-Theodoropoulou I, Kavakou AT, Rigoutsou E, Antoniou E. Infant sleep difficulties at the 6th week and the 12th month postpartum: what is their relationship with maternal mental health and other perinatal factors? Mater Sociomed 2021 Mar;33(1):21-25. [doi: 10.5455/msm.2021.33.21-25] [Medline: 34012345]
18. Lin X, Zhai R, Mo J, Sun J, Chen P, Huang Y. How do maternal emotion and sleep conditions affect infant sleep: a prospective cohort study. BMC Pregnancy Childbirth 2022 Mar 23;22(1):237. [doi: 10.1186/s12884-022-04504-6] [Medline: 35321658]
19. AlSaad R, Youssef A, Kashani S, et al. Multimodal large language models for women's reproductive mental health. Arch Womens Ment Health 2025 Dec;28(6):1369-1384. [doi: 10.1007/s00737-025-01633-7] [Medline: 41148324]
20. Wang X, de Groot ER, Tataranno ML, et al. Machine learning-derived active sleep as an early predictor of white matter development in preterm infants. J Neurosci 2024 Jan 31;44(5):e1024232023. [doi: 10.1523/JNEUROSCI.1024-23.2023] [Medline: 38124010]
21. Werth J, Radha M, Andriessen P, Aarts RM, Long X. Deep learning approach for ECG-based automatic sleep state classification in preterm infants. Biomed Signal Process Control 2020 Feb;56:101663. [doi: 10.1016/j.bspc.2019.101663]
22. Sentner T, Wang X, de Groot ER, et al. The Sleep Well Baby project: an automated real-time sleep-wake state prediction algorithm in preterm infants. Sleep 2022 Oct 10;45(10):zsac143. [doi: 10.1093/sleep/zsac143] [Medline: 35749799]
23. Chang KC, Hasegawa-Johnson M, McElwain NL, Islam B. Classification of infant sleep/wake states: cross-attention among large scale pretrained transformer networks using audio, ECG, and IMU data. Presented at: 2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC); Oct 31 to Nov 3, 2023. [doi: 10.1109/APSIPAASC58517.2023.10317201]
24. Huang X, Huang M. Classification and recognition of infant sleeping positions based on dual model feature fusion. 2024 Jul 19 Presented at: AIAHPC '24: Proceedings of the 2024 4th International Conference on Artificial Intelligence, Automation and High Performance Computing; Jul 19-21, 2024; Zhuhai, China p. 57-63 URL: https://dl.acm.org/doi/proceedings/10.1145/3690931 [doi: 10.1145/3690931.3690941]
25. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry 1987 Jun;150:782-786. [doi: 10.1192/bjp.150.6.782] [Medline: 3651732]
26. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
27. Ayers S, Wright DB, Thornton A. Development of a measure of postpartum PTSD: the City Birth Trauma Scale. Front Psychiatry 2018;9:409. [doi: 10.3389/fpsyt.2018.00409] [Medline: 30279664]
28. Sadeh A. A brief screening questionnaire for infant sleep problems: validation and findings for an Internet sample. Pediatrics 2004 Jun;113(6):e570-e577. [doi: 10.1542/peds.113.6.e570] [Medline: 15173539]

29. Galland BC, Taylor BJ, Elder DE, Herbison P. Normal sleep patterns in infants and children: a systematic review of observational studies. Sleep Med Rev 2012 Jun;16(3):213-222. [doi: 10.1016/j.smrv.2011.06.001] [Medline: 21784676]
30. Sadeh A, Mindell JA, Luedtke K, Wiegand B. Sleep and sleep ecology in the first 3 years: a web-based study. J Sleep Res 2009 Mar;18(1):60-73. [doi: 10.1111/j.1365-2869.2008.00699.x] [Medline: 19021850]
31. Teti DM, Crosby B. Maternal depressive symptoms, dysfunctional cognitions, and infant night waking: the role of maternal nighttime behavior. Child Dev 2012;83(3):939-953. [doi: 10.1111/j.1467-8624.2012.01760.x] [Medline: 22506917]
32. Tikotzky L. Postpartum maternal sleep, maternal depressive symptoms and self-perceived mother-infant emotional relationship. Behav Sleep Med 2016;14(1):5-22. [doi: 10.1080/15402002.2014.940111] [Medline: 25127316]
33. Ran-Peled D, Horwitz A, Finkelstein O, Atzaba-Poria N, Tikotzky L. A longitudinal study of postpartum maternal sleep and sensitivity: examining depressive symptoms and social support as moderators. J Sleep Res 2025 Oct;34(5):e70056. [doi: 10.1111/jsr.70056] [Medline: 40170203]
34. Rawan-alsaad/ai_maternalmentalhealth_infantsleep. GitHub. URL: https://github.com/rawan-alsaad/AI_MaternalMentalHealth_InfantSleep [accessed 2026-01-21]

Abbreviations

BISQ: Brief Infant Sleep Questionnaire
CB-PTSD: childbirth-related post-traumatic stress disorder
CBTS: City Birth Trauma Scale
ECG: electrocardiogram
EPDS: Edinburgh Postnatal Depression Scale
HADS-A: Hospital Anxiety and Depression Scale-Anxiety Subscale
LGBM: light gradient boosting machine
ML: machine learning
MLP: multilayer perceptron
MMH: maternal mental health
PR-AUC: precision-recall area under the curve
SHAP: Shapley additive explanations
SMOTE: synthetic minority oversampling technique
XGBoost: extreme gradient boosting

Edited by S Badawy; submitted 12.Jun.2025; peer-reviewed by AMA Elhenidy, PL Quah; revised version received 30.Nov.2025; accepted 24.Dec.2025; published 03.Feb.2026.

Please cite as:
AlSaad R, Burjaq R, AlAbdulla M, Abd-alrazaq A, Sheikh J, Thomas R
Predicting Infant Sleep Patterns From Postpartum Maternal Mental Health Measures: Machine Learning Approach
JMIR Pediatr Parent 2026;9:e78937
URL: https://pediatrics.jmir.org/2026/1/e78937
doi:10.2196/78937

Review Paper

Evaluating Mobile Information Apps for Parents of Preterm Infants After Hospital Discharge: Systematic App Review

Martine Jeukens-Visser[1,2], PhD; Monique Flierman[1,2,3], MSc; Eline Möller[1,2], PhD; Renate Giezeman[1,2], MSc; Raoul Engelbert[1,3], Prof Dr; Daniël Bossen[3], PhD

[1]Department of Rehabilitation Medicine, University of Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands

[2]Amsterdam Reproduction and Development, Amsterdam, The Netherlands

[3]Faculty of Health, Sport and Physical Activity, Amsterdam University of Applied Sciences, Amsterdam, The Netherlands

Corresponding Author:
Daniël Bossen, PhD
Faculty of Health, Sport and Physical Activity
Amsterdam University of Applied Sciences
Tafelbergweg 51
Amsterdam, 1105 BD
The Netherlands
Phone: 31 205954111
Email: d.bossen@hva.nl

Abstract

Background: After hospital discharge, parents of preterm infants need accessible and reliable information to gain confidence and skills in their child-caring abilities and parental autonomy. Parental need for information after hospital discharge includes topics related to prematurity, such as crying, feeding, sleeping, infant care, general health, and neuromotor development. However, parents report difficulty in finding and understanding this information. Mobile apps have the potential to improve information provision.

Objective: The aim of this systematic app review was to (1) identify mobile apps for parents of preterm infants targeting the period after hospital discharge and (2) evaluate the content, quality of the app, and understandability and actionability of the information material.

Methods: We systematically searched for apps in the Apple App Store, Google Play Store, and Google, along with a literature search using PubMed. Multiple keywords were used (eg, “preterm baby,” “app,” and “home”). Apps were included when they provided information for parents on topics and content related to prematurity after hospital discharge. To examine app content related to the postdischarge period, apps were reviewed, and topics were identified. The Mobile App Rating Scale (MARS) was used to measure the app’s quality, and the Patient Education Materials Assessment Tool for Audiovisual Materials (PEMAT-AV) was used to measure the understandability and actionability of the information material.

Results: After the initial search, the titles and descriptions of 196 apps were screened for eligibility. Eventually, 9 English apps were included in the review. Information related to the postdischarge period constituted only a small part of the app’s content. Most commonly addressed topics related to the period at home were vaccinations, follow-up, feeding, and using home oxygen. Using the MARS, only one of the 9 apps received a good score for overall quality (“MyPreemie app”; Graham’s Foundation), and 7 apps received an acceptable score. Only 4 apps scored high on understandability of the PEMAT-AV, and 6 apps scored high on actionability. No Dutch apps were identified.

Conclusions: The current availability of mobile information apps for parents of preterm infants targeting the period after hospital discharge is limited. A total of 9 English apps were identified, which contained a small portion related to the postdischarge period. This content is not comprehensive for the postdischarge period: topics indicated as relevant by parents, such as crying in preterm infants, diaper change in preterm, or parental well-being after preterm birth, are often missing. The overall quality of the apps is only acceptable. Although the reliability of the information was close to good, the understandability of the apps was moderate. Recommendations for future app development include more relevant and understandable information related to the postdischarge period, which meets the demand of parents of preterm infants.

(JMIR Pediatr Parent 2026;9:e67085) doi:10.2196/67085

KEYWORDS

preterm; mHealth; mobile information applications; review; infant; infancy; neonate; neonatal; newborn; premature; health information; education; information quality; mobile health; app

Introduction

Yearly, 13.4 million babies are born preterm (PT; <37 weeks of gestation), of which 2 million infants are born very preterm (VPT; <32 weeks of gestation) [1]. VPT infants are frequently discharged from the hospital before their term date [2], and parents often feel inadequately prepared to take their vulnerable infant home [3]. Without the continuous professional support of the neonatal intensive care unit (NICU), parents lack confidence or competence in infant caregiving. Parents report practical, emotional, and financial challenges at home and the need for practical support regarding baby caregiving tasks, feeding, medication, or managing unexpected health issues [4]. The uncertainty about the health, growth, and neuromotor development of their VPT infant can heighten parental anxiety and negatively affect parents' caregiving behavior [5]. Parental confidence and competence in caring for their VPT infant can be increased by professional support and tailored information, and is thus an important approach to improve parental and infant outcomes [6,7]. Caring for a VPT infant after hospital discharge can be more demanding for parents than caring for a typically developing infant. VPT infants show different behaviors, such as reduced activity, alertness, and responsiveness, that require specific parenting skills to interpret their baby's cues [4]. Parents therefore require information on common topics specifically targeting prematurity, such as crying (how to comfort a preterm infant), feeding (how and when to transition to solid foods), sleeping (recognize pattern of sleep and fatigue in their baby), infant regulation (help their baby to regulate), infant care (diaper change in a very small infant), general health (when to contact a pediatrician), and neuromotor development (differences in milestones between term and preterm infants) [5,8]. This information, specific to premature infants, is, however, not available on the internet [6]. Parents appreciate that general parenting websites provide accessible information on newborn topics such as feeding and digestion, but the content is perceived as less appropriate for parents with a VPT infant [6]. Therefore, practical and tailored information is necessary to increase knowledge and skills to support parents to feel confident in taking care of their preterm baby at home. To accommodate their underlying emotional needs, parents prefer information that is strength-based and confirming or reassuring in their caregiving [8].

Almost all parents in the NICU use their smartphones to search for information regarding prematurity on the internet [9]. For instance, in the Netherlands, information is provided by the Dutch parent organization (Care4neo) [10]. Facilitated by the ease of use, the 24/7 availability, and the ability to make information attractive, mobile health (mHealth) apps are promising tools to provide health information [11]. In general, parental knowledge about infant development is associated with better caregiving behavior and improved infant development [12]. Mobile apps have the potential to improve parental well-being and parenting in the perinatal period [13]. mHealth apps vary in quality, but many are of moderate quality or out of date [11,14]. A previous review on information apps targeting parents with an infant who was still admitted to the NICU showed that only a quarter of the apps for parents were considered of good quality [15]. For optimal support, parental needs for information should be incorporated in the content of the app [16]. Parents have ongoing information needs, but what they want to know changes over time [8]. After discharge home, when hospital staff support is lacking, parents need different information to feel competent in their caregiving than during their initial hospital stay.

For health care professionals and parents, it is important to be able to use high-quality apps, include engagement, functionality, aesthetics, and information quality. Therefore, the information content of mHealth applications needs to be understandable for all parents, irrespective of health literacy levels. Since preterm birth has been consistently associated with lower socioeconomic status [17], low health literacy is also a prevalent issue in parents of VPT infants. Health literacy refers to the skills needed to function effectively in the health care environment [18], and low health literacy is associated with poorer use of health care services and poorer health outcomes [18]. Parents with lower health literacy may encounter difficulties in obtaining, processing, using, and interpreting information in mHealth applications [19]. To benefit from mHealth, parents require digital health literacy skills, such as using digital devices, searching for and understanding information, and evaluating the validity of the information [20]. However, to date, little is known about the quality, understandability, and actionability of available mHealth apps designed to support parents of VPT infants after hospital discharge. Therefore, the aim of this app review was (1) to identify mobile information apps for parents of preterm infants targeting the period after hospital discharge and (2) to evaluate the content, quality of the app, and understandability and actionability of the information material.

Methods

Study Design

This systematic review of mobile apps followed systematic review methodology adhering to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) standards [21] (Multimedia Appendix 1) and published conduct and reporting recommendations for systematic app store reviews [22].

Search Strategy

To ensure the identification of relevant mobile apps, a comprehensive search was conducted, using 4 different strategies. Apps were directly searched in (1) Apple App Store for iOS and (2) Google Play Store for Android. In addition, mobile apps were also searched via (3) Google and (4) PubMed. The search in the app stores and Google was performed in December 2023. The search in PubMed was performed in May

2022. In the first 3 search strategies, keywords were used both in Dutch and in English.

The different search machines implied different search strategies. In the app stores, separate key terms (in Dutch and English) were used in the search field: “preterm baby,” “preemie,” “premature,” “NICU,” and “discharge.”

For the Google internet search, the term “app” was always used and combined with terms “preterm baby,” “preemie,” “prematurity,” NICU,” “Neonatology,” and “incubator.”

The PubMed search combined terms “parent,” “mother,” “father,” or “caregiver” AND “premature birth,” “premature infant,” “preterm,” “prematurity” AND “mobile application*,” “smartphone application*,” “health app*,” “mobile app*.” No MeSH (Medical Subject Headings) terms were used for the PubMed search. The aim of the Google internet and PubMed search was to identify more apps that were subsequently retrieved from one of the App Stores.

The search in the Apple App Store and PubMed provided a certain number of apps and papers. These were all reviewed for eligibility. In the Google Play Store, the search yielded a continuous stream of apps, many of which were not relevant to our inclusion criteria. Therefore, we limited the screening to the first 50 apps that were displayed in the search results. These are typically ordered by relevance and popularity and align with how parents would conduct such a search. In the Google web search, the first 2 pages of results were reviewed to evaluate whether an app for parents of preterm infants was described. The search ended when 2 pages did not contain new hits. Limiting search results is a common practice in app and website reviews, as later results are less likely to be accessed by parents, often align less with the search criteria, and parents are unlikely to continue their search beyond a certain point [11,15,23].

App Selection

Several inclusion and exclusion criteria were used to select the mobile apps (Table 1). The free-of-charge criterion was used because we wanted to ensure that apps were available to all parents regardless of their socioeconomic status, income, or willingness to pay for a mobile app.

Table 1. Inclusion and exclusion criteria for app selection.

Condition	Inclusion criteria	Exclusion criteria
Topic	VPT[a] infants	Typical developing infants
Timing of information	Period after hospital discharge	Only during hospital stay
Language	English or Dutch	Other languages than English or Dutch
Access	No access code	Access code required
Information in the app	Directed to parents	Not directed to parents
Download	App available for download	App not available for download
Charge	App is free of charge	Paid app

[a]VPT: very preterm.

After removing duplicates between the two app stores, the Google and PubMed search, the app descriptions and features were first screened in the Apple App Store or Google Play Store by one researcher (RG) and discussed within the research team for eligibility. Apps that fulfilled the inclusion criteria were then downloaded. Two reviewers (RG and MJ-V) screened the apps for inclusion in the full app review and discussed the eligibility within the research team.

Data Extraction and Quality Assessment

For each app, the following data was collected: name of the app, operating system, developer and its affiliation, language, target population, year of last update, and a brief description of the app. To evaluate the postdischarge hospital content, a list of topics per app was created. To evaluate the quality of the apps and the understandability and actionability of the information material, two independent reviewers (MJ-V and RG) trained themselves to use the Mobile Application Rating Scale (MARS) [24] and the Patient Education Materials Assessment Tool for Audiovisual Materials (PEMAT-A/V) [25]. Thereafter, all included apps were independently evaluated by the two reviewers, and disagreements were resolved until consensus was reached. When no consensus was reached, the research team was involved. For each instrument, a structured data retrieval form was composed, using a spreadsheet in Microsoft Excel.

MARS

The MARS is a tool for assessing the quality of mobile health apps. The MARS consists of 4 objective scales: “engagement” (5 items: fun, interesting, customizable, interactive, and well-targeted to audience), “functionality” (4 items: app functioning, easy to learn, navigation, and gestural design), “aesthetics” (3 items: layout, graphics, and visual appeal), and “information quality” (7 items: accuracy of app description, measurable and achievable goals, quality of information, quantity of information, visual information, credibility, and evidence-based). Each item is rated on a 5-point rating scale, ranging from 1 “inadequate” to 5 “excellent.” Each item has specific descriptions for these rating anchors. Some items have the option “not applicable”. In addition, there is one scale for “subjective quality” (4 items: recommendation of the app, estimated frequency of use, willingness to pay, and overall star rating of the app). The first 3 items are rated on a 5-point scale, and the last item on a 3-point scale. The overall mean score for the 4 objective subscales is calculated, excluding the items rated

as not applicable. The MARS has a high internal consistency (α=.90) and high interrater reliability (intraclass correlation coefficient [ICC]=0.79) [24]. For this study, we used the Dutch version of the MARS [25].

PEMAT-A/V

The PEMAT-A/V is an instrument that assesses the understandability and actionability of audiovisual patient education materials [26]. The PEMAT-A/V consists of 2 scales: understandability (13 items) and actionability (4 items). Understandability is defined as the ability of people from diverse backgrounds with varying levels of health literacy to comprehend educational material and extract key messages. Actionability is defined as the ability of learners to identify what actions can be taken on the basis of educational material information. Understandability includes 19 items evaluating the content, word choice and style, number usage, organization, layout and design, and use of visual aids. Actionability contains four items and evaluates whether the material (1) identifies an action the user can take, (2) the user is directly addressed, (3) breaks down an action into manageable steps, and (4) explains how to use the charts, graphs, tables, or diagrams to take action. Items are rated with “disagree” (0 points) or “agree” (1 point). Some items have the additional option “not applicable.” The PEMAT-A/V is designed to be completed by professionals and helps them select education material that is understandable and actionable. The PEMAT-A/V items are based on other instruments and concepts for developing educational material and are reliable for raters not trained in the use of the PEMAT-A/V. The researchers read the information in the apps and considered each item from a parental perspective, specifically a parent with low health literacy skills. The researchers did have experience with developing information for people with low health literacy skills. The scores for the two scales are calculated as a percentage, ranging from 0-100. A higher score reflects more understandability or actionability. An expert panel established the face and content validity. Interrater reliability was moderate according to Cohen κ (0.50), but with a high absolute agreement of 80% and high agreement when calculated by Gwet agreement coefficient 1 (0.71). Internal consistency was strong (Cronbach α=0.76), and the average item-total correlation=0.62. Construct validation was established based on differences in actionable and poorly actionable material, as well as a strong negative correlation between grade level and both consumer-testing results and PEMAT-A/V scores [26].

Data Analysis

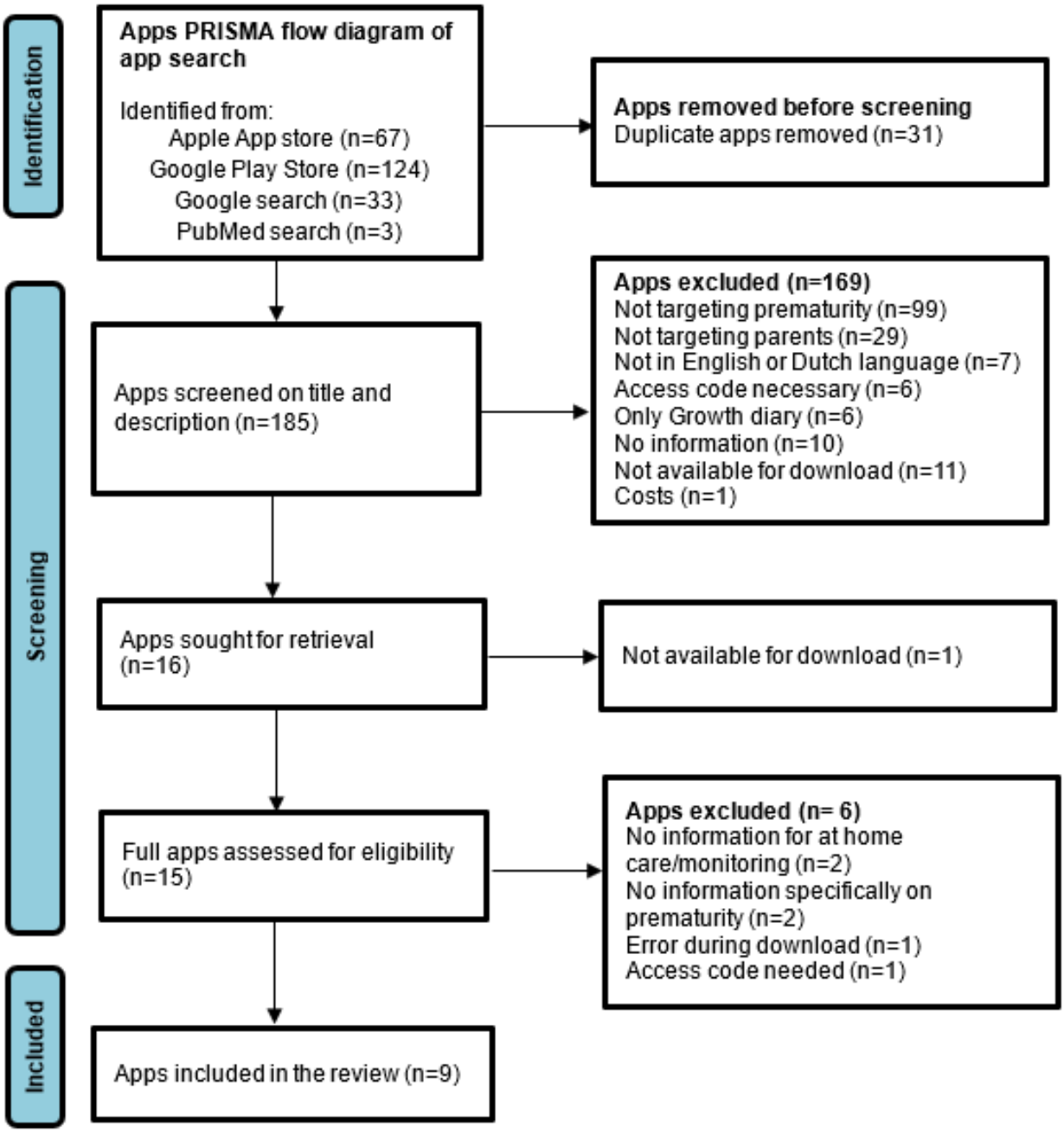
Data analysis was performed using IBM Statistical Package for Social Sciences software (IBM SPSS; version 26). The MARS item scores were averaged for the engagement, functionality, aesthetics, and information subscales. These scores for app quality were then averaged, creating a mean (SD) app quality score. Descriptive statistics were used to summarize the results of the MARS and the PEMAT-A/V. To evaluate consistency between raters, the ICC between the raters was calculated for the MARS and the PEMAT-A/V. Rater agreement was examined by ICC based on a 2-way mixed-effects model. An ICC of <0.50 is considered poor, 0.51-0.75 as moderate, 0.76-0.89 as good, and >0.90 as excellent.

Results

Search Results

The search yielded 191 apps in the Apple and Google Play stores, and additionally 36 apps in Google and PubMed. After removing duplicates, 185 apps remained (Figure 1). Based on the title and description in the app stores, 169 apps were excluded. The majority of the excluded apps did not contain information on preterm-born infants (n=99; 58%), did not target parents but health care professionals (n=29; 17%), or did not contain information, but for instance only growth diaries (n=6; 4%). Only one app was excluded because it was a paid app. A total of 12 apps (6%) that were identified via Google or PubMed could not be retrieved anymore in the app stores. The remaining 16 apps were downloaded and screened for inclusion in the full app review. One app was not available for downloading. Finally, 9 apps fulfilled the inclusion criteria and were included in the final analysis. The majority of the apps were available in both app stores (n=5), 3 apps were only available in the Apple App Store, and 1 app was only available in the Google Play Store. In addition, 2 apps were also described in a scientific paper. One paper describes the content of the MyPreemie app, based on an earlier book, Preemies: the Essential Guide for Parents of Premature Babies, supplemented with new tools [27]. The co-design approach of the Preterm Connect app has been described across 3 settings with different social, economic, and cultural participants [28]. The preliminary findings show similar parental needs, but different preferences across the study populations.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of search for apps for parents of preterm infants after hospital discharge.



Characteristics of the Apps

The apps were developed in the United States of America (n=4; A2, A4, A7, and A9), the United Kingdom (n=2; A3 and A5), Australia (n=1; A6) and New Zealand (n=1; A1), and South Africa (n=1; A8; Table 2). All 9 apps were in English; no Dutch apps fulfilled the inclusion criteria. One app (A2) had information in 25 languages, and another app (A4) was also available in Spanish. The apps were developed by reputable sources, including hospitals (n=4; A1, A3, A6, and A7), nongovernmental organizations (n=3; A2, A4, and A5), and universities (n=2; A8 and A9). The last update of the app varied between 3 weeks and 3 years, with the majority of the apps updated more than a year ago. The size of the apps varied between 7.2 and 102.2 MB. In addition, 3 apps were downloaded more than 1000 times (A1, A2, and A4) and received positive ratings ranging from 4.0 to 4.9 on a scale from 0-5.

Table 2. General characteristics of the apps for parents of preterm infants after hospital discharge.

ID	Name	Operating system	Country	Version	Last update	Target population	Developer	Affiliation	Brief description	Languages	Size (Mb)	Rating	Downloads
A1	Babble NZ Neonatal Family App	Apple iOS; Android	New Zealand	2.7.1	2021	Parent with a baby in the NICU[a].	Neonatal unit at Midcentral Health	Hospital	A reliable source of information about the NICU.	English	7.2	4.9 (n=10)[b]; N/A[c,e]	1000+
A2	Birth & Beyond	Apple iOS; Android	The United States	1.16	2022	Mothers of newborn babies.	Global Health Media Project	NGO[d]	48 videos in 30 languages.	25 languages	10.7	4.8 (n=26)[b]; N/A[e]	10,000+
A3	Family Delivered Neonatal Care (IFDC)	Apple iOS	The United Kingdom	1.1.11	2023	Parents in the NICU.	Imperial College healthcare NHS trust	Hospital	The app offers up-to-date and comprehensive educational material, a developmental timeline, and diary functions to document the neonatal journey.	English	29.2	N/A[e]	N/A
A4	MyPreemie app	Apple iOS; Android	The United States	2.3.1	2020	Families of premature babies.	Graham's Foundation	NGO	Toolkit for the practical and emotional needs of families of premature babies.	English; Spanish	35.4	4.0 (n=182)[b]; 5 (n=1)[e]	10,000+
A5	My Prem Baby	Apple iOS	The United Kingdom	1.8.1	2023	Parents of a premature baby.	Tommy's	NGO	Track and monitor the journey with your premature baby.	English	50.9	N/A[e]	N/A
A6	Neonatal Care and Me	Apple iOS; Android	Australia	N/A	2021	Parents of a baby in the NICU, the special care nursery of pediatrics.	South Western Sydney Local Health District	Hospital	Tools while the baby is in the hospital and beyond, and while being guided by a health professional.	English	102.2	N/A[b]; N/A[e]	N/A
A7	Our Journey in the NICU	Apple iOS	The United States	2.1	2020	Families of children in the NICU.	Phoenix Children's Hospital	Hospital	Identify what families need to know before taking their child home from the hospital.	English	8.2	N/A[e]	N/A

ID	Name	Operating system	Country	Version	Last update	Target population	Developer	Affiliation	Brief description	Languages	Size (Mb)	Rating	Downloads
A8	Preemie Mom Care	Android	South Africa	2.0.5	2020	Mothers of hospitalized premature infants.	UCT Human Computer Interaction lab	University	Provides supportive information to mothers of hospitalized premature infants as they partake in the care of their infant.	English; Afrikaans; Xhosa	24.8	N/A[b]	500+
A9	Preterm-Connect	Apple iOS; Android	The United States	1.8.5	2023	Parents of a preterm baby.	Chih H Wang	University	Connect with other women through forms for preterm birth. Articles and videos about caring for a preterm baby and yourself.	English	63.5	N/A[b]; N/A[e]	50+

[a]NICU: neonatal intensive care unit.

[b]Rating in Google Play.

[c]N/A: not available.

[d]NGO: nongovernmental organization.

[e]Rating in Apple App Store (range 0-5).

App Content

Most of the information in all apps was directed to the period in the NICU. The quantity of information for the posthospital discharge period was limited. Some apps have one “chapter” that covers the postdischarge period (A5, “at home with baby;” A9, “Parenting at home”), whereas other apps have subthemes within a chapter (A4, “Preemie Parenting” and “going home”). Topics that were addressed varied widely between the apps (Table 3). The most common topics that were covered in the apps related to the period post discharge were: vaccinations, follow-up, and recognizing signs of illness. More practical information was provided on feeding, using home oxygen, and sleep (Table 3). Less often, the apps provided information on aspects that were reported as relevant by parents [5,8] as diaper change (A5 and A9), crying (A1, A7, and A9), or parental well-being (A4, A5, and A9).

Table 3. Most common postdischarge topics and functionalities of the apps.

App content	Number of apps	App IDs
Common topics post discharge		
Feeding	7	A2, A3, A4, A5, A6, A8, and A9
Vaccinations	6	A1, A2, A3, A4, A5, and A9
Follow-up	6	A1, A3, A4, A5, A8, and A9
Signs of illness in a baby	5	A1, A2, A7, A8, and A9
Sleep	5	A1, A3, A5, A7, and A9
Home oxygen	4	A3, A4, A5, and A7
Bathing	4	A5, A6, A7, and A9
Functionalities		
Monitoring or tracking (diary, growth, and weight)	3	A4, A5, and A6
Making notes or saving questions	3	A4, A5, and A7
Sharing information	3	A2, A4, and A5
Saving articles	3	A1, A2, and A9
Community groups	1	A9

Besides information provision, the apps also included other functionalities, including monitoring and tracking of infants' weight and height, amount and duration of feeding, or parental mood (A4, A5, A6, and A7; Table 3). The option of making notes was also provided by 4 apps (A3, A4, A7, and A9). Sharing information from the app with others was available in 3 apps (A2, A4, and A5). Community groups were only incorporated in a single app (A9).

Quality of the Apps (MARS)

The interrater reliability of the MARS of the two raters was high (ICC=0.99, CI 0.98-0.99). The overall mean quality (range 0-5) of the 9 apps was 3.4 (SD 0.5; range 2.3-4.3; Table 4). The majority of the apps (n=7) scored between acceptable to good, one app (A7) scored below acceptable, and only one app (A4) scored above good. There was a difference in the ratings between the 4 objective MARS scales. Engagement was rated as poor to acceptable (mean 2.8, SD 0.6; range 1.8-3.4), specifically due to low scores on entertainment, customization, and interactivity. The aesthetics domain was acceptable (mean 3.2, SD 1.0; range 1.7-5). Information quality and functionality were close to good (mean 3.8, SD 0.6; range 2.3-4.3), and (mean 3.9, SD 0.5; range 3.3-4.8), respectively. Several apps received a good score (>4) for information quality (A2, A4, A5, A6, and A9), functionality (A2, A3, A4, and A8), aesthetics (A4 and A6), and overall mean quality (A4). None of the apps received a good score for engagement. The subjective quality (total range 0-18) ranged from 8 to 16, with a mean of 12.6 (SD 3.0). A total of 7 apps (A1, A2, A3, A4, A5, A6, and A9) received a good score for subjective quality.

Understandability and Actionability of the Apps (PEMAT-AV)

The interrater reliability of the PEMAT-AV between the two raters was high (understandability ICC=0.89, 95% CI 0.55-0.98; actionability ICC=0.91, 95% CI 0.59-0.98). The mean understandability of the apps was 78% (SD 12%), ranging from 55% to 100% (Table 4). Only a single app (A2) scored the maximum of 100% for understandability. Lower ratings were obtained when lacking a summary of the information or visual cues to draw attention to key points. The mean actionability was 85% (24%; range 33% to 100%). Lower ratings were obtained when not addressing the user directly or not breaking the action down into manageable, explicit steps. A total of 6 apps (A2, A3, A5, A6, A7, and A9) received the maximum score of 100% for actionability.

Table 4. Quality of the apps and the understandability and actionability of the information material.

ID	Name	MARS[a]						PEMAT A/V[b]	
		A[c]	B[d]	C[e]	D[f]	Mean[g] (SD)	E[h]	U[i]	AC[j]
A1	Babble	3	3.5	3.7	3,8	3.5 (0.3)	14	75	33
A2	Birth & Beyond	2.2	4.8	2	4.2	3.3 (1.2)	10	100	100
A3	IFDC	3.4	4	3.3	3.5	3.6 (0.3)	14	75	100
A4	MyPreemie app	3.2	4.8	5	4.2	4.3 (0.7)	16	82	67
A5	My Prem Baby	3.2	3.8	3	4	3.5 (0.4)	14	55	100
A6	Neonatal Care	3	3.6	4	4.3	3.7 (0.5)	15	83	100
A7	Our Journey in NICU	1.8	3.3	1.7	2.3	2.3 (0.6)	8	75	100
A8	Preemie Mom Care	2.4	4	3.3	3.4	3.3 (0.6)	8	82	67
A9	PretermConnect	3.4	3.5	3	4.2	3.5 (0.4)	14	75	100
	Mean (SD)	2.8 (0.6)	3.9 (0.5)	3.2 (1.0)	3.8 (0.6)	3.4 (0.5)	12.6 (3.0)	78 (12)	85 (24)

[a]MARS: Mobile App Rating Scale.

[b]PEMAT-A/V: Patient Education Materials Assessment Tool for Audiovisual Materials evaluation.

[c]Engagement.

[d]Functionality.

[e]Aesthetics.

[f]Information quality.

[g]Overall mean quality.

[h]Subjective quality.

[i]Understandability.

[j]AC: actionability.

Discussion

Principal Findings

This app review provides insight into the availability, content, quality of the apps, and the understandability and actionability of the information material for parents of preterm infants after hospital discharge. A total of 9 apps were identified that provided information after hospital discharge, but the amount of information on the postdischarge period was limited in all apps. Only one app was of overall good quality, while the mean overall quality was between acceptable and good. The understandability and actionability of the apps were respectively moderate and good.

Although our inclusion criteria focused on the postdischarge period, the apps in this review contained primarily information for the NICU period. The lack of high-quality and understandable apps found in this review is in contrast with the needs of parents of VPT infants after hospital discharge. Parents of VPT infants have reported challenges when they are at home regarding the availability and usability of information [8]. For parents who struggle to seek information, finding an app with appropriate and reliable content will be even more difficult, particularly for those with low health literacy skills [29]. Health care professionals, such as nurses, pediatricians, or pediatric physical therapists, have a responsibility to support parents in their search for relevant and reliable information during their hospital stay. As parental competence was found to decrease after discharge home, it is an important strategy to improve parental confidence in taking care of their VPT infant [30]. When parents and infants are at home, without direct access to a health care professional, apps have the potential to provide health information to parents and can be accessed when and where needed.

Mobile apps can, however, not replace in-person care. Effective use of apps requires guidance from health care providers, as combining digital tools with professional support has been shown to enhance parental confidence [31]. This is even more important for parents with limited health literacy or digital literacy, who are at higher risk of misunderstanding or misapplying information [18,19]. Our findings confirmed that the understandability of many apps is limited, largely due to complex medical terminology and text-heavy formats. This can particularly exclude parents with low health literacy, widening the existing digital divide [32]. Improving understandability, for example, through audio, video, simplified language, and multilingual options, along with professional support, is essential to make apps usable and effective for all parents.

Apps that cover both the period in hospital and after discharge can be beneficial to parents by providing relevant information throughout the different phases. In a previous review of 18 apps in the NICU context [15], only 5 were included in our review, indicating that most NICU apps do not cover topics post-discharge. There was variability in the amount of postdischarge information, the topics, the emphasis within the topics, and how the information was presented. Unfortunately,

the topics do not seem to correspond with the information needs of parents upon discharge [5,8], such as daily infant care, neuromotor development, as well as the impact of prematurity on parents. Instead, most topics are focused on vaccinations, follow-up, and using home oxygen.

Despite the use of Dutch search terms, no Dutch apps were retrieved in the App stores that fulfilled the inclusion and exclusion criteria. The Dutch apps that were found in the Google search were no longer available in the Google Play Store or Apple Store. The majority of the apps evaluated in this review were last updated over one year ago. This lack of updates is in line with a scoping review about problems and barriers related to the use and implementation of apps [33] and, consequently, impedes usability and user experience, which ultimately affects the effectiveness of applications. Apps without active maintenance quickly become outdated due to evolving technology, guidelines, and operating systems [34]. This underscores the necessity of a viable business model and continuous refinement and maintenance after initial development [35]. Sustainable funding for apps is essential, but there are currently few resources available. Partnerships between industry and research may offer a possible solution for some apps.

Only one app had good overall quality, whereas the mean overall quality of the apps was merely acceptable. This is in agreement with an earlier app review, where the mean overall quality was also acceptable [15]. Specifically, aspects within the domains of engagement and aesthetics could be improved. The apps scored particularly low on the engagement domain of the MARS, lower than acceptable. This subscale assesses whether the app is fun, interesting, customizable, interactive, and well-targeted to the audience. Lack of engagement is a common barrier related to the use of mHealth apps and is associated with low adherence [33]. Different functionalities can facilitate parental engagement with an app. A low rating on the Engagement domain suggests improvements are needed. Increasing engagement through entertainment appears not suitable for an app that provides information related to prematurity. However, the app could be customizable or interactive, and should certainly be well-targeted to the audience. If not, this latter aspect would certainly hinder the use of the app. A positive finding from our review was that the domain “information quality” of the apps was close to good. Reliable information is important as it may decrease parental stress and support better caregiving behavior [12]. This also matches the parental needs for reliable information and is probably a result of the reputable sources (hospitals and universities) that developed the apps. This is in contrast with two previous studies in which only 31% and 40% of the websites provided accurate and reliable information for parents of premature babies [6,11].

The understandability of the apps was scored as moderate, largely due to the primarily text-based information, indirect communication with users, and frequent use of medical terminology. In contrast, the app Birth and Beyond (Global Health Media Project) circumvented this problem by using only videos, in multiple languages. During stressful periods, such as hospital discharge, information should be presented in a clear and accessible manner, particularly for parents with low health literacy. For these individuals, the digital divide can be further exacerbated when the information is difficult to comprehend. The hospitals, universities, and nongovernmental organizations create apps with reliable information, but it may not be easily understood by all users. To meet the informational needs of all parents, apps need to be more understandable. Co-design that incorporates both health care professionals’ and parents’ perspectives can enhance app understandability by identifying the preferences and needs of the target group [34]. Reducing text, written at accessible reading age levels, using multiple languages, and incorporating audio and visual formats may improve understandability.

Only 2 papers were retrieved that described the development of an app [27,28], indicating a general lack of transparency about co-creation. None of the 9 apps have been assessed for their impact on parental outcomes. A study on the NICU2HOME app (CF Garfield) [31,36] showed that parental self-efficacy and satisfaction with care improved in parents of preterm infants. This mobile app has not been included in this review, as an access code was required. More research is needed to evaluate the use of apps, parental satisfaction, and the effects of app use on parental outcomes.

Limitations

First, only English and Dutch apps, free and without an access code, were included in the search, thereby possibly missing potential relevant apps. Second, other online resources that provide information to parents, such as websites, were also excluded. Also, progressive web applications were not captured in our search, as these are not available in the searched app stores. Third, the search for apps is time-dependent. Some apps are only available for a short time in the app stores, and replication of the search is therefore difficult. This became clear when apps identified through Google or PubMed were not available in the app stores. During the initial screening of app descriptions and features, followed by a secondary screening for inclusion, app content has been checked to decide whether it also contained information related to the postdischarge period. Fourth, it may be possible that apps have been excluded during the initial screening because the description did not refer to information related to the postdischarge period. However, this information was then likely not substantial and would also not appeal to parents. Fifth, the assessment of the quality of the apps and the understandability and actionability of the information material has been done by the MARS and PEMAT A/V. These are validated tools used by professionals. The researchers were familiar with the parental needs for information [8] and did consider the parental perspective during the evaluation of the apps. However, direct information from parents of a preterm infant has not been taken into account. As parents are the key users, their experiences are most valuable, and their engagement is important to ensure the content meets their needs. A next step would be to include parents to evaluate their experiences with good-quality apps. Finally, although it was evident that information on the postdischarge period was limited, we did not quantify the amount of information provided in the apps. Topics on postdischarge information were identified using a checklist and compared to previously recognized parental needs for information. While the lack of information on the posthospital discharge period was apparent, no specific

measurement was conducted to assess the extent of information for the hospital or home environment. Furthermore, an assessment of the relevance of the topics was also lacking. Future work may establish new methods to incorporate these aspects.

Recommendations

During the post hospital-discharge period, parents of preterm infants need evidence-based, reliable, and practical information. Mobile apps have the potential to offer this information in an accessible way. Currently, few good quality apps exist that contain reliable and understandable information, as the My Preemie app or Preterm Connect. However, more relevant information that matches the needs of parents of VPT infants after hospital discharge is necessary. Future development of digital support tools should also consider solutions that bridge the gap between in-hospital and at-home care by extending access to apps currently limited to the NICU setting. Co-design with parents has been shown to improve the relevance and understandability of health apps [37]. We not only recommend that future apps should be developed or adapted in co-creation with end users, but also that the development process is clearly reported. Research into the use and satisfaction of the parents should establish what information is key for parents, as well as how to deliver this information. In addition, the accessibility and understandability of an app need to be evaluated among parents with a preterm-born infant. The next step would be to evaluate the effect of the information app on parental outcomes as parenting skills, knowledge, and confidence.

Conclusion

The current availability of mobile information apps for parents of preterm infants targeting the period after hospital discharge is limited and not in line with the high parental demand. A total of 9 English apps were identified containing information on the postdischarge period. However, the apps contained limited content for the period at home. The overall quality of the apps was just acceptable, but the information quality was close to good. The understandability of the apps was moderate. Developing apps in co-creation with the end-users to better match their needs and increase the understandability is recommended.

Acknowledgments

No generative artificial intelligence tools have been used in any portion of the manuscript.

Data Availability

The datasets generated and analyzed during this study (ie, the MARS and PEMAT scores from two independent reviewers) are available in the Figshare repository [38]. The mobile apps assessed in this review are publicly accessible via the Apple App Store and Google Play Store.

Funding

This research was financed by Regieorgaan SIA (RAAK.MKB13.004), part of the Netherlands Organization for Scientific Research (NWO).

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA 2020 checklist.
[PDF File (Adobe PDF File), 69 KB - pediatrics_v9i1e67085_app1.pdf]

References

1. Ohuma EO, Moller A, Bradley E, Chakwera S, Hussain-Alkhateeb L, Lewin A, et al. National, regional, and global estimates of preterm birth in 2020, with trends from 2010: a systematic analysis. Lancet 2023;402(10409):1261-1271 [FREE Full text] [doi: 10.1016/S0140-6736(23)00878-4] [Medline: 37805217]
2. Edwards EM, Greenberg LT, Horbar JD, Gagliardi L, Adams M, Berger A, et al. Discharge age and weight for very preterm infants in six countries: 2012-2020. Neonatology 2023;120(2):208-216. [doi: 10.1159/000528013] [Medline: 36649689]
3. Petty J, Whiting L, Green J, Fowler C. Parents' views on preparation to care for extremely premature infants at home. Nurs Child Young People 2018. [doi: 10.7748/ncyp.2018.e1084] [Medline: 29975481]
4. Green J, Fowler C, Petty J, Whiting L. The transition home of extremely premature babies: An integrative review. J Neonatal Nurs 2021;27(1):26-32. [doi: 10.1016/j.jnn.2020.09.011]
5. Davis-Strauss S, Johnson E, Lubbe W. Information and support needs of parents with premature infants: an integrative review. J Early Interv 2020;43(3):199-220. [doi: 10.1177/1053815120957734]
6. Alderdice F, Gargan P, McCall E, Franck L. Online information for parents caring for their premature baby at home: a focus group study and systematic web search. Health Expect 2018;21(4):741-751 [FREE Full text] [doi: 10.1111/hex.12670] [Medline: 29380916]

7. Setiawan J, Mannix T, Sweet L. Understanding the effects of neonatal early discharge on parents: a literature review. J Perinat Neonatal Nurs 2019;33(2):170-188. [doi: 10.1097/JPN.0000000000000395] [Medline: 31021943]
8. Flierman M, Bossen D, de Boer R, Vriend E, van Nes F, van Kaam A, et al. Parents' information needs during the first year at home with their very premature born child; a qualitative study. PEC Innov 2024;4:100270 [FREE Full text] [doi: 10.1016/j.pecinn.2024.100270] [Medline: 38495319]
9. Orr T, Campbell-Yeo M, Benoit B, Hewitt B, Stinson J, McGrath P. Smartphone and internet preferences of parents: information needs and desired involvement in infant care and pain management in the NICU. Adv Neonatal Care 2017;17(2):131-138. [doi: 10.1097/ANC.0000000000000349] [Medline: 27759567]
10. Care4neo. 2025. URL: https://www.care4neo.nl/couveuseouders [accessed 2025-12-11]
11. Dol J, Richardson B, Boates T, Campbell-Yeo M. Learning to parent from Google? Evaluation of available online health evidence for parents of preterm infants requiring neonatal intensive care. Health Informatics J 2019;25(4):1265-1277 [FREE Full text] [doi: 10.1177/1460458217752564] [Medline: 29361867]
12. Leung CYY, Suskind DL. What parents know matters: parental knowledge at birth predicts caregiving behaviors at 9 months. J Pediatr 2020;221:72-80. [doi: 10.1016/j.jpeds.2019.12.021] [Medline: 31987656]
13. Chua JYX, Shorey S. Effectiveness of mobile application-based perinatal interventions in improving parenting outcomes: a systematic review. Midwifery 2022;114:103457 [FREE Full text] [doi: 10.1016/j.midw.2022.103457] [Medline: 35985142]
14. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
15. Richardson B, Dol J, Rutledge K, Monaghan J, Orovec A, Howie K, et al. Evaluation of mobile apps targeted to parents of infants in the neonatal intensive care unit: systematic app review. JMIR Mhealth Uhealth 2019;7(4):e11620 [FREE Full text] [doi: 10.2196/11620] [Medline: 30985282]
16. Spence CM, Stuyvenberg CL, Kane AE, Burnsed J, Dusing SC. Parent experiences in the NICU and transition to home. Int J Environ Res Public Health 2023;20(11):6050 [FREE Full text] [doi: 10.3390/ijerph20116050] [Medline: 37297654]
17. Granés L, Torà-Rocamora I, Palacio M, De la Torre L, Llupià A. Maternal educational level and preterm birth: exploring inequalities in a hospital-based cohort study. PLoS One 2023;18(4):e0283901. [doi: 10.1371/journal.pone.0283901] [Medline: 37018293]
18. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
19. Enlow E, Gray M, Wallace-Keeshen S, D'Agostino J, Abbasi S, Lorch S. Health literacy of parents of very preterm infants at NICU admission and discharge: a prospective cohort study. J Perinatol 2019;39(6):866-875. [doi: 10.1038/s41372-019-0340-y] [Medline: 30770882]
20. Donelle L, Hiebert B, Hall J. An investigation of mHealth and digital health literacy among new parents during COVID-19. Front Digit Health 2023;5:1212694 [FREE Full text] [doi: 10.3389/fdgth.2023.1212694] [Medline: 38333052]
21. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015;4(1):1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
22. Grainger R, Devan H, Sangelaji B, Hay-Smith J. Issues in reporting of systematic review methods in health app-focused reviews: a scoping review. Health Informatics J 2020;26(4):2930-2945 [FREE Full text] [doi: 10.1177/1460458220952917] [Medline: 32914696]
23. Hall M, Challacombe F, Curran C, Shennan A, Story L. Googling preterm prelabour rupture of the membranes: a systematic review of patient information available on the internet. BJOG 2023;130(11):1298-1305. [doi: 10.1111/1471-0528.17498] [Medline: 37077130]
24. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
25. Tency I, Van Hecke A, Coorevits P. Assessing the quality of pregnancy apps through development and validation of the Dutch version of the Mobile Application Rating Scale (MARS). 2019 Presented at: European Congress of Intrapartum Care; 2019; Turin p. A.
26. Shoemaker SJ, Wolf MS, Brach C. Development of the patient education materials assessment tool (PEMAT): a new measure of understandability and actionability for print and audiovisual patient information. Patient Educ Couns 2014;96(3):395-403 [FREE Full text] [doi: 10.1016/j.pec.2014.05.027] [Medline: 24973195]
27. Doron MW, Trenti-Paroli E, Linden DW. Supporting parents in the NICU: A new app from the US, ‘MyPreemie’. J Neonatal Nurs 2013;19(6):303-307. [doi: 10.1016/j.jnn.2013.08.005]
28. Jani S, Nguyen A, Abraham Z, Scala M, Blumenfeld Y, Morton J, et al. PretermConnect: Leveraging mobile technology to mitigate social disadvantage in the NICU and beyond. Semin Perinatol 2021;45(4):151413 [FREE Full text] [doi: 10.1016/j.semperi.2021.151413] [Medline: 33888330]

29. Lee H, Jin S, Henning-Smith C, Lee J, Lee J. Role of health literacy in health-related information-seeking behavior online: cross-sectional study. J Med Internet Res 2021;23(1):e14088 [FREE Full text] [doi: 10.2196/14088] [Medline: 33502332]
30. Polizzi C, Perricone G, Morales MR, Burgio S. A study of maternal competence in preterm birth condition, during the transition from hospital to home: an early intervention program's proposal. Int J Environ Res Public Health 2021;18(16):8670 [FREE Full text] [doi: 10.3390/ijerph18168670] [Medline: 34444418]
31. Garfield CF, Kerrigan E, Christie R, Jackson KL, Lee YS. A mobile health intervention to support parenting self-efficacy in the neonatal intensive care unit from admission to home. J Pediatr 2022;244:92-100. [doi: 10.1016/j.jpeds.2022.01.004] [Medline: 35033562]
32. Estacio EV, Whittle R, Protheroe J. The digital divide: examining socio-demographic factors associated with health literacy, access and use of internet to seek health information. J Health Psychol 2019;24(12):1668-1675. [doi: 10.1177/1359105317695429] [Medline: 28810415]
33. Giebel GD, Speckemeier C, Abels C, Plescher F, Börchers K, Wasem J, et al. Problems and barriers related to the use of digital health applications: scoping review. J Med Internet Res 2023;25:e43808 [FREE Full text] [doi: 10.2196/43808] [Medline: 37171838]
34. Noorbergen TJ, Adam MTP, Teubner T, Collins CE. Using co-design in mobile health system development: a qualitative study with experts in co-design and mobile health system development. JMIR Mhealth Uhealth 2021;9(11):e27896 [FREE Full text] [doi: 10.2196/27896] [Medline: 34757323]
35. van Gemert-Pijnen JEWC, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
36. Garfield CF, Lee YS, Kim HN, Rutsohn J, Kahn JY, Mustanski B, et al. Supporting parents of premature infants transitioning from the NICU to home: a pilot randomized control trial of a smartphone application. Internet Interv 2016;4(Pt 2):131-137 [FREE Full text] [doi: 10.1016/j.invent.2016.05.004] [Medline: 27990350]
37. Huang Z, Benyoucef M. A systematic literature review of mobile application usability: addressing the design perspective. Univ Access Inf Soc 2022;22(3):715-735. [doi: 10.1007/s10209-022-00903-w]
38. The MARS and PEMAT scores from two independent reviewers. Figshare. 2025. URL: https://figshare.com/s/646475d0d75cdf995771 [accessed 2025-12-15]

Abbreviations

ICC: intraclass correlation coefficient
MARS: Mobile App Rating Scale
MeSH: Medical Subject Headings
mHealth: mobile health
NICU: neonatal intensive care unit
PEMAT-AV: Patient Education Materials Assessment Tool for Audiovisual Materials
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PT: preterm
VPT: very preterm

Edited by J zhang; submitted 01.Oct.2024; peer-reviewed by I Tency, A Perez-Navarro; comments to author 29.May.2025; revised version received 12.Nov.2025; accepted 08.Dec.2025; published 19.Jan.2026.

Please cite as:
Jeukens-Visser M, Flierman M, Möller E, Giezeman R, Engelbert R, Bossen D
Evaluating Mobile Information Apps for Parents of Preterm Infants After Hospital Discharge: Systematic App Review
JMIR Pediatr Parent 2026;9:e67085
URL: https://pediatrics.jmir.org/2026/1/e67085
doi:10.2196/67085
PMID:

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org