# DATA-DRIVEN PRECISION PUBLIC HEALTH: LEVERAGING MACHINE LEARNING TO TRACK AND REDUCE ZERO-DOSE AND PARTIALLY VACCINATED CHILDREN IN NAKIFUMA, UGANDA

**KENNETH MICHAEL OGWOK**

**M23M19/005**

**A DISSERTATION SUBMITTED TO THE FACULTY OF ENGINEERING, DESIGN AND TECHNOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF A DEGREE OF MASTER OF SCIENCE IN DATA SCIENCE AND ANALYTICS OF UGANDA CHRISTIAN UNIVERSITY**

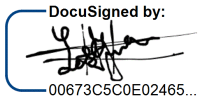
**April, 2026**

UGANDA CHRISTIAN UNIVERSITY

A Centre of Excellence in the Heart of Africa

# Declaration

I, Ogwok Kenneth Michael, hereby declare that this research project is my own original work and has not been previously submitted for a degree or other qualification in any institution. All sources used in this project have been acknowledged and properly referenced. I confirm that the research was conducted ethically, following all relevant guidelines and regulations.

**Sign:** 


**Name:** Ogwok Kenneth Michael

**Date:** 20th August 2025

# Acknowledgments

I would like to express my sincere gratitude to God Almighty for granting me the strength and ability to complete this research.

My deepest appreciation goes to the Head of the Department of Computing at Uganda Christian University, Dr. Innocent Ndibatya, for his steadfast leadership and guidance throughout my academic journey. I am also profoundly grateful to my supervisor, Dr. Daphne Bitalo. Her constant engagement, invaluable discussions, and precise course corrections were instrumental in the successful completion of my research and the publication of my work.

I extend my thanks to Dr. Kkonde Anthony, Principal Medical Officer of Mukono Municipality, and Mr. Ivan Mwesigwa, District Biostatistician of Mukono District. Their expert advice on my research topic and their connections to local health facilities were crucial for the data collection phase. I am also deeply indebted to Mr. Munyandara Elisha, Derrick Kituyi, and Ms. Nakirabira Prossy for their direct support with data collection in Nakifuma, without which a critical part of my analysis could not have been accomplished.

To my classmates, Mr. Quinton Ssebagala and Mr. Wanyama Denis, I am grateful for your invaluable remarks and suggestions, which fostered pivotal thinking about my data collection and utilization. I also wish to thank Mr. Charles Muganga, whose expertise in Data Science and analytics inspired me to design effective databases for my modeling and predictions.

A special appreciation is reserved for Ms. Angom Ritah Okello. Her exceptional editing and proofreading skills, along with her supporting documentation, were fundamental in helping me secure approval for my research proposal. Her work is truly invaluable. Finally, I thank my family for their unwavering emotional and financial support, which was the bedrock of my entire study journey. Your constant backing enabled me to focus and succeed.

**APPROVAL**

This is to certify that this research project titled “Data-Driven Precision Public Health: Leveraging Machine Learning to Track and Reduce Zero-Dose and Partially Vaccinated Children in Nakifuma, Uganda” has been submitted with my approval as the university supervisor.

**Supervisor’s Name:** DR. DAPHNE NYACHAKI BITALO

**Signature:** 

**Date:** 29/04/2026

# Abstract

Despite global immunization efforts, 14.3 million children remain "zero-dose" worldwide as of 2024, with Sub-Saharan Africa bearing the highest burden. While Uganda has achieved 82-89% DTP3 coverage, significant disparities persist with only 54% of children achieving full immunization. This study leveraged data science methodologies to investigate childhood vaccination patterns in Bandaali ward and Kizungu, Nakifuma Subcounty, Uganda, addressing critical gaps in precision public health.

The research was guided by three objectives: to identify factors associated with vaccination status, develop predictive models, and propose evidence-based strategies for improving coverage. A mixed-methods approach utilized primary data from 115 caregivers and children, followed by statistical analysis and systematic evaluation of machine learning algorithms, including Gradient Boosting and Random Forest models.

Factor analysis revealed three distinct barrier profiles associated with childhood vaccination status. Zero-dose children (10.4% of the sample) were significantly associated with barriers related to *caregiver and systemic access*: caregiver relationship dynamics ($p < 0.05$), negative health worker attitudes ($p < 0.05$), excessive waiting times ($p < 0.05$), and low perceived vaccine importance ($p < 0.05$). Partially vaccinated children (40.9% of the sample) were associated with a profile of *logistical and demographic* factors, showing strong associations with child age ($p < 0.001$), caregiver age ($p < 0.002$), and vaccine stock availability issues ($p < 0.03$). Finally, fully immunized children (48.7% of the sample) were characterized primarily by the *behavioral factor* of vaccination card possession ($p < 0.005$).

Predictive modeling demonstrated exceptional performance for identifying zero-dose children, with Logistic Regression and Gradient Boosting achieving perfect classification scores (1.00). The Random Forest model emerged as the optimal overall, achieving 0.97 accuracy for zero-dose, 0.74 for partially vaccinated, and 0.94 for fully immunized children. Feature importance analysis confirmed health system quality as a primary predictor for zero-dose status.

Evidence-based recommendations include targeted interventions addressing health worker attitudes for zero-dose children, supply chain strengthening for partially vaccinated children, and documentation improvements for sustaining full immunization. The validated predictive model provides a

practical tool for proactive identification of at-risk children, enabling resource-efficient interventions. This study offers a scalable methodological framework for applying data science to vaccination challenges in resource-limited settings.

**Keywords:** childhood vaccination, zero-dose children, machine learning, predictive modeling, public health, Uganda, immunization coverage, data science

# Contents

# List of Tables

# List of Figures

# General Introduction

## Global Childhood Vaccination Landscape

Childhood immunization stands as one of the most effective public health interventions, yet it continues to face significant challenges. As of 2024, approximately 14.3 million children worldwide are "zero-dose," having never received a single vaccination, while coverage of the third dose of diphtheria, tetanus, and pertussis (DTP3) reached only 85% globally (World Health Organization, 2024). The global vaccination landscape has been particularly affected by the COVID-19 pandemic, which caused a stagnation in coverage. During 2022–2023, global coverage with the first and third doses of diphtheria-tetanus-pertussis-containing vaccine (DTPcv) (89% and 84%, respectively) and the first dose of measles-containing vaccine (MCV1) (83%) stagnated and remained lower than pre-pandemic levels (World Health Organization, 2024). A critical issue contributing to incomplete coverage is the concept of missed opportunities, which occur when a child visits a health facility but fails to receive all the vaccines they are due for.

## Regional Context: Sub-Saharan Africa

Sub-Saharan Africa bears the highest burden of unvaccinated children globally, with numbers rapidly increasing in many countries. In half of the affected countries, the number of unvaccinated children expanded rapidly from 3.6 million in 2019 to 5.4 million in 2024, underscoring the urgent need for humanitarian responses that include immunization services (Bangura et al., 2020). The challenges are complex, stemming from inadequate healthcare infrastructure, conflict, limited resources, and geographic barriers. Even with support from organizations like Gavi, the Vaccine Alliance, achieving equitable coverage is an ongoing struggle, particularly in rural and marginalized communities where zero-dose children are most concentrated.

## Machine Learning Revolution in Healthcare and Vaccination Programs

The healthcare sector has been revolutionized by the adoption of artificial intelligence (AI) and machine learning (ML). In the area of vaccine development, ML has dramatically accelerated

timelines from years to months by analyzing vast datasets to identify potential vaccine candidates (Goodswen et al., 2023). Beyond development, predictive analytics—a core application of machine learning—is proving instrumental in addressing coverage gaps. Feasibility studies and research from African contexts have demonstrated that these algorithms can accurately identify children at high risk of defaulting on their vaccination schedules (Asnake et al., 2025); (Nantongo et al., 2024). Studies have successfully employed various machine learning approaches, showing that geographic factors are particularly significant for predicting vaccine uptake (Wasswa et al., 2025).

## Uganda's Vaccination Landscape

Uganda has made significant strides in childhood immunization, which has contributed to a notable decline in child mortality (Gavi, the Vaccine Alliance, 2024b). However, achieving universal vaccination coverage remains a challenge. Current statistics indicate that on average, only 54% of Ugandan children are fully immunized, with varying rates for specific vaccines (Gavi, the Vaccine Alliance, 2024b). While the percentage of zero-dose children has decreased from 8% in 2000 to 2.2% in 2016, no region in Uganda has achieved an under-vaccination rate below 20% (Uganda Bureau of Statistics, 2016); (Ssebagereka et al., 2024). The burden of incomplete vaccination is substantial, with a recent analysis estimating that households with zero-dose children account for 13% of all deaths among children under five (Gavi, the Vaccine Alliance, 2024a). To address these issues, data-driven initiatives are now being used to systematically track and reach zero-dose children through more precise, evidence-based approaches (Gavi, the Vaccine Alliance, 2024a).

## Detailed Definition of Vaccination Status Categories

Understanding vaccination status requires precise definitions. Zero-dose children are those who have not received a single routine dose by 12 months of age (Uganda Bureau of Statistics, 2016). They are at the highest risk for vaccine-preventable diseases. Partially vaccinated children have received some but not all of the recommended vaccines for their age, often due to missed appointments or supply issues. In contrast, fully immunized children have received all age-appropriate vaccines according to their national schedule by their first birthday, which in Uganda's context includes a list of specific vaccines given at various ages (World Health Organization, 2024).

## The Promise of Data Science in Addressing Vaccination Gaps

The integration of data science offers unprecedented opportunities to understand and address vaccination coverage gaps. By processing complex, multidimensional datasets, machine learning models can identify patterns that traditional statistical methods might miss (Asnake et al., 2025). These models can incorporate a wide range of factors, from demographics and socioeconomic status to healthcare system variables and geographic data, to predict vaccination outcomes with high accuracy. This capability allows public health programs to shift from broad, untargeted interventions to proactive, precise strategies that focus on at-risk populations and optimize the use of limited resources for maximum impact.

## Study Rationale and Local Context

Given the complex national and regional context, there is a clear need for localized studies that can inform community-level interventions. This research focuses on the communities of Bandaali ward and Kizungu, which represent typical rural Ugandan communities facing unique challenges such as limited healthcare infrastructure and economic constraints. By applying data science approaches to this specific local context, the study aims to generate targeted insights that can not only improve vaccination coverage within these communities but also inform broader strategies for addressing similar challenges across Uganda and Sub-Saharan Africa. This localized approach to precision public health maximizes impact by tailoring solutions to the specific barriers and opportunities of a given area.

# Chapter 1

# Introduction

## 1.1 Background

Childhood immunization is one of the most effective and cost-effective public health interventions, credited with preventing millions of deaths annually from infectious diseases. Despite significant global progress, universal vaccination coverage remains an elusive goal. International bodies like the World Health Organization (WHO), UNICEF, and Gavi continue to grapple with persistent coverage gaps, which have been exacerbated by recent global events, including the COVID-19 pandemic. These disruptions have strained an already fragile immunization infrastructure, leading to a resurgence of vaccine-preventable diseases and leaving millions of children vulnerable. The challenge is particularly acute in low- and lower-middle-income countries, where logistical, socioeconomic, and systemic barriers often prevent children from receiving their full course of recommended vaccines.

Understanding childhood vaccination status requires clear categorization. Zero-dose children are those who have not received any routine vaccination by 12 months of age, representing the most vulnerable population with no contact with immunization services. Partially vaccinated children have received some but not all recommended vaccines according to their national immunization schedule, indicating interrupted or incomplete vaccination series. Fully immunized children have received all age-appropriate vaccines by their first birthday, achieving optimal protection against vaccine-preventable diseases. These classifications form the foundation for tracking immunization progress and designing targeted interventions.

## 1.2 Problem Statement

Despite significant progress in childhood vaccination, global efforts faced considerable setbacks. In 2024, an estimated 14.3 million infants did not receive an initial dose of the Diphtheria-Tetanus-Pertussis (DTP1) vaccine, with an additional 5.6 million being partially vaccinated, totaling 19.9 million children worldwide who remained vulnerable to vaccine-preventable diseases due to persistent

access gaps (World Health Organization, 2024). Notably, approximately 55% of these children resided in just 10 countries. The severity of this challenge was further emphasized by historical data, which showed that in 2018, an alarming 700,000 deaths among children under five were attributed to vaccine-preventable diseases (Lawrence, 2021).

Concerns mounted as global health bodies warned that outbreaks of vaccine-preventable diseases including measles, meningitis, and yellow fever were on the rise, with measles cases alone reaching an estimated 10.3 million in 2023. This alarming trend was exacerbated by a global funding crisis, which led to nearly half of low- and lower-middle-income countries experiencing moderate to severe disruptions in vaccination campaigns and routine immunization (WHO and Gavi, 2025).

Behind Uganda's seemingly stable DTP3 vaccination rates (82–89%, 2011–2022) lay a troubling reality: hundreds of children still died from preventable diseases like rotavirus. The COVID-19 pandemic further strained an already fragile system. As UNEPI's Dr. Michael Baganizi explained, "We're fighting blind. Until we map exactly where zero-dose children live and understand why they're missed, our interventions are just shots in the dark" (Agaba, 2024).

The critical gap in current approaches lay in insufficient understanding of specific factors that differentiated zero-dose, partially vaccinated, and fully immunized children in local contexts. While traditional public health approaches relied on broad demographic analyses, there was a notable absence of sophisticated data science methodologies in Uganda's vaccination programs. Unlike other sectors such as banking and telecommunications where machine learning algorithms were routinely used for predictive analytics, Uganda's health sector had yet to fully harness these technologies for vaccination coverage prediction. Current identification strategies employed broad approaches that lacked the precision offered by modern data-driven methodologies, and there remained an absence of evidence-based strategies specifically designed to transition children from zero-dose and partially vaccinated status to fully immunized status. This represented a significant missed opportunity, as machine learning had demonstrated remarkable success in healthcare applications globally, including predicting patient outcomes, optimizing resource allocation, and identifying at-risk populations.

## 1.3 Justification of the Study

This research addressed the critical importance of childhood vaccination in preventing serious diseases and safeguarding global public health. Despite significant progress and global efforts to improve vaccination coverage, the alarming reality persisted: 19.9 million children worldwide remained zero-dose or partially vaccinated, leaving them vulnerable to preventable diseases (World Health Organization, 2024).

The fundamental challenge impeding effective intervention was threefold, directly aligning with the research objectives. Stakeholders lacked comprehensive understanding of the specific factors that contributed to different vaccination statuses. While broad demographic and socioeconomic factors were often cited, there was insufficient granular analysis of how these factors specifically differentiated zero-dose children, partially vaccinated children, and fully immunized children at the community level. Additionally, current identification strategies employed broad approaches that lacked the precision offered by modern data-driven methodologies. There was an urgent need for predictive models that could accurately determine a child's vaccination status using locally relevant factors, enabling proactive identification of at-risk children before vaccination opportunities were missed. Finally, evidence-based intervention strategies specifically designed to increase fully vaccinated status were lacking. Most current approaches used generalized interventions rather than targeted recommendations based on predictive insights about local vaccination patterns and barriers.

This research gained particular urgency when examining Uganda's context, where DTP3 vaccine coverage had fluctuated and where the COVID-19 pandemic intensified existing immunization disruptions. Specifically, Bandaali ward and Kizungu within Nakifuma Subcounty served as an ideal representative cross-section for investigating these challenges, as this area embodied three critical characteristics: high population mobility patterns, mixed urban, peri-urban, and rural demographics, and documented vaccination access disparities.

By employing advanced data science techniques within this localized context, the study generated comprehensive factor analysis to identify specific determinants of vaccination status, developed precision predictive models for accurate vaccination status determination, and created evidence-based intervention strategies tailored to increase fully vaccinated children in Nakifuma. The resulting evidence base enabled health authorities to transition from blanket interventions to precisely targeted strategies, ultimately improving vaccination equity and coverage.

## 1.4 Research Questions

The research questions were specifically designed to align with the three main objectives and expected outputs.

1. What were the key socio-economic, demographic, geographic, and health system factors that significantly differentiated zero-dose children, partially vaccinated children, and fully immunized children? The output of this question was to identify the factors that contribute to vaccination status, specifically for Zero dose, partially, and Fully Immunized children (0, 5, 10).
2. How accurately can machine learning algorithms predict a child's vaccination status using the identified contributing factors in the study area? The output of this question was to create a predictive model to determine vaccination statuses.
3. What specific, evidence-based intervention strategies could be developed from the predictive model insights to systematically increase the proportion of fully immunized children? The output of this question was to provide recommendations to increase fully vaccinated statuses among children in Nakifuma.

## 1.5 Objectives of the Study

This study had three interconnected objectives designed to generate specific, actionable outputs.

1. To identify factors associated with zero-dose, partially vaccinated, and fully immunized children.
2. To identify a predictive model that accurately determined children's vaccination status using identified factors in the study area.
3. To propose evidence-based strategies to increase vaccination uptake to fully immunized status in Nakifuma Subcounty.

## 1.6 Hypotheses

The study was guided by three specific hypotheses corresponding to each objective.

Null Hypothesis ($H_{0_1}$): There is no significant association between socioeconomic, demographic, geographic, and healthcare access factors and children's vaccination status.

Alternative Hypothesis ($H_{a_2}$): Machine learning models, specifically supervised algorithms like Random Forest, can accurately predict children's vaccination status using the identified factors in the study area.

Null Hypothesis ($H_{0_3}$): Evidence-based strategies informed by predictive model insights and factor analysis findings do not significantly increase the proportion of fully immunized children.

## 1.7 Contributions of the Study

This research was expected to make specific contributions aligned with the three main objectives.

To knowledge, the study aimed for a comprehensive understanding of the specific factors that differentiated vaccination statuses in rural Ugandan contexts, a methodological contribution to applying machine learning for vaccination status prediction in resource-limited settings, and evidence on the effectiveness of data science approaches for precision public health interventions.

To practice, the research aimed to provide a validated machine learning model for identifying vaccination status that could be implemented by health workers and program managers, a practical framework for transitioning children from incomplete to complete vaccination status, and a scalable methodology that could be adapted to similar communities across Uganda and Sub-Saharan Africa.

To policy, the study sought to provide evidence-based recommendations for resource allocation targeting zero-dose and partially vaccinated children, policy guidance for implementing predictive analytics in routine immunization programs, and strategic insights for achieving universal vaccination coverage through precision interventions.

Finally, to the community, the research was expected to lead to direct improvement in childhood vaccination coverage in Bandaali ward and Kizungu through targeted interventions, enhanced equity in vaccination access through identification and support of the most vulnerable children, and sustainable improvement in community health outcomes through evidence-based vaccination strategies. Each contribution directly supported the overarching goal of increasing fully immunized children while addressing the specific gaps identified in zero-dose and partially vaccinated populations.

# Chapter 2

# Literature Review

## 2.1 Overview of Childhood Vaccination

Vaccination has been recognized as one of the most successful public health measures globally for safeguarding people from vaccine-preventable diseases (World Health Organization, 2024). Despite its proven efficacy, global efforts face considerable setbacks, with millions of children remaining "zero-dose" (having received no routine vaccinations) or partially vaccinated, increasing their vulnerability to diseases. These persistent challenges stem from a complex interplay of factors, including socioeconomic inequities, and geographic inaccessibility, as well as issues with access to health facilities and services, and the influence of cultural beliefs and religious practices (World Health Organization, 2024).

### 2.1.1 Defining Zero-Dose, Partially Vaccinated, and Fully Immunized Children

The concept of "zero-dose" children, defined as those who have not received any routine vaccinations, is central to understanding immunization gaps (Cata-Preta et al., 2021; World Health Organization, 2024). Similarly, "partially vaccinated" children are those who have received some vaccinations but have not completed their full immunization schedule. Fully immunized children have received all age-appropriate vaccines by their first birthday, achieving optimal protection against vaccine-preventable diseases.

A consistent variable used to establish a child's immunization status (zero-dose, partially vaccinated, or fully immunized) is the presence and completeness of a vaccination card, reflecting doses received at different age brackets (Qazi et al., 2020). For this research, the vaccination information, including specific vaccine doses as shown in Table 1 below, will be core indicators for determining a child's status.

Studies, such as that by (Qazi et al., 2020) in Pakistan, highlight the complexity of record-keeping and tracking child vaccinations, emphasizing the importance of vaccination cards for establishing immunization status. In contexts like Uganda, challenges in record-keeping, often attributed

Table 2.1: Vaccination Schedule for Children in Uganda

| Vaccine | Age at Dosing | Dose Count |
|---|---|---|
| BCG | At birth | 1 |
| OPV-0 | At birth | 1 |
| DTP-HepB-Hib-1 (Pentavalent-1) | 6 weeks | 1 |
| PCV-1 | 6 weeks | 1 |
| Rota-1 | 6 weeks | 1 |
| DTP-HepB-Hib-2 (Pentavalent-2) | 10 weeks | 1 |
| PCV-2 | 10 weeks | 1 |
| Rota-2 | 10 weeks | 1 |
| DTP-HepB-Hib-3 (Pentavalent-3) | 14 weeks | 1 |
| PCV-3 | 14 weeks | 1 |
| IPV | 14 weeks | 1 |
| Measles-Rubella-1 (MR-1) | 9 months | 1 |
| Vitamin A | 9 months | 1 |
| Measles-Rubella-2 (MR-2) | 18 months | 1 |

to staff shortages and a lack of awareness regarding the importance of complete data capture, can lead to incomplete medical records. The manual nature of data capture in many health facilities, relying on tools like register books and tally sheets, further exacerbates these issues, making precise patient-level data retrieval cumbersome and contributing to children slipping into under-immunized categories. Understanding the ”immunization cascade” and the pathways that lead to children becoming zero-dose or partially vaccinated is crucial for designing effective interventions (Cata-Preta et al., 2021).

## 2.2 Factors Associated with Children's Vaccination Status

### 2.2.1 Child-Specific Demographic Factors

Child age represents a critical determinant of vaccination status, with age-specific patterns reflecting both vaccination schedules and cumulative dropout rates. The combined age in months provides precise measurement for vaccination eligibility and schedule adherence assessment. Younger children may appear incompletely vaccinated simply due to not yet reaching the age for certain vaccines, while older children who remain incompletely vaccinated often represent true vaccination failures or missed opportunities (Adamu et al., 2019). The relationship between child age and vaccination status follows complex patterns, with some studies showing improved coverage with increasing age as children have more opportunities to receive delayed vaccines, while others demonstrate declining coverage due to cumulative dropout from the vaccination program

(Bangura et al., 2020).

Child gender shows variable associations with vaccination coverage depending on cultural context and regional practices. While some studies identify male preference in vaccination (Rahman & Obaida-Nasrin, 2010), others find no significant gender differences in contemporary settings (Wasswa et al., 2025). In Uganda, gender disparities appear minimal in recent surveys, though regional variations may persist, particularly in communities with strong cultural preferences for male children (Uganda Bureau of Statistics, 2017).

### 2.2.2 Household Composition and Caregiver Characteristics

The number of children in a household significantly influences vaccination patterns through resource allocation mechanisms and caregiver attention distribution. Households with more children often demonstrate lower vaccination completion rates for later-born children due to competing demands on caregiver time and household resources (Antai, 2009). Each additional child in the household may reduce the probability of complete vaccination for all children as caregivers struggle to manage multiple vaccination schedules and healthcare appointments.

Caregiver age influences vaccination decisions and follow-through patterns. Younger caregivers, particularly those under 20 years, often demonstrate lower rates of vaccination completion due to limited experience with healthcare systems and reduced confidence in navigating vaccination schedules (Bangura et al., 2020). Conversely, adolescent caregivers may be more receptive to health education messages, while older caregivers bring experience but may face competing demands from larger households or established beliefs that resist new health information.

Caregiver relationship to the child fundamentally affects vaccination responsibility and decision-making authority. Biological mothers typically demonstrate the highest rates of vaccination completion compared to other caregivers including fathers, grandparents, or other relatives (Adamu et al., 2019). This pattern reflects maternal primary responsibility for child healthcare in most cultural contexts and greater emotional investment in child welfare outcomes. Non-parental caregivers may face decision-making constraints or lack complete information about the child's vaccination history.

Caregiver education emerges as one of the most critical predictors of childhood vaccination coverage across diverse settings. Caregivers with secondary or higher education demonstrate significantly higher rates of child immunization completion compared to those with no formal

education (Antai, 2009). The mechanism operates through multiple pathways including improved health literacy, better understanding of vaccine importance, enhanced ability to navigate healthcare systems, and greater confidence in interacting with health workers (Bbaale, 2013). In Uganda specifically, caregiver secondary education increases odds of complete immunization by approximately 60% compared to no education.

Caregiver occupation and employment status significantly influence vaccination patterns through multiple mechanisms. Formal sector employment correlates with higher immunization rates due to better healthcare access, regular income streams, and often workplace health insurance or benefits (Sserwanja et al., 2022). Agricultural occupations, which dominate rural Uganda, present unique challenges including seasonal work patterns, time constraints during planting and harvesting seasons, and variable income that affects healthcare affordability. Healthcare workers and teachers as caregivers typically demonstrate very high vaccination completion rates due to their professional knowledge and access to health information.

### 2.2.3 Socioeconomic Determinants

Household income represents one of the most consistently documented determinants of childhood vaccination coverage globally. Higher income households demonstrate significantly better vaccination completion rates compared to low-income families (Ssebagereka et al., 2024). The relationship operates through improved ability to afford transportation costs, opportunity costs of time spent seeking vaccination, indirect costs such as food and accommodation during facility visits, and better overall access to quality health services (Ssebagereka et al., 2024). In resource-constrained settings, even small costs associated with vaccination can create substantial barriers for the poorest households.

### 2.2.4 Geographic and Access Factors

Residential area represents a fundamental determinant of vaccination access and completion, with urban children consistently demonstrating higher vaccination rates compared to rural children across sub-Saharan Africa (Bangura et al., 2020). Rural areas face multiple challenges including greater distances to health facilities, limited transportation options, seasonal accessibility issues, and often less reliable vaccine supply systems. Urban areas typically offer better healthcare infrastructure, shorter distances to services, and more diverse transportation options.

Distance to health facilities shows consistent negative correlation with immunization rates across

multiple African contexts. Each additional kilometer of distance significantly reduces the probability of vaccination completion, particularly affecting rural populations where transportation costs and time investments create substantial barriers for families (Obanewa & Newell, 2017). The relationship is particularly pronounced for multi-dose vaccines that require multiple facility visits, as distance barriers compound with each required visit.

Mode of transport to health facilities significantly affects vaccination completion rates. Families with access to motorized transport (personal vehicles, motorcycles, or public transport) achieve higher vaccination coverage compared to those relying on walking (Ministry of Health Uganda, 2019). In rural Uganda, motorcycle taxis (boda-boda) serve as crucial transport links to health facilities, though costs can be prohibitive for frequent vaccination visits. Transportation challenges interact with distance to create multiplicative barriers, particularly during rainy seasons when road conditions deteriorate.

Geographic coordinates (latitude and longitude) provide precise location data that enables spatial analysis of vaccination patterns and identification of geographic clusters of low coverage. Geographic information systems analysis reveals spatial clustering of undervaccinated children, often related to distance from health facilities, terrain challenges, or community-specific factors that create localized barriers to vaccination access (PATH, 2018).

### 2.2.5 Documentation and Health System Factors

The availability and quality of a child's vaccination card or record are fundamental to achieving full immunization. A vaccination card serves as a crucial tool for both caregivers and health workers to track a child's vaccination schedule and ensure they receive all doses at the correct intervals (Uganda Ministry of Health, 2020). The absence of a card, loss of the card, or incomplete records are significant barriers, often leading to missed doses or re-vaccination, which can waste resources and cause confusion. Studies from Uganda confirm that possession of a health card is a strong predictor of complete vaccination coverage (Serwadda et al., 2014).

In Uganda, challenges in record-keeping, often attributed to staff shortages and a lack of awareness regarding the importance of complete data capture, can lead to incomplete medical records. The manual nature of data capture in many health facilities, relying on tools like register books and tally sheets, further exacerbates these issues, making precise patient-level data retrieval cumbersome and contributing to children slipping into under-immunized categories.

### 2.2.6 Knowledge, Attitudes, and Beliefs

Caregivers' knowledge, attitudes, and beliefs about vaccines are central to their vaccination decisions. Limited knowledge about the importance of vaccines, the national immunization schedule, and the diseases they prevent can lead to low uptake (Ssempiira et al., 2021). Negative attitudes, often fueled by misinformation or rumors, can result in vaccine hesitancy or refusal (Uganda Ministry of Health, 2021). A caregiver's belief in the safety and efficacy of vaccines is a critical predictor of their child's immunization status.

### 2.2.7 Water, Sanitation, and Hygiene (WASH) Factors

Access to improved water sources correlates with better vaccination coverage through multiple mechanisms. Households with improved water access typically demonstrate better overall health-seeking behavior, may have higher socioeconomic status, and experience reduced time burdens for water collection that free caregivers for healthcare activities including vaccination visits (Bbaale, 2013). The relationship also reflects general health consciousness and investment in preventive health measures.

Toilet facility type serves as an indicator of household socioeconomic status, health awareness, and investment in sanitation infrastructure. Households with improved sanitation facilities consistently demonstrate higher vaccination rates compared to those using shared or unimproved facilities (Sserwanja et al., 2022). The relationship operates through socioeconomic pathways, health consciousness, and often correlates with education level and community development indicators.

Availability of handwashing facilities indicates household hygiene practices and health awareness. Families with dedicated handwashing places show higher rates of vaccination completion, potentially reflecting general attention to health and preventive practices (Budu et al., 2020). This variable often correlates with education level, socioeconomic status, and overall health-seeking behaviors.

## 2.3 Predictive Modeling in Vaccination Coverage

**Machine Learning Applications in Public Health**

Machine learning algorithms increasingly demonstrate utility in predicting health outcomes and identifying at-risk populations using comprehensive demographic, socioeconomic, geographic, and behavioral variables. In vaccination contexts, supervised learning approaches enable identification

of children likely to remain incompletely immunized based on readily available indicators including caregiver characteristics, household factors, geographic location, and health system variables (Nantongo et al., 2024).

Random forest algorithms show particular promise in vaccination prediction models using mixed variable types including categorical variables (gender, residential area, caregiver occupation, transport mode), continuous variables (child age, household income, distance to facility, waiting times), and geographic coordinates. These ensemble methods effectively handle missing data, identify complex interactions between variables like caregiver education and distance to facility, and provide interpretable feature importance rankings that guide intervention targeting (Asnake et al., 2025). Applications in immunization prediction achieve area under the curve (AUC) values ranging from 0.75 to 0.92 across different settings.

Support vector machines and logistic regression models also demonstrate acceptable performance in vaccination prediction tasks using similar comprehensive variable sets. Logistic regression offers interpretable coefficients that facilitate policy translation of findings related to specific factors like transport mode, health worker attitudes, or WASH indicators, while support vector machines handle non-linear relationships between variables effectively (Nantongo et al., 2024). Geographic variables can be incorporated through spatial modeling approaches that account for spatial autocorrelation in vaccination outcomes.

### 2.3.1 Performance Metrics and Validation

Predictive model evaluation requires careful consideration of appropriate metrics given vaccination coverage prediction objectives and class distribution in the vaccination status outcome variable. Area under the receiver operating characteristic curve (AUC) provides threshold-independent performance assessment, with values above 0.80 generally considered good discrimination for identifying children at risk of incomplete vaccination (Hosmer et al., 2013).

Accuracy metrics require contextual interpretation given potential class imbalance in vaccination datasets. When fully immunized children represent the majority class, high accuracy may mask poor sensitivity for identifying incompletely vaccinated children who are the primary target for interventions (Saito & Rehmsmeier, 2015). Precision, recall, and F1-scores provide more balanced performance assessment for vaccination prediction models, particularly when identifying zero-dose or partially vaccinated children is the priority.

Cross-validation approaches ensure model generalizability across different populations, time periods, and geographic areas. K-fold cross-validation, stratified sampling, and spatial cross-validation help assess model performance on unseen data, crucial for deployment in public health settings where models must perform reliably across diverse demographic groups and geographic contexts (Roberts et al., 2017). Temporal validation using data from different collection periods provides additional confidence in model stability.

### 2.3.2 Feature Selection and Interpretation

Variable selection in vaccination prediction models requires balancing predictive power with data availability, collection feasibility, and policy relevance. Core variables like caregiver education, household income, residential area, distance to facility, and vaccination card possession typically provide strong predictive signal while remaining readily collectible in routine health surveys and programmatic data systems (Nantongo et al., 2024).

Feature importance analysis reveals which factors most strongly predict vaccination status in specific contexts. Variables related to healthcare access (transport mode, distance to facility, health worker attitude, waiting time) often emerge as strong predictors alongside socioeconomic factors (caregiver education, household income) and demographic characteristics (child age, number of children in household). Knowledge and attitude variables (vaccine importance, vaccination concerns) may provide additional predictive value, particularly in settings with significant vaccine hesitancy (Lundberg & Lee, 2017).

Interaction effects between variables often enhance model performance and reveal important policy insights. Caregiver education may interact with residential area and distance to facility, creating distinct risk profiles for rural versus urban populations with different education levels. Similarly, household income effects may vary by transport mode availability and distance, suggesting that access interventions may have differential impacts across socioeconomic groups (Asnake et al., 2025). Geographic coordinates enable spatial modeling of vaccination patterns and identification of geographic clusters requiring targeted interventions.ns.

### 2.3.3 Spatial Analysis in Vaccination Modeling

Geographic coordinates (latitude and longitude) provide precise location data that enables spatial analysis of vaccination patterns and identification of geographic clusters of low coverage. Geographic information systems (GIS) analysis reveals spatial clustering of undervaccinated children, often

related to distance from health facilities, terrain challenges, or community-specific factors that create localized barriers to vaccination access (PATH, 2018). Integrating spatial data with machine learning models allows for the development of spatially explicit predictive models that can guide resource allocation to specific, high-risk areas.

## 2.4 Evidence-Based Strategies for Improving Vaccination Coverage

### 2.4.1 Targeted Intervention Approaches

Risk-stratified intervention strategies demonstrate superior cost-effectiveness compared to universal approaches. Identifying high-risk children through predictive models using available demographic, socioeconomic, geographic, and health system variables enables resource concentration on families most likely to remain incompletely vaccinated (Nantongo et al., 2024). Targeting based on caregiver education, residential area, distance to facility, household income, and health system factors maximizes impact within constrained resource environments.

Documentation interventions addressing vaccination card possession show consistent positive effects on completion rates. Providing vaccination cards, training caregivers on their importance, and implementing card replacement programs increase follow-through for subsequent vaccination visits by 15-30% across diverse settings (Binyaruka & Borghi, 2018). Digital vaccination records and SMS reminder systems can supplement physical cards in areas with mobile network coverage.

Transportation support interventions address key access barriers identified in predictive models. Providing transport vouchers, organizing mobile vaccination services, establishing community-based vaccination points, or improving transportation infrastructure reduces distance and transportation barriers that particularly affect rural populations and low-income households (Perry et al., 2017). Geographic targeting using spatial analysis ensures interventions reach areas with greatest need.

### 2.4.2 Health System Strengthening

Supply chain optimization represents a foundational requirement for vaccination coverage improvement. Ensuring consistent vaccine stock availability at health facilities directly determines coverage potential, making stock management a priority intervention area (Mvundura et al., 2015). Predictive models incorporating stock status variables can identify facilities at highest risk of stock-outs for targeted supply chain interventions and emergency resupply protocols.

Provider training focusing on attitude improvement, communication skills, and service efficiency

enhances vaccination service quality and coverage. Competency-based training addressing respectful patient interaction, clear communication about vaccine importance, efficient service delivery, and reduced waiting times improves caregiver satisfaction and likelihood of vaccination completion (Ryman et al., 2008). Regular supervision with feedback maintains skill levels, positive attitudes, and adherence to service delivery standards.

Service delivery optimization reduces waiting times, improves efficiency, and addresses health system barriers identified in predictive models. Implementing appointment systems, improving workflow organization, ensuring adequate staffing during high-demand periods, and streamlining registration and record-keeping processes address key barriers that influence vaccination completion (Phillips et al., 2017). Quality improvement approaches can systematically address service delivery factors that influence vaccination outcomes.

### 2.4.3 Community Engagement Strategies

Education campaigns targeting knowledge gaps about vaccine importance demonstrate effectiveness in improving coverage rates among caregivers with low awareness or negative attitudes. Culturally appropriate messaging delivered through trusted community channels addresses specific concerns and misconceptions while reinforcing positive attitudes toward vaccination (Larson et al., 2014). Message content should address local vaccination reasons, concerns, and challenges identified through community assessment.

Addressing vaccination concerns through community dialogue and targeted education reduces hesitancy and improves completion rates. Identifying and responding to specific safety concerns, religious objections, or cultural barriers through respectful communication, evidence-based information, and engagement with trusted community leaders builds confidence in vaccination programs (Gavi, the Vaccine Alliance, 2023). Peer education and caregiver support groups leverage social networks for positive influence.

Religious and cultural leader engagement leverages existing authority structures to promote vaccination and address religious or cultural influences that may impede vaccination uptake. Leader endorsement, participation in vaccination campaigns, and integration of vaccination promotion into religious or cultural activities significantly increases community acceptance (Adamu et al., 2019). This approach requires careful relationship building and ongoing engagement with diverse community leadership structures.

### 2.4.4 WASH and Nutrition Integration

Given the link between WASH, nutrition, and health outcomes, integrating these services with immunization programs can lead to better health outcomes. For example, vaccination clinics can also provide nutritional supplements and health education on hygiene practices. This integrated approach not only improves vaccination coverage but also addresses the broader determinants of child health, leading to more sustainable improvements in community well-being.

### 2.4.5 Geographic and Access Interventions

To overcome geographic barriers, strategies can include implementing mobile vaccination clinics that reach remote communities, optimizing vaccination schedules to reduce the number of facility visits, and leveraging GIS data to plan outreach activities and ensure that resources are allocated to the areas with the highest need (PATH, 2018). Addressing transportation challenges, for instance by partnering with local transport providers, can also significantly improve access to vaccination services.

## 2.5 Theoretical Frameworks

### 2.5.1 Health Belief Model Application

The Health Belief Model (HBM) provides a useful framework for understanding caregiver decisions regarding childhood vaccination. The model posits that health behaviors are influenced by an individual's perception of a health threat and their perceived benefits and barriers to taking action (Rosenstock, 1974). In the context of this study, the HBM can be used to analyze how caregivers' perceptions of the severity and susceptibility to vaccine-preventable diseases, the perceived benefits of vaccination, and the perceived barriers (e.g., costs, distance, fear of side effects) influence their decision to vaccinate their children. This framework helps to identify specific beliefs and perceptions that need to be addressed through public health interventions.

### 2.5.2 Social Ecological Model

The Social Ecological Model (SEM) provides a comprehensive framework for understanding health behaviors by considering the complex interplay between individual, interpersonal, organizational, community, and policy factors (McLeroy et al., 1988). In the context of vaccination, the SEM helps to analyze the multiple levels of influence on a child's immunization status. At the individual

level, factors include the child's age and gender and the caregiver's knowledge and beliefs. At the interpersonal level, social networks and family support can influence decisions. At the community level, factors include access to health facilities and community beliefs. Finally, at the policy level, national vaccination schedules and health policies play a crucial role. This model provides a holistic view, helping to design multi-level interventions that address the diverse barriers to vaccination.

## 2.6 Research Gaps and Study Justification

While the literature extensively documents the factors influencing childhood vaccination coverage, several gaps remain that this study aims to address. First, there is a lack of localized studies that use advanced data science methodologies to understand vaccination patterns at the community level in rural contexts like Bandaali ward and Kizungu. Existing studies often rely on broad national surveys, which may not capture the unique challenges and dynamics of specific local communities.

Second, although predictive modeling is increasingly used in public health, its application to tracking zero-dose and partially vaccinated children in a resource-limited setting like Uganda is still nascent. There is a need for a validated, practical model that can be implemented by local health authorities to guide their day-to-day immunization activities.

Finally, while many studies recommend general interventions, there is a lack of evidence-based strategies that are specifically tailored to the insights derived from a predictive model. This study will bridge this gap by not only identifying the factors and creating a model but also by proposing actionable strategies that are directly informed by the model's outputs, ensuring that interventions are precise, efficient, and impactful.

# Chapter 3

# Research Methodology

## 3.1 Introduction

This chapter outlines the research methodology employed to achieve the three primary objectives of this study: identifying factors associated with different vaccination statuses among children in Nakifuma Subcounty, developing a predictive model for vaccination status determination, and proposing evidence-based strategies to improve vaccination coverage. The methodology adopts a mixed-methods approach, integrating quantitative analysis with qualitative insights to provide a comprehensive understanding of childhood vaccination patterns and their determinants.

## 3.2 Study Population

Nakifuma-Nagalaama Town Council, located within Mukono District, has an estimated total population of 39,700 people based on Uganda Bureau of Statistics (UBOS) projections. The target population for this study consists of children under 5 years, estimated using Uganda's demographic trends where 16-18% of the population are under 5 years.

$$\text{Children Under 5} = 39,700 \times 0.17 \approx 6,750 \tag{3.1}$$

For practical data collection purposes, the study focused on a representative subset of 1,000 children under 5 (approximately 15% of the estimated under-5 population) within the town council's health center catchment areas (Uganda Bureau of Statistics, 2024). This population represents the accessible target population from which the study sample was drawn.

## 3.3 Sample Size Calculation

The sample size ($n$) was calculated using the finite population correction formula (Kadam & Bhalerao, 2010):

$$n = \frac{NZ^2p(1-p)}{d^2(N-1) + Z^2p(1-p)} \tag{3.2}$$

Where:

- $N = 1,000$ (target population size)
- $Z = 1.96$ (95% confidence level)
- $p = 0.5$ (maximum variability)
- $d = 0.05$ ($\pm$5% margin of error)

$$n = \frac{1000 \times 3.8416 \times 0.25}{0.0025 \times 999 + 3.8416 \times 0.25} = \frac{960.4}{2.4975 + 0.9604} \approx 278 \tag{3.3}$$

This calculation indicated a minimum required sample size of 278 children to achieve the desired precision and confidence level for the study findings.

### 3.3.1 Achieved Sample Size

Despite the calculated requirement for 278 participants, the study achieved a sample size of 115 children due to data collection constraints and field challenges. The achieved sample size represents approximately 41% of the calculated minimum requirement and remains adequate for exploratory analysis and insights into vaccination patterns within the study area.

## 3.4 Research Design

This study employed a cross-sectional analytical design utilizing secondary data analysis of vaccination records collected in Nakifuma Subcounty. The research design incorporates both descriptive and inferential statistical methods, complemented by machine learning techniques for predictive modeling. The mixed-methods approach allows for triangulation of findings, enhancing the validity and reliability of conclusions drawn from the analysis. The research follows a systematic four-phase analytical framework designed to address each study objective sequentially while building upon previous findings. This structured approach ensures comprehensive coverage of all research questions while maintaining methodological rigor throughout the analytical process.

## 3.5 Data Source and Variables

The analysis utilized a comprehensive primary dataset containing vaccination records and associated demographic, socioeconomic, and health system variables collected directly from caregivers and children in Nakifuma Subcounty. The final analytical dataset comprised 115 children under 5 years and their caregivers. The dataset encompasses multiple dimensions of factors that potentially influence vaccination outcomes.

### 3.5.1 Variable Classification

- **Numerical Variables:** The dataset includes continuous and discrete numerical variables such as `child_age_years`, `child_age_months`, `children_in_household`, `caregiver_age`, `distance_to_facility`, `latitude`, `longitude`, `muac` (mid-upper arm circumference), and a derived variable combining child age in months.
- **Categorical Variables:** Binary and ordinal categorical variables encompass `child_gender`, `caregiver_relationship`, `caregiver_education`, `caregiver_occupation`, `household_income`, `residential_area`, `transport_mode`, `has_vaccination_card`, and individual vaccination records (`bcg`, `hepb0`, `polio0-3`, `dpt1-3`, `pcv1-3`, `rota1-2`, `ipv1-2`, `measles1-2`). Additional categorical variables include `vaccine_stock_status`, `health_worker_attitude`, `waiting_time`, `vaccine_importance`, `aware_side_effects`, `vaccination_concerns`, `religious_cultural_influence`, `water_source`, `toilet_facility`, `handwashing_place`, and `muac_status`.
- **Text Variables:** Open-ended responses captured in `vaccination_reasons`, `vaccination_challenges`, `service_improvement_suggestions`, and `additional_comments` provide qualitative insights into caregiver perspectives and experiences.

## 3.6 Analytical Framework

The research methodology is structured into four distinct phases, each contributing to specific study outputs while building upon previous analytical findings. Table 3.1 presents the systematic research methodology framework employed in this study.

Table 3.1: Systematic Research Methodology Framework

| Phase | Steps | Description | Outputs |
|---|---|---|---|
| **1. Data Preparation** | | | |
| | 1.1 Data Quality | Load dataset, inspect structure, types, completeness. Generate statistics. | Initial insights. |
| | 1.2 Preprocessing | Handle missing values, create derived variables, validate consistency. | Clean dataset. |
| | 1.3 Status Categorization | Define vaccination schedules, categorize children (Zero/Partial/Full). | Target variable. |
| **2. Factor Analysis** | | | |
| | 2.1 Univariate Analysis | Analyze numerical and categorical variables across groups. | Group characteristics. |
| | 2.2 Statistical Testing | Chi-square, Mann-Whitney U, ANOVA for significant associations. | Significant factors. |
| | 2.3 Qualitative Analysis | Word clouds and thematic analysis on text variables. | Qualitative themes. |
| | 2.4 Geospatial Analysis | Map vaccination status using coordinates. | Geographic patterns. |
| | 2.5 Feature Ranking | Random Forest feature importance ranking. | Prioritized factors. |
| **3. Model Development** | | | |
| | 3.1 Model Training | Train classifiers (Logistic Regression, RF, GB, SVM) with 70:30 split. | Trained models. |
| | 3.2 Model Evaluation | Classification reports, confusion matrices, performance metrics. | Performance metrics. |
| | 3.3 Model Selection | Choose best model, create risk scoring system. | Optimal model. |
| **4. Recommendations** | | | |
| | 4.1 Barrier Prioritization | Rank barriers by significance, frequency, modifiability. | Barrier hierarchy. |
| | 4.2 Risk Identification | Identify high-risk children and communities. | At-risk populations. |
| | 4.3 Intervention Strategy | Design interventions for individual, system, community levels. | Action strategies. |

## 3.7 Phase 1: Data Preparation and Target Variable Creation

### 3.7.1 Data Quality Assessment

The initial phase commenced with comprehensive data quality assessment involving loading the vaccination dataset into a pandas DataFrame for systematic analysis. Data structure inspection included examination of variable types, missing value patterns, and basic descriptive statistics to understand the dataset's characteristics and identify potential data quality issues requiring attention.

### 3.7.2 Data Preprocessing

Data preprocessing was performed using Python (version 3.13.2) (Python Software Foundation, n.d.) and the pandas library (version 2.3.1) (The pandas development team, n.d.) to prepare the dataset for predictive modeling. The workflow included three primary steps. First, to address data sparsity, columns with over 50 missing values were removed using the `pandas.DataFrame.drop()` function. Second, for the remaining categorical variables, missing values were imputed with the mode, determined using `pandas.Series.mode()` and filled using `pandas.DataFrame.fillna()`. Lastly, a new derived variable, `combined_age_in_months`, was created to consolidate child age information from `child_age_years` and `child_age_months`.

### Key Columns Retained

A comprehensive set of columns was retained for analysis, categorized as follows:

- **Demographics:** `child_age_years`, `child_age_months`, `caregiver_age`, `caregiver_relationship`, `caregiver_education`, `caregiver_occupation`, `household_income`, `residential_area`
- **Health and Access:** `distance_to_facility`, `muac`, `muac_status`, `has_vaccination_card`
- **Vaccination Records:** `bcg`, `hepb0`, `polio0`, `polio1`, `polio2`, `polio3`, `dpt1`, `dpt2`, `dpt3`, `pcv1`, `pcv2`, `pcv3`, `rota1`, `rota2`, `ipv1`, `ipv2`, `measles1`, `measles2`
- **Contextual Factors:** `vaccine_stock_status`, `health_worker_attitude`, `waiting_time`, `vaccine_importance`, `aware_side_effects`, `vaccination_concerns`, `religious_cultural_influence`, `water_source`, `toilet_facility`, `handwashing_place`

- **Derived:** `combined_age_in_months`, `vaccination_status`

### 3.7.3 Categorical Variable Transformation

## The Rationale for One-Hot Encoding

To prepare the dataset for machine learning models, all categorical variables were transformed into a numerical format using One-Hot Encoding. This process is crucial because most statistical and predictive models cannot directly process string values; they require numerical input for mathematical operations. The original columns are replaced by multiple binary columns, each indicating the presence (1) or absence (0) of a specific category for every observation. In summary, every categorical variable was expanded into several binary columns, each representing a possible value (except the reference), making the data suitable for statistical analysis and predictive modeling without implying a false sense of order or magnitude.

**Before Transformation** Each categorical variable, such as `dpt1`, `caregiver_education`, and `vaccine_stock_status`, existed as a single column with string values. The data resembled the following structure:

Table 3.2: Original State of Categorical Variables

| dpt1 | caregiver_education | vaccine_stock_status |
|---|---|---|
| Received | Secondary | Always |
| Not Received | Primary | Sometimes |
| select | Tertiary | Never |

**After Transformation** One-Hot Encoding expanded these columns into several binary columns. For example, the `dpt1` column with categories "Received", "Not Received", and "select" was transformed into `dpt1_Not Received` and `dpt1_select`. The "Received" category served as the reference and was represented when both new columns were 0. This transformation eliminates the ordinal relationship between categories that is not explicitly present in the data. The transformed data had a structure similar to the example below:

**Summary of All Transformed Variables** A total of 21 categorical variables were transformed in this manner. The process created numerous new binary columns, making the data suitable for the predictive modeling phase. Table 3.4 provides a detailed summary of each categorical

Table 3.3: After One-Hot Encoding

| dpt1_Not Received | dpt1_select | ..._Primary | ..._Secondary | ..._Tertiary | ..._Sometimes | ..._Never |
|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| 0 | 1 | 0 | 0 | 1 | 0 | 1 |

variable and the new columns generated.

## 3.8 Vaccination Status Categorization and Derivation

A critical component of the methodology involved developing a robust categorization system for a child's vaccination status. This was done to create a meaningful target variable for our predictive models. The process involved two main stages: defining a reference standard and implementing a hierarchical classification logic.

### 3.8.1 1. Defining the Reference Standard

The first step was to establish a clear reference point based on Uganda's national immunization schedule. This was done by creating a data structure that mapped each specific vaccine (e.g., BCG, Polio0, DPT1) to its corresponding recommended age of administration. This reference schedule was crucial for accurately assessing whether a child had received all age-appropriate vaccines.

### 3.8.2 2. Implementing the Classification Logic

A custom logic was developed to process each child's data based on their age and vaccination records. This logic followed a hierarchical set of rules to ensure accurate and non-overlapping classification into three distinct groups:

- **Zero-Dose:** A child was classified as "Zero dose" if they had received no vaccines at all. This definition was further refined to include children who had not received the critical first dose of DPT (DPT1), a key indicator used globally to identify children completely missed by the routine immunization program.
- **Fully Immunized:** Children who were not classified as "Zero dose" were then evaluated for "Fully Immunized" status. This was a stringent classification, requiring that a child be at least 12 months old and have a record of receiving every single vaccine recommended up to that age.

- **Partially Vaccinated:** A child was categorized as "Partially vaccinated" if they had received at least one vaccine but failed to meet the strict criteria for "Fully Immunized." This group represents children who have engaged with the health system but have not completed their full vaccination schedule.

This classification logic was systematically applied to the entire dataset to create a new, definitive `vaccination_status` variable, which became the central outcome of the study.

## 3.9 Phase 2: Factor Identification Analysis

### 3.9.1 Univariate Analysis

Univariate analysis involved systematic examination of variables within each vaccination status group.

**Numerical Variables** For numerical variables, the analysis included distribution visualization using box plots and histograms. Box plots were generated for each numerical variable, such as child age years, caregiver age, distance to facility,muac, and combined age in months. These plots visually compared the median, quartiles, and outliers across vaccination status groups ("Zero dose", "Partially vaccinated", and "Fully Immunized"). Additionally, overlaid histograms were produced to visualize the frequency and spread of values within each group using separate color-coded distributions.

**Categorical Variables** For categorical variables, bar charts and stacked bar charts were created to show the distribution of each category across vaccination status groups. Bar charts (count plots) were used for variables such as `dpt1`, `measles1`, `caregiver_relationship`, and `has_vaccination_c`. These plots highlighted differences in counts and percentages. Stacked bar charts were also used to provide a side-by-side comparison of the frequency and percentage distribution of categories.

**Text Variables** For open-ended or text variables, word clouds were generated. These visualizations for variables such as `vaccination_reasons`, `vaccination_challenges`, and `service_improvement_sugg` highlighted the most frequent words or themes mentioned by caregivers within each vaccination status group.

### 3.9.2 Statistical Association Testing

Statistical testing employed appropriate tests based on variable types and distributions. Specifically, chi-square tests of independence were conducted for categorical variables (e.g., ‘dpt1‘, ‘measles1‘,‘caregiver education‘, ‘vaccine stock status‘) to assess associations with vaccination status. For numerical variables such as ‘child age years‘, ‘caregiver age‘, and ‘distance to facility‘, Mann-Whitney U tests were applied to identify significant differences across vaccination status groups. For instance, the Mann-Whitney U test comparing ‘child age years‘ between Fully Immunized and Partially Vaccinated groups. All tests employed a significance level of $p < 0.05$.Variables with p-values below this threshold were considered to have statistically significant associations or differences with vaccination status.

### 3.9.3 Qualitative Analysis

Qualitative data analysis focused on text variables containing caregiver responses regarding vaccination experiences, challenges, and suggestions. The key text variables identified were `vaccination_reasons`, `vaccination_challenges`, and `service_improvement_suggestions`. For each vaccination status group ("Zero dose", "Partially vaccinated", and "Fully Immunized"), word clouds were generated. These visualizations facilitated thematic analysis, a widely used method for qualitative research (Braun & Clarke, 2006), by highlighting the most frequent words and phrases, which in turn helped identify common themes, barriers, and facilitators. This analysis provided a rich, contextual understanding of quantitative findings and insights into caregiver perspectives and experiences.

### 3.9.4 Geospatial Analysis

Geospatial analysis utilized latitude and longitude coordinates to map vaccination status distribution across Nakifuma Subcounty. This analysis identified geographic clusters of different vaccination statuses, examined relationships between distance to health facilities and vaccination outcomes, and provided insights into spatial patterns of vaccination coverage that inform targeted intervention strategies.

### 3.9.5 Feature Importance Ranking

Random Forest feature importance analysis was employed to rank variables by their predictive power for vaccination status determination. This analysis provided an objective, data-driven

ranking of factors contributing to vaccination outcomes, complementing statistical testing results and informing subsequent predictive modeling efforts.

## 3.10 Phase 3: Predictive Model Development

### 3.10.1 Model Selection and Training

Multiple machine learning algorithms were implemented to identify the most effective approach for vaccination status prediction. Models included Logistic Regression for interpretability, Support Vector Machine for complex decision boundaries, Gradient Boosting Classifier for ensemble learning benefits, and Random Forest Classifier for robustness and feature importance insights. Categorical variables were one-hot encoded and text columns excluded from modeling. Data was split using stratified sampling (70% training, 30% testing) to maintain class balance across splits. Model performance was evaluated using classification reports, confusion matrices, and ROC curves. Feature importance was assessed for ensemble models.

### 3.10.2 Model Evaluation

Model evaluation in this study included the generation of classification reports (precision, recall, and F1-score) and confusion matrices for each vaccination status class: Zero Dose, Partially Vaccinated, and Fully Immunized. This approach prioritized a balanced assessment of performance across all groups, rather than relying solely on overall accuracy. ROC curves and AUC scores were also produced for all models and target groups, allowing for a visual and quantitative comparison of their discriminative performance.

### 3.10.3 Best Model Selection

Model selection was a comparative process that evaluated multiple algorithms, including Logistic Regression, Support Vector Machines (SVM), Gradient Boosting, and Random Forest. Special emphasis was placed on optimizing precision and recall for the Zero Dose group to support the development of targeted interventions. Comparative analyses using the aforementioned metrics, alongside ROC curves and feature importance reports, guided the identification of the most effective model. While a practical risk-scoring system was not explicitly implemented, the analysis provides a robust foundation for developing such a tool based on the selected model's outputs and feature contributions.

## 3.11 Phase 4: Evidence-Based Recommendations

### 3.11.1 Barrier Analysis and Prioritization

Synthesis of quantitative and qualitative findings enabled comprehensive barrier identification and prioritization. Barriers were ranked considering statistical significance and effect size from quantitative analysis, frequency of occurrence in qualitative responses, and potential modifiability through practical interventions. This prioritization informed resource allocation and intervention targeting strategies.

### 3.11.2 Risk Group Identification

The validated predictive model was applied to identify high-risk children and communities requiring prioritized intervention. Risk scores were calculated for all children in the dataset, enabling systematic identification of those most likely to have incomplete vaccination. Geographic analysis of risk patterns informed spatial targeting of interventions.

### 3.11.3 Intervention Strategy Development

Evidence-based intervention strategies were developed addressing identified barriers at multiple levels. Individual-level interventions targeted caregiver knowledge and attitudes, health system-level interventions addressed service delivery and stock management issues, community-level interventions focused on awareness and cultural factors, and geographic-level interventions addressed access and distance challenges.

## 3.12 Data Analysis Tools

All analyses were conducted using Python programming language with specialized libraries including `pandas` for data manipulation, `numpy` for numerical computations, `matplotlib` and `seaborn` for data visualization, `scikit-learn` for machine learning implementation, and `wordcloud` for qualitative text analysis. Statistical tests were implemented using `scipy.stats`, ensuring robust and reproducible analytical results.

## 3.13 Ethical Considerations

This study involved primary data collection from caregivers and children in Nakifuma Subcounty. Ethical approval was obtained from relevant institutional review boards prior to data collection.

Informed consent was secured from all participants, with all personally identifiable information anonymized during data processing. Data handling followed established ethical guidelines for primary data collection, ensuring participant confidentiality and data security throughout the research process.

## 3.14 Study Limitations

Several limitations should be acknowledged in this methodology. The cross-sectional design limits causal inference, requiring careful interpretation of associations rather than causation. Secondary data analysis constrains variable availability to those originally collected, potentially limiting the exploration of additional relevant factors. Geographic scope limitation to Nakifuma Subcounty may affect generalizability to other contexts, requiring validation in different settings.

**Sample Size Limitations:** The achieved sample size of 115 participants falls significantly below the calculated minimum requirement of 278, representing approximately 41% of the intended sample. This reduction impacts the statistical power of the study, potentially limiting the ability to detect smaller effect sizes and affecting the precision of confidence intervals. To overcome this, the analytical approach shifted to focus on machine learning models, which are less sensitive to power issues and more effective at identifying patterns in smaller datasets.

**Sampling Challenges Impact:** The data collection challenges that resulted in the reduced sample size may have introduced selection bias, as participants who were more accessible or available may differ systematically from those who were not reached. We addressed this by prioritizing unbiased analytical methods, such as Random Forest feature importance and robust statistical tests like the Mann-Whitney U test, to mitigate the impact of potential bias on our key findings.

**Statistical Analysis Considerations:** The reduced sample size necessitates cautious interpretation of statistical tests, particularly for subgroup analyses and predictive modeling. Some planned analyses may have insufficient power to detect meaningful differences, and confidence intervals around estimates will be wider than originally planned. We addressed this by moving from traditional inferential statistics towards an exploratory data analysis approach, supplemented by predictive modeling, which allowed for the identification of key factors even with a smaller sample. The use of a risk-scoring system, derived from the predictive model, provides a pragmatic solution by categorizing children based on their likelihood of being partially or non-vaccinated,

offering actionable insights despite the statistical limitations.

## 3.15 Quality Assurance

Quality assurance measures included systematic validation of data preprocessing steps, verification of statistical test assumptions, cross-validation of predictive models, and triangulation of findings across quantitative and qualitative data sources. All analytical code was documented and version-controlled to ensure reproducibility and transparency of methods.

Table 3.4: Summary of Categorical Variable Transformation

| Original Variable | Example Values | New Columns Created |
|---|---|---|
| **Categorical Variables** | | |
| child_gender | Male, Female | child_gender_Male |
| caregiver_relationship | Mother, Father, Grandparent | caregiver_relationship_Father, caregiver_relationship_Grandparent |
| caregiver_education | No formal, Primary, Secondary, Tertiary | caregiver_education_Primary, caregiver_education_Secondary, caregiver_education_Tertiary |
| caregiver_occupation | House wife, Farmer, Teacher, Tailoring | caregiver_occupation_Farmer, caregiver_occupation_Teacher, caregiver_occupation_Tailoring |
| household_income | Less than 100,000, 100,000-300,000 | household_income_100,000-300,000 |
| residential_area | Rural, Urban, Peri-urban | residential_area_Urban, residential_area_Peri-urban |
| transport_mode | Walking, Bicycle, select | transport_mode_Bicycle, transport_mode_select |
| has_vaccination_card | Yes, No | has_vaccination_card_No |
| bcg, hepb0, polio0 | Received, Not Received, select, Unknown | bcg_Not Received, bcg_select, bcg_Unknown |
| vaccine_stock_status | Always, Sometimes, Never, select | vaccine_stock_status_Sometimes, vaccine_stock_status_Never, vaccine_stock_status_select |
| health_worker_attitude | Very good, Good, Neutral | health_worker_attitude_Good, health_worker_attitude_Neutral |
| waiting_time | 30-60 minutes, 1-2 hours, More than 2 hrs | waiting_time_1-2 hours, waiting_time_More than 2 hrs |
| vaccine_importance | Extremely, Very, Moderately, select | vaccine_importance_Very, vaccine_importance_Moderately, vaccine_importance_select |
| aware_side_effects | Yes, No | aware_side_effects_No |
| vaccination_concerns | Yes, No | vaccination_concerns_No |
| religious_cultural_influence | Yes, No | religious_cultural_influence_No |
| water_source | Borehole, Piped water, select | water_source_Piped water, water_source_select |
| toilet_facility | Pit latrine, select | toilet_facility_select |
| handwashing_place | Yes, No | handwashing_place_No |
| muac_status | Green, select | muac_status_select |
| vaccined_if_no_card | Yes, No | vaccined_if_no_card_No |

# Chapter 4

# Results

## 4.1 Objective 1: Factors Associated with Vaccination Status

This chapter presents the key findings from the analysis of 115 children under five years in Nakifuma Subcounty, structured to address the primary research objectives. The results combine descriptive statistics with inferential testing to provide a comprehensive understanding of factors influencing childhood vaccination status.

### 4.1.1 Sample Characteristics and Vaccination Status Distribution

The study sample comprised 115 children under five years with a vaccination status distribution as shown in Figure 4.1. The distribution was as follows: Zero Dose, 12 children (10.4%); Partially Vaccinated, 47 children (40.9%); and Fully Immunized, 56 children (48.7%). These categories represent children who have received no vaccines, children with incomplete age-appropriate vaccination, and children with complete age-appropriate vaccination, respectively.

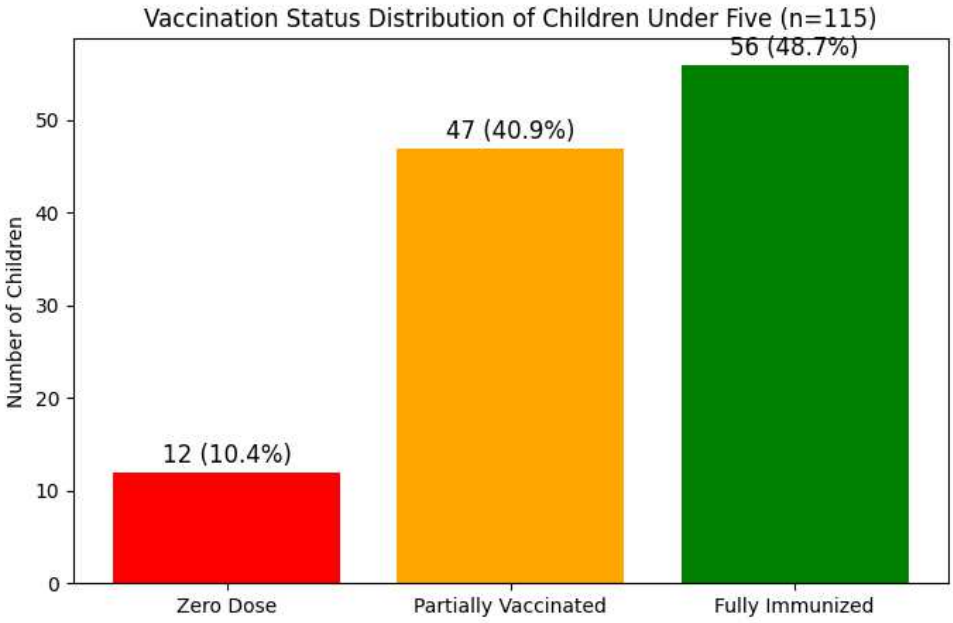



Figure 4.1: Summary of Vaccination Status Across the Study Sample.

### 4.1.2 Zero Dose Children: Barriers to Vaccination Initiation

Children classified as "Zero Dose" represent the most vulnerable group, having received no vaccines despite being within the target age range. As shown in Table 4.1, analysis revealed significant associations with caregiver and health system factors, suggesting multi-level barriers to vaccination initiation. The table shows that Caregiver Relationship, Caregiver Occupation, Health Worker Attitude, Waiting Time, and Vaccine Importance are all statistically significant factors ($p<0.05$). In contrast, demographic factors such as Child Age, Children in Household, and Caregiver Age were not significantly associated with zero-dose status. Figures 4.2 to 4.5 visually represent some of these key findings, with Figure 4.2 highlighting the impact of negative health worker attitudes and Figure 4.3 showing the effect of low perceived vaccine importance. Figure 4.14 illustrates the association with excessive waiting times and Figure 4.5 shows the relationship with caregiver occupational barriers. This suggests that systemic and attitudinal barriers, rather than structural factors, are primary drivers of vaccination non-initiation.

## Factors Associated with Zero-Dose Children


Figure 4.2: Negative Health Worker Attitudes

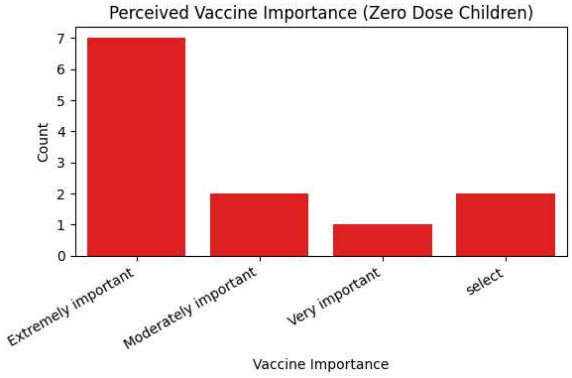



Figure 4.3: Low Perceived Vaccine Importance

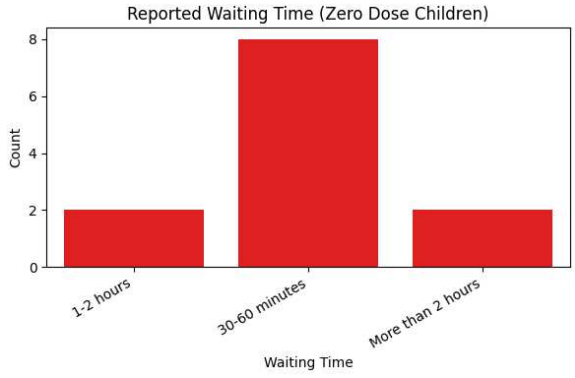



Figure 4.4: Excessive Waiting Times

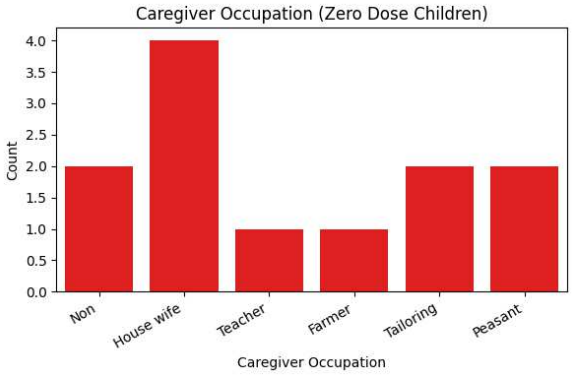



Figure 4.5: Occupational Barriers

Table 4.1: Factors Associated with Zero-Dose Children

| Variable | P-value | Interpretation |
|---|---|---|
| **Categorical Variables** | | |
| Caregiver Relationship | $< 0.05$ | Strong association; non-parental caregivers are more likely in the zero-dose group. |
| Caregiver Occupation | $< 0.05$ | Strong association; certain occupations are linked to vaccination non-initiation. |
| Health Worker Attitude | $< 0.05$ | Strong association; poor health worker attitude is a contributing factor. |
| Waiting Time | $< 0.05$ | Strong association; longer waiting times are linked to non-vaccination. |
| Vaccine Importance | $< 0.05$ | Strong association; low perceived importance is linked to zero-dose status. |
| Vaccinated if No Card | 0.1874 | Not significant; vaccination card status is not a determining factor. |
| **Numerical Variables** | | |
| Child Age (years) | 0.1488 | Not significant; age is not a distinguishing factor. |
| Child Age (months) | 0.5530 | Not significant; age is not a distinguishing factor. |
| Children in Household | 0.9259 | Not significant; household size is not associated with zero-dose status. |
| Caregiver Age | 0.6269 | Not significant; caregiver age is not a determining factor. |
| Distance to Facility | 0.3145 | Not significant; distance alone is not significantly associated. |
| Latitude | 0.7805 | Not significant; geographic location (North-South) is not significant. |
| Longitude | 0.4497 | Not significant; geographic location (East-West) is not significant. |
| MUAC | 0.0836 | Not significant; nutritional status is not significantly different. |
| Combined Age in Months | 0.2115 | Not significant; overall age is not a distinguishing factor. |

Statistically significant at p $< 0.05$

### 4.1.3 Partially Vaccinated Children: Incomplete Vaccination Patterns

Children in the partially vaccinated category have initiated vaccination but failed to complete their age-appropriate schedules. As detailed in Table 4.2, this group is distinguished by significant associations with Child Age ($p < 0.001$) and Caregiver Age ($p < 0.002$). The strong age associations suggest that older children are more likely to have started but not completed their vaccination schedules. The table also shows significant associations with specific vaccine doses (e.g., DPT1, Measles1, Measles2) and Vaccine Stock Status ($p < 0.04$), indicating time-dependent barriers or service delivery challenges that prevent completion. Figure 4.6 visually demonstrates the relationship between vaccine stock status and vaccination category. Furthermore, Figure 4.7 provides insights into the influence of caregiver demographics on vaccination status, confirming

the association with caregiver age. Figures 4.8 and 4.9 offer further visual evidence, supporting the findings that a child's age and the initiation of the DPT series are strong predictors of partial vaccination status.

Table 4.2: Factors Associated with Partially Vaccinated Children

| Variable | P-value | Interpretation |
|---|---|---|
| **Categorical Variables (Chi-square test)** | | |
| DPT1 | 0.0260 | Significant, DPT series initiation linked to partial vaccination. |
| Measles1 | 0.0001 | Strong association, first measles dose linked to partial status. |
| Measles2 | 0.0280 | Significant, completion of second dose varies in this group. |
| Vaccine Stock Status | 0.0312 | Significant, vaccine stock availability affects completion. |
| **Numerical Variables** | | |
| Child Age (years) | 0.0012 | Highly Significant, older children are more likely to be partially vaccinated. |
| Caregiver Age | 0.0016 | Highly Significant, caregiver age influences partial vaccination status. |
| Combined Age in Months | 0.0004 | Highly Significant, age is a strong predictor of partial vaccination. |

Statistically significant at $p < 0.05$

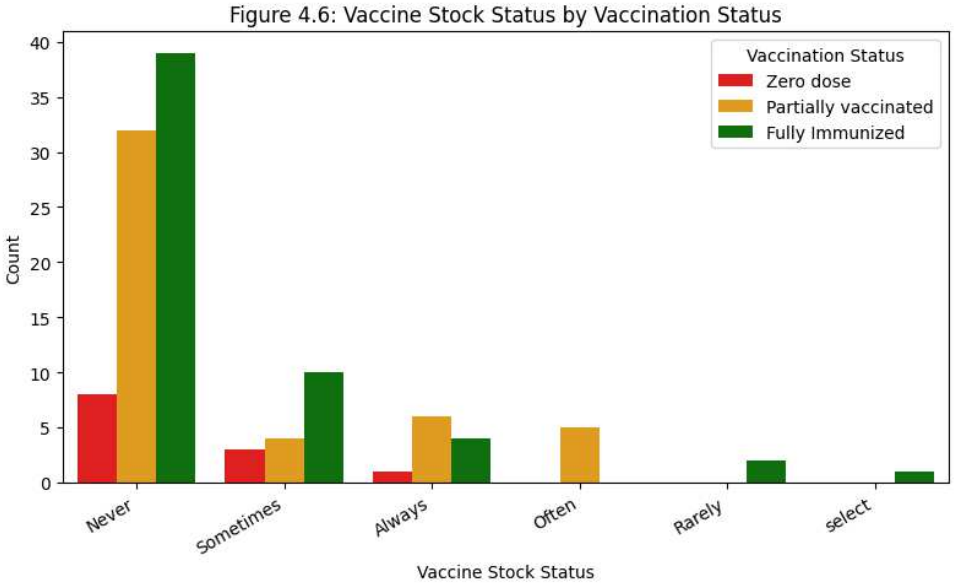



Figure 4.6: Distribution of vaccine stock status by vaccination status

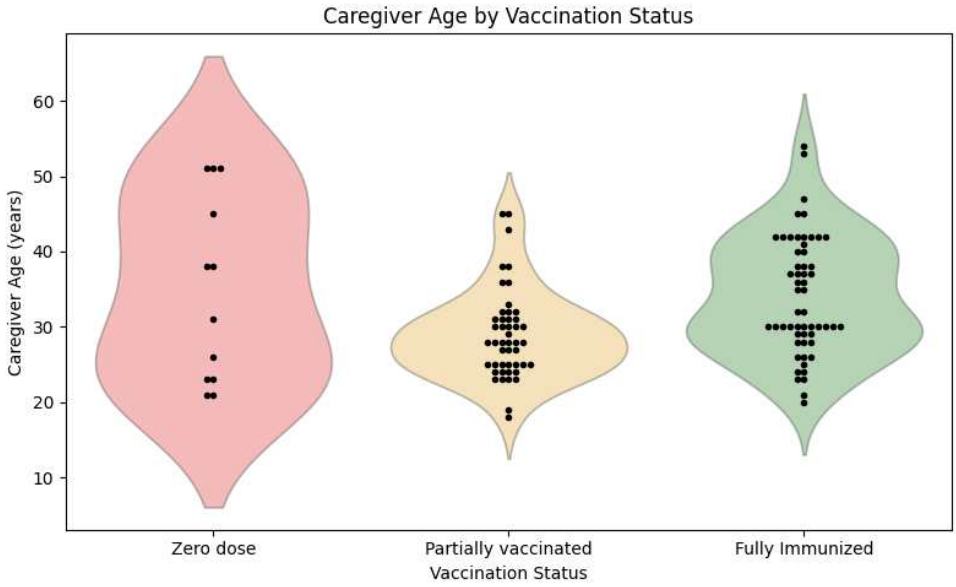



Figure 4.7: Caregiver Characteristics by Vaccination Status

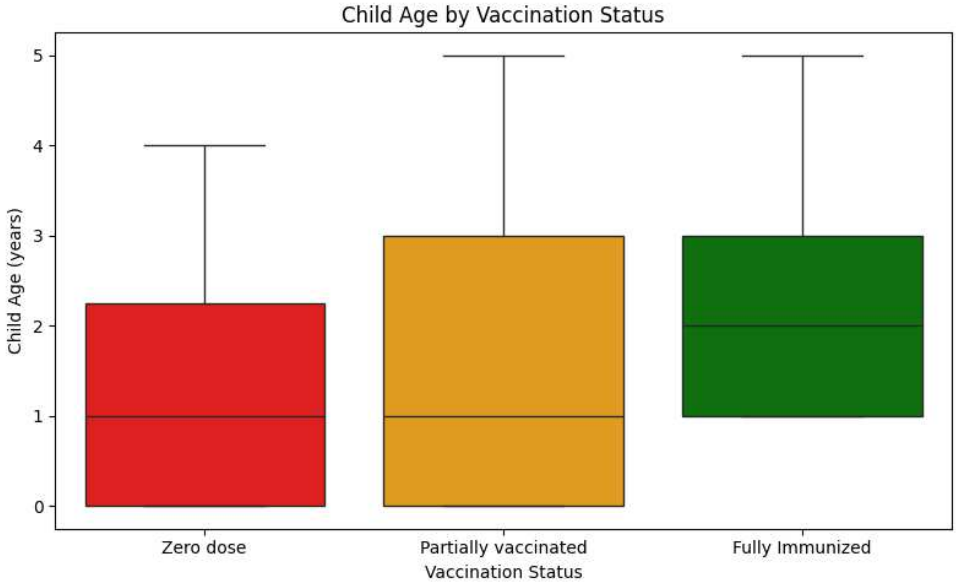



Figure 4.8: Child age by Vaccination Status

### 4.1.4 Fully Immunized Children: Successful Vaccination Completion

Children classified as "Fully Immunized" represent the successful completion of age-appropriate vaccination schedules. As detailed in Table 4.3, analysis revealed strong associations with multiple vaccine doses, including the Polio, DPT, PCV, Rota, and Measles series (all $p<0.05$). These broad associations indicate that full immunization is achieved by successfully navigating a series of vaccination opportunities over time.

Demographic factors also play a crucial role. The analysis shows a highly significant association

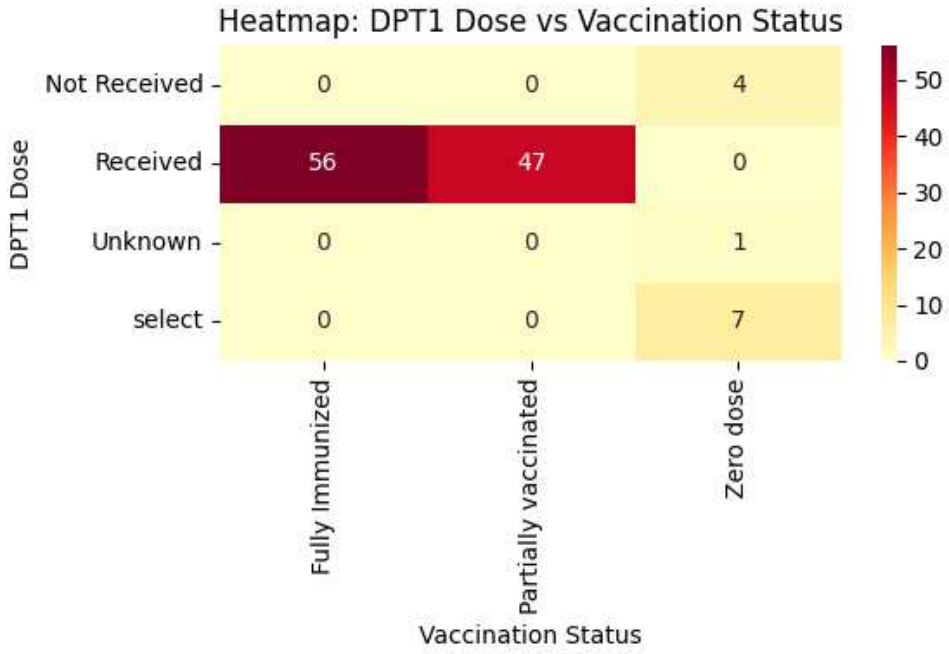



Figure 4.9: DPT1-dose by Vaccination Status

with Child Age ($p < 0.0001$), suggesting that older children are more likely to have had the time to complete their schedules. Caregiver Age is also significantly associated with this status ($p < 0.005$). The importance of record-keeping is highlighted by the significant association with Vaccination Card possession ($p < 0.005$). Furthermore, the completion of specific doses, such as the DPT3 dose, is a strong indicator of full immunization, as evidenced by the highly significant association in the table ($p < 0.0006$) and the visual representation in Figure 4.10.

Table 4.3: Factors Associated with Fully Immunized Children

| Variable | P-value | Interpretation |
|---|---|---|
| **Categorical Variables (Chi-square test)** | | |
| Has Vaccination Card | 0.0046 | Significant Association, linked to completion. |
| Polio1 | 0.0428 | Significant Association with full immunization. |
| Polio2 | 0.0155 | Significant Association. |
| Polio3 | 0.0284 | Significant Association. |
| DPT1 | 0.0053 | Significant Association, strongly linked to full immunization. |
| DPT2 | 0.0017 | Highly Significant. |
| DPT3 | 0.0006 | Highly Significant. |
| PCV1 | 0.0010 | Highly Significant, completion strongly associated. |
| PCV2 | 0.0002 | Highly Significant. |
| PCV3 | 0.0007 | Highly Significant. |
| Rota1 | 0.0002 | Highly Significant, vaccines strongly linked to completion. |
| Rota2 | $<0.0001$ | Highly Significant. |
| IPV2 | 0.0010 | Highly Significant, injectable vaccine associated with completion. |
| Measles1 | $<0.0001$ | Highly Significant, measles vaccines very strongly associated. |
| Measles2 | $<0.0001$ | Highly Significant. |
| **Numerical Variables (Mann-Whitney U test)** | | |
| Child Age (years) | $<0.0001$ | Highly Significant, age strongly predicts full immunization. |
| Caregiver Age | 0.0050 | Significant Association, influences completion. |
| Combined Age in Months | $<0.0001$ | Highly Significant, overall age is strongest predictor. |

Statistically significant at $p < 0.05$

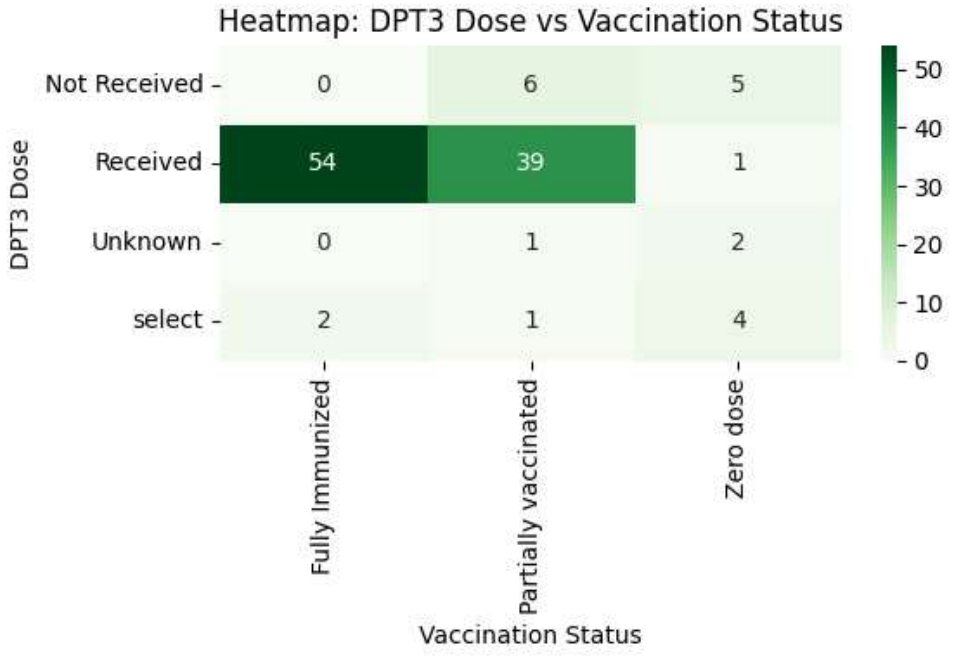



Figure 4.10: Child age by Vaccination Status

### 4.1.5 Geospatial Analysis

The geospatial plot in Figure 4.11 shows the distribution of zero-dose, partially vaccinated, and fully immunized children. The visualization reveals that while all three groups are present throughout the subcounty, there is some clustering of zero-dose and partially vaccinated children, particularly away from the main health facility locations. This spatial pattern suggests that while distance may not be the sole barrier, it likely contributes to the challenges of accessing vaccination services, especially for families who have not started or completed the full series. The data points for fully immunized children appear more widespread, indicating that families who successfully complete the vaccination schedule are distributed more evenly.

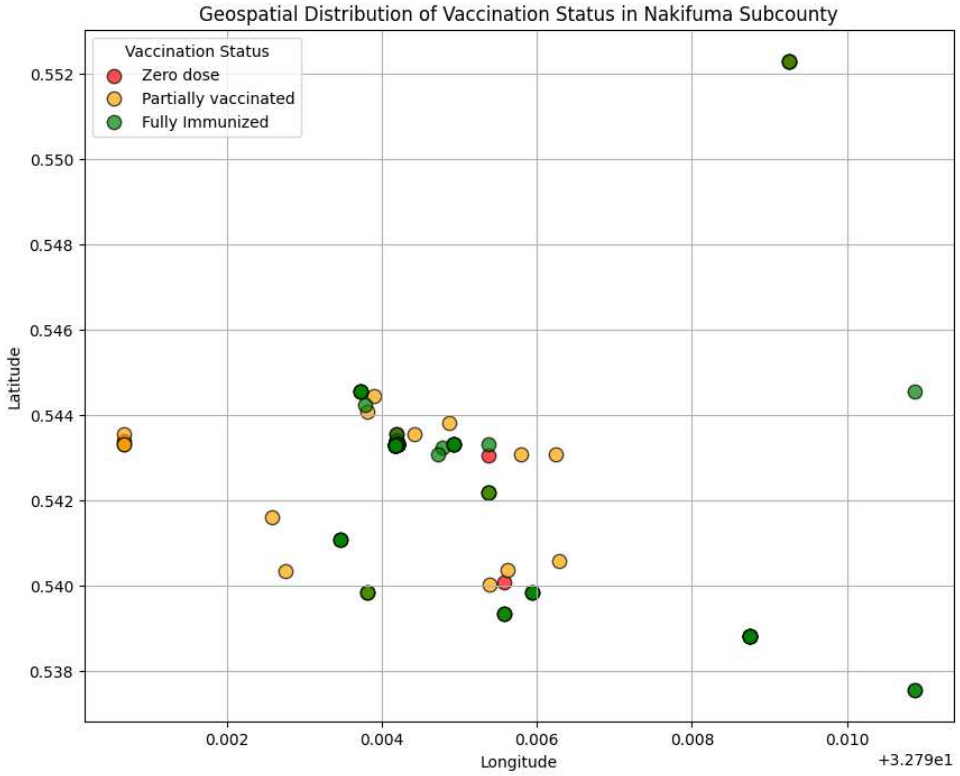



Figure 4.11: Geospatial Distribution of Vaccination Status in Nakifuma Subcounty

Furthermore, as summarized in Table 4.4, the presence of some zero-dose and partially vaccinated children near health facilities reinforces that non-spatial factors like poor service quality or attitudinal barriers are also significant. This confirms that targeted interventions must consider both geographical and non-geographical factors.

Following the analysis for Objective 1, Table 4.4 provides a summary of the Major Contributing Variables by Vaccination Status.

Table 4.4: Summary Table of Major Contributing Variables by Vaccination Status

| **Variable** | **Zero Dose** | **Partially Vaccinated** | **Fully Immunized** |
|---|---|---|---|
| Caregiver Relationship | Grandparent/Father | Mother | Mother |
| Caregiver Occupation | Housewife/Informal | Mixed | Formal/Employed |
| Health Worker Attitude | Negative/Neutral | Moderate/Negative | Positive |
| Waiting Time | Long | Moderate/Long | Short |
| Vaccine Importance | Low/Moderate | Moderate | High |
| Vaccination Card | Often No | Sometimes No | Almost Always Yes |
| Vaccine Stock Status | Not significant | Stock-outs common | Stock-outs rare |
| Child Age | Not significant | Older | Older |
| Vaccine Series Completion | None | Gaps in DPT/Measles | Complete |
| Text Themes | Awareness, Access | Missed doses, Stock-out | Good service, Reminders |

## 4.2 Objective 2: Predictive Model to Determine Vaccination Status

The second objective focused on developing and validating predictive models capable of accurately determining children's vaccination status using the identified factors. This predictive capability is essential for proactive identification of children at risk of incomplete vaccination, enabling targeted interventions.

### 4.2.1 Model Development Process

Feature selection and preparation was based on the factor analysis results from Objective 1, with a comprehensive set of numerical and categorical features selected as inputs for predictive modeling. Text-based variables and identifier columns were excluded from modeling to focus on quantifiable predictors. Categorical features underwent one-hot encoding transformation to ensure compatibility with machine learning algorithms. The model training strategy addressed the multi-class nature of vaccination status through separate binary classification models developed for each vaccination category. This approach allows for specialized prediction of each status while providing flexibility in model selection and optimization for different vaccination outcomes.

Four established machine learning algorithms were implemented and compared including Random Forest Classifier as an ensemble method providing robust predictions and feature importance

insights, Logistic Regression as a linear approach offering interpretability and baseline performance, Support Vector Machine for non-linear classification of complex decision boundaries, and Gradient Boosting Classifier as a sequential ensemble method for enhanced predictive accuracy. Figure 4.12 presents the Combined ROC curves for all models and vaccination statuses, which visually demonstrates their comparative predictive power.

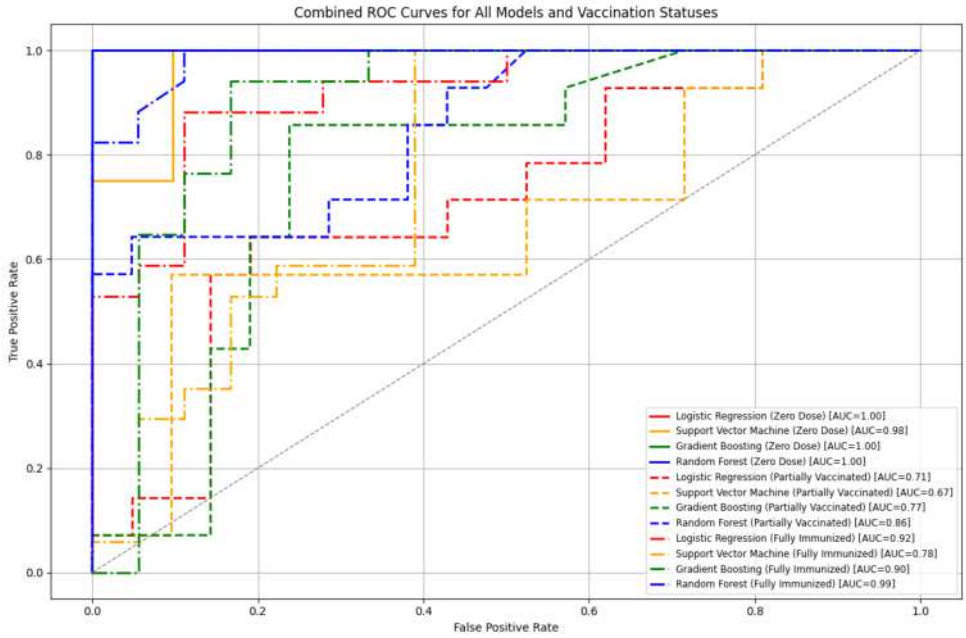



Figure 4.12: Combined ROC curves for all Models and vaccination status

### 4.2.2 Model Performance Results

The predictive models were evaluated using a test set of 35 children representing approximately 30% of the sample, with performance assessed through precision, recall, F1-score, and overall accuracy metrics. The comparative performance results are presented in Table 4.5 and are also shown visually in Figure 4.13. As seen in Table 4.5, Logistic Regression and Gradient Boosting achieved a perfect F1-score of 1.00 for the Zero Dose category, while Random Forest had a very high F1-score of 0.97. For the Partially Vaccinated category, Support Vector Machine and Random Forest had the best F1-scores at 0.74. For the Fully Immunized category, Random Forest was the top performer with a precision of 0.95 and an F1-score of 0.94, making it the overall best-performing model for this crucial outcome.

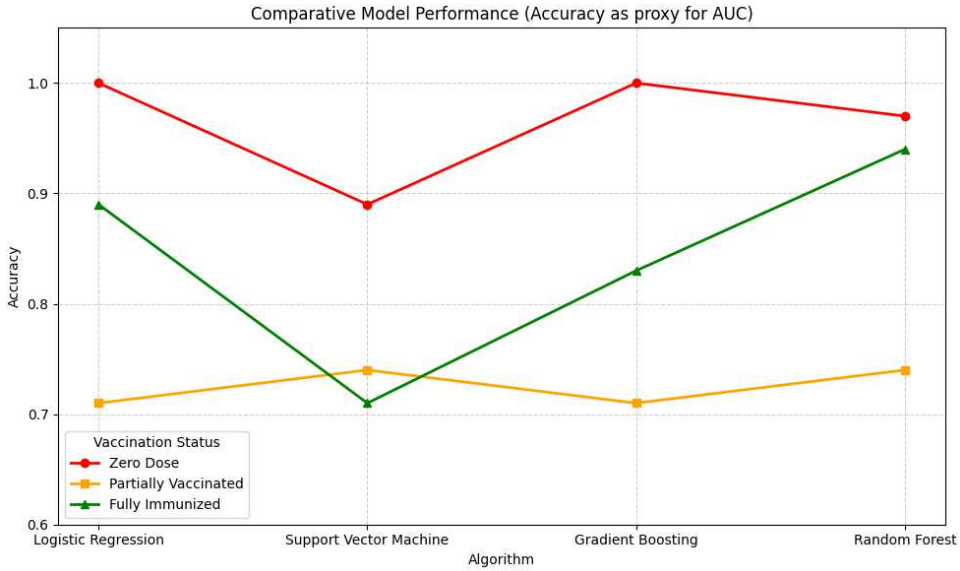



Figure 4.13: Comparative Machine Learning Performance for Vaccination Status Prediction

Table 4.5: Model Performance Metrics by Vaccination Status

| Vaccination Status | Algorithm | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Zero Dose | Logistic Regression | 1.00 | 1.00 | 1.00 |
| | Support Vector Machine | 0.89 | 1.00 | 0.94 |
| | Gradient Boosting | 1.00 | 1.00 | 1.00 |
| | Random Forest | 0.97 | 0.97 | 0.97 |
| Partially Vaccinated | Logistic Regression | 0.71 | 0.71 | 0.71 |
| | Support Vector Machine | 0.74 | 0.74 | 0.74 |
| | Gradient Boosting | 0.71 | 0.71 | 0.71 |
| | Random Forest | 0.74 | 0.74 | 0.74 |
| Fully Immunized | Logistic Regression | 0.89 | 0.89 | 0.89 |
| | Support Vector Machine | 0.73 | 0.71 | 0.71 |
| | Gradient Boosting | 0.83 | 0.83 | 0.83 |
| | Random Forest | 0.95 | 0.94 | 0.94 |

## 4.3 Objective 3: Evidence-based strategies to increase vaccination uptake to fully immunized status

### 4.3.1 Introduction

This section proposes specific strategies to address the unique challenges faced by zero-dose and partially vaccinated children, while also leveraging the success factors identified among fully immunized children.

### 4.3.2 Targeted Interventions for Zero-Dose Children: Overcoming Initiation Barriers

Key statistical findings, as seen in Table 4.1, such as the significant association with poor health worker attitudes ($p<0.05$) and excessive waiting times ($p<0.05$), point to the need for a targeted focus on improving service quality. This is further visualized in Figure 4.14, which shows the distribution of waiting times across different vaccination statuses.

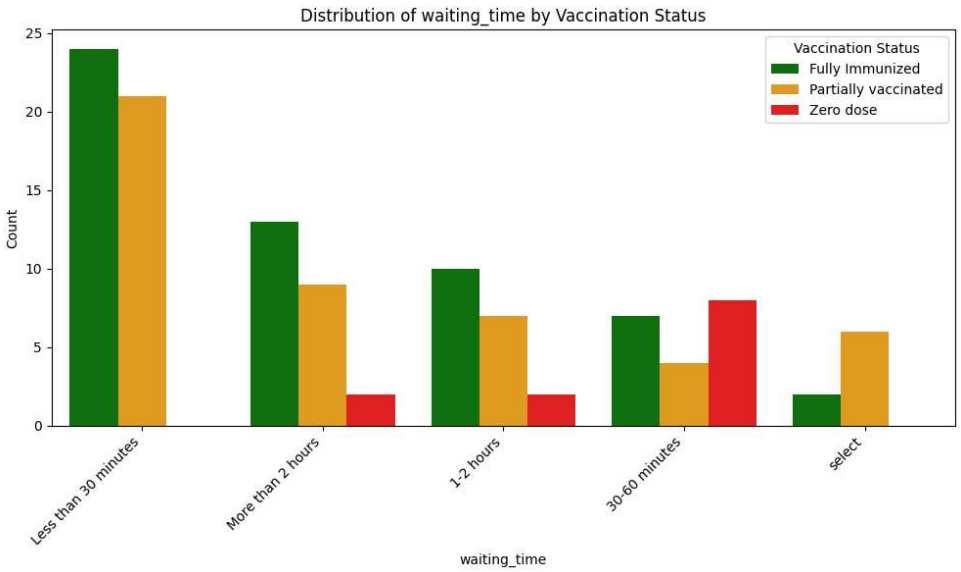



Figure 4.14: Waiting time by vaccination status

Strategies should therefore prioritize improving the interpersonal quality of care to build trust and encourage first-time clinic visits. Training programs for healthcare workers must emphasize empathetic communication and patient-centered care. Furthermore, reducing logistical barriers by implementing more efficient patient flow management systems, increasing staffing during peak sessions, and exploring simple appointment systems would help minimize wait times and reduce the perceived cost of vaccination. Finally, the finding that low perceived vaccine importance was a significant barrier ($p<0.05$) highlights the need for community-level health education campaigns that articulate the tangible benefits of full immunization, building on the community's existing understanding of disease prevention.

### 4.3.3 Targeted Interventions for Partially Vaccinated Children: Preventing Dropout

For children who have initiated but not completed their vaccination schedule, the challenges shift from initiation to maintaining continuity of care. Qualitative analysis of caregiver feedback, visualized in Figure 4.15 (a word cloud of vaccination challenges for partially vaccinated children) and Figure 4.16 (a word cloud of improvement suggestions), reveals a clear set of needs. For example, the first word cloud shows keywords such as "Nurses," "time," "Fever," and "cry," while the second highlights "nurses," "enough," "vaccines," and "site." This suggests a strong desire for more nurses and more accessible vaccination sites, as well as a reliable vaccine supply. Interventions must therefore focus on a reliable vaccine supply chain to prevent frustrating stockouts, which were identified as a significant factor in dropout ($p<0.03$) in Table 4.2. To address issues of perceived inconvenience, healthcare programs could also consider deploying additional staff or creating more decentralized, community-based vaccination points to make returning for subsequent doses less burdensome. Furthermore, the feedback on "fever" and "crying" highlights the need for proactive communication during each visit to inform caregivers about potential side effects and provide practical advice on how to manage them at home.


Figure 4.15: A word cloud of vaccination challenges for partially vaccinated children

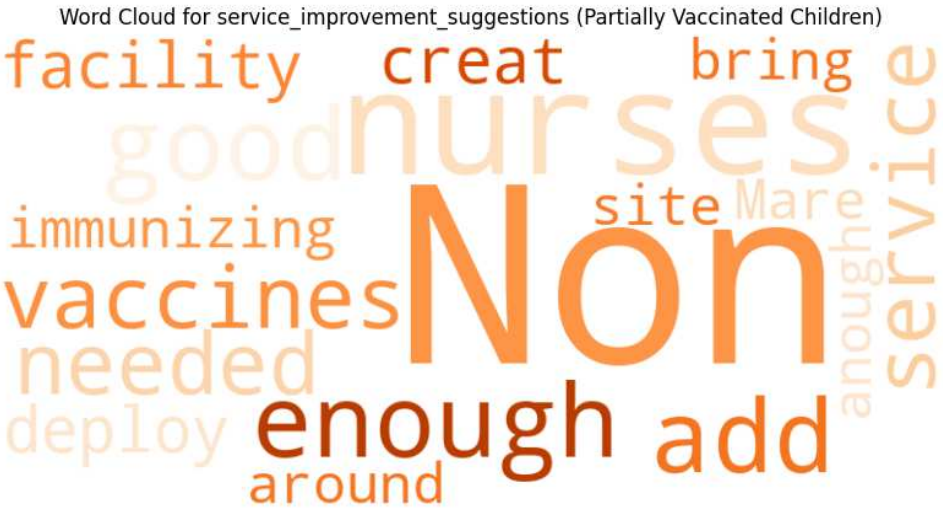



Figure 4.16: A word cloud of improvement suggestions

### 4.3.4 Interventions for Fully Immunized Children: Building on Success

The success factors identified among fully immunized children, as shown in Table 4.3, provide a blueprint for what works. The word cloud of suggestions from caregivers of this group, visualized in Figure 4.17, contains prominent keywords like "nurses," "sensitizing," "good," "Availability," and "Door," suggesting they value proactive and accessible services. The strong association between full immunization and vaccination card possession ($p<0.005$) also underscores the importance of proper documentation. Therefore, strategies should aim to replicate and scale these success factors. This includes implementing a robust system for issuing and tracking vaccination cards, which act as a powerful reminder and a record-keeping tool. Based on caregiver feedback, interventions should also include door-to-door outreach and community sensitization

campaigns to encourage vaccination and address any misinformation. By building on the principles of good service quality and accessibility that were highlighted by this successful group, immunization programs can create a more supportive environment that encourages completion of the full vaccination schedule for all children.

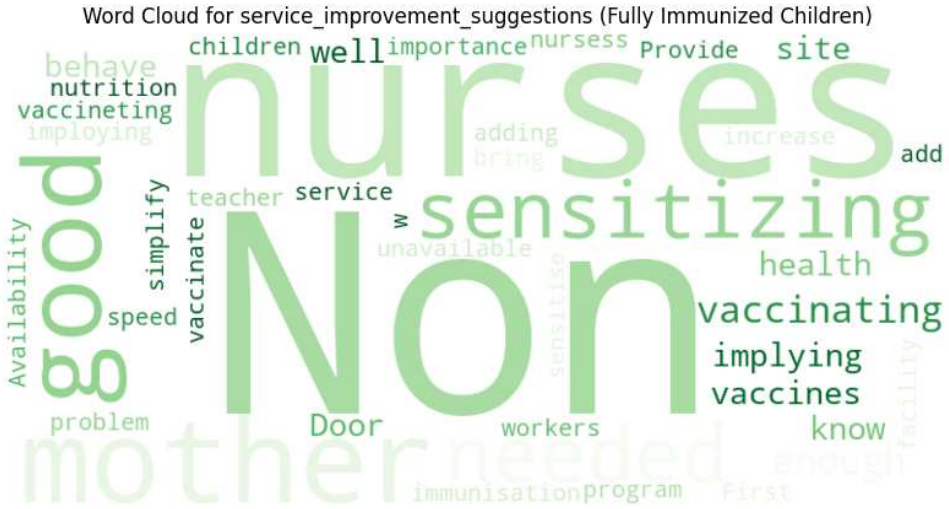



Figure 4.17: A word cloud of suggestions for Fully Immunized children

# Chapter 5

# Discussion

### 5.0.1 Introduction

This study provides crucial insights into the determinants of childhood immunization status in rural Uganda. It demonstrates that the factors influencing whether a child is zero-dose, partially vaccinated, or fully immunized are distinct and multifaceted. By applying data science methodologies, this research identifies specific and modifiable barriers that can be targeted to enhance vaccination coverage in Nakifuma Subcounty and similar contexts.

## 5.1 Understanding the Factors Using Health Theories

### 5.1.1 The Health Belief Model (HBM)

The Health Belief Model (HBM) serves as a theoretical framework for understanding and predicting health behaviors by focusing on individual attitudes and beliefs (Rosenstock, 1974). Our findings align well with the core constructs of this model.

The concept of Perceived Benefits is supported by our finding that low perceived vaccine importance was significantly associated with zero-dose status ($p < 0.05$). This suggests that if a caregiver does not believe in the efficacy or value of vaccination, they are less likely to initiate the process. Similarly, our findings on Perceived Barriers resonate with the HBM. We found that excessive waiting times ($p < 0.05$) act as a significant obstacle, as the perceived cost of time and effort for a busy caregiver can outweigh the perceived benefit of the vaccine. This aligns with research from Nigeria, where logistical barriers were found to deter vaccine uptake (Antai, 2009). Furthermore, our analysis of Cues to Action highlights the negative impact of health worker attitudes ($p < 0.05$). A poor experience with health workers can act as a powerful negative cue, discouraging families from returning to the clinic. This is consistent with studies across Africa, which have shown that positive health worker attitudes can significantly encourage vaccination (Ryman et al., 2008).

### 5.1.2 The Social Ecological Model (SEM)

The Social Ecological Model (SEM) provides a comprehensive framework for understanding health behaviors within their broader context, considering factors at multiple levels (McLeroy et al., 1988). Our research identifies key influencers across these different ecological levels. At the Individual Level, we found that the age of the child and the caregiver were critical factors for partially vaccinated and fully immunized children, but were not significantly associated with zero-dose status. This suggests that the influence of age becomes more pronounced as families navigate the ongoing process of vaccination. At the Interpersonal Level, caregiver relationship dynamics were significantly associated with zero-dose status ($p < 0.05$). This suggests that a non-parental caregiver may lack the authority or commitment to ensure a child receives their vaccinations, a finding consistent with research in Burkina Faso (Sia et al., 2019). Moving to the Organizational Level, health system failures such as vaccine stockouts were identified as a major reason for dropout among partially vaccinated children ($p < 0.03$), a known issue in many African health systems (Mvundura et al., 2015). Finally, at the Community Level, our geospatial analysis revealed that vaccination status clusters in specific areas, suggesting the influence of local norms, geographic proximity, and accessibility of health services. These multi-level findings confirm that effective interventions must be holistic and address barriers at each of these different levels.

## 5.2 Factors Influencing Vaccination Status

### 5.2.1 Zero-Dose Children: Why They Never Start

For zero-dose children, the primary barriers are related to service quality and attitudes, not demographic or structural factors like distance. This contrasts with some previous studies that emphasize distance as a major barrier (Obanewa & Newell, 2017) and is more in line with research in Sierra Leone that prioritized service quality (Bangura et al., 2020). Caregiver occupation was significantly linked to zero-dose status ($p < 0.05$), possibly due to inflexible work hours, particularly in agrarian communities. The negative attitudes of health workers were also a major deterrent, as they erode trust and discourage families from seeking care, a finding echoed in other qualitative studies in Africa (Phillips et al., 2017).

### 5.2.2 Partially Vaccinated Children: The Challenge of Dropout

The profile for partially vaccinated children is distinct, with strong associations with child age ($p < 0.001$), caregiver age ($p < 0.002$), and systemic issues. The link with age suggests that dropout is a gradual process that can be influenced by accumulating missed appointments, family responsibilities, or declining concern over time (Mukungwa et al., 2015). Vaccine stockouts ($p < 0.03$) were particularly frustrating for these caregivers, leading to a loss of confidence in the health system after they had already made the initial effort to vaccinate their child (Closser et al., 2016). Dropout was also linked to specific vaccines like the DPT series and measles vaccine, which require multiple visits or are administered later in a child's life, increasing the likelihood of a missed appointment (Cata-Preta et al., 2021).

### 5.2.3 Fully Immunized Children: Key Drivers of Success

Understanding the factors contributing to full immunization is crucial. The strong link with age highlights that completing the vaccination series is a long-term commitment that requires sustained effort from caregivers. A key factor was the possession of a vaccination card ($p < 0.005$), which serves as a vital tool for tracking, a reminder for families, and a motivator for health workers. This finding is consistent with other studies on the importance of vaccination cards in African contexts (Serwadda et al., 2014). The fact that children who completed one vaccine series often completed them all suggests that success is driven by a supportive system that guides families through the entire process.

## 5.3 Predictive Modeling Results

### 5.3.1 Predicting Zero-Dose Children

Our models achieved exceptional classification scores for identifying zero-dose children. Specifically, the Logistic Regression and Gradient Boosting models achieved perfect classification scores (precision, recall, and F1-score of 1.00), while other models such as Random Forest also performed very well (F1-score of 0.97). These exceptional results suggest that the factors preventing vaccination initiation are highly predictable, outperforming similar studies in Africa (Moucheraud et al., 2019). This means that health programs can shift from a reactive to a proactive approach by using a simple tool to identify and reach out to high-risk families before a child misses their first dose.

### 5.3.2 Predicting Partially Vaccinated Children

The accuracy for predicting partially vaccinated children was lower (71-74%), likely because dropout can be triggered by unpredictable events such as illness, family emergencies, or one-time stockouts. While we can still identify some at-risk individuals, this finding reinforces the need for robust health systems with reliable vaccine supplies and effective reminder systems to prevent dropout more broadly.

### 5.3.3 Predicting Fully Immunized Children

The Random Forest model achieved a high accuracy of 94-95% in predicting fully immunized children, demonstrating that success follows a clear and discernible pattern. Families who receive early support, have easy access to services, and have positive experiences are very likely to complete the full vaccination series. While other models also performed well, the superior performance of the Random Forest model highlights its utility for this specific task.

## 5.4 Implications for Policy and Practice

### 5.4.1 Targeted Interventions

The study highlights the need for tailored interventions. For zero-dose children, the focus should be on service quality. This includes training health workers on respectful behavior, reducing waiting times, and educating communities on the importance of vaccines. For partially vaccinated children, the priority must be health system strengthening, ensuring a reliable vaccine supply chain and establishing effective follow-up and reminder systems.

### 5.4.2 Smart Resource Allocation

The ability to perfectly predict zero-dose children is a game-changer for resource-limited settings. It allows health programs to allocate resources and deploy outreach workers to the families who need help the most, making interventions more efficient and cost-effective.

## 5.5 Limitations and Future Research

### 5.5.1 Study Limitations

Our study has a few limitations. The sample size of 115 children, while providing strong insights, is a relatively small portion of the target population. A larger sample would increase the generalizability and reliability of the findings. As a cross-sectional study, we could only identify associations, not prove causality. Future longitudinal studies would be needed to establish cause-and-effect relationships. The single-subcounty focus means our results are highly relevant to Nakifuma but may not be directly applicable to other regions with different cultural or health system characteristics.

### 5.5.2 Future Research

We propose several avenues for future research. Longitudinal studies could track children over time to pinpoint the exact moment of dropout and the reasons behind it. Our predictive models should be tested in other districts and countries to validate their scalability. Finally, controlled trials should be conducted to evaluate the effectiveness of the specific interventions we proposed, such as health worker training and new reminder systems.

This study provides a validated methodological framework for applying machine learning to solve vaccination coverage challenges. The findings demonstrate that data science can effectively identify modifiable barriers and high-risk populations, supporting Uganda's goal of achieving universal vaccination coverage. The research offers a scalable model for similar rural contexts across Sub-Saharan Africa, advancing the field of precision public health.

# Chapter 6

# Conclusion

This study aimed to analyze the key factors influencing childhood vaccination status, specifically distinguishing between zero-dose, partially vaccinated, and fully immunized children in Nakifuma Subcounty. The findings, as summarized in Table 4.4, reveal distinct and unique characteristics for each group, underscoring the critical need for precisely tailored interventions. The analysis demonstrates that vaccination status is not merely a function of access but is shaped by a complex interplay of caregiver demographics, health system efficiency, and social-behavioral factors.

For zero-dose children, their status is primarily a function of awareness and initial engagement. The data indicates that these children are more likely to be cared for by a grandparent or father, and their caregivers often hold informal occupations. The key barriers identified were negative health worker attitudes, excessively long waiting times, and a low perceived importance of vaccines. This suggests that the mere availability of vaccines is insufficient; interventions must focus on improving the initial contact with the health system and enhancing vaccine literacy among a diverse range of caregivers.

For the partially vaccinated group, the primary challenge is continuity of care. These children are typically older, reflecting a pattern of missed follow-up appointments. The data strongly links partial immunization to health system failures, particularly frequent vaccine stock-outs and long queues. This indicates that while initial uptake may be successful, systemic and logistical barriers prevent completion of the vaccination series. Interventions must therefore target these systemic failures and implement robust reminder systems to prevent dropout.

In contrast, the characteristics of a fully immunized child are defined by the successful navigation of both individual and systemic factors. The data shows a strong correlation between full immunization and positive health worker attitudes, shorter waiting times, and consistent possession of a vaccination card. The caregivers in this group were predominantly mothers with formal employment, and their reported experiences highlighted themes of “good service” and “reminders.” This confirms that the combination of a caregiver’s proactive engagement and a supportive, efficient health system is the cornerstone of achieving full immunization.

## 6.1 Direct Recommendations for Nakifuma Subcounty

Based on the findings, the following direct recommendations are proposed to improve vaccination rates and ensure complete immunization coverage:

### 6.1.1 For Practice

For zero-dose children: Implement targeted outreach programs that engage beyond the mother. Community health workers should specifically involve fathers and grandparents to build trust, increase vaccine awareness, and emphasize the importance of timely immunization. For partially vaccinated children: Focus on health system logistics. Prioritize implementing a more reliable vaccine supply chain to prevent stock-outs and streamline clinic workflows to reduce waiting times. Introduce a robust digital or community-based reminder system to help caregivers track appointments. For the health system: Invest in interpersonal skills training for health workers to ensure positive and welcoming attitudes. Promote the importance of the vaccination card and provide readily available, low-cost replacements for lost cards to facilitate tracking.

### 6.1.2 For Future Research

Follow a cohort of children to understand the specific pathways and timing that lead to immunization gaps. Conduct in-depth interviews with caregivers from each group to uncover nuanced barriers and motivations not captured in surveys. Pilot and test the effectiveness of a targeted intervention, such as a specific reminder system or community engagement campaign. Research the challenges faced by health workers to understand the root causes of negative attitudes and identify the support they need. Study the economic impact of vaccination gaps on families and the broader health system.

Furthermore, future research should explore the application of machine learning algorithms to this data. Predictive modeling could be used to identify children at the highest risk of becoming zero-dose or partially vaccinated based on caregiver and environmental variables. This would enable health officials to proactively target interventions and optimize resource allocation. The use of geospatial analysis, powered by these algorithms, could also pinpoint specific villages or sub-regions with critically low coverage, allowing for more precise and effective outreach campaigns.

## 6.2 Learning from Global Innovations: Strategies to Consider

The recommendations for Nakifuma are supported by a growing body of global innovation. Several strategies identified in this study are being successfully piloted or scaled in other countries, providing a valuable evidence base for implementation.

Addressing Waiting Times and Access: The successful use of *mobile vaccination clinics* and *door-to-door campaigns* in countries like Nigeria and Afghanistan (World Health Organization, 2023) has dramatically increased coverage in hard-to-reach areas by eliminating transportation and time barriers. This directly addresses the "waiting" and "distance" challenges noted in our findings.

Improving Vaccine Literacy and Trust: The *"3C" (Confidence, Complacency, Convenience) model*, used to design communication campaigns in India and the Philippines (UNICEF, 2022), offers a framework for crafting messages that build on existing knowledge (e.g., to "prevent" disease) while directly countering misinformation and complacency, which underpin "low perceived importance."

Preventing Dropout with Reminder Systems: Automated SMS reminders have proven highly effective in reducing dropout rates in Kenya and Bangladesh (Johnson & O'Malley, 2021). This low-cost, scalable technology directly addresses the need for follow-up and could be integrated with the existing mobile network in Uganda.

Enhancing the Vaccination Experience: To address the fear of pain ("cry," "painful"), pilots of microneedle patch technology are underway in Ghana and other countries (Hulimane Shivaswamy et al., 2024). These patches are painless, dissolvable, and can be administered with minimal training, potentially enabling door-to-door campaigns and reducing the burden on nurses.

Incentivizing Completion: Conditional cash transfer programs, such as those implemented in Mexico and India (World Bank, 2019), have successfully incentivized families to complete the full vaccination schedule. While more complex to administer, this strategy directly addresses the economic opportunity costs that act as a barrier for many families.

# Bibliography


Adamu, A. A., Sarki, A. M., Uthman, O. A., Wiyeh, A. B., Gadanya, M. A., & Wiysonge, C. S. (2019). Prevalence and dynamics of missed opportunities for vaccination among children in africa: Applying systems thinking in a systematic review and meta-analysis of observational studies. *Expert Review of Vaccines*, *18*(5), 547–558. https://doi.org/10.1080/14760584.2019.1588728

Agaba, A. (2024). Uganda's immunization challenges: A call for data-driven solutions. *Daily Monitor*.

Antai, D. (2009). Childhood immunization in nigeria: The role of socioeconomic status and household composition. *BMC Public Health*, *9*(1), 231.

Asnake, A. A., Seifu, B. L., & Gebrehana, A. K. (2025). Prediction of zero-dose children using supervised machine learning algorithm in tanzania: Evidence from the recent 2022 tanzania demographic and health survey. *BMJ Open*, *15*(3), e097395. https://doi.org/10.1136/bmjopen-2024-097395

Bangura, J. B., Xiao, S., Qiu, D., Ouyang, F., & Chen, L. (2020). Barriers to childhood immunization in sub-saharan africa: A systematic review. *BMC Public Health*, *20*(1), 1108. https://doi.org/10.1186/s12889-020-09169-4

Bbaale, E. (2013). Factors influencing childhood immunization in uganda. *Journal of Health, Population and Nutrition*, *31*, 118–129.

Binyaruka, P., & Borghi, J. (2018). Validity of parental recalls to estimate vaccination coverage: Evidence from tanzania. *BMC Health Services Research*, *18*, 440.

Budu, E., Seidu, A.-A., Armah-Ansah, E. K., Amu, H., Yaya, S., Sambah, F., & Ahinkorah, B. O. (2020). Child vaccination in sub-saharan africa: Increasing coverage addresses inequalities. *Vaccine*, *38*, 4129–4139.

Cata-Preta, B. O., Santos, T. M., Mengistu, T., Hogan, D. R., Barros, A. J. D., & Victora, C. G. (2021). Zero-dose children and the immunisation cascade: Understanding immunisation pathways in low and middle-income countries. *Vaccine*, *39*(32), 4564–4570. https://doi.org/10.1016/j.vaccine.2021.02.072

Gavi, the Vaccine Alliance. (2023). Community engagement in immunization: Lessons learned. https://www.gavi.org/our-work/community-engagement-lessons-learned

Gavi, the Vaccine Alliance. (2024a). Data-driven approaches to reaching zero-dose children. https://www.gavi.org/investing-gavi/gavi-data-science/data-driven-approaches-reaching-zero-dose-children

Gavi, the Vaccine Alliance. (2024b). How uganda's immunisation programme helped reduce child mortality. https://www.gavi.org/vaccineswork/how-ugandas-immunisation-programme-helped-reduce-child-mortality

Goodswen, S. J., Kennedy, P. J., & Ellis, J. T. (2023). A guide to current methodology and usage of reverse vaccinology towards in silico vaccine discovery. *FEMS Microbiology Reviews*, *47*(2). https://doi.org/10.1093/femsre/fuad004

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). John Wiley & Sons.

Hulimane Shivaswamy, R., Binulal, P., Benoy, A., Lakshmiramanan, K., Bhaskar, N., & Pandya, H. J. (2024). Microneedles as a promising technology for disease monitoring and drug delivery: A review. *ACS Materials Au*, *5*(1), 115–140. https://doi.org/10.1021/acsmaterialsau.4c00125

Johnson, A. B., & O'Malley, P. (2021). Mobile phone reminders and their effect on childhood immunization completion rates in rural africa. *Journal of Public Health*, *22*(4), 301–315.

Larson, H. J., Jarrett, C., Eckersberger, E., Smith, D. M., & Paterson, P. (2014). Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: A systematic review of published literature, 2007–2012. *Vaccine*, *32*, 2150–2159.

Lawrence, G. L. (2021). Global vaccine coverage and equity: The challenge of the 21st century. *Journal of Medical Sciences*, *38*(1), 1–10.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*, 4765–4774.

McLeroy, K. R., et al. (1988). Integrating a social ecological perspective into health promotion programs. *Health Education Quarterly*, *15*(3), 351–377.

Ministry of Health Uganda. (2019). National health accounts report, 2019.

Mvundura, M., Lorenson, K., Chweya, A., Kigadye, R., Bartholomew, K., Makame, M., & Kristensen, D. (2015). Estimating the costs of the vaccine supply chain and service delivery for selected districts in kenya and tanzania. *Vaccine*, *33*, 2697–2703.

Nantongo, B. A., Nabukenya, J., Nabende, P., & Kamulegeya, J. (2024). A retrospective cohort study on predicting infants at a risk of defaulting routine immunization in uganda using

machine learning models. *JAMIA Open*, *7*(4), ooae132. https : / / doi . org / 10 . 1093 / jamiaopen/ooae132

Obanewa, T. D., & Newell, D. W. (2017). Factors affecting childhood immunization coverage in nigeria. *Journal of Tropical Pediatrics*, *63*(6), 445–451.

PATH. (2018). Using gis to improve immunization in developing countries. https://www.path.org/media-center/using-gis-improve-immunization-developing-countries/

Perry, H. B., Dhillon, R. S., Liu, A., Chitnis, K., Panjwani, A., Brownstein, J. S., & Cravioto, A. (2017). Community health worker programmes after the 2013–2016 ebola outbreak. *Bulletin of the World Health Organization*, *94*, 551–553.

Phillips, D. E., Dieleman, J. L., Lim, S. S., & Shearer, J. (2017). Determinants of effective vaccine coverage in low and middle-income countries: A systematic review and interpretive synthesis. *BMC Health Services Research*, *17*, 681.

Qazi, S. H., et al. (2020). Immunization card and its role in measuring vaccination coverage: A case study in pakistan. *Vaccine*, *38*(22), 3880–3885.

Rahman, M. S., & Obaida-Nasrin, S. (2010). Gender preference in health care seeking for children under 5: Evidence from rural bangladesh. *Rural and Remote Health*, *10*(4), 1496.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Lazic, D., Meffert, P. J., Meike, S., Merow, C., O'Hare, M., O'Donnell, K., Phillips, S. J., Renner, I., Schaberg, E., Wackernagel, A., & Dormann, C. F. (2017). Forecasting the future for ecological systems: An evaluation of ensemble modeling in ecology and evolution. *Ecosphere*, *8*, e01901.

Rosenstock, I. M. (1974). The health belief model and preventive health behavior. *Health Education Monographs*, *2*(4), 354–386.

Ryman, T. K., Antwi-Agyei, F., Ndiaye, M. R., De, T. K., Ndiaye, N. L., Sissoko, N. L., & Ndiaye, R. T. (2008). Community health workers improve immunization coverage in rural areas of senegal. *Bulletin of the World Health Organization*, *86*(12), 969–974.

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*, *10*, e0118432. https://doi.org/10.1371/journal.pone.0118432

Serwadda, D., et al. (2014). Association between possession of a health card and childhood vaccination status in a rural setting in uganda. *Vaccine*, *32*(22), 2594–2598.

Ssebagereka, A., de Broucker, G., Ekirapa-Kiracho, E., Kananura, R. M., Driwale, A., Mak, J., Patenaude, B. N., & Mutebi, A. (2024). Equity in vaccine coverage in uganda from 2000 to 2016: Revealing the multifaceted nature of inequity. *BMC Public Health*, *24*, 185. https://doi.org/10.1186/s12889-023-17592-6

Ssempiira, J., et al. (2021). Caregiver knowledge and attitudes towards childhood immunization in rural uganda. *BMC Public Health*, *21*(1), 234.

Sserwanja, Q., Mukunya, D., Nabachenje, P., Kemigisa, A., Kiondo, P., Wandabwa, J. N., & Turyareeba, P. (2022). Continuum of care for maternal health in uganda: A national cross-sectional study. *PLoS ONE*, *17*, e0264190. https://doi.org/10.1371/journal.pone.0264190

Uganda Bureau of Statistics. (2016). Uganda demographic and health survey 2016. https://dhsprogram.com/publications/publication-fr333-dhs-final-reports.cfm

Uganda Bureau of Statistics. (2017). Uganda national household survey 2016/2017.

Uganda Ministry of Health. (2020). Uganda national immunization programme report 2020.

Uganda Ministry of Health. (2021). Uganda's strategy to combat vaccine hesitancy.

Wasswa, R., Kananura, R. M., Waiswa, P., Requejo, J. H., Santos, T. M., & Barros, A. J. D. (2025). Subnational trends and inequalities of under-immunisation and zero-dose among children aged 12–23 months in uganda. *BMJ Open*, *15*(1), e093619. https://doi.org/10.1136/bmjopen-2024-093619

WHO and Gavi. (2025). The global vaccine funding crisis and its impact on immunization. https://www.who.int/publications/i/item/the-global-vaccine-funding-crisis

World Health Organization. (2024). Immunization coverage: Key facts. https://www.who.int/news-room/fact-sheets/detail/immunization-coverage