Statistical Analysis Plan (SAP)

Title: Association between vaccination beliefs and immunization status

CRU/Department/Division/Center	Department of Epidemiology at UNC
Investigators:	
Primary Investigator	Dr. Ross Boyce, MD
Collaborative Lead	Annika Gunderson (PhD Candidate)
Analysis Biostatistician(s)	Taeim Kwon, Yohhan Kumarasinghe, Yuanyuan Yan, Joyce Choe, Qikai Jiang
Subject Matter Expert	\
Version Date	Version 1, 3/27/2025
Project Folder Location	SharePoint: OneDrive > BIOS 841-Sp25 >shared document
Project Goal(s)	Final Presentation and Report for BIOS 841
Submission Deadline(s)	Final Presentation due 4/10/2025 Final Report due 4/24/2025

Investigator Agreement

☐ All statistical analyses included in an abstract or manuscript should reflect the work of the biostatistician(s) listed on this SAP. No changes or additional analyses should be made to the results or findings without discussing with the project biostatistician(s).

☐ All biostatisticians on this SAP should be given sufficient time to review the full presentation, abstract, manuscript, or grant and be included as co-authors on any abstract or manuscript resulting from the analyses.

☐ If substantial additional analysis is necessary or the aims of the project change, a new SAP will need to be developed.

☐I have reviewed the SAP and understand that any changes must be documented.

Acknowledged by: Dr. Ross Boyce and Annika Gunderson

Activity Log

3/25/2025: This is an entirely new SAP before getting into the analysis part for our final presentation and report.

Acronyms	BCG	Bacillus Calmette-Guérin vaccine
	OPT	Oral Polio Vaccine
	PENTA1	Penta (diphtheria, tetanus, pertussis, Haemophilus influenzae type b (Hib) infections, and hepatitis B) vaccine dose #1
	PENTA2	Penta (diphtheria, tetanus, pertussis, Haemophilus influenzae type b (Hib) infections, and hepatitis B) vaccine dose #2
	PENTA3	Penta (diphtheria, tetanus, pertussis, Haemophilus influenzae type b (Hib) infections, and hepatitis B) vaccine dose #3
	ROTA1	Rotavirus dose #1
	ROTA2	Rotavirus dose #2

1 Study Overview

Background/Introduction:

In 2016, a national health survey in Uganda revealed that only 55% of children aged 12-23 months received all recommended vaccinations. The investigators hypothesize that areas with limited access to immunization services in Uganda have lower childhood vaccination coverage compared to regions with full access to these services. The target population for this study consists of households with children aged 12-23 months who reside with a caregiver in Bugoye Sub-county, a small rural town in western Uganda. This cross-sectional study aims to explore the social determinants influencing vaccination coverage, focusing on caregiver attitudes and beliefs regarding vaccination. Data for the study was collected between January 20, 2021, and April 30, 2021.

1.1 Study Aims

1. How are caregiver beliefs and experiences about vaccination associated with the children aged 12-23 months in Bugoye Sub-county, Uganda, receiving all recommended immunizations, after adjusting for demographic factors?
2. Among those with vaccination cards, does the status of individual vaccinations (e.g., Measles, Rota2) differ based on caregiver beliefs and experiences, after adjusting for confounders?
3. If an association is found, which variables may be acting as confounders in this relationship?

1.2 Study Hypotheses

1.3 Primary Hypotheses

1. Status to all recommended immunizations for children ages 12-23 months would be associated with caregiver positive beliefs and experiences about vaccination.
2. Vaccination status to specific individual vaccines (e.g. Mealses, Rota2) would be associated with caregiver positive beliefs and experiences about vaccination.

1.4 Secondary Hypotheses

1. If an association is present, demographic variables may confound this association.

2 Study Population

2.1 Inclusion Criteria

- Household in Bugoye sub-county with child 12-23 months of age who reside with a caregiver

2.2 Exclusion Criteria

- Caregiver does not present at home at time of visit
- Three unsuccessful attempts to obtain finger-prick blood

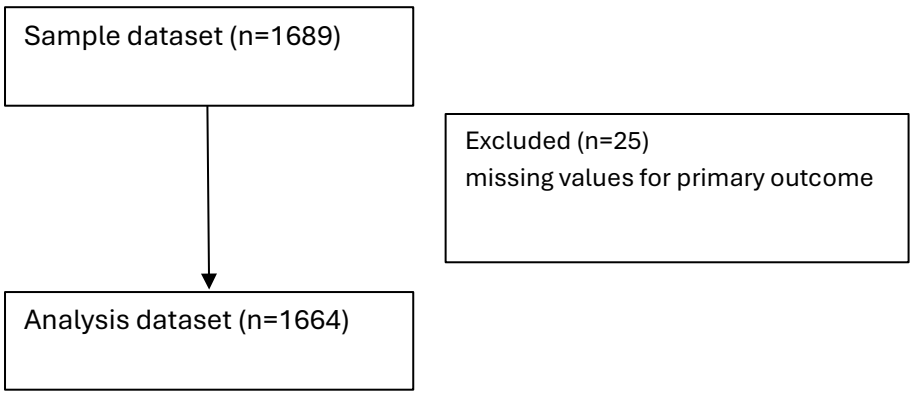
2.3 Data Acquisition

Study design	Cross-sectional census
Study population	Children ages 12–23 months in rural western Uganda who reside with a caregiver
Contact information for team member responsible for data collection/acquisition	Annika Gunderson Agunderson@unc.edu
Study Time	1/20/2021 to 4/30/2021
Data transfer method and date	SharePoint (2025-03-06)
Where dataset is stored	OneDrive: BIOS 841-Sp25->shared document

Data Description:

The Gates Vaccine Access (Phase 3) dataset is a cross-sectional survey of 1,689 children aged 12–23 months in Bugoye sub-county, capturing data from vaccination cards or self-reports when cards were missing. It includes 67 variables covering caregiver information, child characteristics, vaccination beliefs and experiences (predictors), immunization status to eight vaccines (responses), and potential confounders such as caregiver education, marital status, and household demographics. We created a binary variable called vacc_full, indicating full vaccination status based on the 8 vaccine statuses already in the dataset.

Description of patient flow:



3 Outcomes, Exposures, and Additional Variables of Interest

3.1 Primary Outcome(s)

Outcome	**Description**	**Variables and Source**	**Specifications**
Full vaccination status	Whether a child has received all 8	vacc_full	Binary variable; 1 indicates that all eight vaccines have been received and 0 indicates that less than eight

	vaccines or not; for analysis, we created this categorical variable to represent vaccination status for all children according to their vaccination card or self-reported vaccination status		vaccines have been received (according to the child's vaccination card and self-reported vaccination status)

3.2 Secondary Outcome(s)

Outcome	**Description**	**Variables and Source**	**Specifications**
Self-reported vaccination status	Whether a child without a vaccination card received all vaccines or not according to a caregiver's self-reported information	vacc_recall	The vacc_recall variable is a categorical variable that indicates whether a child without a vaccination card has received all (coded as 2), greater than 50% (coded as 1), or at most 50% (coded as 0) of the vaccines
Measles vaccination card status	Whether a child has received the measles vaccine or not	measles	Binary variable; 1 indicates that the vaccine has been received (according to the child's vaccination card) and 0 indicates that the vaccine has not been received (according to the child's vaccination card); the status of children without a vaccination card is recorded as NA
Rotavirus 1, 2 vaccination card status	Whether a child has received the rotavirus vaccines or not	rota1, rota2	Binary variable, 1 indicates that the vaccine has been received (according to the child's vaccination card) and 0 indicates that the vaccine has not been received (according to the child's vaccination card); the status of children without a vaccination card is recorded as NA
BCG vaccination card status	Whether a child has received the	bcg	Binary variable, 1 indicates that the vaccine has been received (according to the child's vaccination card) and 0 indicates that the vaccine has not

	BCG vaccine or not		been received (according to the child's vaccination card); the status of children without a vaccination card is recorded as NA
OPV vaccination card status	Whether a child has received the OPV vaccine or not	opt	Binary variable, 1 indicates that the vaccine has been received (according to the child's vaccination card) and 0 indicates that the vaccine has not been received (according to the child's vaccination card); the status of children without a vaccination card is recorded as NA
PENTA vaccination card status	Whether a child has received the PENTA vaccines or not	penta1, penta2, penta3	Binary variable, 1 indicates that the vaccine has been received (according to the child's vaccination card) and 0 indicates that the vaccine has not been received (according to the child's vaccination card); the status of children without a vaccination card is recorded as NA
Vaccination card	Whether a child's vaccination status was reported on a card or not	card	Binary variable, 1 indicates that vaccines were reported on a card and 0 indicates that vaccines were not reported on a card, and otherwise self-reported);

3.3 Additional Variables of Interest

Variable	**Description**	**Variables and Source**	**Specifications**
Vaccine beliefs	Caregiver beliefs about vaccines (10 predictor variables)	vacc_more, vacc_healthykids, vacc_harmgood, vacc_natimmune, vacc_parentchoice, vacc_2for1, vacc_severe, vacc_refuse, vacc_ses, vacc_guidelines,	Categorical variable; 0 indicates a response of "no" to the question asked, 1 indicates a response of "yes" to the question asked, and 2 indicates a response of "not sure" to the question asked

Vaccine experiences	Caregiver experiences about vaccines (14 predictor variables)	exp_policy, exp_safe, exp_protect, exp_self, exp_seen, exp_delay, exp_defer, exp_infant, exp_loc, exp_when, exp_access, exp_wait, exp_concern, exp_trust,	Categorical variable; 0 indicates a response of "no" to the question asked, 1 indicates a response of "yes" to the question asked, and 2 indicates a response of "not sure" to the question asked
Vaccine location	Where a vaccine was received (1 covariate)	vacc_where	Categorical variables; specifies a value that corresponds to a particular location
Confounding variables	Includes 3 biological measures, 5 childbirth measures,and 7 caregiver information/demographic variables	muac (continuous), child_sex (categorical), age_months (continuous), -- birthplace2 (categorical), birth_status(categorical), bednet (categorical), inpatient (categorical), card (categorical) -- caregiver_sex (categorical), relationship (categorical) education (categorical), marital (categorical), children (continuous), birthplace (continuous), oldest_child (continuous)	Coding differs based on the variable; we will not explain the scheme of each variable due to lack of space, but please reference the data dictionary

4 Statistical Analysis Plan

4.1 Demographic and Clinical Characteristics ("Table 1")

As part of our exploratory data analysis, we will present a table summarizing the demographic and clinical characteristics of the participants. For numerical variables, we will report the mean and standard deviation, while for categorical variables, we will provide the frequency and proportions of each category. We aim to complete this section by March 28th.

4.2 Analyses Plan for Aim 1

In this analysis, a logistic regression model will be fitted to predict the vaccination status of a child, classified as fully vaccinated or not fully vaccinated. The model will include 24 predictors related to vaccination beliefs and experiences, as well as 24 potential confounding variables. The outcome variable is defined based on the number of vaccinations a child has received. Specifically, children who have received all eight recommended vaccines, as indicated on their vaccination card, will be classified as "fully vaccinated." Those who have received fewer than eight vaccines will be classified as "not fully vaccinated." In cases where the vaccination card is lost or damaged, the child's immunization status will be validated using self-reported immunization data.

In terms of the predictors, we will use two approaches and choose one of them that makes more sense for further analysis.

1. Given that there are 10 variables related to vaccination beliefs and 14 variables related to vaccination experiences, backward selection will be used to build a logistic regression model with the most significant predictors.
2. The 10 variables related to vaccination beliefs and the 14 variables related to vaccination experiences will be aggregated into two separate predictors: one representing vaccination beliefs and the other representing vaccination experiences. The two aggregated predictors will capture the overall positive or negative beliefs and experiences, respectively, before fitting the logistic regression model.

We will fit a logistic model with all the variables first and eliminate confounding variables determined from exploratory data analysis from the model one by one to get a final model. Finally, we plan to report adjusted odds ratios and their two-sided 95% confidence intervals, which would be interpreted. We will also check coefficients of the variables in the model to see the direction and strength of the association with the outcome. A hypothesis testing will be based on a type I error of 0.05. We plan to finish this section by April 3rd. Our full model using the second approach for the predictors is shown below.

$$logit\,(\Pr(Vaccination\ Status_i = 1)) = \beta_0 + \beta_1 1(Beliefs_i = 1) + \beta_2 1(Experiences_i = 1) + \sum_{i=0}^{3} \beta_{3i} 1(Education_i = j) + \cdots + \sum_{i=0}^{4} \beta_{22i} 1(Relationship_i = j)$$

4.3 Analyses Plan for Aim 2

A logistic regression model will be fitted with a binary outcome indicating whether a child is vaccinated against a specific virus, such as Measles, or not vaccinated. Additionally, a logistic regression model will be fitted with a binary outcome indicating whether a child's vaccination status was reported on a card or not. The predictors and confounding variables will remain the same as those outlined in the analyses plan for Aim 1. The only difference is that the outcome will be specific to the vaccination type. This analysis is expected to be completed by April 6th.

5 Limitations

Missing Data: A significant percentage of participants (34%) in the census do not have a vaccination card available, which inherently limits the accuracy of our analysis. Furthermore, among participants with a vaccination card, 94.6% received all vaccines, while 0.009% (only one participant) received fewer than 50% of all vaccines. By contrast, among participants without a vaccination card, just 58% recalled receiving all vaccines, and 3.3% (18 participants) received fewer than 50% of all vaccines. Thus, the accuracy of our analyses will be significantly affected by the accuracy of self-reported data, and inaccuracies in self-reported vaccination status are likely to significantly impact our conclusions. There are less than 2% missing in the primary outcome: number of vaccines received. We just removed them from our analysis.

6 Addendum for Additional Analyses

We will evaluate the summary statistics, sample distributions, correlation, outliers, and missing data of all variables for exploratory data analysis. The confounding variables would be examined using principal component analysis (PCA) to avoid multicollinearity before fitting the model. We will conduct sensitivity analyses to test the impact of a) possible inaccuracies in self-reported vaccination status and b) outliers/ influential observations. If time permits, we will visualize a confusion matrix and ROC curve to see the performance of our classification model after splitting the dataset into a train (80%) and test set (20%).

7 Appendix

Statistical Software: We will use R version 4.4.3.

8 References

Boyce RM, Delamater P, Muhindo R et al. Accessible metrics of access: Novel tools to measure immunization coverage in rural sub-Saharan Africa [version 1; peer review: 1 approved, 1 approved with reservations]. Gates Open Res 2019, 3:1540 (https://doi.org/10.12688/gatesopenres.13066.1)

Uganda Bureau of Statistics (UBOS) and ICF. (2018). Uganda Demographic and Health Survey 2016. Kampala, Uganda and Rockville, Maryland, USA: UBOS an