Examining the Association Between BCG Immunization and Leprosy in East Africa: A Case-Control Study

Biostatistics 536 | Fall 2023 | Group 1

Background

The global burden of leprosy or Hansen's Disease, a chronic infectious disease, decreased considerably in the past four decades but remains a major cause of disability.[1] While leprosy trended downward worldwide, cases have recently increased in Africa, particularly East Africa. Some countries in this region are on the World Health Organization's global priority list, accounting for most new cases.[2] Research suggests that the Bacillus Calmette-Guerin (BCG) vaccine offers protection against mycobacterial infections, including leprosy.[3,4] This study aims to investigate whether BCG vaccination against tuberculosis in early childhood was associated with lower odds of leprosy in East Africa.

Study Design

We employed an unmatched case-control study design. Cases consisted of patients aged 5 years or older with a new diagnosis of leprosy, using a standard case definition of the disease. Population-based controls were aged 5 years or older and leprosy-free, and identified through a cross-sectional survey questionnaire. A total of 274 cases and 1,096 controls were included in this study.

Statistical Methods

Characteristics of the study population were summarized using absolute counts and corresponding proportions. For our main confirmatory analysis, we fit an adjusted multivariable logistic regression model to evaluate the association between BCG vaccination in early childhood and new leprosy diagnosis. The exposure, BCG vaccination, was a binary variable. Participants were considered vaccinated when a vaccination scar was present and unvaccinated when absent, coded as "1" and "0", respectively. The outcome was a new diagnosis of leprosy, also a binary variable, coded as "1" if present and "0" if absent. We selected confounding variables *a priori* using subject matter knowledge;[5] they are presented in a directed acyclic graph, see Figure 1, and described in detail in Table 1. In our main analysis, we adjusted for the following: sex (male or female); categorized age in years (5-9, 10-14, 15-19, 20-24,

25-29, 30-44, or 45 and older) as indicator variables; housing type (brick, sunbrick, wattle, or temporary) as indicator variables; and level of formal education (none, 1-5 years, 6-8 years, or secondary/tertiary level) as indicator variables. We used complete-case analysis, excluding participants with missing data. We estimated p-values and 95% confidence intervals (CI) with robust standard errors using a Wald's approach. To check the robustness of our findings, we refit the main analysis model excluding participants aged >=30 years (i.e., 135 cases and 307 controls). All tests were two-sided and we set an alpha level of 0.05.

Results

A total of 47 (17.2%) cases and 500 (45.6%) controls were vaccinated with BCG. Sociodemographic characteristics of cases and controls were similar (Table 2). In our main analysis, the odds of leprosy diagnosis for those who received a BCG vaccination was 0.35 (95% CI: 0.23 - 0.52; $p < 0.01$) times the odds of those who did not receive a BCG vaccination, adjusted for age, sex, housing type, and education. Our estimate remained similar in additional analysis. After excluding patients aged >=30 years, the odds of leprosy diagnosis for those who received a BCG vaccination was 0.39 (95% CI: 0.25 - 0.60; $p < 0.01$) times the odds of those who did not receive a BCG vaccination, adjusted for age, sex, housing type, and education.

Discussion

We found an inverse association between BCG vaccination in early childhood and leprosy diagnosis among participants in East Africa. This study has numerous strengths: we captured new leprosy cases and minimized the possibility of length-biased sampling. Our research used real-world data from a regional, representative health survey, with information on key factors such as age, sex, housing type, and education. Further, we used an efficient study design to evaluate a relevant public health issue that should be confirmed in future studies. This study also presents several limitations. The potential misclassification of exposure status remains for BCG vaccination. Vaccinated participants who do not develop a BCG scar (a marker of impaired immune response) may be classified as unvaccinated and are at higher risk of developing leprosy.[6] In keeping, we did not have data on participants' immune status. BCG vaccination is contraindicated in immunocompromised patients, who are at a higher risk of developing leprosy.[7] Thus, our results may overestimate the association between BCG and leprosy. Finally, we did not measure variables such as income, comorbidities, or contact with patients living with leprosy. Thus, unmeasured confounding may be a threat to the validity of our findings. In conclusion, our study found an inverse association between BCG and leprosy in East Africa, suggesting that BCG vaccination early in life may be a protective factor for leprosy. Future investigations confirming our findings in additional settings and in dissimilar populations remain warranted.

APPENDIX

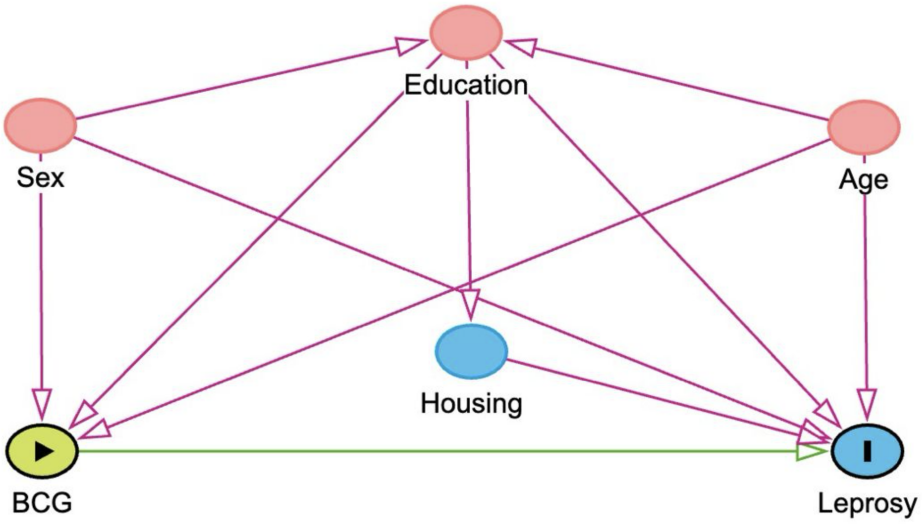

Figure 1: Directed Acyclic Graph (DAG)

Table 1: Rationale for Variable Selection in Primary Regression Model

Variable	Role	Rationale/Justification	Include/Exclude
age *Categorized age in years*	Confounder	Using prior knowledge, we believe age is associated with the exposure and outcome. We also recognize that it is not on the causal pathway. We constructed a directed acyclic graph (DAG) to illustrate the role of age in the relationship and found that not adjusting for the variable would result in a backdoor path. See Figure 1. This variable is part of a minimally sufficient set	Include
sex *Participant's sex*	Confounder	Using prior knowledge, we believe sex is associated with exposure and outcome. We also recognize that it is not on the causal pathway. The DAG in Figure 1 revealed an unblocked path from exposure to outcome if we did not adjust for sex. This variable is part of a minimally sufficient set.	Include
school *Level of formal school, categorized*	Confounder	Using prior knowledge, we believe that school is associated with exposure and outcome. We also recognize that it is not on the causal pathway. The DAG illustrated school as a collider along the path from exposure to outcome. See Figure 1. This variable is part of a minimally sufficient set.	Include
house *Type of housing of the participant*	Precision Variable	Using prior knowledge, we believe house is associated with the outcome. As a proxy for socioeconomic status, house may influence the spread of the bacteria following prolonged and close contact. While the DAG revealed that the variable was not necessary for the minimally sufficient set, we believe it will increase the precision of the estimate for BCG vaccination. See Figure 1	Include

Table 2: Summary Statistics for Case-Control Study Demographics

	Disease Outcome Group	
	Control	Case
n	1096	274
Sex = Female (%)	562 (51.3)	162 (59.1)
Age (%)		
5-9	263 (24.0)	34 (12.4)
10-14	207 (18.9)	35 (12.8)
15-19	167 (15.2)	21 (7.7)
20-24	95 (8.7)	18 (6.6)
25-29	57 (5.2)	31 (11.3)
30-44	135 (12.3)	64 (23.4)
45+	172 (15.7)	71 (25.9)
School (%)		
No Schooling	198 (19.3)	84 (34.4)
1-5 Years	486 (47.3)	120 (49.2)
6-8 Years	311 (30.3)	39 (16.0)
Secondary / Teritary Schooling	32 (3.1)	1 (0.4)
Housing (%)		
Brick	207 (20.7)	33 (13.3)
Sunbrick	237 (23.7)	58 (23.3)
Wattle	528 (52.9)	151 (60.6)
Temporary	26 (2.6)	7 (2.8)

REFERENCES

1, 2. World Health Organization. 2022. "Global Leprosy (Hansen Disease) Update, 2021: Moving Towards Interruption of Transmission." 36. *Weekly Epidemiological Record*. Vol. 97.

3. Nepomuceno Gondim Costa Lima L, Leticia Pinto Paz J, do Perpetuo Socorro Correa Amador Silvestre M, Siqueira Moura L, Perini Furlaneto I, Valeria Batista Lima K. 2020. "BCG Vaccination Status, Age, and Gender as Risk Factors for Leprosy in Endemic Areas in the Brazilian Amazon." *Infect Disease Reports* 12 (3): 97–104.

4. Zodpey SP, Shrikhande SN, Salodkar AD, Maldhure BR, Kulkarni SW. 1998. "Effectiveness of Bacillus Calmette-Guerin (BCG) Vaccination in the Prevention of Leprosy; a Case-Finding Control Study in Nagpur, India." *The International Journal of Leprosy and Other Mycobacterial Diseases* 66 (3): 309–15.

5. VanderWeele TJ. 2019. "Principles of Confounder Selection." *European Journal of Epidemiology* 34 (3): 211–19.

6, 7. Lobo, N., Brooks, N.A., Zlotta, A.R. et al. 2021. "100 Years of Bacillus Calmette–Guérin Immunotherapy: From Cattle to COVID-19." *Nature Reviews Urology* 18: 611–22.

Code Appendix

```
knitr::opts_chunk$set(echo = TRUE)

library(knitr)
library(tidyverse)
library(dplyr)
library(rigr)
library(tableone)
library(haven)
library(kableExtra)

lep <- read_dta("leprosyCCstudy.dta")
glimpse(lep)
min(lep$sex)

########################################
######### REGRESSION MODELS ############
########################################


###### READ IN DATA #######

lep <- read_dta("leprosyCCstudy.dta")

####### ONLY USE COMPLETE CASE #######

lep_complete <- na.omit(lep)

###### ASSESS RELEVANT VARIABLES ######
# Note: school appear ordinal; house is nominal

min(lep$age)
min(lep$school)
min(lep$house)
min(lep$sex)
```

```
##### AS FACTOR DESIRED CAT VARIABLES #####

#lep_complete$sex <- as.factor(lep_complete$sex)
#lep_complete$house <- as.factor(lep_complete$house)
lep_complete$age <- as.factor(lep_complete$age)
lep_complete$school <- as.factor(lep_complete$school)

############### FULL MODEL ###############

model_full <- regress("odds", D ~ BCG + sex + age + school + house,
                  data = lep_complete)


############ AGE ADJUSTED MODEL ############

# Omit 30+
lep_younger <- filter(lep_complete, age != 6 & age != 7)

model_young <- regress("odds", D ~ BCG + sex + age + school + house,
                       data = lep_younger)




########################################
############ DAG FIGURE ############
########################################

knitr::include_graphics('Group1_DAG.png')


####################################
####### JUSTIFICATION TABLE ########
####################################

data <- data.frame(
```

```
  Variable = c("\\textbf{age}\\newline \\textit{Categorized age in years}", "\\textbf{sex}\\newline \\te
  Role = c("Confounder", "Confounder", "Confounder", "Precision Variable"),
  Rationale_Justification = c(
    "Using prior knowledge, we believe age is associated with the exposure and outcome. We also recogni
    "Using prior knowledge, we believe sex is associated with exposure and outcome. We also recognize th
    "Using prior knowledge, we believe that school is associated with exposure and outcome. We also rec
    "Using prior knowledge, we believe house is associated with the outcome. As a proxy for socioeconom
  ),
  Include_Exclude = c("Include", "Include", "Include", "Include")
)

##### ADJUST MAX OF CHARACTERS PER LINE  #####
data$Rationale_Justification <- stringr::str_wrap(data$Rationale_Justification, width = 90)

##### NICE TABLE W/ ADJUSTED COLUMNS #####
kable(data, format = "latex", booktabs = TRUE,
      col.names = c("Variable", "Role", "Rationale/Justification", "Include/Exclude"),
      align = 'l',
      escape = FALSE,
      caption = "Rationale for Variable Selection in Primary Regression Model") %>%
  kable_styling(full_width = F) %>%
  column_spec(1, width = "2.5cm") %>%
  column_spec(2, width = "2cm") %>%
  column_spec(3, width = "8cm") %>%
  column_spec(4, width = "2.5cm") %>%
  row_spec(0:nrow(data), extra_latex_after = "\\addlinespace")
#%>% kable_styling(latex_options = c('HOLD_position', 'scale_down'))


####################################
############ TABLE TWO #############
####################################

# CAPITALIZE AND CREATE TABLE WORTHY TITLES
lep <- lep %>%
  rename(Sex = sex,
```

```
        Age = age,
        School = school,
        Housing = house) %>%
  mutate(
    BCG = factor(BCG, levels = c(1,2),
                 labels = c("BCG Vaccination Scar Absent",
                            "BCG Vaccination Scar Present")),
    D = factor(D, levels = c(0,1), labels = c("Control", "Case")),
    Sex = factor(Sex, levels = c(1,2), labels = c("Male", "Female")),
    Age = factor(Age, levels = c(1,2,3,4,5,6,7),
                 labels = c("5-9","10-14", "15-19", "20-24", "25-29", "30-44", "45+")),
    School = factor(School, levels = c(1,2,3,4),
                    labels = c("No Schooling", "1-5 Years", "6-8 Years", "
                               Secondary / Teritary Schooling")),
    Housing = factor(Housing, levels=c(1,2,3,4),
                     labels = c("Brick", "Sunbrick", "Wattle", "Temporary"))
  )

# IDENTIFY VARIABLES AND CAT VARIABLES
vars <- c("Sex", "Age", "School", "Housing")
catvars <- c("Sex", "Age", "School", "Housing")

# CREATE TABLE1
table1 <- CreateTableOne(vars = vars, data = lep, strata = "D", factorVars = catvars,
                         test = FALSE) #, addOverall = TRUE

# CONVERT TO DATAFRAME
table1_df <- as.data.frame(print(table1, printToggle = FALSE))

# PRINT FANCY KABLE TABLE 1
kable(table1_df ,format = 'latex', align = 'c', vline = '',booktabs = TRUE,
      caption ="Summary Statistics for Case-Control Study Demographics") %>%
  add_indent((1:nrow(table1_df))[substr(rownames(table1_df), 1, 1) == ' ']) %>%
  add_header_above(c(' ' = 1,'Disease Outcome Group' = 2),bold = TRUE)
  # %>% kable_styling(latex_options = c('HOLD_position', 'scale_down')) #
```