Group 4

Data Analysis Project

1. Background

Leprosy is a disease caused by bacterial infection which can lead to long-term nerve damage if left untreated. The bacterium that causes leprosy is closely related to the bacterium that causes tuberculosis. The purpose of this study is to assess whether BCG vaccination against tuberculosis in East Africa also has a protective effect against leprosy. BCG vaccination status was assessed in both cases and controls by the presence or absence of the characteristic BCG vaccination scar. The investigative team also collected data on the age, sex, years of education, and type of housing of each participant.

2a. Study Design

This is an unmatched case-control study to determine if BCG vaccination against tuberculosis in early childhood protects against leprosy in East Africa. Cases were defined as patients with newly diagnosed leprosy, and controls were identified from a cross-sectional survey of the population. There were a total of 274 (20.0%) cases, and 1096 (80.0%) controls, for a total sample size of 1370 participants.

2b. Statistical Methods

To evaluate if BCG vaccination protected against leprosy in East Africa, we used a logistic regression model to calculate unweighted adjusted odds ratios and 95% confidence intervals using robust standard errors comparing vaccination status and disease status. The exposure, BCG vaccination, was used as a binary variable (0 = no vaccination, 1 = vaccination). The outcome, leprosy, was also coded as a binary variable (0 = no leprosy, 1 = leprosy). From Table 1, type of housing for both case and control groups seem to have very similar distributions with that of the population, so type of housing will not be adjusted in the model. We identified age and sex as possible confounders. As illustrated in Figure 1, we expect age to be causally and non-monotonically associated with BCG vaccination due to vaccination schedule, and to low availability of BCG vaccination at the time when participants over 30 years of age were young children. We also expect age to be causally associated with leprosy status due to the increase of lifetime exposure risk increasing over time. We expect sex to be causally associated with both BCG vaccination and tuberculosis status for social reasons. While we also expect education to be associated with both BCG vaccination and tuberculosis status, we expect both age and sex to have a causal influence on years of education, and so do not adjust for it in our analysis. Age was adjusted for as a dummy indicator variable, with each dummy variable representing age 5 to 9 (reference group), 10 to 14, 15 to 19, 20 to 24, 25 to 29, 30 to 44, and

above 45. Sex was coded as a binary variable with 0 = males and 1 = females. The model is presented as below:

$$logit(P(Leprosy)) = \alpha + \beta * BCG + \gamma_1 * age_{10-14} + \gamma_2 * age_{15-19} + \gamma_3 * age_{20-24} + \gamma_4 * age_{25-29} + \gamma_5 * age_{30-44} + \gamma_6 * age_{45+} + \gamma_7 * sex$$

We would use a Wald test with robust standard error estimates and with 1 degree of freedom to test whether the likelihood of leprosy infection is equal between people getting or not getting a BCG vaccination (null hypothesis: $\beta = 0$), with the significance level set to 0.05.

3. Results

According to Table 1, there were 547 (39.9%) participants who had the vaccine, and 823 (60.1%) of participants who did not have the BCG vaccine. Of the 274 cases who had leprosy, 47 (17.2%) were vaccinated and 227 (82.8%) were not, and of the 1096 controls, 500 (45.6%) were vaccinated and 596 (54.4%) were not (Table 1). The percentage of vaccination in the case group was substantially lower than that of the control group. Other than the type of housing, covariates showed an obvious difference in case group and control group.

By fitting the logistic regression model as stated in the method, we estimated that the odds ratio of leprosy infection for those who got the BCG vaccination, adjusted for age group and sex with dummy indicator variables, was 0.34 (95% CI using robust standard error: 0.23 - 0.49, P < 0.0001) as compared to those who did not get the BCG vaccination (Table 2). A Wald test was performed to test whether the odds ratio was equal to one, and based on the P-value, we concluded that there was very strong evidence against the null hypothesis.

4. Discussion

Our results indicated that participants with tuberculosis had substantially lower odds of having a history of BCG vaccination than participants without tuberculosis. Although this was an observational study and not sufficient to establish a causal relationship, this result is suggestive of a protective effect of BCG vaccination against leprosy. Strengths of the study include exposure classification based on BCG scar, which eliminates the potential for recall bias, a large enough sample size for a well-powered analysis, and the completeness of the dataset for relevant variables. Fundamentally, the interpretation of this study is limited by the study design. Since there are a number of interacting social factors that could potentially act as confounders on the relationship between BCG vaccination and tuberculosis, it is difficult to be confident that confounding is sufficiently accounted for in the analysis. However, it is not feasible to investigate this research question using a randomized controlled trial.

5. Tables and Figures

Table 1. Descriptive statistics of participants by case and control group

	Case (N=274)	Control (N=1096)	Overall (N=1370)
BCG vaccination scar			
Yes	47 (17.2%)	500 (45.6%)	547 (39.9%)
No	227 (82.8%)	596 (54.4%)	823 (60.1%)
BCG vaccination for age 30+			
Yes	1 (0.4%)	25 (2.3%)	26 (1.9%)
No	134 (48.9%)	282 (25.7%)	416 (30.4%)
Under 30	139 (50.7%)	789 (72.0%)	928 (67.7%)
Age category			
5-9	34 (12.4%)	263 (24.0%)	297 (21.7%)
10-14	35 (12.8%)	207 (18.9%)	242 (17.7%)
15-19	21 (7.7%)	167 (15.2%)	188 (13.7%)
20-24	18 (6.6%)	95 (8.7%)	113 (8.2%)
25-29	31 (11.3%)	57 (5.2%)	88 (6.4%)
30-44	64 (23.4%)	135 (12.3%)	199 (14.5%)
45+	71 (25.9%)	172 (15.7%)	243 (17.7%)
Sex			
male	112 (40.9%)	534 (48.7%)	646 (47.2%)
female	162 (59.1%)	562 (51.3%)	724 (52.8%)
Schooling level			
none	84 (30.7%)	198 (18.1%)	282 (20.6%)
1-5yrs	120 (43.8%)	486 (44.3%)	606 (44.2%)
6-8yrs	39 (14.2%)	311 (28.4%)	350 (25.5%)
sec/tert	1 (0.4%)	32 (2.9%)	33 (2.4%)
Missing	30 (10.9%)	69 (6.3%)	99 (7.2%)
Housing type			
brick	33 (12.0%)	207 (18.9%)	240 (17.5%)
sunbrick	58 (21.2%)	237 (21.6%)	295 (21.5%)
wattle	151 (55.1%)	528 (48.2%)	679 (49.6%)
temp	7 (2.6%)	26 (2.4%)	33 (2.4%)
Missing	25 (9.1%)	98 (8.9%)	123 (9.0%)

Figure 1. Directed acyclic graph illustrating leprosy study variables

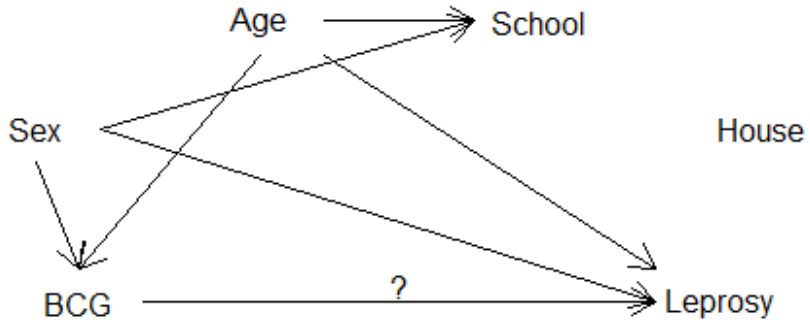

Table 2. Odds ratio estimates, 95% CI, and P-value for study covariates

	Estimate	95% CI using robust SE	DF	P-Value
BCG(yes)	0.34	0.23, 0.49	1	< 0.0001
Age			6	0.0005
10-14	1.34	0.79, 2.26	1	0.27
15-19	1.22	0.67, 2.22	1	0.53
20-24	1.61	0.86, 2.99	1	0.13
25-29	3.29	1.83, 5.93	1	0.0001
30-44	2.38	1.47, 3.85	1	0.0004
45+	2.08	1.31, 3.32	1	0.0021
Sex(female)	1.22	0.92, 1.61	1	0.18

Code

```
library(tidyverse)
knitr::opts_chunk$set(echo = F)
## Preparation
setwd("C:/Users/second/Desktop/BIOST 536/DataAnalysisProject")
load("leprosyCCstudy.Rdata")
head(leprosyCCstudy)
## table1
library(table1)
lep.table <- leprosyCCstudy
lep.table$BCG <- factor(lep.table$BCG, levels = c("yes","no"), labels = c("Yes","No"))
label(lep.table$BCG) <- "BCG vaccination scar"
label(lep.table$age) <- "Age category"
label(lep.table$sex) <- "Sex"
label(lep.table$school) <- "Schooling level"
label(lep.table$house) <- "Housing type"
lep.table$D <- factor(lep.table$D, levels = c(1,0), labels = c("Case","Control"))

lep.table <- lep.table %>%
  mutate(
    bcg30 = case_when(
      age %in% c("30-44", "45+") ~ as.character(BCG),
      TRUE ~ "under 30"
    )
  )
lep.table$bcg30 <- factor(lep.table$bcg30,
                          levels = c("Yes","No","under 30"),
                          labels = c("Yes","No", "Under 30"))
label(lep.table$bcg30) <- "BCG vaccination for age 30+"

table1(~BCG+bcg30++age+sex+school+house|D, data = lep.table)
## dag
library(dagR)
dag1 <- dag.init(outcome=NULL,exposure=NULL, # these are never used
                 covs=c(1,1,1,1), # one element for each covariate (1=standard, 2=unmeasured)
                 arcs=c(2,-1, # Age -> Leprosy
                        2,0, # Age -> BCG
                        2,3, # Age -> School
                        1,0, # Sex -> BCG
                        1,-1, # Sex -> Leprosy
                        1,3 # Sex -> School

                        ),
                 xgap = 0.1, ygap = 0.08, len = 0.15,
                 x.name = "BCG",y.name = "Leprosy", # name of exposure and outcome
                 cov.names = c("Sex","Age","School","House"),
                 symbols = c("BCG", "Sex","Age","School","House","Leprosy"))

dag.draw(dag1, legend = FALSE)
# Analysis
library(rigr)
regress("odds", D ~ BCG+age+sex, data = leprosyCCstudy)
```