Statistical Analysis Plan for SCOPE

Dan Weinberger

2022-09-22

## Overview

The SCOPE study enrolled adults 60 years of age and older in greater New Haven over the course of 2 years (October 2020-September 2022). Individuals were followed longitudinally over a 10 week period (6 visits, 2 weeks apart). At each study visit, individuals provided a saliva sample, and at the final visit, they provided a urine sample. They also completed a questionnaire about their activities and contacts with others. A unique aspect of this study is that subjects were enrolled as pairs living in the same household, and these two people were the only members of te household.

## Study objectives

The goals of the study are to quantify the acquisition and clearance rates of pneumococcus among older adults living in a community setting, to identify risk factors for colonization, and to quantify within-household transmission of pneumococcus.

## Laboratory methods

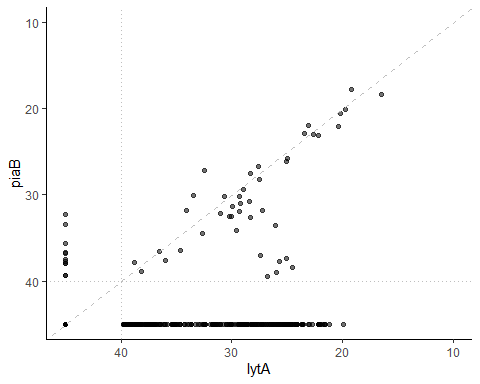
In the laboratory, saliva was culture enriched and tested for *lytA* and *piaB with qPCR.* A Ct value of <40 was considered to be positive. Because piaB is a highly specific marker for pneumococcus and because piaB was more sensitive than the lytA assay at low concentration, we considered those who were positive for piaB to be positive for pneumococcus. To reduce the possibility of false positivity, any samples in the range of 35-40 Ct were retested to confirm the positive. In the case of discrepant results, they were tested a 3rd time as a tie breaker.

## Analysis plan

### Descriptive statistics

* lytA vs piaB Ct values

## Warning: Removed 2 rows containing missing values (geom\_point).



* Percent of tests that were positive, overall

## # A tibble: 1 x 3  
## N\_piaB N\_test `Percent piab`  
## <dbl> <int> <dbl>  
## 1 54 1088 4.96

- and by year

## # A tibble: 2 x 4  
## season N\_piaB N\_test Pct\_piab  
## <chr> <dbl> <int> <dbl>  
## 1 S1 37 567 6.5  
## 2 S2 17 521 3.3

* Percent of people who were positive (period prevalence)

## `summarise()` has grouped output by 'ID'. You can override using the `.groups`  
## argument.

## # A tibble: 2 x 4  
## season N\_people N\_pos `Percent Positive People`  
## <chr> <int> <dbl> <dbl>  
## 1 S1 95 20 21.1  
## 2 S2 89 10 11.2

* Percent of households that were positive
* ## `summarise()` has grouped output by 'Household'. You can override using the  
  ## `.groups` argument.
* ## # A tibble: 2 x 4  
  ## season N\_households N\_pos `Percent Positive HH`  
  ## <chr> <int> <dbl> <dbl>  
  ## 1 S1 48 15 31.2  
  ## 2 S2 45 8 17.8
* Percent positive and average Ct, by sex

## `summarise()` has grouped output by 'season'. You can override using the  
## `.groups` argument.

## # A tibble: 6 x 5  
## # Groups: season [2]  
## season Gender N\_piaB N\_test Pct\_piab  
## <chr> <chr> <dbl> <int> <dbl>  
## 1 S1 F 16 270 5.9  
## 2 S1 M 18 254 7.1  
## 3 S1 <NA> 3 43 7   
## 4 S2 F 14 262 5.3  
## 5 S2 M 3 252 1.2  
## 6 S2 <NA> 0 6 0

* Percent positive by calendar month
* Percent of individuals with contacts, by age of contacts
* Activity with family?

## `summarise()` has grouped output by 'season'. You can override using the  
## `.groups` argument.

## # A tibble: 5 x 5  
## # Groups: season [2]  
## season Gender N\_Acitivity\_family N Pct  
## <chr> <chr> <dbl> <int> <dbl>  
## 1 S1 F 39 242 16.1  
## 2 S1 M 35 231 15.2  
## 3 S2 F 99 190 52.1  
## 4 S2 M 93 183 50.8  
## 5 S2 <NA> 2 4 50

contact <5 years

## `summarise()` has grouped output by 'season'. You can override using the  
## `.groups` argument.

## # A tibble: 5 x 5  
## # Groups: season [2]  
## season Gender N\_contact\_u5 N Pct  
## <chr> <chr> <dbl> <int> <dbl>  
## 1 S1 F 47 268 17.5  
## 2 S1 M 32 252 12.7  
## 3 S2 F 41 190 21.6  
## 4 S2 M 34 183 18.6  
## 5 S2 <NA> 0 4 0

Child contact <12 m

## `summarise()` has grouped output by 'season'. You can override using the  
## `.groups` argument.

## # A tibble: 5 x 5  
## # Groups: season [2]  
## season Gender N\_contact\_u12m N Pct  
## <chr> <chr> <dbl> <int> <dbl>  
## 1 S1 F 27 268 10.1   
## 2 S1 M 14 252 5.56  
## 3 S2 F 18 190 9.47  
## 4 S2 M 11 183 6.01  
## 5 S2 <NA> 0 4 0

Child contact 12-23m

## `summarise()` has grouped output by 'season'. You can override using the  
## `.groups` argument.

## # A tibble: 5 x 5  
## # Groups: season [2]  
## season Gender N\_contact\_12\_23m N Pct  
## <chr> <chr> <dbl> <int> <dbl>  
## 1 S1 F 8 268 2.99  
## 2 S1 M 6 252 2.38  
## 3 S2 F 23 190 12.1   
## 4 S2 M 21 183 11.5   
## 5 S2 <NA> 0 4 0

Child contact 24-59 m

## `summarise()` has grouped output by 'season'. You can override using the  
## `.groups` argument.

## # A tibble: 5 x 5  
## # Groups: season [2]  
## season Gender N\_contact\_24\_59m N Pct  
## <chr> <chr> <dbl> <int> <dbl>  
## 1 S1 F 20 268 7.46  
## 2 S1 M 19 252 7.54  
## 3 S2 F 20 190 10.5   
## 4 S2 M 14 183 7.65  
## 5 S2 <NA> 0 4 0

Child contact 5-9y

## `summarise()` has grouped output by 'season'. You can override using the  
## `.groups` argument.

## # A tibble: 5 x 5  
## # Groups: season [2]  
## season Gender N\_contact\_5\_9y N Pct  
## <chr> <chr> <dbl> <int> <dbl>  
## 1 S1 F 52 268 19.4  
## 2 S1 M 30 252 11.9  
## 3 S2 F 35 190 18.4  
## 4 S2 M 34 183 18.6  
## 5 S2 <NA> 0 4 0

Child contact >10y

## `summarise()` has grouped output by 'season'. You can override using the  
## `.groups` argument.

## # A tibble: 5 x 5  
## # Groups: season [2]  
## season Gender N\_contact\_10y\_older N Pct  
## <chr> <chr> <dbl> <int> <dbl>  
## 1 S1 F 38 268 14.2  
## 2 S1 M 29 252 11.5  
## 3 S2 F 42 190 22.1  
## 4 S2 M 32 183 17.5  
## 5 S2 <NA> 1 4 25

* Proportion positive, stratified by contacts
  + Contact with children <5 years
  + Contact with children <1 year, 1-2 years, 2-4 years, 5-9 years, 10 years and older
  + Test with Fisher’s Exact Test (p<0.05)

### Regression models

Evaluate correlates of pneumococcal positivity.

* Multivariate logistic regression to identify relevant age groups

mod1 <- glm(piab\_pos ~ child\_contact\_u12m + child\_contact\_13\_23m + child\_contact\_24\_59m + child\_contact\_5\_10y + child\_contact\_over10y , family='binomial')

Generalized estimating equation to account for correlation within individuals

mod1.gee <- gee(piab\_pos ~ child\_contact\_u12m+ child\_contact\_13\_23m + child\_contact\_24\_59m + child\_contact\_5\_10y + child\_contact\_over10y,id=mod.df$ID, data=mod.df, family=binomial(link='log'), corstr="exchangeable")

### Longitudinal data analysis

Use a Markov transition model with 2 states (colonized, uncolonized) to quantify acquisition and clearance rates and correlates of both. The model below estimates acquisition and clearance rates, and teh effect of recent contact with a child <12m on acquisition rates

msm.mod1 <- msm( state~ day , subject=ID , qmatrix=q1 , data=l2,covariates = list("1-2" = ~ child\_contact\_u12m))

Effect of positive household contact. This quantifies transmission within households.

msm.mod1 <- msm( state~ day , subject=ID , qmatrix=q1 , data=l2,covariates = list("1-2" = ~ pos\_hh))

## Notes and challenges

* The analyses will all be done initially based on piaB positivity. Given the low rate of carriage in adults, this is probably fine to assume that if a person is positive at multiple time points, or multiple people in the household are positive, it is the same serotype. (the former has been confirmed–we have 1 person positive for 15B/C at all time points). Once serotype data are available, we can more precisely define duration and transmission, if there is discordance within individuals or households
* The evaluations of household transmission are likely underpowered due to low levels of carriage and the fact that when one household member is positive, the other individual is often also positive at the same time point.