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 ech 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 acquistion 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

* Percent of tests that were positive, overall and by year
* Percent of people who were positive (period prevalence)
* Percent of households that were positive
* Percent positive and average Ct, by sex
* Percent positive by calendar month
* Percent of individuals with contacts, by age of contacts
* 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)
* 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.