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

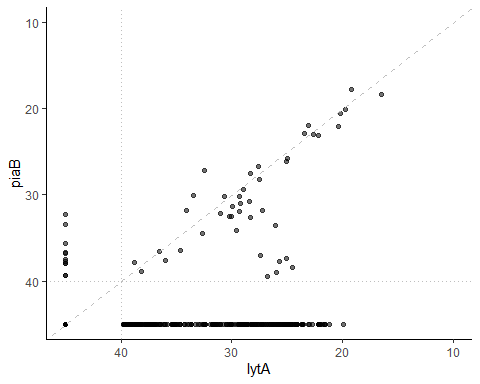
In the laboratory, saliva was culture enriched and tested for *lytA* and *piaB with qPCR.* A Ct value of <40 was considered to be positive. Because piaB is a highly specific marker for pneumococcus and because piaB was more sensitive than the lytA assay at low concentration, we considered those who were 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

| N\_piaB | N\_test | Percent piab |
| --- | --- | --- |
| 54 | 1,088 | 4.96 |

- and by year

| season | N\_piaB | N\_test | Pct\_piab |
| --- | --- | --- | --- |
| S1 | 37 | 567 | 6.5 |
| S2 | 17 | 521 | 3.3 |

* Percent of people who were positive (period prevalence)

| season | N\_people | N\_pos | Percent Positive People |
| --- | --- | --- | --- |
| S1 | 95 | 20 | 21.1 |
| S2 | 89 | 10 | 11.2 |

* Percent of households that were positive

| season | N\_households | N\_pos | Percent Positive HH |
| --- | --- | --- | --- |
| S1 | 48 | 15 | 31.2 |
| S2 | 45 | 8 | 17.8 |

* Percent positive and average Ct, by sex

| Gender | season | N\_piaB | N\_test | Pct\_piab |
| --- | --- | --- | --- | --- |
| F |  |  |  |  |
|  | S1 | 18 | 294 | 6.1 |
|  | S2 | 14 | 268 | 5.2 |
| M |  |  |  |  |
|  | S1 | 19 | 277 | 6.9 |
|  | S2 | 3 | 252 | 1.2 |

* Percent positive by symptoms

| recent\_nasal | N\_piaB | N\_test | Pct\_piab |
| --- | --- | --- | --- |
| No |  |  |  |
|  | 46 | 882 | 5.2 |
| Yes |  |  |  |
|  | 4 | 110 | 3.6 |
|  |  |  |  |
|  | 4 | 99 | 4.0 |

| recent\_runny\_nose | N\_piaB | N\_test | Pct\_piab |
| --- | --- | --- | --- |
| No |  |  |  |
|  | 45 | 893 | 5.0 |
| Yes |  |  |  |
|  | 5 | 99 | 5.1 |
|  |  |  |  |
|  | 4 | 99 | 4.0 |

Recent Abx

| recent\_abx | N\_piaB | N\_test | Pct\_piab |
| --- | --- | --- | --- |
| No |  |  |  |
|  | 45 | 848 | 5.3 |
| Yes |  |  |  |
|  | 2 | 66 | 3.0 |
|  |  |  |  |
|  | 7 | 177 | 4.0 |

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

| Gender | season | N\_Activity\_family | N | Pct |
| --- | --- | --- | --- | --- |
| F |  |  |  |  |
|  | S1 | 38 | 216 | 17.59 |
|  | S2 | 133 | 252 | 52.78 |
| M |  |  |  |  |
|  | S1 | 33 | 200 | 16.50 |
|  | S2 | 118 | 236 | 50.00 |

contact <5 years

| Gender | season | N\_contact\_u5 | N | Pct |
| --- | --- | --- | --- | --- |
| F |  |  |  |  |
|  | S1 | 47 | 257 | 18.29 |
|  | S2 | 55 | 252 | 21.83 |
| M |  |  |  |  |
|  | S1 | 32 | 231 | 13.85 |
|  | S2 | 48 | 236 | 20.34 |

Child contact <12 m

| Gender | season | N\_contact\_u12m | N | Pct |
| --- | --- | --- | --- | --- |
| F |  |  |  |  |
|  | S1 | 27 | 257 | 10.5 |
|  | S2 | 26 | 252 | 10.3 |
| M |  |  |  |  |
|  | S1 | 14 | 231 | 6.1 |
|  | S2 | 17 | 236 | 7.2 |

Child contact 12-23m

| Gender | season | N\_contact\_12\_23m | N | Pct |
| --- | --- | --- | --- | --- |
| F |  |  |  |  |
|  | S1 | 9 | 257 | 3.50 |
|  | S2 | 29 | 252 | 11.51 |
| M |  |  |  |  |
|  | S1 | 6 | 231 | 2.60 |
|  | S2 | 29 | 236 | 12.29 |

Child contact 24-59 m

| Gender | season | N\_contact\_24\_59m | N | Pct |
| --- | --- | --- | --- | --- |
| F |  |  |  |  |
|  | S1 | 20 | 257 | 7.78 |
|  | S2 | 28 | 252 | 11.11 |
| M |  |  |  |  |
|  | S1 | 19 | 231 | 8.23 |
|  | S2 | 21 | 236 | 8.90 |

Child contact 5-9y

| Gender | season | N\_contact\_5\_9y | N | Pct |
| --- | --- | --- | --- | --- |
| F |  |  |  |  |
|  | S1 | 51 | 257 | 19.84 |
|  | S2 | 52 | 252 | 20.63 |
| M |  |  |  |  |
|  | S1 | 30 | 231 | 12.99 |
|  | S2 | 42 | 236 | 17.80 |

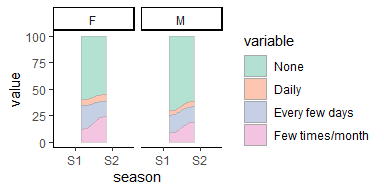
Child contact >10y

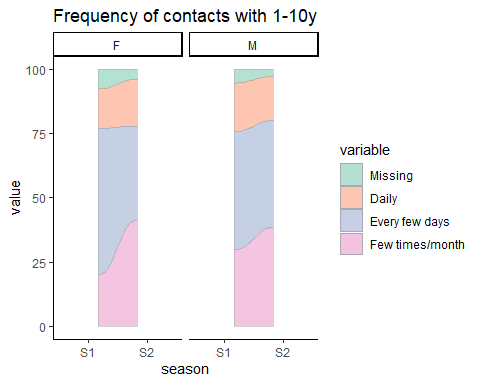
| Gender | season | N\_contact\_10y\_older | N | Pct |
| --- | --- | --- | --- | --- |
| F |  |  |  |  |
|  | S1 | 39 | 257 | 15.18 |
|  | S2 | 55 | 252 | 21.83 |
| M |  |  |  |  |
|  | S1 | 28 | 231 | 12.12 |
|  | S2 | 37 | 236 | 15.68 |

How does intensity of child contact vary by season

0= none reported, 1=Daily; 2=every few days; 3= once or twice a month

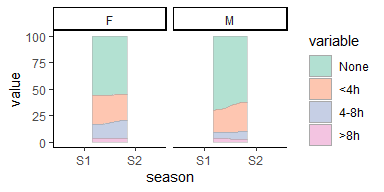
| Gender | season | child\_contact\_often | N | Percent |
| --- | --- | --- | --- | --- |
| F |  |  |  |  |
|  | S1 |  |  |  |
|  |  | 0 | 156 | 60.23 |
|  |  | 1 | 14 | 5.41 |
|  |  | 2 | 58 | 22.39 |
|  |  | 3 | 31 | 11.97 |
|  | S2 |  |  |  |
|  |  | 0 | 138 | 54.98 |
|  |  | 1 | 16 | 6.37 |
|  |  | 2 | 36 | 14.34 |
|  |  | 3 | 61 | 24.30 |
| M |  |  |  |  |
|  | S1 |  |  |  |
|  |  | 0 | 166 | 70.64 |
|  |  | 1 | 9 | 3.83 |
|  |  | 2 | 40 | 17.02 |
|  |  | 3 | 20 | 8.51 |
|  | S2 |  |  |  |
|  |  | 0 | 145 | 61.44 |
|  |  | 1 | 12 | 5.08 |
|  |  | 2 | 35 | 14.83 |
|  |  | 3 | 44 | 18.64 |

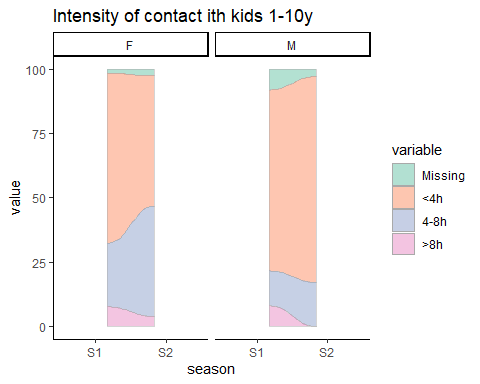


Just among those who have contact with kids 1-10y 

Hours with kids? 1= < 4 hours (just a morning or afternoon or evening) 2= 4-8 hours (full day) 3= 8+ hours (longer visit such as day care or overnight)

| Gender | season | child\_contact\_hours | N | Percent |
| --- | --- | --- | --- | --- |
| F |  |  |  |  |
|  | S1 |  |  |  |
|  |  | 0 | 146 | 56.37 |
|  |  | 1 | 71 | 27.41 |
|  |  | 2 | 34 | 13.13 |
|  |  | 3 | 8 | 3.09 |
|  | S2 |  |  |  |
|  |  | 0 | 138 | 54.98 |
|  |  | 1 | 62 | 24.70 |
|  |  | 2 | 42 | 16.73 |
|  |  | 3 | 9 | 3.59 |
| M |  |  |  |  |
|  | S1 |  |  |  |
|  |  | 0 | 164 | 69.79 |
|  |  | 1 | 50 | 21.28 |
|  |  | 2 | 13 | 5.53 |
|  |  | 3 | 8 | 3.40 |
|  | S2 |  |  |  |
|  |  | 0 | 147 | 62.29 |
|  |  | 1 | 66 | 27.97 |
|  |  | 2 | 17 | 7.20 |
|  |  | 3 | 6 | 2.54 |



And just for kids with contact with 1-10 year olds 

* Does probability of both people being positive differ by season?

0= neither positive; 1= 1 person positive, 2=both positive

| season | N\_pos\_time | N |
| --- | --- | --- |
| S1 |  |  |
|  | 0 | 253 |
|  | 1 | 31 |
|  | 2 | 3 |
| S2 |  |  |
|  | 0 | 248 |
|  | 1 | 13 |
|  | 2 | 2 |

* Proportion positive, stratified by contacts
  + Contact with children <5 years

Between seasons, among people without reported child contact, prevalence is much lower; it is similar among those with a reported contact

| child\_contact\_u5 | season | N | N\_pos | Pct\_Pos |
| --- | --- | --- | --- | --- |
| 0 |  |  |  |  |
|  | S1 | 409 | 19 | 4.6 |
|  | S2 | 385 | 7 | 1.8 |
| 1 |  |  |  |  |
|  | S1 | 79 | 13 | 16.5 |
|  | S2 | 103 | 10 | 9.7 |
| 9,999 |  |  |  |  |
|  | S1 | 83 | 5 | 6.0 |
|  | S2 | 32 | 0 | 0.0 |

- Contact with children \<1 year, 1-2 years, 2-4 years, 5-9 years, 10 years and older  
  
<12m

| child\_contact\_u12m | season | N | N\_pos | Pct\_Pos |
| --- | --- | --- | --- | --- |
| 0 |  |  |  |  |
|  | S1 | 447 | 28 | 6.3 |
|  | S2 | 445 | 12 | 2.7 |
| 1 |  |  |  |  |
|  | S1 | 41 | 4 | 9.8 |
|  | S2 | 43 | 5 | 11.6 |
| 9,999 |  |  |  |  |
|  | S1 | 83 | 5 | 6.0 |
|  | S2 | 32 | 0 | 0.0 |

12-23m

| child\_contact\_13\_23m | N | N\_pos | Pct\_Pos |
| --- | --- | --- | --- |
| 0 |  |  |  |
|  | 903 | 41 | 4.5 |
| 1 |  |  |  |
|  | 73 | 8 | 11.0 |
| 9,999 |  |  |  |
|  | 115 | 5 | 4.3 |

| child\_contact\_24\_59m | N | N\_pos | Pct\_Pos |
| --- | --- | --- | --- |
| 0 |  |  |  |
|  | 888 | 35 | 3.9 |
| 1 |  |  |  |
|  | 88 | 14 | 15.9 |
| 9,999 |  |  |  |
|  | 115 | 5 | 4.3 |

| child\_contact\_5\_10y | N | N\_pos | Pct\_Pos |
| --- | --- | --- | --- |
| 0 |  |  |  |
|  | 801 | 27 | 3.4 |
| 1 |  |  |  |
|  | 175 | 22 | 12.6 |
| 9,999 |  |  |  |
|  | 115 | 5 | 4.3 |

| child\_contact\_over10y | N | N\_pos | Pct\_Pos |
| --- | --- | --- | --- |
| 0 |  |  |  |
|  | 817 | 41 | 5.0 |
| 1 |  |  |  |
|  | 159 | 8 | 5.0 |
| 9,999 |  |  |  |
|  | 115 | 5 | 4.3 |

intensity of contact

| child\_contact\_hours | N | N\_pos | Pct\_Pos |
| --- | --- | --- | --- |
| 1 | 253 | 28 | 11.1 |
| 2 | 106 | 7 | 6.6 |
| 3 | 31 | 1 | 3.2 |
| 999 | 701 | 18 | 2.6 |

frequency of contact vs pneumo positivity

| child\_contact\_often | N | N\_pos | Pct\_Pos |
| --- | --- | --- | --- |
| 1 | 51 | 8 | 15.7 |
| 2 | 172 | 21 | 12.2 |
| 3 | 157 | 6 | 3.8 |
| 999 | 711 | 19 | 2.7 |

Prevalence by sex. Seems that men and women had similar prevalence in Season 1; Women had similar prevalence again in Season 2, men had very little carriage.

| Gender | season | N | N\_pos | Pct\_Pos |
| --- | --- | --- | --- | --- |
| F |  |  |  |  |
|  | S1 | 294 | 18 | 6.1 |
|  | S2 | 268 | 14 | 5.2 |
| M |  |  |  |  |
|  | S1 | 277 | 19 | 6.9 |
|  | S2 | 252 | 3 | 1.2 |

Prevalence by sex and child contact (small numbers)

| child\_contact\_u10 | Gender | N | N\_pos | Pct\_Pos |
| --- | --- | --- | --- | --- |
| 0 |  |  |  |  |
|  | F |  |  |  |
|  |  | 352 | 11 | 3.1 |
|  | M |  |  |  |
|  |  | 348 | 7 | 2.0 |
| 1 |  |  |  |  |
|  | F |  |  |  |
|  |  | 157 | 18 | 11.5 |
|  | M |  |  |  |
|  |  | 119 | 13 | 10.9 |
| 9,999 |  |  |  |  |
|  | F |  |  |  |
|  |  | 53 | 3 | 5.7 |
|  | M |  |  |  |
|  |  | 62 | 2 | 3.2 |

- 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', data=pcr\_survey)  
  
summary(mod1)

##   
## Call:  
## glm(formula = piab\_pos ~ child\_contact\_often + N\_child\_contact\_1\_10,   
## family = "binomial", data = pcr\_survey)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.6158 -0.2792 -0.2241 -0.2241 2.7191   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.691e+00 3.854e-01 -4.388 1.14e-05 \*\*\*  
## child\_contact\_often2 -2.929e-01 4.501e-01 -0.651 0.51523   
## child\_contact\_often3 -1.534e+00 5.673e-01 -2.705 0.00683 \*\*   
## child\_contact\_often999 -1.980e+00 4.601e-01 -4.304 1.68e-05 \*\*\*  
## N\_child\_contact\_1\_10 1.392e-05 1.706e-05 0.816 0.41437   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 429.92 on 1090 degrees of freedom  
## Residual deviance: 397.42 on 1086 degrees of freedom  
## AIC: 407.42  
##   
## Number of Fisher Scoring iterations: 6

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=pcr\_survey$ID, data=pcr\_survey, 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

mod.df <- pcr\_survey %>%  
 filter(child\_contact\_13\_23m!=9999 & !is.na(piab\_pos)) %>%  
 mutate(ID=as.factor(ID))  
  
mod.df$child\_contact\_u5 <-   
 mod.df$child\_contact\_u12m + mod.df$child\_contact\_13\_23m + mod.df$child\_contact\_24\_59m  
  
mod.df$child\_contact\_1\_10 <- mod.df$child\_contact\_13\_23m + mod.df$child\_contact\_24\_59m +  
mod.df$child\_contact\_5\_10y  
  
  
#N contacts  
mod.df$child\_contact\_1\_5 <- mod.df$child\_contact\_13\_23m + mod.df$child\_contact\_24\_59m  
  
  
l1 <- mod.df  
l1$day <- 0  
l1$day[l1$time==2] <- 14  
l1$day[l1$time==3] <- 14\*2  
l1$day[l1$time==4] <- 14\*3  
l1$day[l1$time==5] <- 14\*4  
l1$day[l1$time==6] <- 14\*5  
  
l1$state <- NA  
l1$state[l1$piab\_pos==0 & !is.na(l1$piab\_pos)] <- 1  
l1$state[l1$piab\_pos==1 & !is.na(l1$piab\_pos)] <- 2  
#l1$state <- as.factor(l1$state)  
  
  
q1 <- rbind(c( 0,0.01),  
 c(0.01,0)  
)  
  
l1 <- l1[order(l1$ID, l1$day),]  
  
l2 <- l1[, c('state','child\_contact\_u5','child\_contact\_1\_5','child\_contact','Gender','child\_contact\_u12m','child\_contact\_13\_23m',"child\_contact\_24\_59m","child\_contact\_5\_10y",'child\_contact\_over10y','child\_contact\_1\_10', 'ID','day')]  
  
msm.mod1 <- msm( state~ day , subject=ID , qmatrix=q1 , data=l2,covariates = list("1-2" = ~ child\_contact\_u12m))  
msm.mod1  
  
msm.mod1b <- msm( state~ day , subject=ID , qmatrix=q1 , data=l2,covariates = list("1-2" = ~ child\_contact\_1\_5))  
msm.mod1b  
  
  
msm.mod1b <- msm( state~ day , subject=ID , qmatrix=q1 , data=l2,covariates = list("1-2" = ~ child\_contact\_u5))  
msm.mod1b  
  
msm.mod1c <- msm( state~ day , subject=ID , qmatrix=q1 , data=l2,covariates = list("1-2" = ~ child\_contact\_1\_10))  
msm.mod1c  
  
#Or dichotomize contacts--less clear. but prob need a better way to quantify N contacts (not possible based on way data recorded; only know N age categories)  
l2$child\_contact\_1\_10\_dic <- if\_else(l2$child\_contact\_1\_10>1,1,0)  
  
msm.mod1d <- msm( state~ day , subject=ID , qmatrix=q1 , data=l2,covariates = list("1-2" = ~ child\_contact\_1\_10\_dic))  
msm.mod1d

Does duration or acquisition rate vary by sex? No.

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

mod.df <- pcr\_survey %>%  
 filter( !is.na(piab\_pos)) %>%  
 mutate(ID=as.factor(ID))  
  
recent\_hh\_col <- mod.df %>%  
 group\_by(Household,time ) %>%  
 summarize(piab\_pos=max(piab\_pos)) %>%  
 mutate(time=time+1) %>%  
 rename(piab\_pos\_HH\_prev = piab\_pos)  
  
recent\_ID\_col <- mod.df %>%  
 group\_by(Household,ID,time ) %>%  
 summarize(piab\_pos=sum(piab\_pos)) %>%  
 mutate(time=time+1) %>%  
 rename(piab\_pos\_ID\_prev = piab\_pos)  
  
#colonized\_partner\_prev is defined based on a positive household without positive individual on previous time  
recent\_col <- merge(recent\_hh\_col,recent\_ID\_col, by=c('Household', 'time') ) %>%  
 mutate(colonized\_partner\_prev = piab\_pos\_HH\_prev - piab\_pos\_ID\_prev)  
  
mod.df <- merge(mod.df,recent\_col, by=c('ID','Household', 'time'))   
  
#N contacts  
mod.df$child\_contact\_1\_10 <- mod.df$child\_contact\_13\_23m + mod.df$child\_contact\_24\_59m +mod.df$child\_contact\_5\_10y  
  
  
l1 <- mod.df  
l1$day <- 0  
l1$day[l1$time==2] <- 14  
l1$day[l1$time==3] <- 14\*2  
l1$day[l1$time==4] <- 14\*3  
l1$day[l1$time==5] <- 14\*4  
l1$day[l1$time==6] <- 14\*5  
  
l1$state <- NA  
l1$state[l1$piab\_pos==0 & !is.na(l1$piab\_pos)] <- 1  
l1$state[l1$piab\_pos==1 & !is.na(l1$piab\_pos)] <- 2  
#l1$state <- as.factor(l1$state)  
  
  
q1 <- rbind(c( 0,0.01),  
 c(0.01,0)  
)  
  
l1 <- l1[order(l1$ID, l1$day),]  
  
l2 <- l1[, c('state','Gender','child\_contact\_1\_10', 'ID','day','colonized\_partner\_prev')]   
  
table(l2$state, l2$colonized\_partner\_prev)  
  
msm.mod1 <- msm( state~ day , subject=ID , qmatrix=q1 , data=l2,covariates = list("1-2" = ~ colonized\_partner\_prev ))  
  
msm.mod1

## 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.