

# TB Vaccination

PDF Version available [here](#).

## 0. Data creation

We construct three datasets: `peopleNoVax`, `peopleWithVax`, and `peopleTB`.

`peopleNoVax` is a dataset that spans 17 years. Each year, 10000 people are added to it. Each year, a person has a 5% annual risk of being infected with LTBI. Once a person is infected with LTBI, they have a 2.5% annual risk of being infected with ATBI.

`peopleWithVax` is a dataset that spans 17 years. Each year, 10000 people are added to it. For the first 10 years, a person has a 5% annual risk of being infected with LTBI. Subsequently they have a 2.5% annual risk of being infected with LTBI. Once a person is infected with LTBI, they have a 2.5% annual risk of being infected with ATBI.

`peopleTB` is a dataset that spans 17 years. Each year, 10000 people are added to it. Each year, a person has a 5% annual risk of being infected with ATBI.

```
library(data.table)
library(pomp)
library(ggplot2)
library(rstan)

rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())

#everyone enters on jan 1
#results are for dec 31
newPeople <- data.table(id=1:10000,
                         year=2000,
                         age=1,
                         timeNoVax=0,
                         timeWithVax=0,
                         vaccinated=0,
                         ltbi=0,
                         atbi=0,
                         ageAtTB=0)

if(FALSE){
  file.remove("data/peopleNoVax.RDS")
  file.remove("data/peopleWithVax.RDS")
  file.remove("data/peopleTB.RDS")
  file.remove("results/noVaxLTBI.RDS")
  file.remove("results/withVaxLTBI.RDS")
  file.remove("results/ATBI.RDS")
  file.remove("results/noVaxATBI.RDS")
}

pomp::bake("data/peopleNoVax.RDS", {
  set.seed(4)
  peopleNoVax <- vector("list", length=17)
```

```

peopleNoVax[[1]] <- newPeople[1,] [-1,]
for(i in 1:17){
  # NO VACCINATIONS
  if(i!=1) peopleNoVax[[i]] <- copy(peopleNoVax[[i-1]])
  peopleNoVax[[i]][,age:=age+1]
  peopleNoVax[[i]] <- rbind(peopleNoVax[[i]],newPeople)
  peopleNoVax[[i]][,year:=2000+i]

  N <- nrow(peopleNoVax[[i]])
  peopleNoVax[[i]][sample(1:N,size=round(N*0.050)),ltbi:=1]
  peopleNoVax[[i]][,timeNoVax:=timeNoVax+1]

  peopleNoVax[[i]][,atbi:=0] # people with ATBI go to the doctor and get cured last year
  Nltbi <- sum(peopleNoVax[[i]]$ltbi==1 & peopleNoVax[[i]]$ageAtTB==0)
  peopleNoVax[[i]][sample(
    which(peopleNoVax[[i]]$ltbi==1 & peopleNoVax[[i]]$ageAtTB==0),
    size=round(Nltbi*0.025)),
    atbi:=1]
  peopleNoVax[[i]][atbi==1,ageAtTB:=age]
}
peopleNoVax
}) -> peopleNoVax

pomp::bake("data/peopleWithVax.RDS",{
  set.seed(4)
  peopleWithVax <- vector("list",length=17)
  peopleWithVax[[1]] <- newPeople[1,] [-1,]
  for(i in 1:17){
    # WITH VACCINATIONS, BUT STOPPING THEM AFTERWARDS
    if(i!=1) peopleWithVax[[i]] <- copy(peopleWithVax[[i-1]])
    peopleWithVax[[i]][,age:=age+1]
    peopleWithVax[[i]] <- rbind(peopleWithVax[[i]],newPeople)
    peopleWithVax[[i]][,year:=2000+i]
    N <- nrow(peopleWithVax[[i]])
    if(i>=10){
      peopleWithVax[[i]][sample(1:N,size=round(N*0.050)),ltbi:=1]
      peopleWithVax[[i]][,timeNoVax:=timeNoVax+1]
    } else {
      peopleWithVax[[i]][sample(1:N,size=round(N*0.025)),ltbi:=1]
      peopleWithVax[[i]][,timeWithVax:=timeWithVax+1]
    }

    Nltbi <- sum(peopleWithVax[[i]]$ltbi==1 & peopleWithVax[[i]]$ageAtTB==0)
    peopleWithVax[[i]][atbi==1,ageAtTB:=age]
    peopleWithVax[[i]][,atbi:=0] # people with ATBI go to the doctor and get cured last year
    peopleWithVax[[i]][sample(
      which(peopleWithVax[[i]]$ltbi==1 & peopleWithVax[[i]]$ageAtTB==0),
      size=round(Nltbi*0.025)),
      atbi:=1]
  }
  peopleWithVax
}) -> peopleWithVax

```

```

pomp::bake("data/peopleTB.RDS", {
  set.seed(4)
  peopleTB <- vector("list", length=17)
  peopleTB[[1]] <- newPeople[1,][-1,]
  for(i in 1:17){
    # NO VACCINATIONS
    if(i!=1) peopleTB[[i]] <- copy(peopleTB[[i-1]])
    peopleTB[[i]][,age:=age+1]
    peopleTB[[i]] <- rbind(peopleTB[[i]], newPeople)
    peopleTB[[i]][,year:=2000+i]

    peopleTB[[i]][,timeNoVax:=timeNoVax+1]

    peopleTB[[i]][,atbi:=0] # people with ATBI go to the doctor and get cured last year
    N <- sum(peopleTB[[i]]$ageAtTB==0)
    peopleTB[[i]][sample(which(peopleTB[[i]]$ageAtTB==0), size=round(N*0.05)), atbi:=1]
    peopleTB[[i]][atbi==1, ageAtTB:=age]
  }
  peopleTB
}) -> peopleTB

```

## 1. Detecting LTBI without vaccination

We will try to model the annual risk of LTBI in the dataset `peopleNoVax`, where LTBI is only measured at the end of the dataset (i.e. after 17 years).

For a person  $i$ , who has spent  $T_i$  years at risk, their probability of not having LTBI is:

$$Pr(Y_i = 0|p, T_i) = (1 - p)^{T_i}$$

And their corresponding probability of having LTBI is:

$$Pr(Y_i = 1|p, T_i) = 1 - Pr(Y_i = 0|p, T_i)$$

Leaving us with a likelihood function of:

$$L(p|Y, T) = \prod_{i=1}^n [1 - (1 - p)^{T_i}]^{y_i} \times [(1 - p)^{T_i}]^{1-y_i}$$

```

stanData <- copy(peopleNoVax[[17]])
stanData <- stanData[, .(R=.N), by=.(timeNoVax, timeWithVax, ltb)]
data = list(N=nrow(stanData),
           R=stanData$R,
           y=stanData$ltb*stanData$R,
           timeNoVax=stanData$timeNoVax)

stan_code =
data {
  int<lower=0> N;
  int R[N];
  int y[N];
  vector[N] timeNoVax;

```

```

}

parameters {
  real<lower=0,upper=0.5> thetaNoVax;
}
model {
  thetaNoVax ~ beta(0.05, 1);

  for (n in 1:N){
    y[n] ~ binomial(R[n],1-((1-thetaNoVax)^timeNoVax[n]));
  }
}

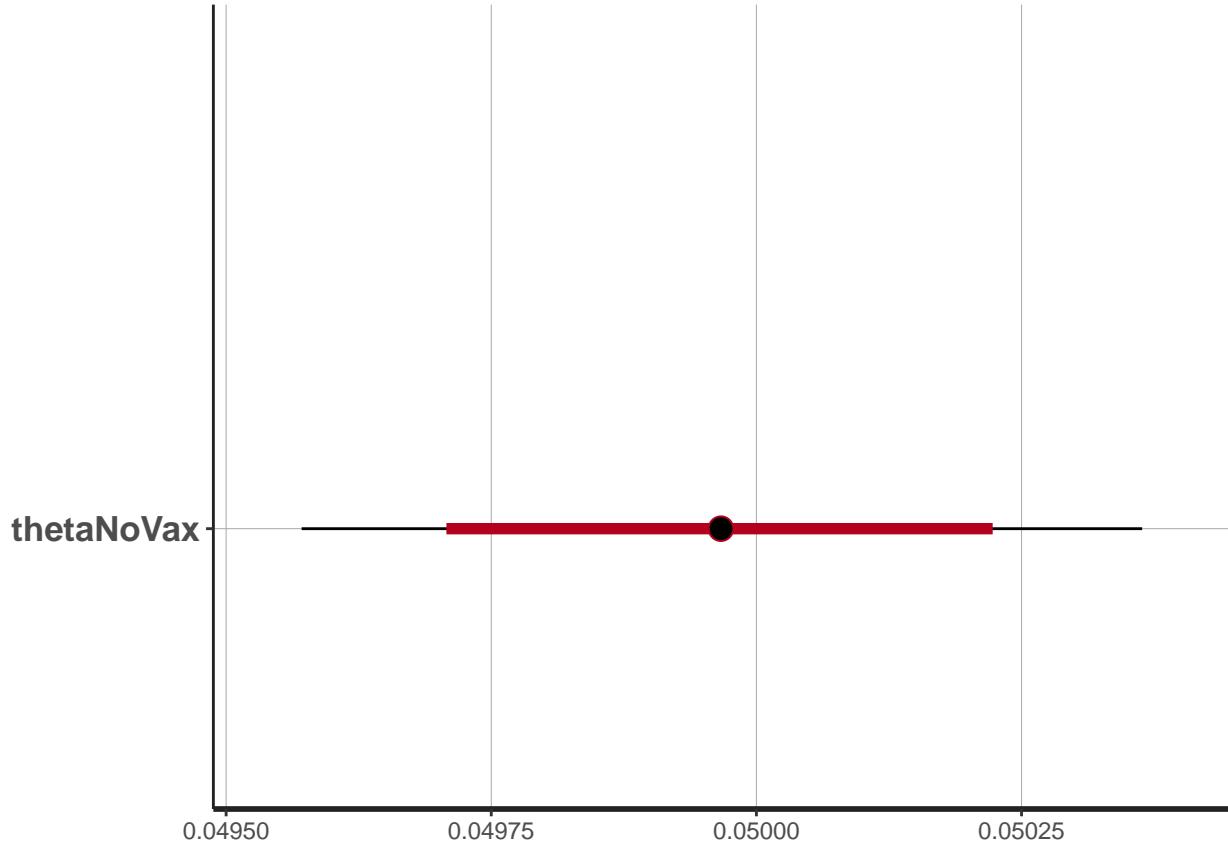
pomp::bake("results/noVaxLTBI.RDS",{
  stan(model_code=stan_code,
        model_name="noVaxLTBI",
        data=data,
        iter=10000, chains=4, init=0, seed=4)
}) -> fit
summary(fit)$summary

##               mean       se_mean         sd      2.5%
## thetaNoVax  4.996660e-02 2.324218e-06 0.0002020466 4.957123e-02
## lp__        -9.904536e+04 8.661057e-03 0.7103202531 -9.904734e+04
##             25%       50%       75%      97.5%
## thetaNoVax  4.982977e-02 4.996657e-02 5.010166e-02 5.036368e-02
## lp__        -9.904552e+04 -9.904509e+04 -9.904492e+04 -9.904487e+04
##             n_eff     Rhat
## thetaNoVax 7556.996 1.000427
## lp__        6726.151 1.000013

plot(fit)

## ci_level: 0.8 (80% intervals)
## outer_level: 0.95 (95% intervals)

```



## 2. Detecting LTBI with vaccination

We will try to model the annual risk of LTBI in the dataset `peopleWithVax`, where LTBI is only measured at the end of the dataset (i.e. after 17 years), however, the risk of LTBI has changed due to the cessation of vaccination.

For a person  $i$ , who has spent  $TW_i$  years at risk when vaccinations were occurring, and  $TN_i$  years at risk when no vaccinations were occurring, their probability of not having LTBI is:

$$Pr(Y_i = 0|p_w, p_n, TW_i, TN_i) = (1 - p_w)^{TW_i} \times (1 - p_n)^{TN_i}$$

And the probability of person  $i$  having LTBI is:

$$Pr(Y_i = 1|p_w, p_n, TW_i, TN_i) = 1 - Pr(Y_i = 0|p_w, p_n, TW_i, TN_i)$$

Leaving us with a likelihood function of:

$$L(p_w, p_n | Y, TW, TN) = \prod_{i=1}^n [1 - (1 - p_w)^{TW_i} \times (1 - p_n)^{TN_i}]^{y_i} \times [(1 - p_w)^{TW_i} \times (1 - p_n)^{TN_i}]^{1-y_i}$$

```
stanData <- copy(peopleWithVax[[17]])
stanData <- stanData[, .(R=.N), by=.(timeNoVax,timeWithVax,ltbi)]
data = list(N=nrow(stanData),
           R=stanData$R,
```

```

y=stanData$ltbi*stanData$R,
timeNoVax=stanData$timeNoVax,
timeWithVax=stanData$timeWithVax
)

stan_code =
data {
  int<lower=0> N;
  int R[N];
  int y[N];
  vector[N] timeNoVax;
  vector[N] timeWithVax;
}
parameters {
  real<lower=0,upper=0.5> thetaNoVax;
  real<lower=0,upper=0.5> thetaWithVax;
}
model {
  thetaWithVax ~ beta(0.05, 1);
  thetaNoVax ~ beta(0.05, 1);

  for (n in 1:N){
    y[n] ~ binomial(R[n], 1-((1-thetaWithVax)^timeWithVax[n])*((1-thetaNoVax)^timeNoVax[n]));
  }
}

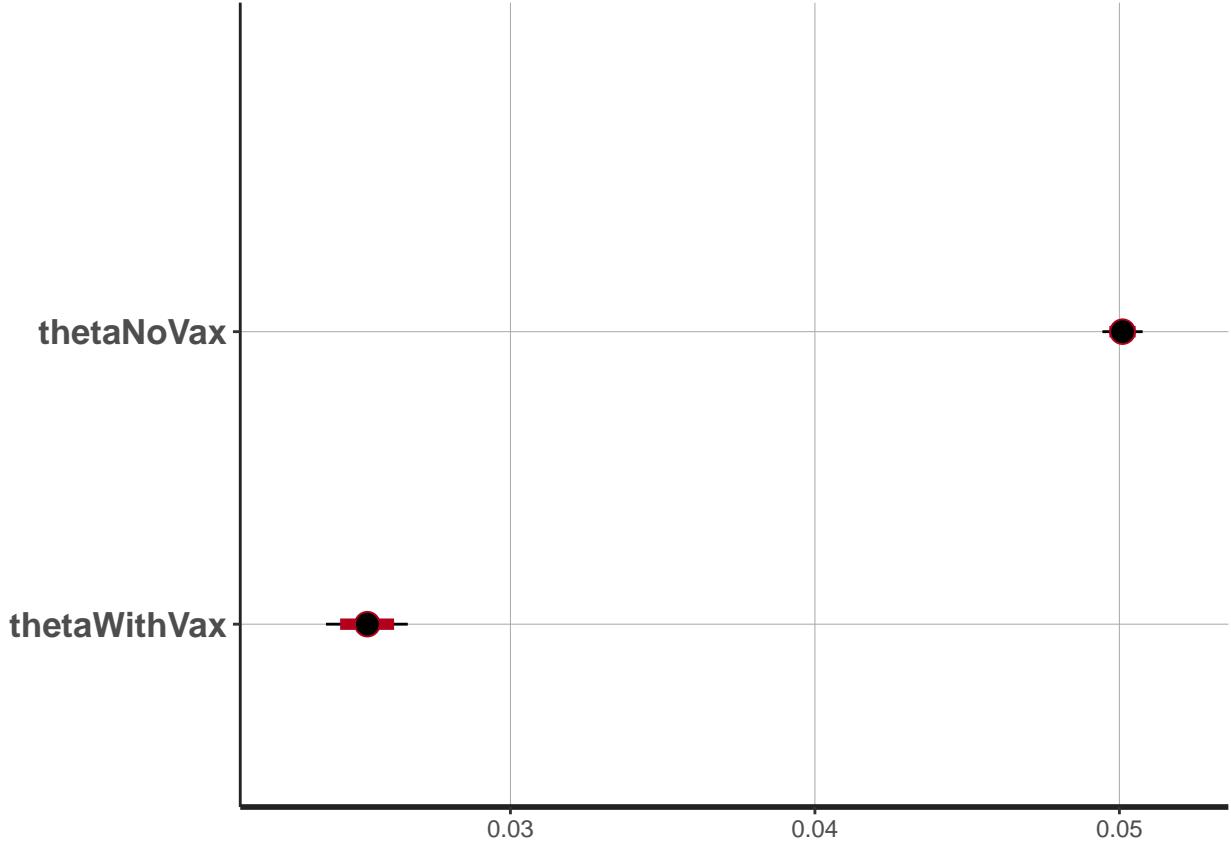
pomp:::bake("results/withVaxLTBI.RDS",{
  stan(model_code=stan_code,
        model_name="withVaxLTBI",
        data=data,
        iter=10000, chains=4, init=0, seed=4)
}) -> fit
summary(fit)$summary

##               mean      se_mean       sd      2.5%
## thetaNoVax   5.010210e-02 4.030704e-06 0.0003370411 4.944198e-02
## thetaWithVax 2.529498e-02 8.147072e-06 0.0006895297 2.394224e-02
## lp__
##          -9.863473e+04 1.153416e-02 1.0014517034 -9.863745e+04
##          25%           50%           75%         97.5%
## thetaNoVax   4.987414e-02 5.010327e-02 5.032944e-02 5.076345e-02
## thetaWithVax 2.482474e-02 2.530020e-02 2.576694e-02 2.662851e-02
## lp__
##          -9.863512e+04 -9.863443e+04 -9.863402e+04 -9.863375e+04
##          n_eff      Rhat
## thetaNoVax   6992.041 1.000197
## thetaWithVax 7163.130 1.000373
## lp__
##          7538.559 1.000411

plot(fit)

## ci_level: 0.8 (80% intervals)
## outer_level: 0.95 (95% intervals)

```



### 3. Detecting straight ATBI without vaccination

We will try to model the annual risk of ATBI in the dataset `peopleTB`, where ATBI is measured at the end of each year. This is an example of survival analysis, as the people who were not infected with ATBI will have a right-censored time at the end of the followup.

Standard survival analysis definitions apply here. The probability of an individual's surviving (i.e. not acquiring ATBI) till time  $t$  is given by the survivor function:

$$S(t) = 1 - F(t) = P(T > t)$$

The likelihood function for parameters  $\lambda$ , given data  $D$  is:

$$L(\lambda|D) = \prod_{i=1}^n f(y_i|\lambda)_i^v \times S(y_i|\lambda)^{(1-v_i)}$$

Where  $v_i = 1$  if person  $i$  is uncensored (0 if censored), and  $f$  is the density distribution for  $y_i$  (time until ATBI).

In our dataset, for person  $i$ , the probability that they will get ATBI after  $y_i$  years is distributed as a geometric distribution:

$$f(Y_i = y_i|p) = (1-p)^{(y_i-1)} \times p$$

The survival function can thus be defined as:

$$S(y_i|p) = (1 - p)^{y_i}$$

```

cases <- list()
for(i in 1:17){
  cases[[i]] <- peopleTB[[i]][atbi==1]
}
cases <- rbindlist(cases)
cases[,timeNoVax:=timeNoVax+1]
controls <- peopleTB[[17]][ageAtTB==0]

cases <- cases[,.R=.N],by=.(timeNoVax,timeWithVax,atbi)]
controls <- controls[,.R=.N],by=.(timeNoVax,timeWithVax,atbi)]

stanData <- rbind(controls,cases)
#stanData <- cases
data = list(N=nrow(stanData),
            R=stanData$R,
            y=stanData$atbi*stanData$R,
            timeNoVax=stanData$timeNoVax)

stan_code = "
data {
  int<lower=0> N;
  int timeNoVax[N];
  int R[N];
  int y[N];
}
parameters {
  real<lower=0,upper=0.5> atbiNoVax;
}
model {
  atbiNoVax ~ beta(0.05, 1);

  {
    int replicates;
    int personTime;
    real prob_no_atbi;
    real prob_atbi_in_year;

    for (n in 1:N){
      prob_no_atbi = (1-atbiNoVax)^timeNoVax[n];
      prob_atbi_in_year = (1-atbiNoVax)^(timeNoVax[n]-1)*atbiNoVax;
      replicates = R[n];

      if( y[n]==0 ){
        target += R[n]*log(prob_no_atbi);
      } else {
        target += binomial_lpmf(y[n] | R[n], prob_atbi_in_year);
      }
    }
  }
}
"

```

```

""

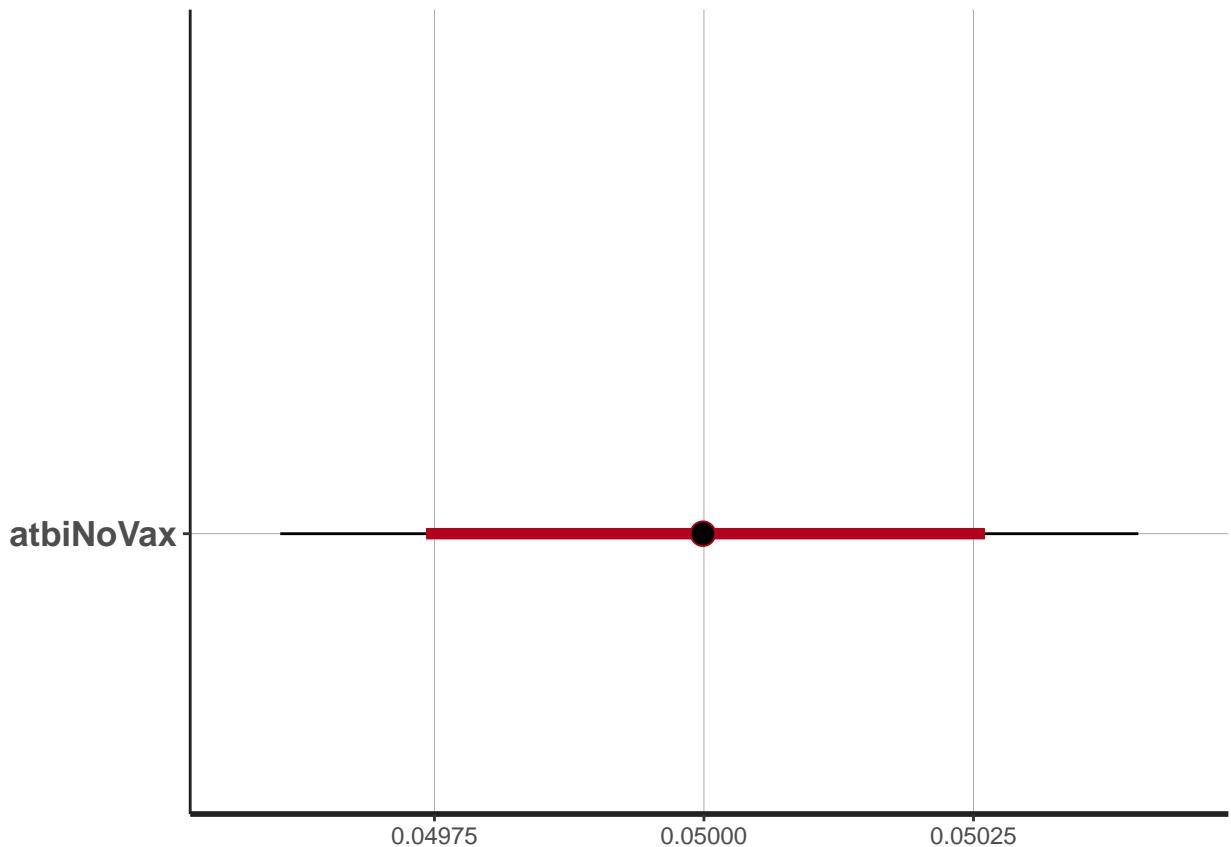
pomp::bake("results/ATBI.RDS",{
  stan(model_code=stan_code,
    model_name="ATBI",
    data=data,
    iter=10000, chains=4, init=0, seed=4)
}) -> fit
summary(fit)$summary

##               mean        se_mean         sd      2.5%
## atbiNoVax  5.000067e-02 2.451507e-06 0.0002021065 4.960702e-02
## lp__       -2.360101e+05 8.109510e-03 0.7305285894 -2.360122e+05
##             25%          50%          75%      97.5%   n_eff
## atbiNoVax  4.986375e-02 4.999907e-02 5.013366e-02 5.040290e-02 6796.637
## lp__       -2.360103e+05 -2.360098e+05 -2.360097e+05 -2.360096e+05 8114.938
##             Rhat
## atbiNoVax 1.000848
## lp__      1.000348

plot(fit)

## ci_level: 0.8 (80% intervals)
## outer_level: 0.95 (95% intervals)

```



## 4. Detecting ATBI following LTBI without vaccination

Using the dataset `peopleNoVax`, we try to estimate the annual risk of LTBI and ATBI, given that we can only observe ATBI in the year that it occurs. This is an extension of the previous models.

In our dataset, for person  $i$ , the probability that they will get LTBI after  $l_i$  years is distributed as a geometric distribution:

$$f(L_i = l_i | p_L) = (1 - p_L)^{(l_i - 1)} \times p_L$$

where  $l_i = 1, \dots, \infty$ .

In our dataset, for person  $i$ , the probability that they will get ATBI  $a_i$  years after being infected with LTBI (where the year of LTBI acquirement counts as a year) is distributed as a geometric distribution:

$$f(A_i = a_i | p_A) = (1 - p_A)^{(a_i - 1)} \times p_A$$

where  $a_i = 1, \dots, \infty$ .

We can put these together, and say that for person  $i$ , the probability that they will get ATBI after  $y_i$  years is equal to  $l_i + a_i$  and is distributed as a geometric distribution:

$$f(Y_i = y_i | p_L, p_A) = \sum_{t=1}^{y_i} f(L_i = t | p_L) \times f(A_i = y_i - t + 1 | p_A)$$

$$f(Y_i = y_i | p_L, p_A) = \sum_{t=1}^{y_i} (1 - p_L)^{(t-1)} \times p_L \times (1 - p_A)^{(y_i - t)} \times p_A$$

The survival function can thus be defined as:

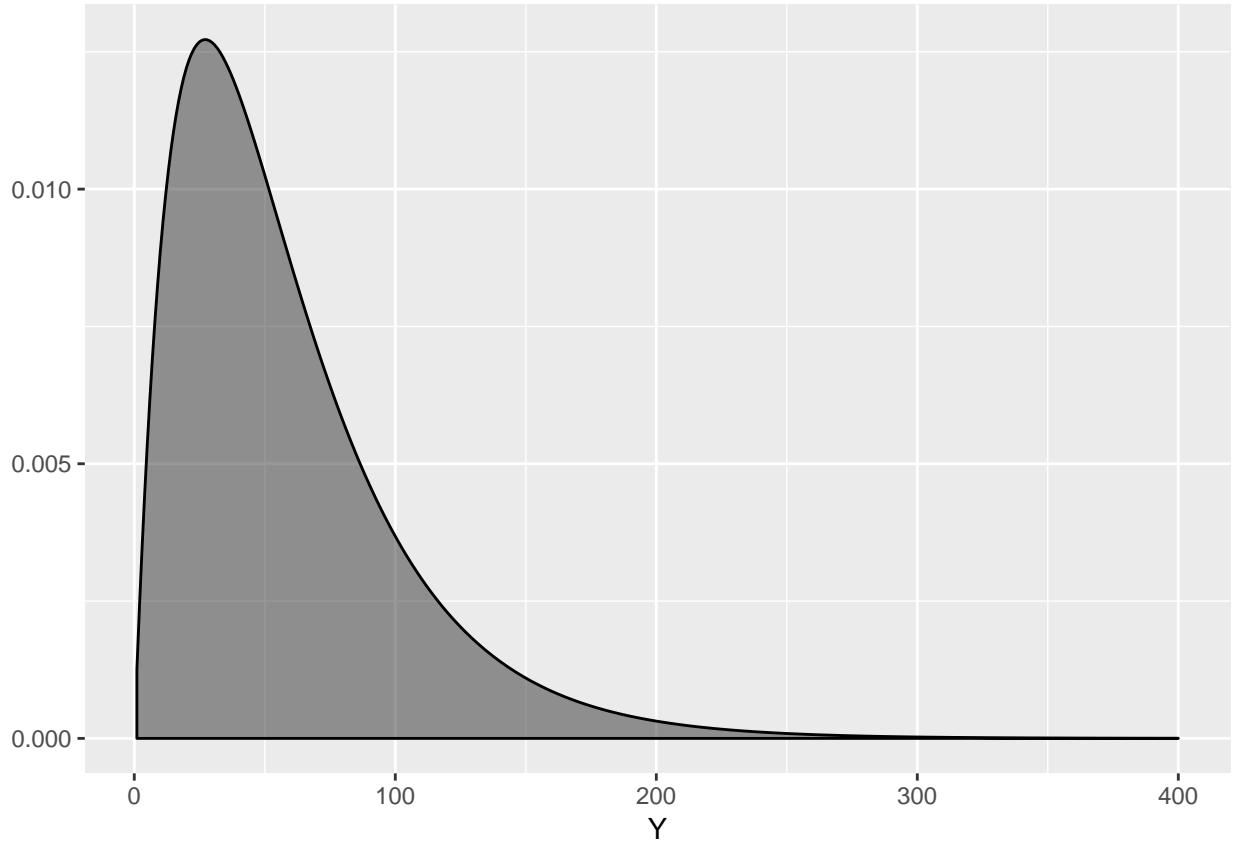
$$S(y_i | p_L, p_A) = 1 - \sum_{t=1}^{y_i} f(Y_i = t | p_L, p_A)$$

We can also see the distribution of Y using simulations:

```
pl <- 0.05
pa <- 0.025
TS <- ts <- 400

f <- c() # ATBI IN FIRST 10 YRS
f[1]=0
for(i in 1:TS){
  ts <- i
  t <- 1:(ts)
  f[i] <- sum(((1-pl)^(t-1))*pl*((1-pa)^(ts-t))*pa)
}

pd <- data.frame(Y=1:TS,p=f)
q <- ggplot(pd,aes(x=Y,ymax=p,ymin=0))
q <- q + geom_ribbon(alpha=0.5,colour="black")
q
```



We also note that it sums to 1:

```
sum(f)
## [1] 0.999922
```

We can now try to fit this using stan:

```
#file.remove("results/noVaxATBI.RDS")
cases <- list()
for(i in 1:17){
  cases[[i]] <- peopleNoVax[[i]][atbi==1]
}
cases <- rbindlist(cases)
controls <- peopleNoVax[[17]][ageAtTB==0]

cases <- cases[,.(R=.N),by=.(timeNoVax,timeWithVax,atbi)]
controls <- controls[,.(R=.N),by=.(timeNoVax,timeWithVax,atbi)]

stanData <- rbind(controls,cases)
stanData[timeNoVax==1]

##      timeNoVax timeWithVax atbi      R
## 1:          1          0    0 9987
## 2:          1          0    1  193
#stanData <- cases
data = list(N=nrow(stanData),
           R=stanData$R,
```

```

ATB1status=stanData$atbi,
timeNoVax=stanData$timeNoVax)

stan_code =
data {
  int<lower=0> N;
  int timeNoVax[N];
  int R[N];
  int ATB1status[N];
}
parameters {
  real<lower=0,upper=0.5> ltbiNoVax;
  real<lower=0,upper=0.5> atbiNoVax;
}
model {
real ps;
ltbiNoVax ~ beta(0.05, 1);
atbiNoVax ~ beta(0.05, 1);

{
  int replicates;
  int personTime;
  real prob_atbi_in_year;
  real survival;
  real p[201];

  for(q in 1:17){
    prob_atbi_in_year=0;
    for(t in 1:q){
      prob_atbi_in_year = prob_atbi_in_year + ((1-ltbiNoVax)^(t-1))*ltbiNoVax*((1-atbiNoVax)^(q-t))*atbiNoVax;
    }
    p[q]=prob_atbi_in_year;
  }

  for (n in 1:N){
    personTime = timeNoVax[n];
    replicates = R[n];

    if( ATB1status[n]==1 ){
      target += replicates*log(p[personTime]);
    } else {
      survival=0;
      for(q in 1:personTime){
        survival=survival+p[q];
      }
      target += replicates*log(1-survival);
    }
  }
}

"

```

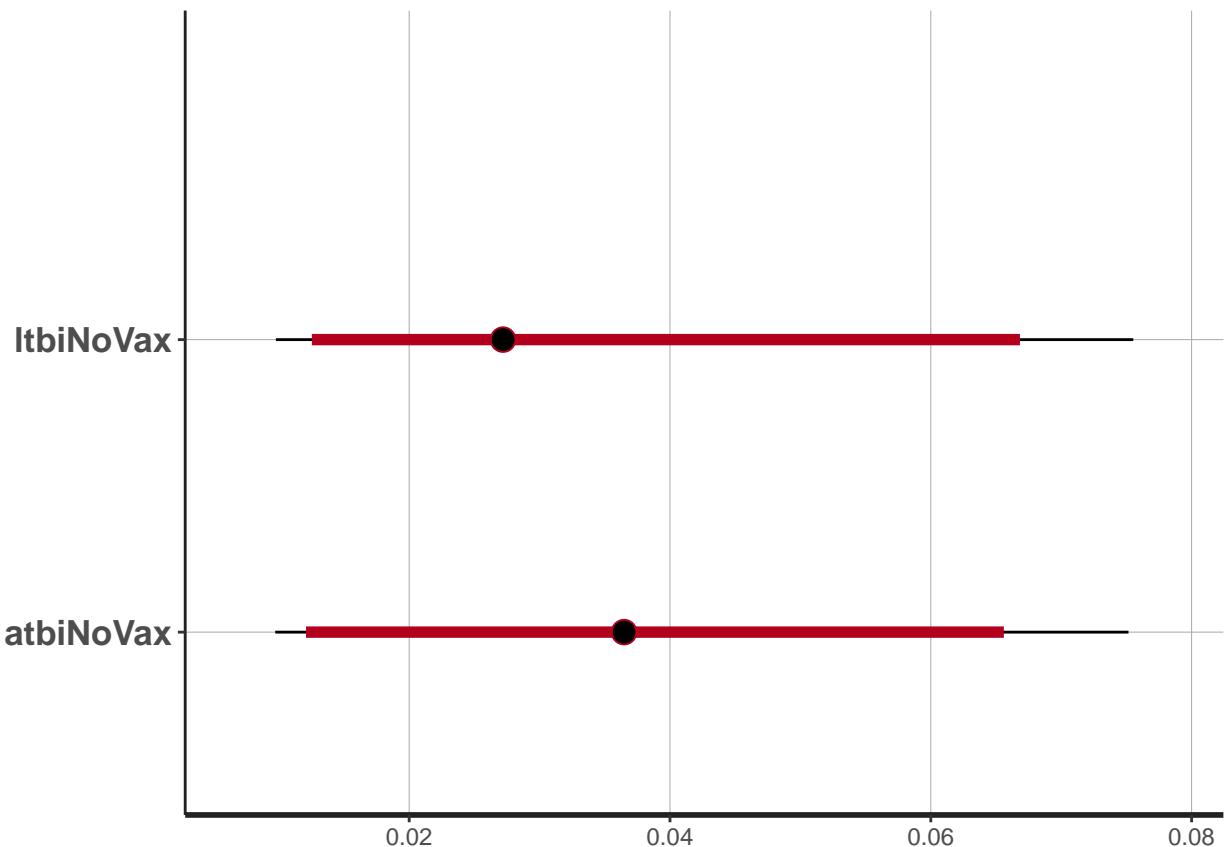
```

pomp::bake("results/noVaxATBI.RDS", {
  stan(model_code=stan_code,
        model_name="noVaxATBI",
        data=data,
        iter=2000, chains=4, init=0, seed=4)
}) -> fit
summary(fit)$summary

##               mean      se_mean       sd      2.5%      25%
## ltbiNoVax  3.558139e-02 0.001659797 0.02172076 9.780369e-03 1.565057e-02
## atbiNoVax  3.718116e-02 0.001610692 0.02141395 9.732980e-03 1.609325e-02
## lp__      -5.365275e+04 0.034156061 0.96333352 -5.365531e+04 -5.365311e+04
##             50%      75%     97.5%   n_eff   Rhat
## ltbiNoVax  2.722512e-02 5.580675e-02 7.552955e-02 171.2538 1.007132
## atbiNoVax  3.645770e-02 5.677942e-02 7.515658e-02 176.7538 1.010552
## lp__      -5.365253e+04 -5.365206e+04 -5.365175e+04 795.4590 1.002649
plot(fit)

## ci_level: 0.8 (80% intervals)
## outer_level: 0.95 (95% intervals)

```



Unfortunately, this does not meet our expected output of  $\text{ltbiNoVax}=0.05$  and  $\text{atbiNoVax}=0.025$