Supplemental Trichomonas vaginalis mathematical modeling information

1 Modeling framework

Our modeling framework accounts account for how symptoms differentially affect both recovery. Those with an asymptomatic infection will be less likely to be tested for infection. This framework mathematically describes the natural history of infection, what is represented by prevalence estimates, and population sizes. These equations and descriptions follow.

1.1 TV Natural history

The simplest possible model for this situation elaborates upon the general SIS model. We assume three possible states of infection: uninfected (U), asymptomatic infected (A), and symptomatic infected (S). All people must be in one of these three states. We consider four mechanisms: 1) infection, 2) recovery as a result of natural clearance, and 3) recovery as a result of symptomatic treatment seeking. We assume there is no background screening for TV. These mechanisms and corresponding equations are described in more detail below.

1.1.1 Infection

Uninfected people acquire infection at rate λ, also known as the force of infection. Here, we make a simplyfying assumption, that the force of infection is constant, which is consistent with our assumption of steaedy state dynamics overall described in more detail below. A proportion of newly infected people β develop symptomatic infection, and a complementary proportion $(1-\beta)$ develop asymptomatic infection.

1.1.2 Recovery as a result of natural clearance

Infected people (regardless of symptoms) can recover from infection due to natural clearance. This occurs at rate ψ, which is the inverse duration of time to natural clearance.

1.1.3 Recovery as a result of symptomatic treatment seeking

Those with symptomatic infection are likely to seek medical care at a rapid rate. Assuming a perfect test (i.e., 100% sensitivity and specificity), and assuming all who test positive are treated effectively (no treatment failure) the rate of symptomatic treatment seeking τ is equivalent to the rate of recovery as a result of this process.

1.1.4 TV natural history equations

These mechanisms are described mathematically in the following differential equations:

$$\begin{aligned}\frac{dU}{dt} &= -\lambda U + \psi A + (\psi + \tau)S \\ \frac{dA}{dt} &= \lambda U \beta - \psi A \\ \frac{dS}{dt} &= \lambda U(1-\beta) - (-\psi + \tau)S\end{aligned} \tag{1}$$

1.2 Point prevalence equation

Assuming perfect diagnostic testing and the natural history equations above, the point prevalence of infection (P) includes both the symptomatic and asymptomatically infected in the numerator:

$$P = \frac{A+S}{A+S+U} \tag{2}$$

1.3 Population size equation

Population size for a given subpopulation is summarized as:

$$N = A + S + U \tag{3}$$

1.4 Description of state space and parameters

Symbol	Description
λ	Force of infection (rate)
β	Proportion of new TV infections that are asymptomatic
ψ	TV Natural clearance rate
τ	Symptom related treatment seeking rate
N	Population size
U	Number of people who are uninfected and susceptible to infection
S	Number of people with symptomatic infection
A	Number of people with asymptomatic infection
P	Point prevalence

$$\begin{aligned}\frac{dU}{dt} &= -\lambda U + \psi A + (\psi + \tau)S \\ \frac{dA}{dt} &= \lambda U \beta - \psi A \\ \frac{dS}{dt} &= \lambda U(1-\beta) - (-\psi + \tau)S \\ N &= A + S + U \\ P &= \frac{A+S}{A+S+U}\end{aligned} \tag{4}$$

2 TV equation solving

TV's solution utilizes all five equations below to estimate annual incidence. We are able to solve for four unknowns from this system; this means we can not only solve for the three state space variables, and the force of infection.

$$\begin{aligned}
\frac{dU}{dt} &= -\lambda U + \psi A + (\psi + \tau)S \\
\frac{dA}{dt} &= \lambda U \beta - \psi A \\
\frac{dS}{dt} &= \lambda U(1-\beta) - (-\psi + \tau)S \\
N &= A + S + U \\
P &= \frac{A+S}{A+S+U}
\end{aligned} \tag{5}$$

We use symbolic algebra in Python to solve these systems of equations for their steady state values. By solving for the state variable formulations (i.e., U, A, and S) as well as force of infection formulation (λ), we are able to derive steady state solutions for the annual number of incident infections (λU).

Solving for steady state values of the natural history equations requires assuming that the change over time in each state (A, S, and U) is zero; thus, these differential equations are set to zero, implying no change, while all others remain unchanged.

Below is python code to initialize the size equations desribed in words above.

```
In [36]: from sympy.interactive import printing
         printing.init_printing(use_latex=True)
         from sympy import Eq, solve_linear_system, Matrix, Symbol
         import sympy as sp
         import math
         ##########################################
         eq1=sp.Function('eq1')
         eq2=sp.Function('eq2')
         eq3=sp.Function('eq3')
         eq4=sp.Function('eq4')
         eq5=sp.Function('eq5')

         #DEFINE STATE VARIABLES
         A,U,S,N, P=sp.symbols('A, U, S, N, P')
         #DEFINE MODEL PARAMETERS
         LAMBDA, BETA, TAU, PSI = sp.symbols('lambda, beta, tau, psi')

         eq1 = Eq((TAU+PSI)*S + (PSI)*A - LAMBDA*U)
         eq2 = Eq(-(PSI)*A + LAMBDA*BETA      *U )
         eq3 = Eq(-(TAU+PSI)*S + LAMBDA*(1 - BETA)*U )
         eq4= Eq(A+S+U, N)
         eq5= Eq((S+A)/(S+A+U), P)
         display(eq1, eq2, eq3, eq4, eq5)
```

$$A\psi + S(\psi + \tau) - U\lambda = 0$$

$$-A\psi + U\beta\lambda = 0$$

$$S(-\psi - \tau) + U\lambda(-\beta + 1) = 0$$

$$A + S + U = N$$

$$\frac{A + S}{A + S + U} = P$$

We use the sympy function solve to find the solutions as shown below. These solutions are then used to estimate incidence (further below).

```
In [37]: #solve TV equation system (Prev  available)
         sol_tv = sp.solve((       eq1, eq3,  eq4,eq5), (U,A, S,  LAMBDA))# prevalence available
         display(sol_tv)

         solution= sol_tv
         tv_inc= sp.simplify(solution[0][0]*solution[0][3]) # annual number of incident infectio
         display(tv_inc)
```

$$\left[\left(N(-P+1), \quad \frac{NP\beta(\psi+\tau)}{\beta\tau+\psi}, \quad -\frac{NP\psi(\beta-1)}{\beta\tau+\psi}, \quad -\frac{P\psi(\psi+\tau)}{(P-1)(\beta\tau+\psi)}\right)\right]$$

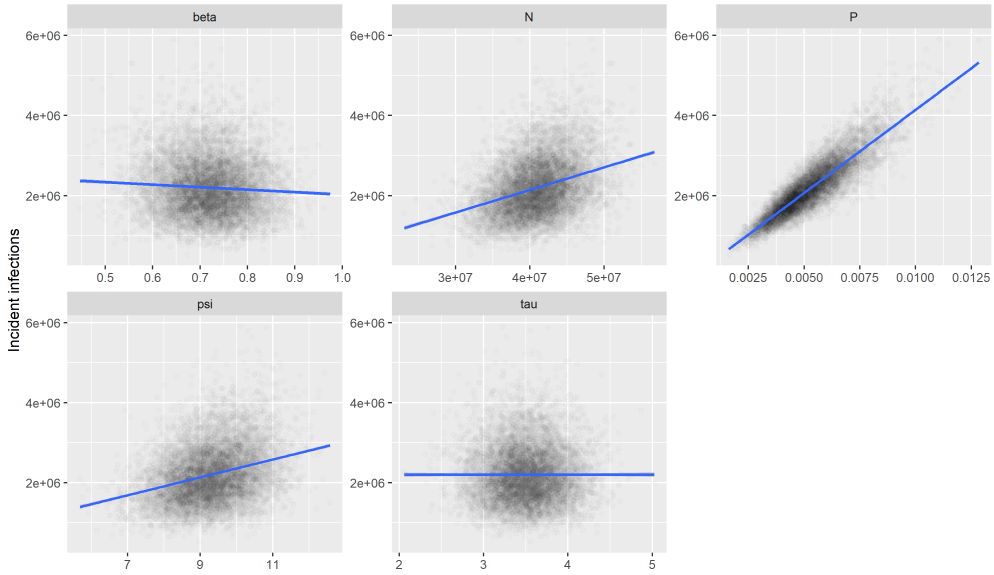
$$\frac{NP\psi(\psi+\tau)}{\beta\tau+\psi}$$

Thus the annual number of incident TV infections is given by : $\frac{NP\psi(\psi+\tau)}{\beta\tau+\psi}$

3 Correlation analysis

We describe how each input parameter affects incidence and prevalence by using Monte Carlo simulation to randomally generate 10,000 parameter sets. Except for prevalence, all parameter values in each set are determined by randomally sampling from a normal distribution with mean equal to the mean value specific to a given subpopulation, and standard deviation equal to 10% of the mean. This standard deviation value is not meant to mimic any real situation, but rather to have a similar magnitude of dispersion for each parameter. Because prevalence is so small, sampling from a normal distribution would result in negative prevlaence values; as a result, rather than sampling from a normal, we instead sample from a beta distribution with shape parameters informed by NHANES based on mean and unweighted sample size. We generate a scatterplot of each input parameter value against TV incidence to visualize this effect.

3.1 Scaterplot of model estimated TV incidence and each input parameter



Three parameters have positive effects on TV incidence: population size N, natural clearance ψ, and most strongly point prevalence P. The asymptomatic proportion has a small negative effect on TV incidence. The symptomatic treatment seeking rate τ has an extremely small positive effect on TV incidence.