Mathematical Modeling Supplement

Estimates of the Prevalence and Incidence of Chlamydia and Gonorrhea Among US Men and Women, 2018

1 Modeling framework

Our modeling framework accounts for how Chlamydia trachomatis (CT) and Neisseria gonorrhea (GC) symptoms differentially affect both recovery and being reported as a diagnosed case . Those with an asymptomatic infection will be less likely to be tested for infection, and thus less likely to be a reported case. This framework mathematically describes the natural history of infection, how cases are reported, what is represented by prevalence estimates, and population sizes. These equations and descriptions follow.

1.1 CT/GC 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, 3) recovery as a result of background screening, and 4) recovery as a result of symptomatic treatment seeking. These 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 simplifying assumption, that the force of infection is constant, which is consistent with our assumption of steady 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 Recovery as a result of background screening

Recovery from background screening can occur in both the symptomatically and asymptomatically infected. Background screening is meant to mimic the screening in the absence of symptoms that all people within a given subpopulation experience; this is akin to an annual checkup medical visit. Specific subpopulations will have screening rates specific to them. Within a subpopulation, the screening rate is homogeneous, and averaged across all who comprise this subpopulation. 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 screening the asymptomatically infected people (σ) is the rate of recovery in this group. Those who are symptomatically infected may also experience background screening, in addition to their symptom-related treatment seeking, though it is likely that the rate of background screening is much slower than the other rate. As a result, the symptomatically infected people also recover at the rate at which they are screened (σ).

1.1.5 CT/GC natural history equations

These mechanisms are described mathematically in the following differential equations:

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

1.2 Point prevalence equation

Assuming perfect diagnostic testing and the natural history equations above, the point prevalence of infection (P) includes both the symptomatically 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:

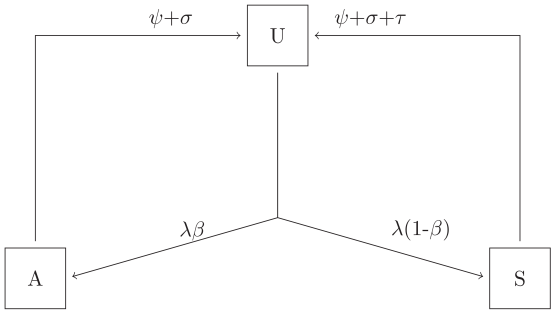
$$N = A + S + U \tag{3}$$

1.4 Case reporting equation

Cases are reported as they are diagnosed, either from symptomatic treatment seeking or background screening. Everyone receives background screening at the same rate, in contrast with symptomatic treatment seeking, which symptomatically infected receive. The below equation assumes only a proportion of cases (ρ) are reported; it also assumes perfect testing. The number of cases reported over a year (K) may be summarized as:

$$K = \rho(\sigma(A + S) + \tau S) \tag{4}$$

1.5 Description of state space and parameters



Symbol	Description
λ	Force of infection (rate)
β	Proportion of new chlamydial infections that are asymptomatic
ψ	Natural chlamydial clearance rate
σ	Background screening rate
τ	Symptom related treatment seeking rate
ρ	Reporting fraction
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
K	Case report number (over a specified period)
P	Point prevalence

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

2 CT equation solving

Chlamydia's solution utilizes all six equations below to estimate annual incidence. We are able to solve for five unknowns from this system; this means we can not only solve for the three state space variables, and the force of infection, but also the proportion of new infections that are asymptomatic (β).

$$
\begin{aligned}
\frac{dU}{dt} &= -\lambda U + (\psi+\sigma)A + (\psi+\sigma+\tau)S \\
\frac{dA}{dt} &= \lambda U\beta - (\psi+\sigma)A \\
\frac{dS}{dt} &= \lambda U(1-\beta) - (-\psi+\sigma+\tau)S \\
N &= A+S+U \\
P &= \frac{A+S}{A+S+U} \\
K &= \rho(\sigma(A+S)+\tau S)
\end{aligned} \tag{6}
$$

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 described above.

```
In [21]: 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')
         eq6=sp.Function('eq6')

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

         eq1 = Eq((TAU+SIGMA+PSI)*S + (PSI+SIGMA)*A - LAMBDA*U)
         eq2 = Eq(-(PSI+SIGMA)*A + LAMBDA*BETA      *U )
         eq3 = Eq(-(TAU+SIGMA+PSI)*S + LAMBDA*(1 - BETA)*U )
         eq4= Eq(A+S+U, N)
         eq5= Eq(RHO*((SIGMA+TAU)*S +SIGMA*A)          , K)
         eq6= Eq((S+A)/(S+A+U), P)
         display(eq1, eq2, eq3, eq4, eq5, eq6)
```

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

$$A\left(-\psi-\sigma\right)+U\beta\lambda=0$$

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

$$A+S+U=N$$

$$\rho\left(A\sigma+S\left(\sigma+\tau\right)\right)=K$$

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

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

```
In [22]: #solve CT equation system (Prev and CaseReports available)
         sol_ct = sp.solve((       eq1, eq3,  eq4,eq5, eq6), (U,A, S,  LAMBDA, BETA))
         display(sol_ct)

         solution= sol_ct
         ct_inc= sp.simplify(solution[0][0]*solution[0][3]) # annual incident infections
         display(ct_inc)
         #display(sp.latex(ct_inc))
```

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

$$\frac{K}{\rho}+NP\psi$$

Thus the annual number of incident CT infections is given by : $\frac{K}{\rho}+NP\psi$.

3 GC equation solving

Gonorrhea's solution utilizes the five equations below to estimate annual incidence and point prevalence. Note, that here prevalence is an output rather than used as an additional input equation.

For GC, NHANES prevalence data are no longer available but case reports are. As a result, we can only solve for 4 unknowns (rather than 5 for CT): the three state variables plus the force of infection (λ).

$$
\begin{aligned}
\frac{dU}{dt} &= -\lambda U + (\psi + \sigma)A + (\psi + \sigma + \tau)S \\
\frac{dA}{dt} &= \lambda U \beta - (\psi + \sigma)A \\
\frac{dS}{dt} &= \lambda U(1 - \beta) - (-\psi + \sigma + \tau)S \\
N &= A + S + U \\
K &= \rho(\sigma(A + S) + \tau S)
\end{aligned} \tag{7}
$$

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) and the point prevalence of infection $\frac{A+S}{A+S+U}$.

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 described above.

```
In [23]: 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,  K=sp.symbols('A, U, S, N,  K')
         #DEFINE MODEL PARAMETERS
         LAMBDA, BETA, SIGMA, TAU, PSI,  RHO = sp.symbols('lambda, beta, sigma, tau, psi,  rho')

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

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

$$A\left(-\psi - \sigma\right) + U\beta\lambda = 0$$

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

$$A+S+U=N$$

$$\rho\left(A\sigma+S\left(\sigma+\tau\right)\right)=K$$

We use the sympy function solve to find the solutions as shown below. These solutions are then used to estimate incidence and prevalence (further below). We are able to formulate prevalence from the state variables where $P=\frac{A+S}{A+S+U}$.

```
In [24]: #solve GC equation system (no Prev / ONLY case reports available)
         sol_gc = sp.solve((     eq1, eq3, eq4, eq5), (U,A,S,  LAMBDA))
         display(sol_gc)

         solution= sol_gc
         gc_inc  = sp.simplify(solution[0][0]*solution[0][3]) # annual incident infections
         gc_prev = sp.factor((solution[0][1]+solution[0][2])/ (solution[0][0]+solution[0][1]+sol

         display(gc_inc)
         display(gc_prev)
```

$$\left[\left(\frac{K\beta\tau+K\psi+K\sigma+N\beta\psi\rho\tau-N\psi\rho\sigma-N\psi\rho\tau-N\rho\sigma^2-N\rho\sigma\tau}{\rho\left(\beta\psi\tau-\psi\sigma-\psi\tau-\sigma^2-\sigma\tau\right)},\right.\right.$$

$$-\frac{K\beta\left(\psi+\sigma+\tau\right)}{\rho\left(\beta\psi\tau-\psi\sigma-\psi\tau-\sigma^2-\sigma\tau\right)},$$

$$\frac{K\left(\beta-1\right)\left(\psi+\sigma\right)}{\rho\left(\beta\psi\tau-\psi\sigma-\psi\tau-\sigma^2-\sigma\tau\right)},$$

$$\left.\left.-\frac{K\left(\psi+\sigma\right)\left(\psi+\sigma+\tau\right)}{K\beta\tau+K\psi+K\sigma+N\beta\psi\rho\tau-N\psi\rho\sigma-N\psi\rho\tau-N\rho\sigma^2-N\rho\sigma\tau}\right)\right]$$

$$\frac{K\left(\psi+\sigma\right)\left(\psi+\sigma+\tau\right)}{\rho\left(-\beta\psi\tau+\psi\sigma+\psi\tau+\sigma^2+\sigma\tau\right)}$$

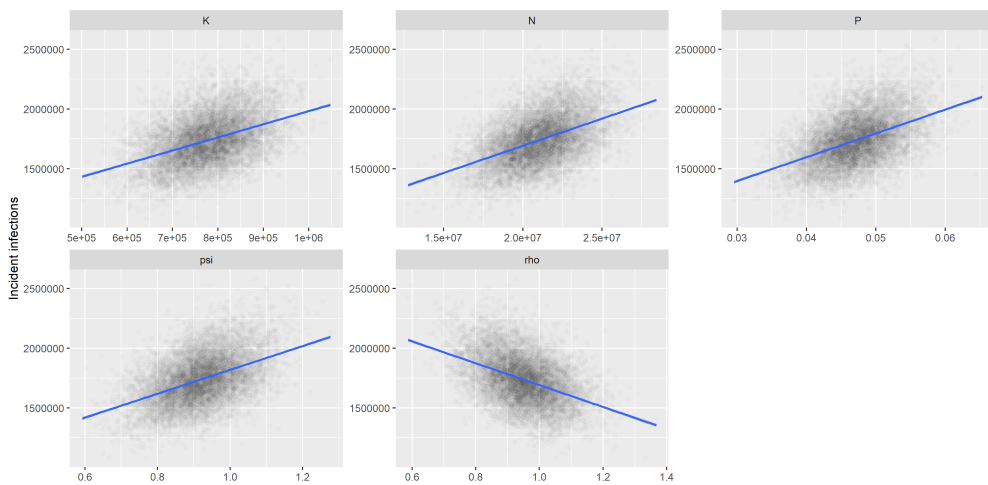
$$-\frac{K\left(\beta\tau+\psi+\sigma\right)}{N\rho\left(\beta\psi\tau-\psi\sigma-\psi\tau-\sigma^2-\sigma\tau\right)}$$

Thus, the annual number of incident gonococcal infections is formulated as: $\frac{K(\psi+\sigma)(\psi+\sigma+\tau)}{\rho(-\beta\psi\tau+\psi\sigma+\psi\tau+\sigma^2+\sigma\tau)}$, and the point prevalence of gonococcal infection is formulated as: $-\frac{K(\beta\tau+\psi+\sigma)}{N\rho(\beta\psi\tau-\psi\sigma-\psi\tau-\sigma^2-\sigma\tau)}$.

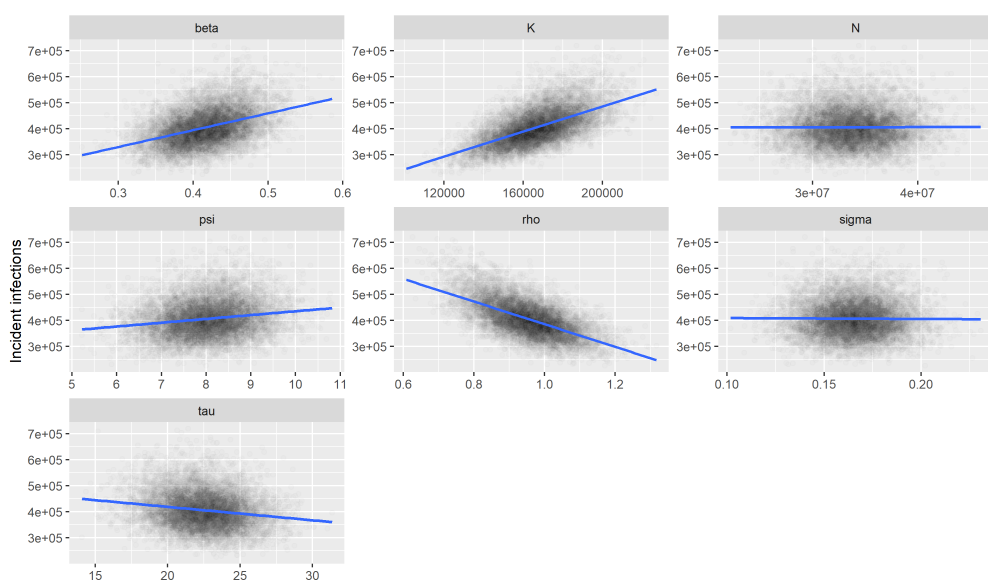
4 Correlation analysis

We describe how each input parameter affects incidence and prevalence by using Monte Carlo simulation to randomly generate 10,000 parameter sets. Parameter values in each set are determined by randomly 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. We generate a scatterplot of each input parameter value against incidence (and prevalence for gonorrhea) to visualize this effect. We also summarize the effects of these analyses in tabular form below.

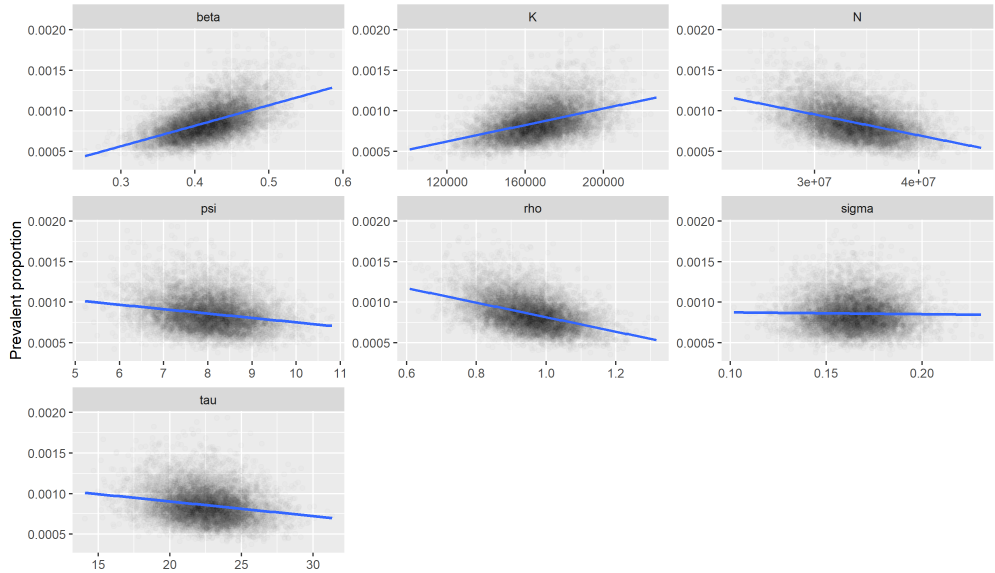
4.1 Scaterplot of model estimated CT incidence and each input parameter



4.2 Scatterplot of model estimated GC incidence and each input parameter



4.3 Scatterplot of model estimated GC prevalence and each input parameter



4.4 Table summarizing CT/GC incidence and prevalence parameter correlations

Parameter	CT-incidence	GC-incidence	GC-prevalence
N	+	0	-
K	+	+	+
ψ	+	+	-
ρ	-	-	-
P	+	NA	NA
σ	NA	-	-
τ	NA	-	-
β	NA	+	+

Two parameters have similar effects on CT incidence, GC incidence, and GC prevalence. Increases in case reports leads to an increased estimates in all three measures while increased case reporting fraction (i.e., a greater percentage of diagnosed cases being reported) leads to decreases in all three measures.

Increased natural clearance leads to higher estimated incidence (for both CT and GC) but lower estimated GC prevalence. Faster clearance, means that more incident cases go undiagnosed or observed in NHANES, thus requiring higher incidence to sustain a given level of case reports or NHANES prevalence. Increased asymptomatic infection leads to higher estimated GC incidence and prevalence. The more asymptomatic infection there is, the more infections go undiagnosed, thus leading to a higher level of incidence and prevalence needed in order to sustain a given level of case reports. The effect of background screening (σ) is quite small, but negative on both GC prevalence and incidence.