Syphilis transmission model readme

Last updated 28 September 2017 by ART

Folders

Cluster: code for running the model on Odyssey (Harvard cluster)

Data: source data files used to generate R inputs used in the model

Documentation: documents related to the project

Model\_code: R code

Model\_outputs: R calibration outputs are stored here

R\_inputs: text files containing calibration data and model priors that are read into the model

Cluster

syph\_mcmc.R: R script to run the MCMC calibration and save outputs

syph.mcmc.slurm: script to run the model on the cluster

Data

LA\_syph\_data.xlsx: file provided by Louisiana containing data on syphilis trends

MA\_HIV\_estimates: file provided by Massachusetts containing data on HIV incidence and prevalence (2013-2015)

MA\_syph\_data.xlsx: file provided by Massachusetts containing data on syphilis trends

NSFG\_estimates.xlsx: results of analysis of 2011-2013 NSFG for 20-44y age group (age mixing and ever had sex)

parameters.xlsx: spreadsheet containing prior estimates for model parameters. First tab contains calculations, second tab is saved as .txt file to import into R

project\_info.docx: info about project, including contact details for Louisiana and Massachusetts Departments of Public Health

state\_early\_v\_late\_syph.xlsx: state-levels estimates of early and late latent syphilis (not by age or sex)

Documentation

figures.pptx: model figures

immunity\_params.docx: description of issues around partial immunity after treatment and options for calibration

Model\_workflow.docx: flowchart showing how the various functions are integrated into the model calibration in R

state\_data.pdf: plots of data from Louisiana and Massachusetts

syph\_all-hands\_2017.pptx: presentation from PPML All Hands meeting in May 2017

syph\_current\_status.docx: current status of the project

syph\_MS\_draft.docx: draft of manuscript (mainly methods at this point)

syph\_NEEMA\_Oct\_2017.pptx: draft presentation for NEEMA in October 2017

syph\_param\_tables.docx: tables of parameters for manuscript

syph\_project\_overview.docx: project overview sent to contacts in LA and MA

syph\_tech\_app.pdf: technical appendix

syph\_tech\_appendix: folder containing latex files for technical appendix

syphilis\_refs.Data: references for endnote

syphilis\_refs.enl: endnote library

Model\_code

aging.fun.R: function to calculate aging parameters based on specified age categories

bezier.funs.R: function to calculate annual rates/probabilities for time-varying parameters

initial.files.R: loads the required input text files and performs required data prep

load.start.conditions.R: initializes the model, including start conditions

load.start.conditions.interv.R: initializes the model, including start conditions for a model with an intervention period after the calibration period (currently set to 10 years)

mcmc.fit.function.R: sample from posterior distributions to compare model outputs to calibration targets

mcmc.fit.interv.function.R: sample from posterior distributions to compare model outputs with different interventions

mcmc.funs.R: functions used to run MCMC

mixing.fun.R: function used to calculate and balance contact matrix

model.likelihood.fun.R function that updates model parameters, runs the model, and calculates likelihood for a given parameter set

model.out.fun.R: function that runs the model for a given parameter set and prints outputs used by likelihood function

model.pred.fun.R: contains function to generate model outputs required for calibration

model.pred.interv.fun.R: contains function to generate model outputs under different interventions

model.priors.fun.R: function to calculate prior likelihood

myMCMC.R: modified mcmcMH function from fitR package that ensures covariance matrix is positive definite

population.size.fun.R: function to calculate population sizes for different model strata

syph\_input\_data\_plot.R: plots the input data from Louisiana and Massachusetts

syph\_model\_interv\_code.R: file used to run the model under different intervention scenarios (still a work in progress)

syph\_model\_main\_code.R: this is the main file used to run the model and save and visualize outputs. It calls the various functions described here.

syph\_model\_outputs.R: visualize model outputs after running calibration and save to pdf

syph\_trans\_model\_interv.cpp: syphilis transmission model for implementing interventions; allows for different screening rates in those with a prior treated infection

syph\_trans\_model.cpp: syphilis transmission model code

R\_inputs

\*\*see file\_descriptions.docx for more details about input files\*\*

age\_nsfg.txt: age assortativity; calibration target

case\_subpop\_rr\_LA.txt: relative risk of being a reported case by sex, age, and race/ethnicity in Louisiana (LA); calibration target

case\_subpop\_rr\_MA.txt: relative risk of being a reported case by sex, age, and race/ethnicity in Massachusetts (MA); calibration target

diag\_age\_sex\_rate\_LA.txt: reported early syphilis cases by age and sex for LA; calibration target

diag\_age\_sex\_rate\_MA.txt: reported early syphilis by age and sex for MA; calibration target

file\_descriptions.docx: summary of R inputs with details and sources

msm\_hiv\_proportions\_LA.txt: proportion of MSM early syphilis cases with HIV co-infection, LA; calibration target

msm\_hiv\_proportions\_MA.txt: proportion of MSM early syphilis cases with HIV co-infection, MA; calibration target

p\_early\_LA.txt: proportion of all LA syphilis cases that are early (primary, secondary, or early latent); calibration target

p\_early\_MA.txt: proportion of all MA syphilis cases that are early (primary, secondary, or early latent); calibration target

priors.txt: probability distribution parameters for model priors

stage\_diag\_LA.txt: proportion of LA reported early syphilis cases diagnosed in the secondary or early latent stage, by age and sex; calibration target

stage\_diag\_MA.txt: proportion of MA reported early syphilis cases diagnosed in the secondary or early latent stage, by age and sex; calibration target

subpop\_diag\_LA.txt: reported LA early syphilis cases by age, sex, and subpopulation; calibration target

subpop\_diag\_MA.txt: reported MA early syphilis cases by age, sex, and subpopulation; calibration target

theta\_LA.rda: starting values for parameters for Louisiana model

theta\_MA.rda: starting values for parameters for Massachusetts model

Running the model for calibration

The calibration requires installation of the fitR package. This can be done in R using the devtools package:

install.packages('devtools')

library('devtools')

install\_github('sbfnk/fitR')

1. To run the model without changing anything, open the file: ‘syph\_model\_main\_code.R’.
2. Update the working directory and ensure that all of the source functions are stored in a folder named ‘Model code’, and input files are stored in a folder named ‘R\_inputs’.
3. Set the city you are interested in, either ‘LA’ for Louisiana or ‘MA’ for Massachusetts.
4. You can use the ‘prediction.epi’ function (saved in ‘model.pred.function.R’) to run a single instance of the model (without calculating likelihoods).
5. Run the code to load start conditions and use the ‘my\_mcmc’ function to run the MCMC algorithm. Note that this code is provided here mainly for troubleshooting. For longer runs, you can use the Harvard cluster (Odyssey). There are scripts in the ‘Cluster’ folder that can be used for running on the cluster. Multiple longer chains can be merged to get enough samples for calibration.

Visualizing calibration results

1. Ensure that MCMC outputs are stored in the ‘Model\_outputs’ folder in the directory.
2. Follow the instructions for loading outputs from MCMC and visualizing calibration targets in ‘syph\_model\_main\_code.R’.
3. Assign the name of the calibration trace file to ‘file.out’.
4. Update the desired burn and thin values and run the code to save the outputs in a pdf file (saved to Model\_outputs’ folder).

Predictions under different screening scenarios

\*\*\* This is still under development \*\*\*