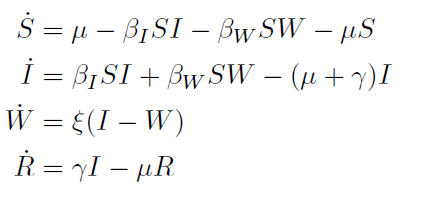
Cholera and many waterborne diseases exhibit multiple pathways of infection, which can be modeled (for example) as direct and indirect transmission. A major public health issue for waterborne diseases involves understanding the modes of transmission in order to improve control and prevention strategies (see e.g. Hartley 2006). An important epidemiological question is therefore: given data for an outbreak, can we determine the role and relative importance of direct (human-mediated) vs. environmental/waterborne routes of transmission?

To examine this question, we will use the SIWR model developed by Tien and Earn (2010), shown in Figure 1. We will combine this model with modified data from a recent cholera outbreak.

The scaled SIWR model is given by the following equations:



Where

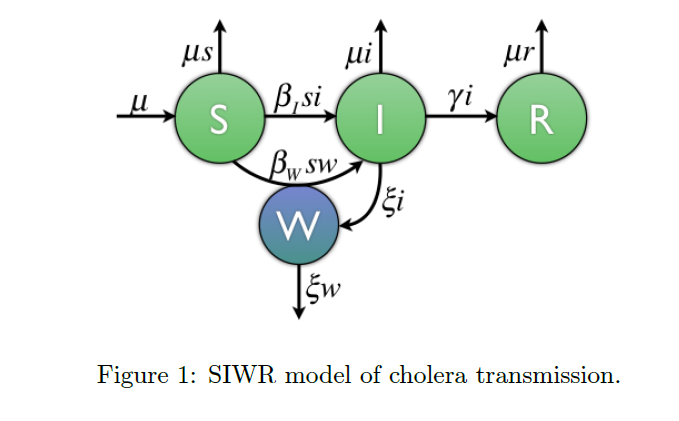
* S, I, and R are the fractions of the population who are susceptible, infectious, and recovered
* W is a scaled version of the concentration of bacteria in the water
* and are the transmission parameters for direct (human-human) and indirect (environmental) cholera transmission
* is the pathogen decay rate in the water
* is the recovery rate
* is the birth/death parameter for the population

Since we are considering a short-term outbreak (less than one year), it is reasonable to assume that the effects of births and deaths are negligible, so we set = 0. In addition, the recovery time for cholera is reasonably well known, so we can = 0.25 based on previous work (Tuite 2011) (i.e. we don't need to estimate this).

*Data & Measurement Equation*: Data from a recent outbreak in Angola is given on the course web site. To connect the model with the data, we will use the following measurement equation: y = I/k, where 1/k is a combination of the reporting rate, the asymptomatic rate, and the total population size.

*Estimation*: For fitting, we'll use ordinary least squares (OLS) for now, i.e. 

Try another cost (objective) function also. They can give quite different answer for the parameter estimates, so it can be interesting to see.



1. **SIWR Model Simulation**. Write code to simulate the SIWR model and plot both the data set provided and the measurement equation *y = I/k* (i.e. plot both the data and y in one graph vs. time). Use the following parameter values: = = 0:75, = 0:01, *k = 1/89193*. For initial conditions, we can determine them from the data by noticing that if *y = I/k*, then , i.e. we can approximate *I(0)* by the first data point times k. Since the data begins early in the epidemic, we can take *R(0) = 0*, and let *S(0) = 1 - I(0)*, since the sum of the fractions of the population in *S, I*, and *R* must sum to 1. Lastly, let *W(0) = 0*.
2. **Parameter Estimation**. Write code to estimate the model parameters ; ; , and *k* using the data set provided. The parameters and will remain fixed (not to be fitted). Use the parameter values from 1) as starting values and the initial conditions. Try different optimization routines (depending on what are available in preferred programming language: MatLab, R, Python, etc).

Plot the cholera data together with your model using the parameter estimates you found. Be sure to plot the data as circles (`o' in the plot function) and your model simulation as a line so that you can compare your model with the data easily.

References:

Hartley, D., Morris, J., Smith, D., 2006. Hyperinfectivity: a critical element in the ability of V. cholerae to cause epidemics? PLoS Med. 3, 63–69

Tien JH, Earn DJ. Multiple transmission pathways and disease dynamics in a waterborne pathogen model. Bull Math Biol 2010, 72:1506-33