If there are no data describing the distributions we need, we can use literature-reported average values and confidence intervals to extrapolate such distributions.

We need:

* Distribution of transmission rate by day (can use multiple distributions to model differences by day, or the averaged distribution across all days of infection).
  + If we use an averaged distribution across all days of infection, we also need to know transmission rate trends per day: by how much does transmission increase for the first days of infection period until maximum value; by how much does it decrease during the last days of infection period until 0?
* Distribution of death probability by day (can use multiple distributions to model differences by day, or the averaged distribution across all days of infection).
  + If we use an averaged distribution across all days of infection, we also need to know death probability trends per day: by how much does death probability increase for the first days of infection period until maximum value; by how much does it decrease during the last days of infection period until 0?
* Distribution of duration of infection for the population
* Distribution of reinfection probability for the population
* Distribution of vaccine efficacy for the population
* Prevalence of disease that leads to outbreak (may consider sampling from some distribution of prevalence at the district level to assign different values by district)
  + Do we want to use NFHS data to model those most probable to have infection at start? If so, need more data to model this.
* Are we modeling individuals ages 1-5? We should update the contact matrix accordingly, and find the right data for it
  + We can consider incorporating multiple other contact matrices using NFHS data for attributes besides age, and using the contact matrices to create a more comprehensive interaction matrix for the population
* Vaccinations:
  + What proportion of population gets vaccinated when vaccinations are available? (Maybe we don't need this data, we might instead show how effective vaccinations are by frequency of administration)
  + Can total vaccination by day be modeled by exponential decay or some other function? (we currently estimate using weak exponential decay)
  + Do we want to model booster vaccinations? If so, how much does booster add to total vaccine efficacy?
* Antibiotics: (I am not entirely confident that below encompasses all data we need as we have not tried implementing this yet)
  + What proportion of infected are treated with antibiotics?
  + What is the probability of administration of each variant of antibiotic?
  + What is the probability that antibiotic will be effective (for each variant)?
  + How does effective antibiotic treatment affect transmission rate and death probability by day of administration?
  + If antibiotic fails, what is the probability that the infected individual receives a different antibiotic?
  + By how much does effective antibiotic shorten infection period (preferably distribution of values by antibiotic variant)?