Simulator Project – Falsification Detection

# Description and Goals

This project aims to simulate pharmaceutical supply chains in a low- or middle-income country (LMIC) that are subject to insertions of substandard and/or falsified products (hereby referred to as SF products). The supply chain of a single pharmaceutical product (e.g., amoxicillin) is characterized through a network of nodes and arcs, where falsifier product originates from a single falsifier source node. Entities such as importers or outlets may possibly obtain product from these falsifier nodes under different circumstances or configurations. Drug regulatory agencies (DRAs) have the ability to sample and test products from different entities according to testing policies and procurement budgets.

The general aim of this project is summarized as follows: “Under what types of network configurations and falsification contexts can different testing policies most ably detect the underlying structures of SF distribution?” Falsification structures need to be ascertained using the following pieces of information:

* Sample collection results
  + Importer name
  + Batch in-country arrival date
  + Pass/Fail diagnostic reading
* Reports of stockouts at different entities – either when arriving to collect a sample, or by other means

Summarily, this project will explore some of the following questions:

* Can the location of “bad actors” be detected in the network, and with what degree of confidence? Under what circumstances?
* How do the testing data appear under temporal network changes, i.e., the occurrence of unfit batches of a product that is otherwise fit for consumption?
* How can sampling data and supply chain structure be integrated into decision-making policies?
* What happens as the network expands/contracts in complexity?
* Which testing policies are most/least effective (at all)?

# Simulation Model Schematic

(See “Falsification\_Simulator.py” for the Python code running this model.) The simulation model utilizes a graph-based structure to represent the pharmaceutical supply chain. The primary objects are “entity nodes,” signifying manufacturers, importers, and pharmaceutical outlets, and “drug packets,” signifying some number of single dosages of a particular pharmaceutical product (e.g., a round of amoxicillin). Entity nodes are classified as one of the following:

1. **Root Nodes:** These nodes signify the global market supply of the product and can only be accessed by importer/intermediate nodes. For the purposes of the simulation, only two root nodes are needed: one to generate acceptable product, and the other to generate SF product. As we are only interested in the national supply system of a particular country and not the global market as a whole, we assume that root nodes always have enough supply to meet orders placed with them.
2. **Intermediate Nodes:** These nodes signify the national importers for the product. They procure from either of the two root nodes, and supply pharmacy outlets (end nodes).
3. **End Nodes:** These nodes signify the pharmacy outlets from where consumers purchase products. Upon deciding to place an order for more product, end nodes reference a ranked preference list of intermediate nodes. Should the most preferred node be stocked out of product, the end node attempts to procure product from the next node on the list, and so on. Upon exhausting this list, end nodes will delay for a period of time before moving to procure directly from the root node supplying unacceptable product.

Intermediate and end nodes follow fixed-order-quantity inventory policies, looking to order a set amount once the available inventory drop below a set reorder point. Orders are delayed according to lead times that are end-node dependent. Demand schedules at the end nodes are generated according to a decided distribution.

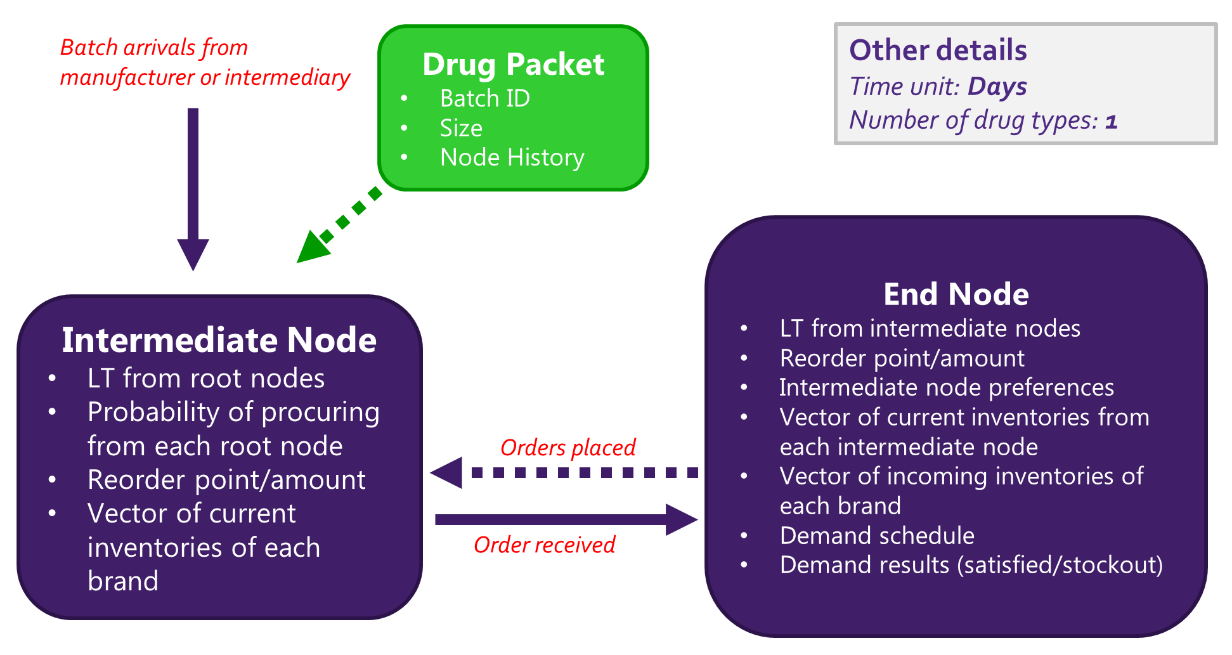
Product sampling (similar to post-market surveillance) is potentially conducted at end nodes during each day of the simulation, according to a preordained testing policy, limited by an entered sampling budget. On a scheduled day, an end node is “visited,” a single package of product is purchased, an audit of the inventory at the end node is compiled, and the product is tested according to an established diagnostic sensitivity and specificity. The inventory audit during the visit notes the intermediate node sources for all available inventory at the end node, with product from the root node supplying SF product “hidden” from the audit (yet still available for purchase and testing).

The simulation timeline is as follows:

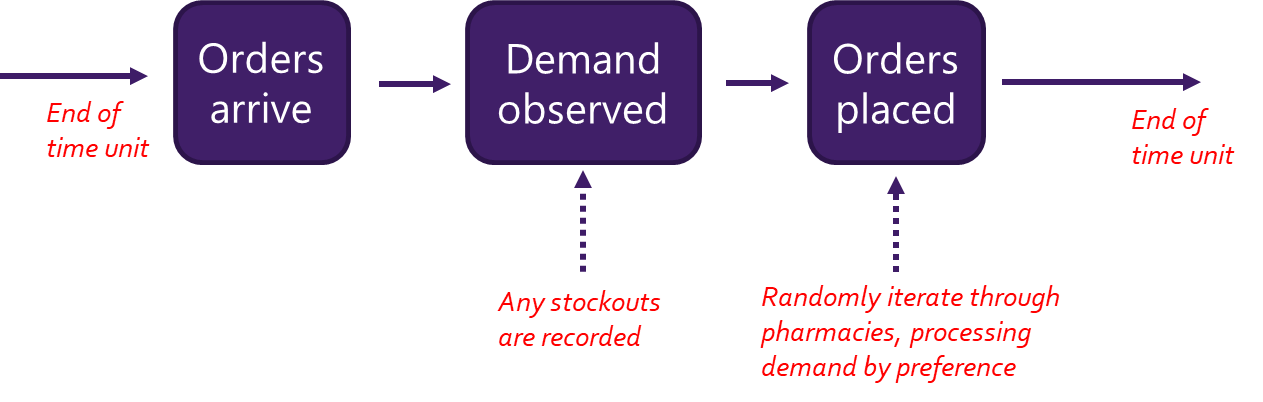
1. Input files are read into the system. These input files include the following:
   * The node list, containing each node and its key characteristics
   * Arc preference transition matrix, containing the ranked preference of each end node for each intermediate node
   * (Optional) Arc-dependent reorder points and lead times, in case these quantities change for end nodes with different intermediate nodes
2. Key simulation parameters are set, including the following:
   * Number of simulation days
   * Sampling budget
   * Testing policy to use
   * Diagnostic tool sensitivity and specificity
   * Whether to use prior established simulation warm-up runs
   * Probabilities that each importer procures from the root node supplying SF products
3. Respective objects are constructed via the read input files
4. Testing schedules are initialized
5. During each day of the simulation, the following process takes place:
   * Intermediate nodes process incoming orders
   * End nodes process incoming orders
   * End nodes process day’s demand and update consumption records
   * Sampling of end nodes is conducted per the testing schedule
   * End nodes make orders if reorder points reached
   * Intermediate nodes make orders if reorder points reached
   * Testing schedules (if dynamic) are updated
6. Once all days have completed, statistics are compiled, and outputs are printed and/or stored

The following graphics depict the different structural logic of each element of the simulation.

### Overview of the flow of products (“Drug Packets”)



### Timeline at the pharmacy outlet (end node) level



## Model Parameters

The following is a list of parameters that can be modified within the model:

* Reorder points and reorder quantities for intermediate and end nodes (i.e., internal policies)
* Lead times and global stockout rates (i.e. external factors)
* Underlying network structures
  + Number of importers that can be accessed by each outlet
  + Number of entities engaging with the falsifier node
  + Overall size/complexity
* Trigger sensitivities for entities opting to source from the falsifier node
* Testing policies and budgets

# Base Model

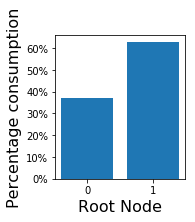
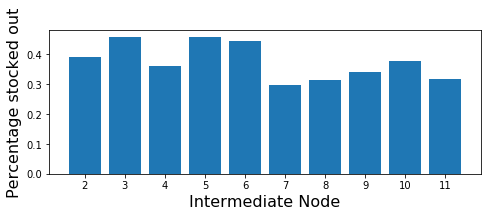
This section details the assumptions made in generating a base model intended to capture the dynamics of a small-to-medium sized country (our inspiration is the national private supply chain of Liberia). These various assumptions may be adjusted for different configurations and scenarios:

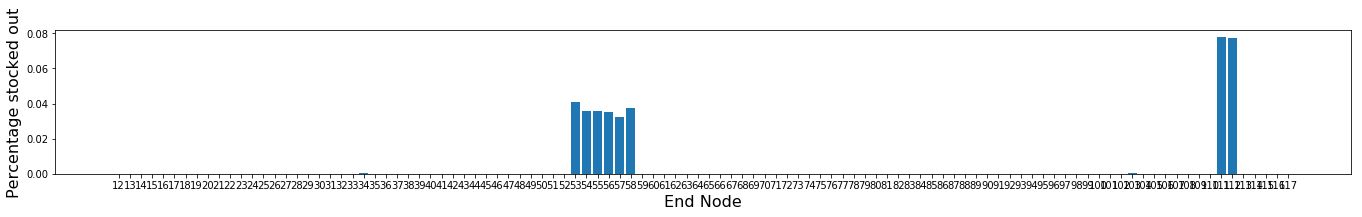
* **1 SF product node; 1 global manufacturer/supplier**: These are the two source nodes for all drug packet objects. Intermediate nodes (importers) may source from either node, while end nodes (pharmaceutical outlets) source from either intermediate nodes or the falsifier node.
* **10 intermediate nodes**: During a September 2019 visit to Liberia, it was established that there are roughly 10 major private importers of pharmaceuticals that would provide the large majority of drugs for the private supply chain (especially antibiotics). This value could be adjusted for looking at different pharmaceutical classes (doxycycline, ampicillin, etc.).
* **106 end nodes**: The September 2019 visit also revealed there to be 10 registered pharmacies in Bong County in central Liberia. Using population estimates, we extrapolate this knowledge to other counties presuming a similar pharmacy-per-capita ratio, which results in 106 total pharmacies.
  + Ideally, we will be able to use the exact number of registered outlets at some point.
* **Average demand per region**: The 2018 WHO Report on Surveillance of Antibiotic Consumption shows each of the surveilled African nations (4 total, excluding Liberia) consuming around 5 daily dosage XXx per 1000 inhabitants of beta-lactam antibacterials/penicillins. Amoxicillin falls under this umbrella and constitutes the most commonly consumed antibiotic in many countries, including Liberia (amoxicillin signifies about 25% of this category in Burundi). Without differentiating, and assuming that we are looking at all beta-lactam antibiotics in our simulation, we use an assumption of 5 days per treatment/blister pack to approximate a 1 blister pack/1000 inhabitants per day in our base model.
  + These approximations can be **greatly improved** upon receiving importer information with respect to specific drugs, like amoxicillin, doxycycline, etc.
* **Average demand per pharmacy**: Once the demand per region is established, we can use the estimated pharmacy numbers to establish how much average demand should be experienced by each pharmacy. We assume pharmacies in a particular region see the same average demand. Again, more realistic demand values can be inserted with the requisite data.
* **Demand variability**: Currently, the model uses Poisson distributions at each end node via the established average demands.
* **Lead time from Monrovia per region**: These values are best guesses from personal knowledge of round trips to and from each county of Liberia to the capital, Monrovia, and currently exhibit no variability.
* **Inventory policies**: Reorder points are set at each end node to satisfy a 90% service rate and the reorder levels are set equal to 15 days of mean demand. For intermediate nodes we use a 90% service rate relative to first-preference customers, and 60 days of mean demand.

# Simulation output examples

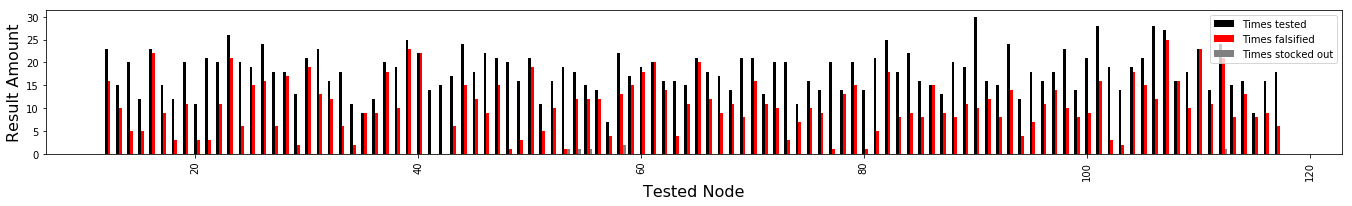
The following is a collection of sample output charts:

*Root node consumption and intermediate/end node stock-out percentages:*

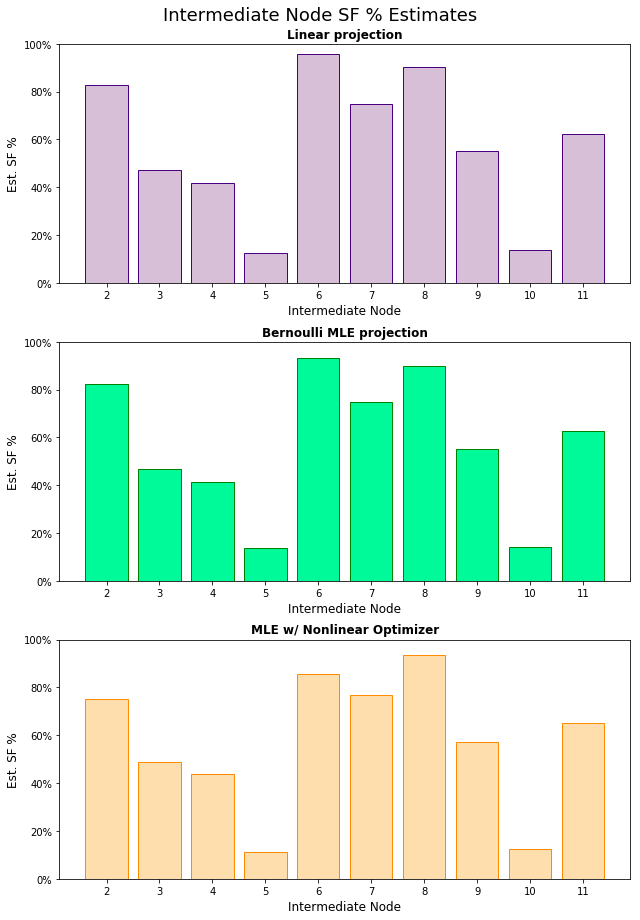
 



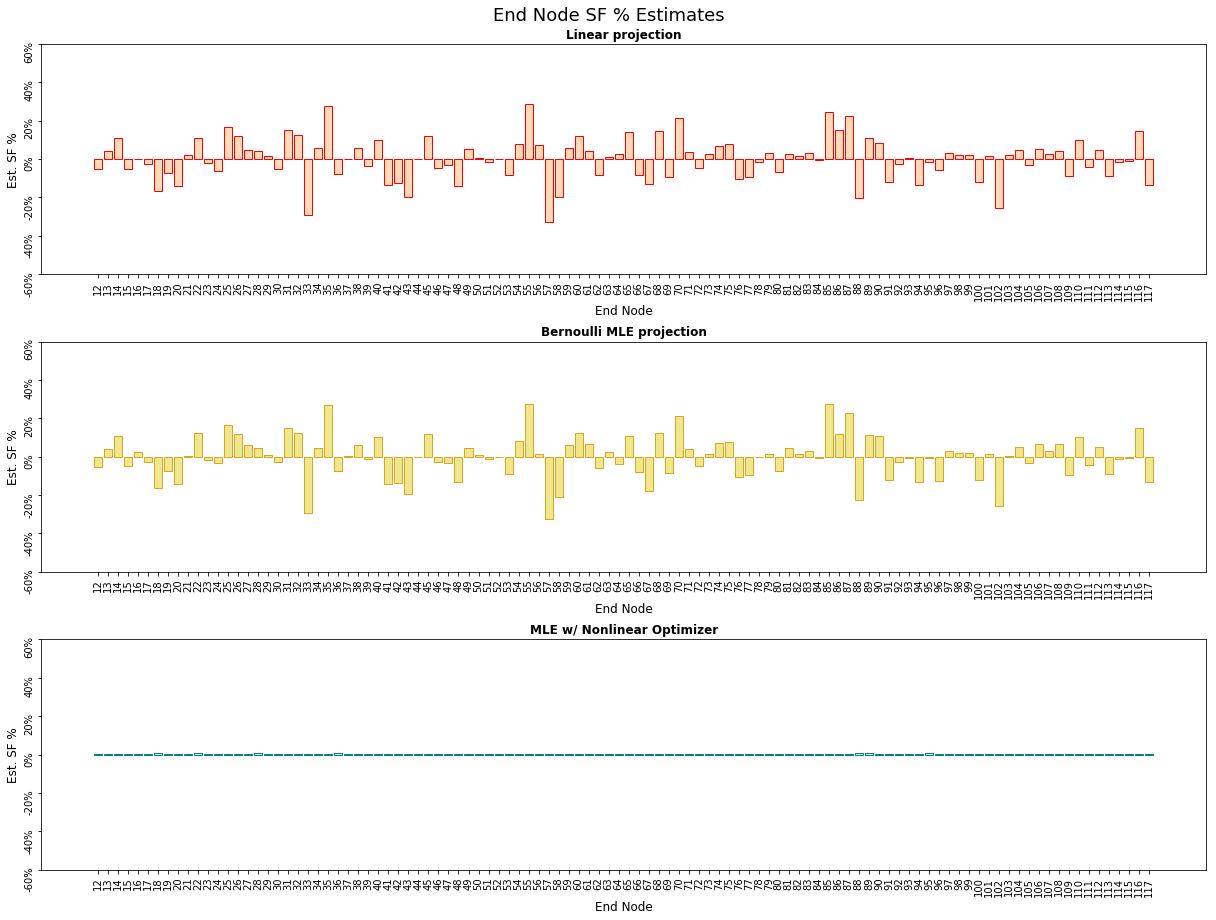
*End-node testing results:*



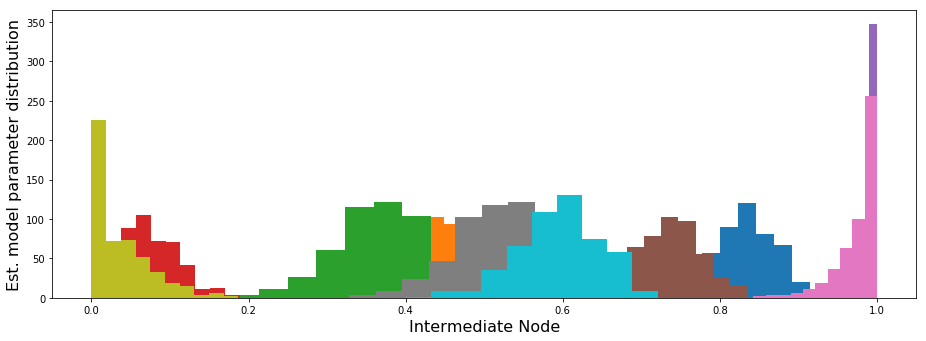
*Intermediate node SF estimates, using testing data:*

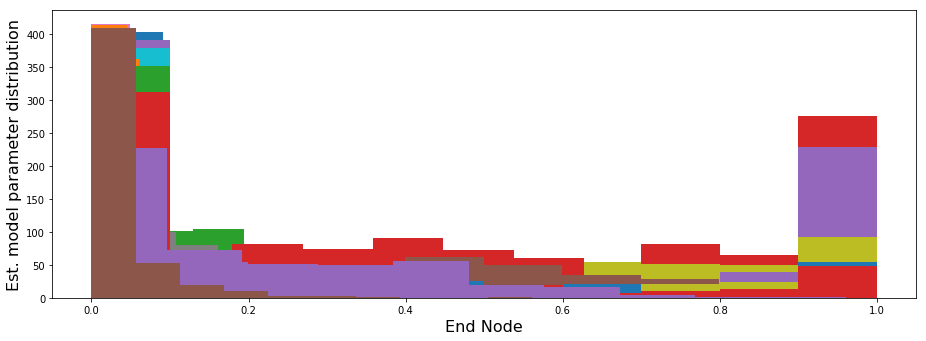


*End node SF estimates, derived from intermediate node estimates and data:*



*Samples from estimated posterior distribution of SF rates, using the No-U-Turn-Sampler:*





# Building this model elsewhere

The principal drivers of this simulation model are the sampling data collected at different points in the system and an idea of the overall supply chain structure. Heightened information around each of these drivers will result in a more useful model. Although assumptions for any piece of data can help bridge gaps, the following is a list of different pieces of information that would help the simulation best emulate reality:

* **Supply chain structure**
  + Importer data
    - Inventory policies
    - Lead time to order reception
    - Size + date of shipments received
    - Demand data
    - Stockout prevalence
    - Procurement criteria for different manufacturers
  + Outlet data
    - Inventory policies
    - Lead time to order reception
    - Size + date of shipments received
    - Demand data
    - Stockout prevalence
    - Procurement criteria for different importers
    - Regional disease prevalence
* **Detection capacity**
  + Potential sampling plans
  + Regulatory actions
  + Sensitivity/specificity of screening diagnostic