Memo for reshuffling locations to generate random spatial distributions

Qinyun

February 18, 2020

The context of ABM project: the study focuses on 198 zip code areas within IL (1383 zip codes in total across IL). We have all MOUDs within IL available (three different types of MOUD: MOUD - Buprenorphine, MOUD - Methadone, MOUD - Naltrexone).

The research question is the how spatial distribution of MOUDs affects health outcomes (with the help of agent-based modeling).

This memo records different approaches to generate random spatial distribution of MOUD locations. (Precisely speaking, we don’t want complete randomness. We want to have complete randomness conditional on the distribution of at-risk population/need for MOUDs.)

For both approaches, we assume that the total amount of resources is limited. That is, within IL, we only have resources to support the amount of MOUDs that already exists: 447 MOUD – Buprenorphine, 81 MOUD – Methadone, and 141 MOUD – Naltrexone. The reshuffling is only about relocating these locations.

Approach 1 (R code see: ABM\_reshuffle; results is named as reshuffle 1, Scenario 1)

We first separate IL into two areas: study area that includes 198 zip codes and other area. The study area has 333 MOUD – Buprenorphine, 62 MOUD – Methadone, and 100 MOUD – Naltrexone locations. In total that gives us 495 points. We then generate 495 random locations (using runifpoint in spatstat package) within the window of the study area. Then we randomly labeled these 495 locations as one of MOUDs (Buprenorphine, Methadone, or Naltrexone), and we keep the distributions of three medications unchanged. That is, within the study area, after reshuffling, we still have 333 MOUD – Buprenorphine, 62 MOUD – Methadone, and 100 MOUD – Naltrexone locations. We do exactly same thing for other area that is outside study area.

The rationale for Approach 1 is: the selection of study area partially depends on consideration of population at risk. By reshuffling separately in and outside study area, we take into account this factor of population at risk for spatial distribution of MOUDs for reshuffling.

Approach 2 (R code see: ABM\_reshuffle\_needbasedapproach; results is named as reshuffle 2, Scenario 2)

The rationale for Approach 2 is that we want to better consider population at risk when reshuffling MOUDs. Specifically, we first calculate the population at risk (population from age 18 to 39) for each zip code within IL (in total 1383 zip codes). Then we assign each type of MOUD proportional to the population at risk for each zip code: the more population at risk, the more MOUD. For this step, since the number of MOUD is pretty small (447, 81, 141) while we have 1381 zip code areas, we use the “Largest remainder method” (<https://en.wikipedia.org/wiki/Largest_remainder_method>) to make sure that for each zip code, an integer number of MOUD (each type) is assigned. As such, we get how many MOUD (each type) is assigned to each zip code. For example, for zip code area 60002, 1 MOUD – Buprenorphine, 0 MOUD – Methadone, and 0 MOUD – Naltrexone has been assigned. Finally, we generate the assigned number of random locations (again, using runifpoint in spatstat package) within the window of each zip code area, and we randomly labeled these locations as one of MOUDs (Buprenorphine, Methadone, or Naltrexone) while keeping the distribution of three medications unchanged. For example, for 60002, one location/point is randomly generated, and this location is assigned as MOUD – Buprenorphine. Note that in Approach 2, the total number of MOUD locations within the study area might change from real situation.

Other possibilities that have not been realized yet: we relax the constraint of fixed number of locations/points. Instead, we use the Poisson distribution and only fix the expectation/mean (the expected intensity will be based on the real situation). That is, we may have number of MOUDs that are different from 447 MOUD – Buprenorphine, 81 MOUD – Methadone, and 141 MOUD – Naltrexone across IL. Again, we can either reshuffle within study area and other area (like Approach 1), or within each zip code area (like Approach 2). We can also define “local need” using actual health outcomes instead of at-risk population.

In the perfect scenario, I think we should generate a distribution of the health outcome that we care about based on repeated reshuffling locations. That is, each time we reshuffle, we calculate the corresponding health outcome. We repeat this, say 1000 times. Then we have a distribution of the health outcomes and see where the observed/real health outcome is in the distribution. Like what we do in statistical inference.

[Note after talking with Jonathan: we have two dimensions that we need to deal with: (1) sensitivity to different parameters in the ABM (2) uncertainty in reshuffling. We might see if the results are sensitive to one dimension and then for another.]