**Hospital Arrivals / Occupancy Forecast Model**

**Motivation**

It is useful to be able to forecast hospital occupancy, arrivals, discharges, transfers, etc.

**High-level Idea**

For currently admitted patients, we need to be able to forecast when they might be discharged (or transfer). We also need to be able to forecast arrivals, and then forecast when those arrivals would be discharged.

**Arrivals**

When we initially developed the arrival forecast aspect, we tried many ML algorithms in R. Then we tried a time series decomposition approach, and it was as good as the best ML algorithm. As we didn’t have an ML server, we went with the time series approach.

To get a better forecast, we break out arrivals into separate patient streams. Some of these streams, such as ED, can’t be accurately predicted, as they are inherently random. A time series approach works better than any other approach we’ve tried. Nursing evaluated our model against other predictions, including the HOM’s prediction (I.e., expert opinion), and our model outperformed all other models.

For patient streams using the time series, we take a moving average based on day of week. Each weekday gets its own average, and weekends and observed holidays get lumped together and share an average.

As we worked on the second version of the model, we updated some of the patient streams to be based on data in Epic. Surgical cases are scheduled more than 30 days out, so this can be used. There is a certain rate of cancellations as well as of add-ons, and these rates change depending on how far in the future it is. I created a scaling factor that made the forecast look reasonable. I have been recording predictions in order to come up with a more rigorous scaling factor. It will still be necessary but can and should be based on harder numbers.

Similarly, most OB patients come in for prenatal care at UIHC. When they do this, they enter an expected delivery date. While these dates aren’t that accurate, on average it can be expected that early deliveries will cancel out late deliveries, and vice versa. But again, there needs to be further analysis of cancellations - here, meaning patients who see OB for prenatal care but then deliver elsewhere. Perhaps we could look for patients with 3 or more prenatal visit dates?

OB is also interesting in that there are OB patients who come to deliver at UIHC without any prenatal care at UIHC. We treat these patients as a separate stream and use the time series approach.

Neonates (I.e., babies) are based on the OB arrival numbers, multiplied by a derived average number of babies per delivery.

**Occupancy / Length of Stay**

Next, we consider how long patients will stay at UIHC. Let P(d) denote the probability that a patient is still admitted on day d of their admission (with d=0 being their day of admission). It’s helpful to divide patients into different types, to get more representative probability distributions. So, let P(d,j) be the probability that a patient of type j is still admitted on day d of their admission. To derive these probability distributions, we look at historical admissions. For each patient type j, we count the total number of patients, and for each day of admission d, we count the number of patients still admitted. We then divide the daily number of patients by the total number of patients and get a probability of being admitted on day d.

The probability distribution for each patient type j looks like this:

**Applying to Currently Admitted Patients**

We can then apply P(d,j) to currently admitted patients to forecast future occupancy. It’s extremely difficult to predict when one individual patient will discharge. This would require a lot of clinical data which is not currently in the model. Certainly, something to look into for future work. The approach is to instead apply the probability model to all currently admitted patients. So, if you’re currently admitted today, depending on the probability distribution, you might contribute 0.9 to the occupancy number for tomorrow, 0.8 for the day after, and 0.65 for the day after that. When you add this up across all patients, it gives a good forecast of the overall occupancy number.

Now, how do we deal with the fact that currently admitted patients are generally not on the first day of their admission? We can use conditional probability for this. Let A mean that a patient is admitted on day 3 and let B mean that a patient is admitted on day 4.

E.g., the probability you’re admitted on day 4 given that you’re already here on day three is

P(4|3) = P(4) / P(3).

**Applying to Arrival Forecast**

Things get interesting when you apply the probability distribution to the various patient stream arrival forecasts. In addition to having an arrival forecast for each patient stream for each day in the future, each of those forecasts gets scaffolded out with its own applied “still admitted” probability distribution. Here is an illustration:

**A Note on Patient Type**

The way we’re splitting patients into different types could be really improved. Especially Surgical - there must be a better way than procedure name… We wanted to use DRG, as this is how CMS groups patients for length of stay estimates. However, only about 50% of currently admitted patients have any DRG (working or final) assigned. If this could be improved, that would be a great data point to plug into.

One of the applications of this forecast model is to drive a “what if” scenario tool. Ideally, patient types would align with the different “levers” available to the hospital, in terms of what streams to temporarily turn off. More work with users / admin would be needed to make sure this aligns in a meaningful way. The way we modeled patient types makes this kind of work actually pretty straightforward - there just needs to be better alignment in terms of what useful patient types would be.

**Future Directions**

Discharges

One way to forecast discharges (without getting into clinical ML models) would be to compare the summed predicted occupancy for the day in question and subtract that from the summed predicted occupancy for the previous day, excluding new arrivals.

Hospital Location

The single biggest and most common request we’ve received on this project since rolling out version 2 is to include hospital. There are many problems with this. First, combined with patient type, this will reduce data sets to too small of a size, leading to unreliable numbers. Second, patients don’t always get assigned to the right unit, because that unit is already full. Any forecast model at the unit level will predict how we manage our patient load, not what the load itself is. This seems like the wrong idea. Furthermore, the more detailed your forecast, the less accurate it will be. There is a concept called pooled variance, which says that the larger your denominator is, the less variable your numbers will be, and vice versa.

What we can and should do is report by Care Area (CA) and Care Area Specialty (CAS) (groupers 15 & 16). Generally, patients will be put at least in the right CA & CAS. Furthermore, this level of detail will still be helpful to admin, and won’t cause as many issues with data set sizes or pooled variability.

Transfers

Once you introduce location, you have to deal with transfers. This gets tricky. Some ideas…

Markov approach: use a continuous Markov chain methodology to determine how frequently patients get transferred between different CA / CASs.

Probability approach: extend P(d,j) to be P(d,j,l), where l is location. Then, you would determine the historical probability of a patient of type j being in location l on day d of their admission. Conditional probability for currently admitted patients gets a bit tricky here. Can you assume that patients only stay the same or go to lower levels of care? Can you assume that any patient at any level of care is equally likely to go to any level of care within a given Care Area? Can you ignore individual patients and treat them as a collective probability distribution? Here is one possible formulation:

Hourly Forecast

Another common request is to have an hourly forecast. Due to the small size of data sets, and the large number of scaffolded combinations, not to mention the data refresh cadence, this is a tricky request. One idea would be to come up with an overall hourly distribution of arrivals, by day of week, as well as an overall hourly distribution of discharges, by day of week. Then, you can apply that to the day’s high-level forecasts, and get an hourly forecast.

Arrivals vs Admissions

An important question is whether we should be considering arrivals or admissions. Do you count an ED admission when they’re admitted to inpatient, when they arrive at the ED, or when the decision to admit is made? I would vote for using decision to admit, but it depends on what the question is. This same idea applies across most patient streams (e.g., Direct Admits). How do you include this backlog / waiting list demand without completely skewing everything else?