Codebook/notes for mind\_exploratory work:

Adt\_long:

Covid\_visits\_start\_dt

Adt\_event\_dt

Adt\_ed\_admission\_dt

Adt\_inpatient\_dt

Adt\_discharge\_dt

All above with “\_date” are the date versions

Covid\_long:

Covid\_pcr\_dt

Covid\_pcr\_date (date version)

Covid\_pos\_logic

Micro\_long:

Admit\_dt

Specimen\_collected\_dt

Abnormal\_result\_logical

Patients:

Deceased\_logical

Deceased\_df

Hospice\_logical

Hospice\_dt

Covid\_pos\_ever

Covid\_pos\_dt

Adt\_long\_patients:

Dur\_covid\_to\_inpatient

Covid\_admission

Covid\_inpatient\_admission\_dt

Covid\_inpatient\_discharge\_dt

Covid\_admission\_time\_in\_hosp

**9/27/21**

Tal presents:

* Create metwork model
* Try to estimate parameters of model
* Try to estimate likelihood of specific patients to be colonized.
* “inference over different trajectories” – he considers this to be “not good enough” to tell if things have converged. With “synthetic data” you can check in (I think that’s what he means?) but not with real data….
* Looking at number of observations does not tell you how cases are distributed. Eg, one case in each ward – would imply that they mostly came in through importation. On the other hand, if all are in one ward, suggests transmission.
* “entropy” as an indicator in his model: how many ways could it have been rearranged, normalize to see how much entropy does each observation add to the system.
* Next steps: find best indicator of clustering; non-homogenous observation rate (by ward, by status, by identifiers); individual-level inference. Consider: what’s the probability of being tested if they’re negative? What’s the probability of being tested if positive? May be different (a little hard to tell how you’d figure this out).
* Baseline period: 11 years of baseline data. A little hard getting this consistent btw tal and Jaime. This would be followed by a two stage process Sen designed: estimate parameters guiding importation, enviro colonization, person to person transmission, then do this across the entire domain. It’s a stochastic simulation. Some ppl will be randomly designed to be already colonized, etc. Then take this to the macro level and see “how many ppl in 4 wk period have been observed to be infected?” – use this only to adjust parameters – iterative filter to estimate these parameters. THEN, ultimately go back and do individual level estimation of probability of infection.
* “synthetic true simulation” – he assigned some “real” parameters, and then is going back and running to try to estimate the parameters. “this is the estimation problem for a dynamic system”.
* “observations don’t tell the entire story” – Tal’s point.
* Maybe we could glean some information from the level of clustering. What are the types of observations that we could leverage to try to make our estimated parameters better? Looks like entropy is similarly correlated to the variable as observations is.
* AC’s comment: they use MRSA data, but there’s a lot of variability within MRSA. Jeff says: that’s why we have the sequencing – way to id when transmission is happening and when that’s really not what is happening.
* Jeff says: first order of business is just to try to get this model to run. Sen says: one of the most important aspects is the observational model.
* In this model: we’re looking just at mrsa. We could similarly say, if, eg, urine and blood variants tend to be different – we could make a model with just the for the variant of interest. Might be sparse, but could be important depending
* Notes on the network: we’re new to the network.

**11/3/21**

**MIND network meeting:**

U Utah – vaccination modeling is their common theme

Vaccination against C Difficile infection: effects in healthcare and community populations – Damon Toth

* Candidate vaccine does not target organism directly
* Transmission reduction depends on role of symptomatic vs asymptomatic carriers
* C difficile spores that contaminate surfaces are difficult to clean
* How much could/would this vaccine impact transmission in real world setting
* 10 facility agent based model – looks like metapopulation model to me
* Vaccine is targeted at toxins, assumed to act on progression-to-infection only (delta, in their model)
* Their paper was published last year, but did not consider a community transmission component.
* New question introduced in this presentation: could toxoid vaccine provide an additional benefit beyond what they estimated before by impacting community transmission
* Puzzling result they got: why did the vaccine’s reduction in community transmission not produce a drop in CDI
  + Answer: most community cases don’t progress to CDI/

Alex Beams – dose response in SIR models

* P = 1 – (1-r) ^d
  + D = number of viruses
  + Calculate probability that none of the viruses will succeed, then take 1 minus that….
  + P = 1 – e^(-r delta) – if poisson assumption for dose size.
  + Oy
* Models of vaccine: Leaky model – more realistic – vaccine reduces virion success; All-or-nothing model – assume vaccine either works or fails.
  + In leaky model, infection still depends on the dose size, even if you’re vaccinated
  + More breakthrough infections expected under the “leaky” model
* “simplest model for how dose response could work”

12/7/21

Notes from Columbia MIND Meeting:

Anne-Catrin mentioned: Pseudomonas is the trickiest.

AC is struggling to spend money for covid research

16000 for 96 samples…..

Looking at emergence of variance among immunocompromised hosts

Evolution of virus among immunocompromised patients.

For Iota variant – they know where it arose – in a patient with HIV. In a preprint they’re working on – sequence in August and November

Thought is that it doesn’t reassort, but long term evolution in individual. Addumulation of mutation in individuals, but it usually doesn’t spread. She can’t rule out that this individual passed it to someone else, and then it evolved,

Trying to get Jaime to do modelling of Omicron – how much of a suppression in near term mortality can omicron offer, since it seems to have a lower fatality rate – if it protects you against Omicron and Delta.

Wan Yang and Jeff put together something – delta, replication dynamics are accelerated and reaches higher titer (?), something about neutralization should account for delay in immune response.

Bill Hanish – “Transmission properties are more important than immune evasion”

Jeff says: people say we need to vaccinate to prvent emergence of new variants, he doesn’t see logic in that, he thinks it just applies a selective pressure (hmm….)

**12/20/21**

Notes from meeting with AC and Jeff about possible nursing home model:

Questions I had before meeting (didn’t pose all of these):

* What have you observed regarding nhs and infections?
* What particular pathogens are of concern here?
* What is availability of data from nursing homes?
* What solutions are you imagining could be feasible?
* Existing research on nursing homes
* Existing nursing home-related interventions

A few silly meeting notes:

Covid: Jeff wonders if they should close schools

AC and J bonded over Brahms: both think Brahms = “foolish”

Meeting notes sent to AC and Jeff (sent 12/29):

* Lingsheng’s current work on nursing homes, covid cases and AMROs:
  + Are people coming to the hospital coming fro nursing homes? Did they previously have covid?
  + Looking at covid diagnosis (yes/no) and subsequent readmission
* It can be complicated to ascertain time of covid diagnosis: sometimes time of admission < time of positive test; sometime there are notes in the chart that indicate covid diagnosis
* Primary question: are nursing homes amplifying MDRO in covid patients?
* Population that goes between nursing homes and hospitals:
  + Population that goes to nursing homes/LTACHs is very diverse: often people with underlying chronic conditions
  + The population of people going back and forth (between hospitals and nursing homes) is 3000-4000
  + There are long stretches of time with very few covid patients
  + Readmissions are somewhat high, since the US hospital system is very concerned about minimizing length of stay
* Are infections spreading between hospitals and nursing homes a concern outside of the covid context? – yes; > 20% of KPC infections are in people coming from nursing homes.
* As an exercise, we could look at covid transmission within a subset of nursing homes
* Next steps:
  + Anne-Catrin will share data with Emma and Jeff
  + We will keep Lingsheng involved in any work on this topic

**1/5/22**

Work notes:

* Working on setting up dataframe for individual pathogen analysis using the new plan we discussed in meetings: rather than matching people only with individuals who haven’t had any other infection, try including “have they had a prior infection?” as a component of the propensity score model on which they are matched
* First try: started making variables in matching\_df including “time of pathogen of interest”, “first time of pathogen of interest”, “prior infection” by looking at whether the first infection was prior to the first infection with the pathogen of interest.
* However: I think this will pose a problem for including in the matching model. It’s clear whether individuals who HAVE an infection with the “pathogen of interest” have had a prior infection, but complex to set a time cutoff for potential controls (ie, would we consider any other infection during their stay to be a “prior infection” with another pathogen?)
* **Possible solution:** incorporate “other bacterial infection at the same site” as a time-varying covariate – after all, they have infections with another pathogen at certain times
* Remaining questions: how to set this up; do we have a time when we consider the other infection “resolved” ? Or do we just make the start date the onset date of infection, and not have a time when the other infection ends? Will be complicated by the fact that some infections are tested for repeatedly, and others may not be – may be presumed that they’ll last a certain length of time, based on course of antibiotics or whatever.
* Look into options for this tomorrow? And also look into incorporating ICU variable in the same way.
* Steps for tomorrow:
  + Find out how data needs to be set up to do a time-varying cox ph model
  + Sketch out data setup that we’ll need
  + Figure out how to get these two variables (prior infections; ICU variable) into the format to use for matching.

**1/6/21**

Work notes:

* Helpful article summing up the ones I read previously: <https://onlinelibrary.wiley.com/doi/full/10.1002/sim.8533>
* In article, look at section 5.2.3 for a description of Bo Lu’s 2005 paper that my analysis is based on.
* It looks like the key component is: “Using standard software for Cox proportional hazard regression with time-dependent covariates, the propensity is estimated for each patient (on the log hazard scale) taking the value 𝜸ˆ′𝑍𝑖(𝑠𝑗) at the initiation of the jth experiment. Controls are selected from among those who remain at risk, Ri(sj)=1, eligible, ℰ𝑖(𝑠𝑗)=1 and were not treated in a previous experiment, Si≥sj, by minimizing the distance function 𝛿={𝜸ˆ′[𝑍𝑖(𝑠𝑗)−𝑍𝑗(𝑠𝑗)]}2 . Within each strata (j=1,…,nS), matching proceeds as with a cross-sectional study.”
* So, I think I will need to run a separate cox analysis for each timepoint when an infection occurred, and then conduct matching.
* Lessons on cox models with time varying covariates:
  + <https://mathweb.ucsd.edu/~rxu/math284/slect7.pdf>
  + <https://cran.r-project.org/web/packages/survival/vignettes/timedep.pdf>
  + <https://stats.stackexchange.com/questions/210609/how-to-use-time-dependent-covariates-with-cox-regression-in-r>

**1/7/22**

Work notes:

* I think I can set up time-dependent covariates with the tmerge function in the survival package. <https://stat.ethz.ch/R-manual/R-devel/library/survival/html/tmerge.html> Also described in second link from yesterday.
* <https://daynebatten.com/2016/01/tmerge-customer-churn-data/>

**1/10/22**

* Challenges/questions about incorporating previous infection as a time-varying covariate: Do we include all infections together, or somehow include individual pathogens? What is the time span during which we consider someone to be infected? Maybe people only have a test or two, so “end of infection” is challenging to ascertain. Do we just consider them infected from the time that the infection starts onward? Or do we calculate an estimated end of infection that’s some kind of standard for each type of pathogen, and then use either that or the final positive test for the infection, whichever is longer?
* Plan: first, start with lumping all pathogens together, and with just the onset date – no end date.

**1/25/22**

Not sure of date, adding this retroactively:

* Jeff suggests we can’t just consider people exposed forever after detection of an infection (fair enough). Instead, consider five categorical variables: the day of detection of another infection, the day after, etc through 4 days after. There are a few problems I can think of with this, but I’ll try it anyway.

**2/24/22**

* Working on: setting up according to Jeff’s instructions
* Question here: do we need to worry about multiple overlapping infections "resetting the clock" kind of? ie, if someone gets infection a on day 0 and then infection b on day .5, will we be accidentally creating a "one day out" line on day 1, when it shouldn't be until day 1.5..? check this by checking how many people really have multiple infections.
* Struggled with this question for quite a while.
* One possible solution: Since we’re setting up days after infection as separate variables, we could just peacefully allow these contradictions to exist. Ie, since they’re separate variables, you’re allowed to be “during” and infection and also be “3 days post” another infection, for instance, at the same time.
* This is more complicated to understand, but I think it might actually be easier to set up.
* In other words, as I put in the R file: how to handle: see above. I set up separate variables, so it's possible to be on "infection lag 0" and simultaneously on "infection lag 2" for instance. they're no longer mutually exclusive.