**­2/15/20**

Notes:

* Research Qs I’m assigned to look at:
  + **How has COVID mortality changed over time? And in association with bacterial infection?**
    - Basic viz, glm model with penalized spline for time? Once I have bacterial infection data, include that and (more importantly) interaction term?
  + **How has covid patient time in hospital changed? What features are predictive of covid patient time in hospital (eg conditions, treatment, AMRO).** 
    - Considering multiple stays?
    - Change based on admission day I guess?
    - How is AMRO defined?
* Looking at covid care dictionary: what are relevant datasets and variables for each of these questions?
  + ADT = admission discharge transfer
  + Adt\_summary: one row per patient. This is probably the key one.
    - Related variables: Mrn (mrn number), adt\_ed\_admission (ed admission date/time), adt\_inpatient\_x (inpatient admission date/time), adt\_discharge\_x (discharge date/time), adt\_time\_inpatient\_1 (????), deceased\_final, deceased\_date
  + Problem: this dataset has all patient medical history, problem list, and encounter diagnosis. The summary doesn’t include bacterial infections but maybe the long problem list does.
  + Current meds:
    - Related variables: antibiotics\_ordered
* Questions:
  + Where exactly are we looking for the bacterial infection? It’s not in the problem list according to the covid care data dictionary, but I think that only includes items in the summary file. Maybe it’s in the long problem list files? Alternately, are we looking at the meds prescribed to consider this? A: see below, Jason’s going to be sharing with us this week.
* Steps to take:
  + Basic data viz
  + Look at mortality first, since we don’t have micro data.

MIND meeting notes:

* People at meeting: Anne-Catrin Uhlemann, Sen Pei, Claire Xiao, Shenglin Liu– calling in from China, Jason Zucker, Anne-Catrin and Jason are clinicians at Columbia med center.
* Grant includes modelling and also covid stuff (in a supplement), including statistical analyses. Part of the process of doing all of this.
* Want to do a deep data dive to see “what’s there” – what can we understand through statistical analyses, etc, which will be used to inform future mathematical analyses.
* Sen: associate research scientist, working w jeff last 5 years, flu and healthcare associated infections.
* Anne-Catrin: adult inf. Disease specialist, looking at bact. Resist infections
* Jason – adult and peds inf disease doctor
* Sasi: programmer with Jeff’s dept
* Claire: 1st yr biostats masters student,
* Shenglin: 2nd yr biostats student.
* Expect update from Jason next week.
* “microdata”
* Bacterial data – upcoming from Jason
* Anne-Catrin comments: Questions on colonization and infection: less black and white, somewhat bleeds over into clinical arena, may need additional chart reviews, how to classify. Standard: two ID specialists review, then a tie breaker if they disagree.

**2/18/20**

Notes:

* Don’t have micro data yet, still
* Are those who were discharged to hospice considered in mortality calculation, if we don’t have data on death?
* This is very depressing
* What’s the difference btw adt\_first\_inpatient\_encounter\_date.x and the same.y?

Meeting with Jeff:

* What should I expect about how AMRO is defined?
* He also doesn’t know how those data are going to look – for AMRO.
* Admission data – not all covid patients necessarily.
* Critical that I pull out the covid patients –
* Won’t have time prior to covid to see if something changed (ie for non-covid patients before vs during covid)
* Hospice: what time fragments I can break that up into. (?)
* Can break this down by facility
* Can do the same thing for bacterial infection vs not.
* Maybe just descriptive statistics.
* Predictive features: more formal regression framework.

**3/1/21**

To-do:

* Merge covid test data with mortality data. Look first at if the patient EVER was diagnosed with covid and if they died – for now, disregard covid-related or non-covid-related cause of death.
* Visualize covid mortality data and non-covid mortality data
* Next, look at bacterial infection data. Make list of questions. Maybe merge with rest of data, looking at mortality for covid and non-covid patients stratified by EVER having

Questions for meeting:

* Do we care when a patient was diagnosed with covid when looking at mortality? I’ve tabulated the simplest possible thing which is deaths in people who have EVER had a pos covid test (that we know of) vs mortality of ppl who NEVER did
* Who is in the covid tests wide dataset who is not in the adt dataset??
* Looking at change over time, what time are we most interested in: date of positive covid pcr, date of ed admission, etc?
* Are we interested in bacterial infections that occurred ANYTIME in patients who EVER had covid tests, or in bacterial infections that coincide with covid positive..?
* Can you talk through the sensitivities data, and what we should be looking for?
  + Is it just common sense like “>100,000” or “10,000< x < 100,000” or “positive” is positive, and other things (“none” or blank or “negative”) are negative? Or are there other details we should know?

**3/1/21**

MIND meeting:

* Meaning of sensitivities file: S = sensitive, R = resistant, “<=” means sensitive “>” eans resistant, no > or < sign also means sensitive. They will go through this data and provide a simpler translation for us
* Sasi made very detailed graphs, including organism data (organized by fungus, gram positive, gram negative)
* Site dependency: site of sample collection can mean different things (like infection vs colonization). Blood is most reliable, then respiratory, urine needs more “curation”.
  + Consider first pass with blood, urine and resp and aldo with just blood? Or maybe just blood first?
* There’s some data on tests that’s for ppl not hospitalized, I basically don’t need to use/worry about that. From cornell, for example – probs in there by accident.
* Focus is on people admitted for sars-cov2 and then contracted bacterial infection, but there could be some of the opposite. Might want to just look at both.
* AMRO = anti microbial resistant organisms
* Look at ALL infections initially, not just drug resistant ones.

**3/4/21**

Done:

* Imported and merged in micro data
* Switched to date to adt\_event\_time\_1 which seems to be available for all individuals

To Do with notes from meeting with Jeff:

* Not sure why non-covid graph is going through 2022…?
* Create similar graphs over time of mortality for covid patients who have and have not ever had a bacterial infection – try both based on blood only, and based on blood, resp and urine -------------------------
* Tabulate when bacterial infections took place – before, during after covid infection.
* Try the same but disaggregating by whether bacterial infection took place before, during or after covid infection (or covid infection related admission)
* Consider plots of mortality over time but considering only deaths of covid patients while they had covid – since what I have is simplified to patients who EVER had covid
  + Is there even a notable population of people who had covid, recovered, then subsequently died in our sample?
  + How to disaggregate: Discharge and readmission (they were fine), vs negative test following positive (tabulate and see if anybody)
* Basic tabulation of mortality among covid patients with/without bacterial infections (will be biased towards not having infection -> not dying based on length of time in hospital, I assume, because those who die more quickly won’t have time to get an infection)
* If I want to do a formal model for this first question, outcome would be death (or hospice too?), covariates would include bacterial infection (binary), time variable with a penalized spline to allow for non-linearity (?), interaction btw bacterial infection and time. 🡨-- Ultimately this would make sense.
* Anne-Catrin told him in the summer that rate of bacterial infection among covid patient swas low and most lung problems were bacterial pneumonia, but those who did have bacterial inf ewere twice as likely to die.

4/15/20

**Meeting with Jeff and Sasi:**

Could expand on #3 to look at different groups of patients

An appropriate was to organize this would be to look at predictants: survival (time to mortality or time to discharge), mortality (predictors), time in hospital, bacterial infections (can think of as modulators, can think of outcomes in their own right)

There’s a there there in survival time to mortality analysis…….

5/4/21

**Meeting with Jeff and Sasi:**

* A lot of covid-specific outcome topics have already been covered by Jason, AC and team.
* Will want to refocus efforts to focus more on AMRO and bacterial infection outcomes, rather than covid
* Shenglin and Heujin are already doing more AMRO focused work
* Tal and Jaime are building dynamic models
* Sasi says: it makes sense to do analyses that include both covid and AMRO
* We’ll want to make sure we don’t overlap too much with Huejin, also.
* MIND network: interested in treatment and resource use change
* Think about what questions to pose given the context of what we’ve seen in the database to this point.

5/6/21

Notes to self:

Work over the last week:

* Re-organizing R file. Worked on creating a better system for importing data, naming files, data cleaning, and merging (still some work on this)
* ID-ed some discrepancies in the data, put relevant notes in teams folder.
* Brainstormed research ideas considering notes from meeting above

To do on Monday:

* Continue to fix up R file a little bit, make standard for dates vs datetimes (datetimes, I guess? Retain full info as long as possible?
* Make graph using notes from Sasi – horizontal error bar of time in hospital/ covid diagnoses and bacterial infection results (only ~ 150 ppl, should be doable.)
* Look at notes from before about what comparisons AC would like to see, and start working on them.
* Quantify time in hospital in some different ways, as preliminary exploration for the questions about that
* Make a list with some questions for AC and Jason
* Do a little more brainstorming about what questions would be good to answer, and be prepared to talk about that at meeting
* Put questions, plots, etc in a powerpoint

**5/13/21**

Meeting with Jeff:

* Research Q to look at clarification:
  + Once diagnosed, how long are they there either until discharge or resolution of infection?
  + What are the things (predictors I guess) that indicate faster resolution?
  + Some considerations:
    - What counts as resolution?: could just be – first negative test, basically.
    - Need to think about different species
    - Likewise different locations of infection
    - Might have limiting returns bc of limited power once we disaggregate all of these.
* We still have incomplete both sensitivity data and culture data
* Some of the next steps for me to do:
  + Look at data by site and species categories.
  + We want to get a sense of, basically, how much data we have
  + Also, what’s the distribution of time spans for each
  + See if data is well enough kept to use it: we want to check about both consistency of data and sample size
  + Literature reading is a good idea both because it is the scholarly thing to do, and because it would be good to know if other have already done this same work, basically.
* Mind network: network of modelers working on nosocomial (hospital) transmission. Mostly focused on agent-based modelling, there will be some meta population modeling eventually too.
* Thoughts about dissertation research: need
* Eventually we will pull out some dynamical models in the next few weeks.

**5/14/21**

Done today:

* Continued working on cleaning up code so that it matches my new data cleaning structure. I still think this structure will probably save time and code, yay!
* Question: why do we have so many NAs for covid\_pos\_ever in the micro dataset when we merge in the adt data? We have fewer NAs in the patients dataset, so we must have a bunch of micro data for people who don't exist in the adt data. Maybe check if these all occurred earlier than the others – ie before micro testing began?

**5/19/21**

Notes from work today:

* Fixed issue where some ppl with covid had NA as their covid\_pos\_dt in the patients dataset.
* 2nd issue: adt\_long\_patients has missing deceased status for ~4000 patients – why?
  + Seems like adt\_long\_patients has a bunch of ppl (almost 3000) who have F for adt\_dataset, indicating that they don’t appear in the adt\_wide\_dataset which was used to create the patients dataset. Why?? Does adt\_wide not match adt\_long??
  + Adding section to troubleshoot discrepencies btw adt\_wide and adt\_long datasets
  + No obvious difference btw people in the adt\_wide and adt\_long
* May want to either ask about discrepancy btw adt\_wide and adt\_long, or just switch to using adt\_long only.

**5/24/21**

Ideas on how to merge:

* Merge micro and adt by mrn, then, filter for specimen collected date falling between adt date (entry date?) and discharge date. That should leave one obs (I think?) per micro test, paired with the adt visit that it occurs during. Then, discharge date should be the “end of stay” / end of time in hospital

MIND meeting:

* Claire’s work:
  + considered gram negative, gram positive, and fungus. There’s not much sample size for the fungus data, particularly for resistant
  + Used univariate criteria for inclusion in model
  + Interactions were not significant
  + Time in hospital: put in 4 categories (from entry to exit time..) quantiles to account for non-linearity
  + Pr> chisq criteria for model checking (look into this?)
  + AC commented: more resistance in men v women (typically?), gram positive: staph and entero- should be considered separately, but gram negative it’s ok to summarize. Don’t need to make a model with all organisms
* Anne-Catrin’s comments:
  + Large # of urine e-coli positive on admission, can often have pos culture contamination
  + Would be useful to look at covid vs non-covid
  + Generally blood results are robust, with the exception of staph-epi, which needs multiple cultures to diagnose. So, we should basically look at blood infections. Urine we should emphasize least, and resp can also be complicated by infection vs colonization issues.
  + Blood infections should have a clear test before discharge, so we can look at that for end of infection, but it’s a little complicated because sometimes they don’t send the tests.
  + They’ve done some work on duration of bacteremia, but the covid vs non-covid angle would be new.
  + She can send a paper (?)

**6/7/21**

Meeting with AC, Jason and team:

* Comments from AC:
  + She’s very confident that lapses of 3-7 days in positive tests would still be the same infection
  + For longer lapses, we’d need to look into whether it’s treatment failure or a new infection.
  + Type of bacteremia: staph infections usually treated for 6-8 weeks, other types for 7-14 days.
  + Break out diagnostic chart of lapses in infection by organism (done)
  + Staph epi: would be useful to see (done during meeting, could add to power point.)
  + Staph hominis and staph capitus are “coag negative” infections like staph epi, also could just be contamination. Staph epi -> can be skin colonization
  + We can talk more about where to put cutoff for “ambiguous” for each organism
  + Staph epi: would need to last for 3 days to be a real infection.
  + Look into: do these occur (not sure what she was saying here – positive tests, I guess?) while on treatment or not?

**6/10/21**

Meeting with Jeff:

* KS test for differences in distributions
* For pairs of bacterial infections and covid infections – look at questions I added to slides:
  + Second category: continuous stay or multiple inpatient stays? Ie, were they infected with covid in the hospital
  + Same for final group: same stay or not?
  + Plot distribution of time after covid identification that bacterial infection was identified – was it picked up, or just about order in which tests took place?

**6/15/21**

Youtube intro to bacteria: (<https://www.youtube.com/watch?v=2iysq1kkTSk>)

* Gram pos vs gram negative: different staining has to do with structural diff in cell membrane and wall. Negative: membrane, wall, then another membrane; positive: membrane, then wall
* Shapes: spirilla (spiral), bacilli (rod shaped), cocci (spherical)
* Motility: flagella
* Metabolic diversity: photo-autotroph, chemo-autotroph, photo-heterotroph, chemo-heterotroph. Bacteria exist in all four
* Diversity in oxygen needs: obligate aerobe, facultative anaerobe, obligate anaerobe.
* Bacterial reproduction: binary fission (not much diversity); bacterial conjugation: exchange plasmid piece through conjugation tube (sometimes get slightly different version of plasmid, leading to diversity); bacterial transduction (movement of bacterial dna from one bacterium to another via a viral vector)
* Plasmids: extra non-essential pieces of dna

**6/16/21**

MIND grantee meeting:

* Seth Blumberg presentation:
  + Francis Coll worked on transmission of MRSA in England
  + “accurate modeling considers both strength and heterogeneity of infection
  + Targeting superspreaders….
  + At NYU theyre sequencing all colonized isolates
  + Someone mentioned there had been reports of reduced MRSA incidence/transmission during the pandemic
* Praachi Das: covid transmission model
  + Aggregated movement data to the ZCTA level using “Safe Graph” data from cell phones
  + Population density was not significantly associated with case risk
  + Someone mentioned that the CMI (not sure of acronym) delphi (adelphi?) group had a catalogue of publicly available data that would be useful.

**6/17/21**

MIND grantee meeting day 2:

* David Rasmussen: SARS-CoV-2 fitness variations
  + Future work: prediction of covid strains.
  + Check out pre-print!: <https://www.biorxiv.org/content/10.1101/2020.12.14.422739v2>
* Gregory Madden:
  + Decreased MRSA rates during covid
  + Did NOT seem to have to do with changes in network, so must have had to do with hygiene interventions, such as: modest improvements in hand hygiene, masks, certain patients are thought to produce a cloud of aerosolized MRSA, so could have helped with that
  + To look up: NHSN guidelines
  + Daniel Sewell – asked a question about visiting policies
* Tal Robin:
  + Took a bunch of screenshots. Take a look at these!
  + Suzanne Lenhart – aseked a bunch of questions about equations
* Rachel Slayton – organizer person
* Ben Haaland etc (Damon, Michael?): not sure exactly, study in hospital? (duh, they all were), computational advantages of not using mcmc.

**6/21/21**

MIND meeting:

* To do for next time (from Jason): days from covid to bacterial infection without three likely contaminants (ie, staph epi, staph hominis, staph capitis)
* AC will want to break this down further later
* He thinks infection on day 0 would probably be staph epi
* No real reason to think that covid would prolong bacterial infections. Focus instead on whether bacterial infections would prolong hospital stays for covid patients

**6/24/21**

Meeting with Jeff:

* MIND work:
  + Quantification of: how much longer were they in the hospital? What factors are associated with that length?
  + Does it vary by pathogen?
  + Covid, while a short infection, wasn’t putting people in hospitals for short amounts of times
  + Eventually look at: are they more likely to have cytokine storm, etc
  + Start simply: look at lengths of stay for covid patients with vs without bacterial infections, Use anova test or KS test to compare. THEN, look at resistant vs non-resistant infections. THEN, look at specific organisms. THEN, consider building out a model that predicts duration of time in hospital using factors like infection, pre-existing conditions, etc, to account for some of the confounding. But start out by plotting and doing tests – anova, ks….
  + Eventually might want to do survival analysis (I pointed out – Sasi may already be doing this)
* Future reading and research:
  + He’ll invite me to weekly meetings. Next call is Wednesday at 2:00.
  + Start reading: Utah and Iowa groups in MIND, also NCSU, Mark Lipsitch at Harvard. Look up MIND network and papers. M Samore is the PI at Utah, also maybe Vittoria Kalitza (sp?) who is somewhere in Paris.
  + While reading, take notes! Write my own summaries, organize them. Organize by type of study – math modeling, inference, exploratory, tech, etc. Consider study design and methods.
  + Modeling text book recommendation: Keeling and Rohani.
  + Bio textbook recommendation – intro bio – Keaton and Gould (could take out of library)

**6/30/21**

MIND weekly meeting:

* Gary Lin (I think)
  + Something about geographic mobility and covid spread using safegraph
  + Knowledge gaps/opportunities include: Movement around low income neighborhoods, synthetic populations (cool!), mentioned area deprivation index (ADI) scale of how disadvantaged a population is.
  + Chris Hoover – made some comment, similar work I guess?
* Theresa Sheets: 4th year grad student at U Utah in math
  + Covid transmission in households with/without children
  + How do children change transmission dynamics?
  + Data from Jan-Apr 2021
  + HH with children: lower within household transmission, slightly lower importation

**7/17 ish**

Call with Tory about TAing:

* Keep in mind: extent to which I want/need real training
* PhDs in climate and health: her, Sebastian, Maggie, Misbath
* 3rd year: good time to double up on TAing
* TAing Jeff’s class -a good amount of work. Maybe more manageable in person
* GIS: easiest to TA, motoriously.
* Greg or Darby’s classes – also probably pretty easy. Paper writing
* Consider taking my own classes pass/fail
* She took: epi 5 (or others did?), data science

Call with Jeff about TAing and MIND analyses

* For his class: typically student doesn’t teach. If there was a topic that had gelled for me I could. Student grades HWs, NOT the exams though
* Other fall options: wait on GIS.
* Modelling notes: matching – would have to make pairs in some kind of “heuristic way”, he’s also interested in is there a difference in length of time in hosp among covid paitents with bacterial infections vs non covid patients with bacterial infections
* Survival analyses: look at what factors contribute to time in hospital
* Could also estimate from a mathematical model…?

**7/18/21 ish**

Notes about propensity score/risk-set matching

* Things to include in score: age, sex, race or combined race ethnicity, zipcode, insurance type, bmi?, site (hospital) procedures? Time varying? Problem list, intubation and extubation, adt\_care\_level
* Steps: 1) simpler model with all time fixed covars:
  + Set up data – cols include mrn, treatment time (or NA), time, age at admission etc etc
  + Cox ph model, then sue this model to generate hazards that will be used as propensity score
  + Compare values between matched and unmatched groups
* 2) More complex model – introduce time-varying covariates.
* By my “covid admission” definition, people with multiple covid admissions are considered to be hospitalized for the whole time from the start of one to the end of the final one
* Some people are rapidly readmitted – are these basically the same admissions (follow up from meeting – Jason says yes, see adt\_summary)
* Idea about how to handle these: make rule about time distance btw admissions:
  + Make variable of distance since last admission
  + Type of last admission? Need to somehow show that covid admission was at the start of the chain
  + Consider covid admission if short time since original covid admit.
* Follow up note: Jason says that I should be looking at the encounter ID to figure this out.

**7/19/21**

MIND meeting:

* Jason and AC recommend ICU stay and BMI as covariates
* Jason: adt summary groups things by encounter ID – good way to differentiate btw visits
* Interesting comment – there are many infections we treat that don’t have any positive cultures, actually.
* They recommend I talk to Monica or Brian who have done propensity score matching in the past
* AC will email around paper about propensity score matching

**7/21/21**

Meeting with Claire:

* Review parsing\_sensitivities document that they sent around (maybe)
* She’s doing her summer internship at a research institute at USC

Meeting with Jeff:

* Sneha is doing a cox hazards model – maybe talk to her?
* He recommends a numerical methods class
* “whole science” – how do you represent equations numerically? Problems “arise if you don’t do certain things”. Look into what class is offered and what language they use.
* Other classes: could consider other quant methods, bio class – maybe in E3B, or epi class (I kinda added that).