A biostatistician's perpective on COVID-19 research in NYC

Usha Govindarajulu, Phd

Center for Biostatistics
Icahn School of Medicine at Mount Sinai
New York, NY

Outline

1. How it all started
2. Data was a mess
3. Being on a team with data scientists
4. Guiding the research
5. Leading the way
6. Results
7. Conclusions

How it all started

- In late February, one of the first COVID patients in NYC came to our hospital
  - Slowly slowly, the city started becoming emptier and emptier….
  - By mid-March, we were given the option to work from home.

- Around late March, after having been home almost two weeks, we were told based on our All Hands on Deck policy, would have to be redeployed to help on the hospital side

- Subsequently I was redeployed to help in Covid-19 research

- I was redeployed to join a team to do research.
  - Redeployed along with a colleague of mine
  - Team consisted of some data scientists and headed by one
  - We were told to handle requests that would need a fast turnaround on analyses
  - We were asked to report to one of the data scientists who would handle the statistical requests and provide the data for us.

Commanding our time

- Were told would need to handle requests at a fast pace
- However, one request came in that took over our time
  - Analyzing a particular drug to understand if it was working amongst Covid patients
  - The goal was to have a fast turnaround so that physicians could make determinations for its use
  - We started working over a weekend and were from then on working days and nights on this one project which I will discuss in more detail

The data was a mess

- The data was being obtained from machine learning methods from the EPIC online medical records system by this particular group of data scientists.
  - Sometimes this was fairly accurate and sometimes it was not.
  - Then it was being put into a Microsoft Teams area for us to download
- I have never merged so many files together in my life!
- It became a daily routine, download daily new updated data and re-merge all the data
- We were told to do it this way and create our own analytic dataset

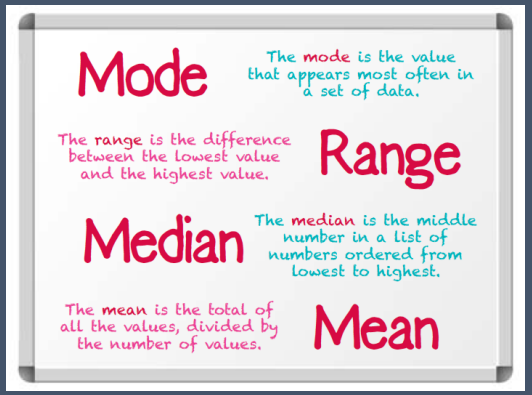
Messy data

- Once we got into the data and trying to understand patterns, we started realizing that we would have to change our set point.

- There was a lot of missing data, especially in measures like height and weight and lab measures
  - Sometimes they had no time to get these measures at all in the COVID unit

  - I did not think the data was missing at random due to the way it was missing
  - But the data scientist in charge thought we should be using multiple imputation
    - I inserted myself in a meeting that I did not believe this would be legitimate given the pattern of the missingness, not missing at random

Summary measures were important



- But there was pressure to jump into analyses and make sense of the data at hand which was constantly changing every single day and numbers were going up (Covid positive patients and number of deaths)
- The investigator, an *infectious disease* doctor, said she was looking for answers fast
- The project dragged out for 4 weeks where we looked at every nuance possible in terms of modeling survival
- We needed a way to deal with these observational data to balance out some confounding between groups
    - Which ended up significantly improving the results

Breaking away from the data scientists

- We started working more closely with the doctor and broke away from the data scientists so that we could better understand her requests and analyze the data faster
  - As we worked with her more closely, we really had a better feel for what was happening with the COVID patients in the hospital
  - We started realizing how important it was to look at *timing*
    - how different it would be if given before ventilation vs after ventilation
  - The things we started finding started influencing some of decisions about how the drug was being used for floor and ICU patients as we began understanding the timing better
  - Results we found would be communicated to her chief of infectious disease who reported directly to the president of the hospital
    - And they were desperately looking for cures from repurposed drugs (hospital and public pressure in general)

Days turned into nights…nights into weeks!

- We would speak with her many times during the day and even as late as 11:30 pm at night!
- Would be working long days, eating dinner at 11 pm
- Friends would call and wonder why I was still working at 9 pm
- Weekends were taken up with the work and had to balance teaching a class at the same time as well as any other requests that came up.
- It was challenging for all the biostatisticians who were assisting investigators all across the board!

Data was not being cleaned

- We then started learning later our analyses were potentially compromised
    - I started learning that data we were trusting from their daily dump of data was not at all accurate when we compared ventilator dates from one set of data to another set of data which they had created
- Finally, an old fashioned approach of having medical students review medical records became a way to validate data that was being pulled from their machine learning methods
- They would obtain the data via their methods but had no process in place for actually cleaning those data or validating it

Many other COVID research needs

- So my “talents” had to be redeployed elsewhere in the hospital even though that particular drug study wasn’t finished
- And there was many other groups in need of survival analysis at our same hospital also doing COVID research
- I then got asked to be with Anesthesiology and another group. Separately, to assist each group with the anti-coagulation protocol analysis amongst different subsets of patients
    - Will next describe that protocol….

Anti-coagulation (AC) protocol:

- A system-wide AC protocol was developed at our hospital after it was discovered that COVID-19 is often associated with increased coagulation of the blood

  - Anticoagulants can help reduce or prevent potential events like deep vein thrombosis, pulmonary embolism and stroke

  - And furthermore could help prevent multi-organ failure

  - Patients were eligible for treatment with therapeutic anticoagulation if they did not have any of the following contraindications:

    - active bleeding or a platelet count <50x 10^3/L on ICU admission
    - a history of heparin induced thrombocytopenia,
    - or a history of stroke within the past year.

Things became busy yet again

- There was a request from the anesthesiologists:
    - They had been repurposed to work in the ICU COVID unit (since no elective surgeries were occurring during that time)
    - They wanted to get out timely results on the AC protocol that had been adopted for patients, mainly for their focus on the ICU patients.
- The pressure was back on just like in the drug study with which I had been involved just prior to that.
    - Of course then I worried about the data
        - Would it be just as messy or better?

The data was neater but still had issues

- I thought, well at least I received what seemed like a neatly organized data set in Excel
  - little did I know what was yet to come……
- They changed and updated the data almost 20 times thereafter!
- Problems came anew:
  - There were data entry errors from a tired resident
  - incorrect information from charts.
  - The main PI ended up always being in the ICU, especially the weekend that we needed to finish the paper.

I could have lost my mind

- Just before we were going to finish the paper and analyses
  - found out that someone else in the hospital had published his results on the anti-coagulation protocol with system-wide data whereas we had been limited to the main Mount Sinai hospital

- Suddenly our work didn't seem as novel or may not be perceived as such

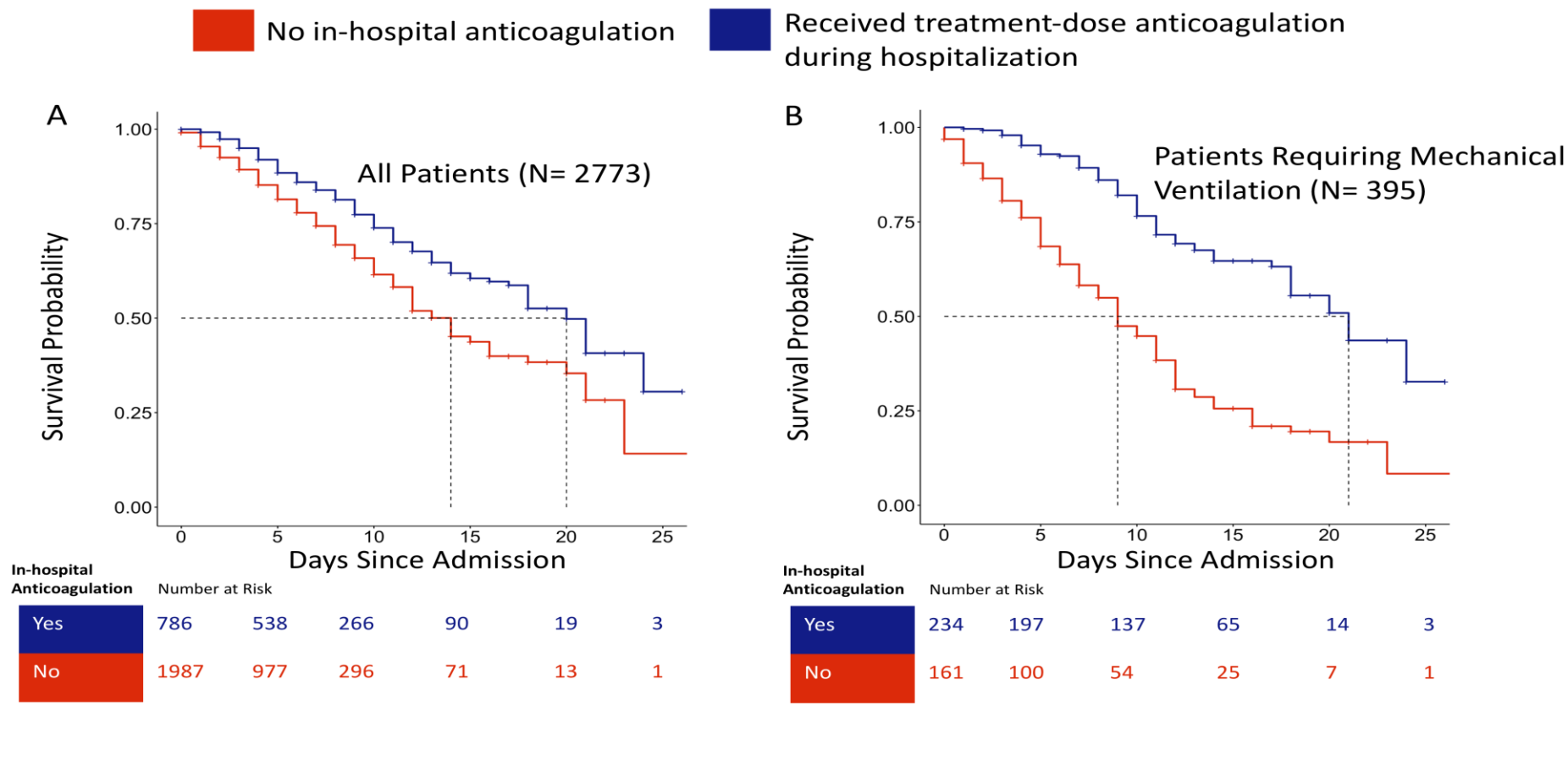
  - Even though we had IRB approval to do our work and there was general agreement about division of labor

  - How were we to move forward with what we had?

The difference in analyses

- The lead physician at the main hospital, had published a letter to the editor in *JACC, Journal of the American College of Cardiology*
- He had done a system-wide analyses of the AC protocol amongst all patients at Mount Sinai and even included an analysis on mechanically ventilated patients or those who had gone to the ICU
    - Our analyses were on those who went to the ICU and the effect of the AC protocol on mortality
        - But ours were only for those at Mount Sinai main hospital
        - These were patients with whom the anesthesiologists had directly taken care and had been directly involved so they felt close to the data and had more granularity

They had overall patients and those in the ICU



Association of Treatment Dose Anticoagulation with In-Hospital Survival Among Hospitalized Patients with COVID-19
Ishan Paranjpe, Valentin Fuster, Anuradha Lala, Adam Russak, Benjamin S. Glicksberg, Matthew A. Levin, Alexander W. Charney, Jagat Narula, Zahi A. Fayad, Emilia Bagiella, Shan Zhao, Girish N. Nadkarni
J Am Coll Cardiol. 2020 May 06. Epublished DOI:10.1016/j.jacc.2020.05.001

Their results:

- In patients who required mechanical ventilation:
  - (N=395), in-hospital mortality was 29.1% with a median survival of 21 days for those treated with AC as compared to 62.7% with a median survival of 9 days in patients who did not receive AC .
- In a multivariate proportional hazards model
  - Found longer duration of AC treatment associated with a reduced risk of mortality for mechanically ventilated
  - adjusted HR of 0.86 per day, 95% confidence interval, 0.82-0.89, $p<0.001$

Our methods

- Data source:
  - retrospective review of all COVID-19 positive adult (age > 18) patients admitted to any of the ICU at The Mount Sinai Hospital between March 1, 2020 and April 11, 2020.

  - All patients were confirmed to have COVID-19 by PCR from nasopharyngeal swab.

  - During the study period, 279 patients were admitted to the ICU and required mechanical ventilation, of which 34 died within five days of admission.

  - 244 patients were included in the analysis:
    - 161 received therapeutic anticoagulation and 83 received prophylactic anticoagulation

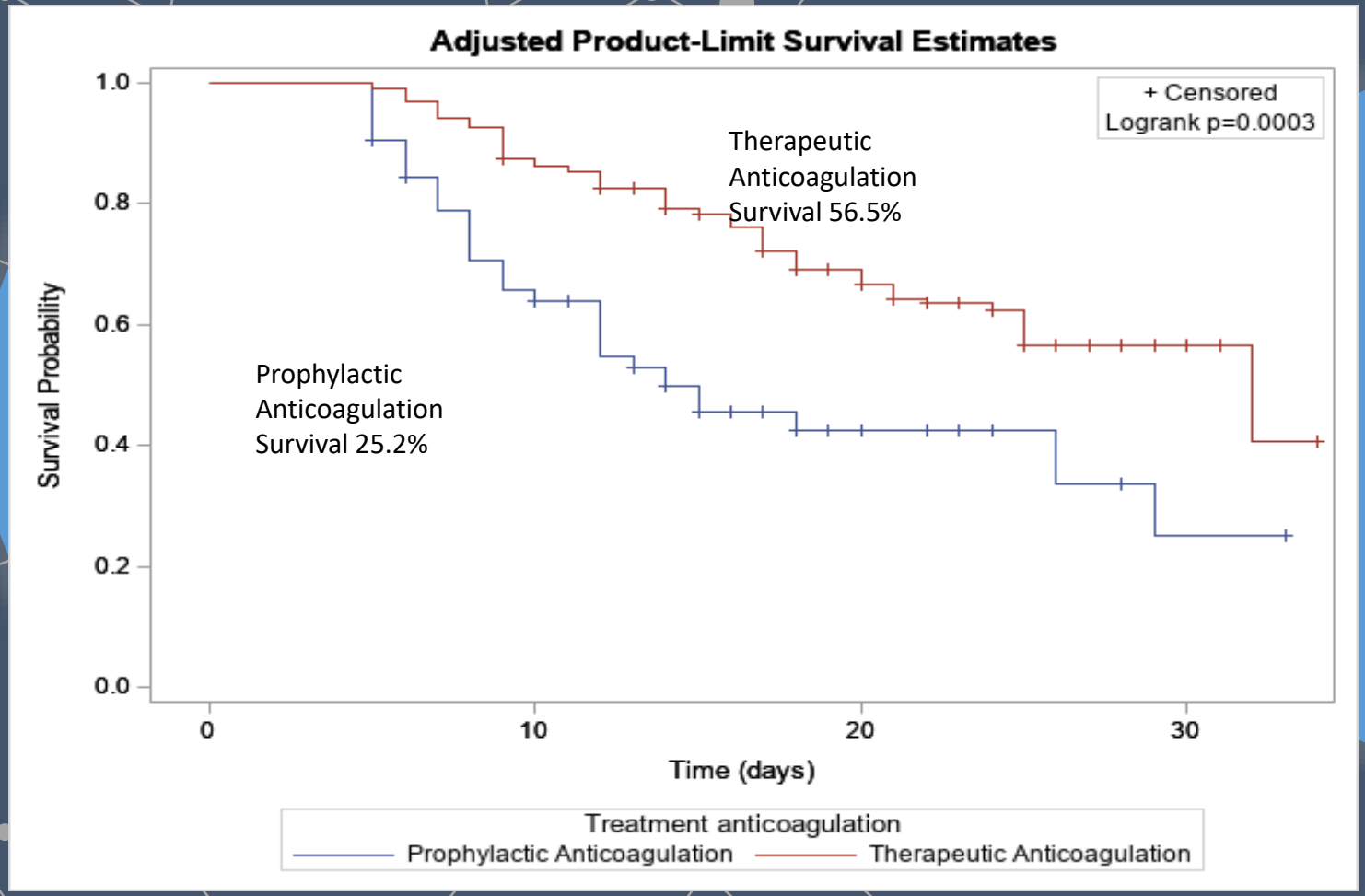
Methods (continued)

- We were interested in a survival analysis of time from ICU admission to date of death or date of discharge
  - to compare those on the therapeutic AC protocol to those on prophylaxis
  - Employed right censoring and truncation of those who died right away.
- Initially we observed the Kaplan-Meier plot and also used propensity-weighted survival for the KM
- We were then also interested in running a Cox proportional hazards regression model to adjust for potential confounders
- The other group had not done anything to balance the groups in their analyses.

Our baseline analyses

- Bleeding complications were similar between the two groups ($p<0.07$) as determined by chi-square test
  - 31.7%: therapeutic and 20.5%: prophylaxis
- Majority of patients in the treatment group received a combination of therapeutic enoxaparin and heparin during their ICU course, rather than a single agent.
- On average, patients received therapeutic anticoagulation for 17 days during their hospitalization.
- Patients in both groups had similar comorbidities including history of chronic kidney disease.

Our PS weighted Kaplan-Meier plot



Our results

- To adjust for baseline differences between the two groups, PS score-weighted Kaplan Meier plot (**previous slide)** was generated
-
    - showed a survival advantage for the anticoagulation treated group (57% vs. 25% survived to 35 days, p <0.001; Log-rank test).

- Patients in the TA group had a longer length of stay with 18 days in the ICU compared to 11 days in the prophylactic group
    - may have reflected increased survival.

- A univariate Cox proportional hazards regression was performed with the group
    - treatment reduced mortality: HR of 0.425 for therapeutic compared to prophylaxis with a 95% C.I: (0.23, 0.78).

Our results

- Then ran a multivariate Cox proportional hazards regression model for AC group with the PS weights
  - adjusted by these covariates:
    - age, gender, history of chronic kidney disease, time-varying creatinine, asthma, concurrent therapies (corticosteroids, tocilizumab), lactate, baseline SOFA score, and time from intubation day
- In the adjusted Cox model, therapeutic anticoagulation for at least five days reduced the rate of death even further [HR 0.209, 95% CI (0.10, 0.46), p <0.001].
- We believe that these findings suggest that the initiation of therapeutic anticoagulation for critically ill patients with COVID-19 who require mechanical ventilation is a beneficial intervention.

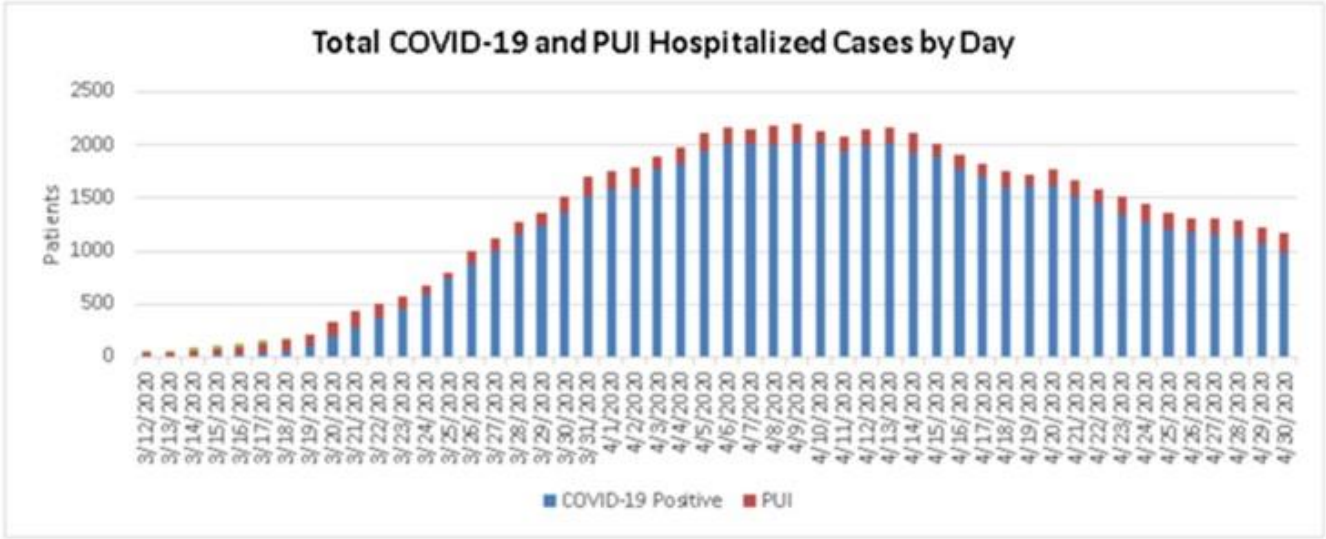
We were cool..

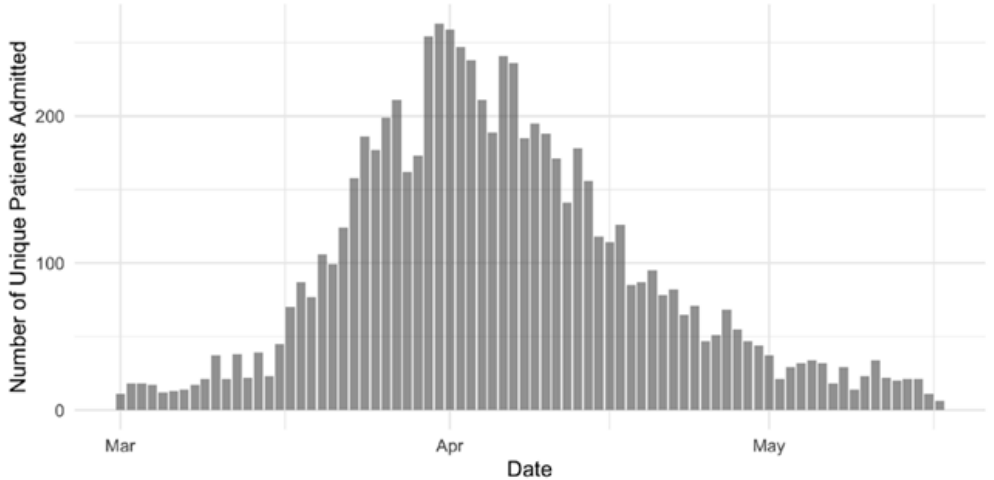
- We stayed institution friendly and let the author of the other paper be part of our paper
- He has been helping for this to be in press in an external journal
- So this is moving forward and continues
- Also his and our research affected what is done at the institution level with the AC protocol for COVID patients in the ICU besides those not admitted to the ICU

The other group…..

- Had opportunity to look at floor patients on the AC protocol
- They had laid out the groundwork, obtained well-curated data after learning from mistakes from drug study of which I had been part of before and they cleaned the data as well
- However they tried to look at system-wide data which now was under the control of the physician chief-of-medicine
- They are trying to figure out what else to do with their analyses…..

Some current hospital statistics earlier and later……





Conclusions

- Yes we were lucky to have a boatload of patient level data from our own hospitals because we are in the epicenter of the COVID-19 pandemic
- handling this messy data and being able to do well balanced analyses have been a challenge for all of us in our group
- Would have been nice to correct for measurement error…didn't even have time to think about that honestly!
- There was literally no time to think…..just had to get results out right away
- These were the challenges we faced and still face
- Will moving on to assist in leading a registry of COVID patients with data collection and future analyses

Discussions

- How best to get good analyses out in a timely fashion?
- How to have consensus among different groups in one institution doing research so there is no overlap?
- How to get best data driven results with best statistical methods?
- How to avoid the messy data trap?


Thank you all for taking the time out of your busy day to listen to this talk!
Stay safe!