



Machine Learning-Enabled Personalized Medicine for Rapidly Evolving Infectious Diseases

UNC-CH: [Jonathan Schisler](#), Matthew Wolfgang, Robert Hagan,
& Benjamin Sines

SAS: [Scott McClain](#), Xi Jiang, Ella Pavlechko, Julia Florou-
Moreno, Zhaleh Hosseini, & Varunraj Valsaraj

Rejuvenate Biomed: Gabriel Risa

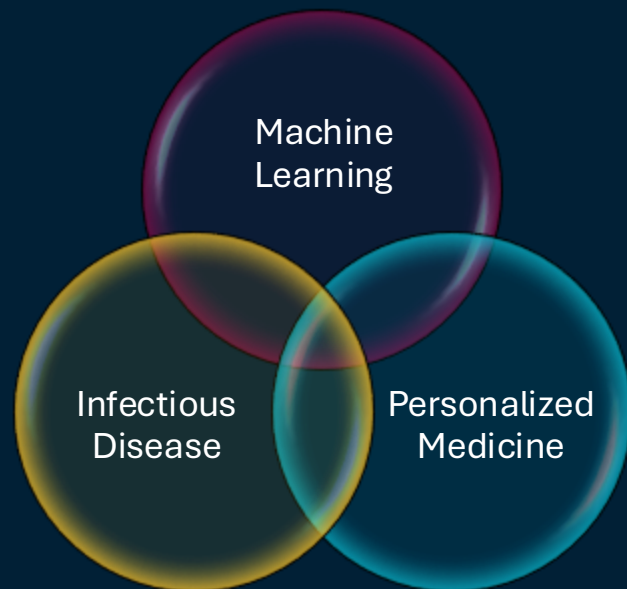
From Data to Decisions

Introduction

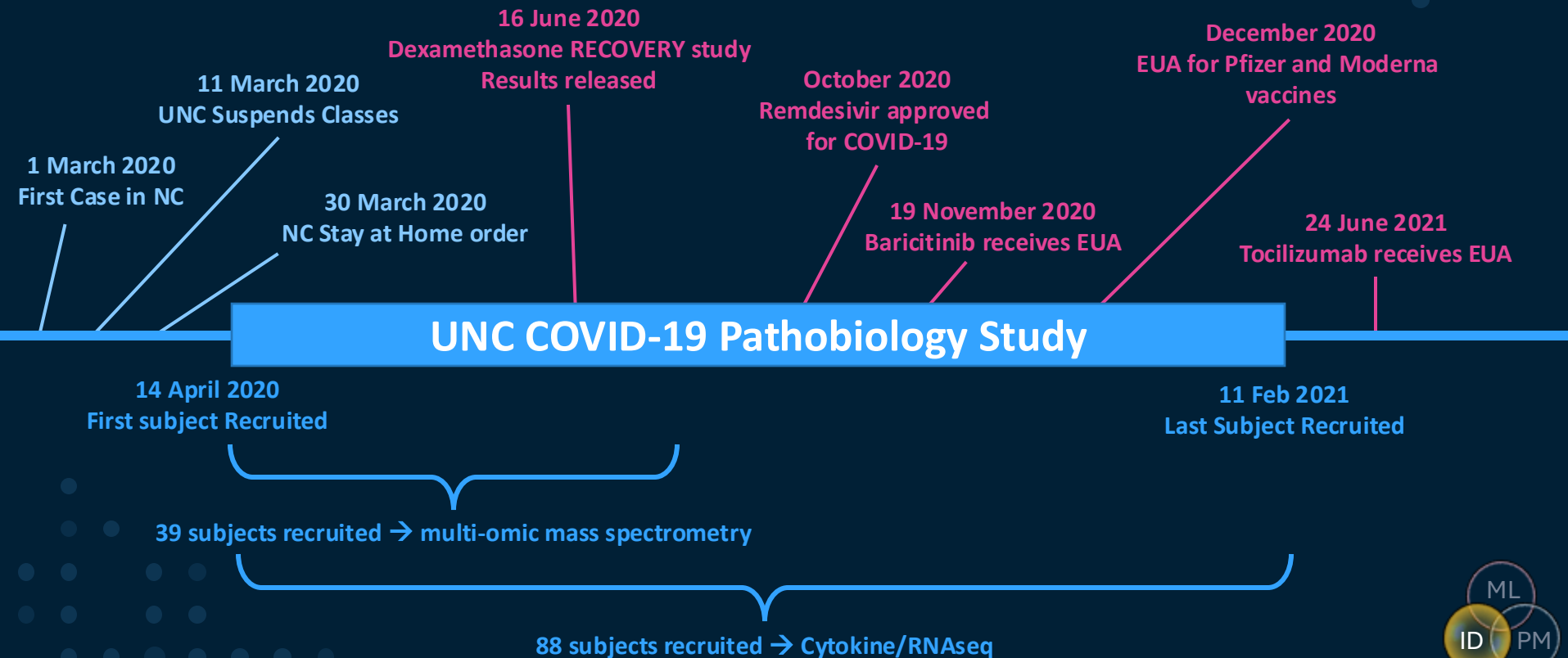
- **Challenge:** Urgent need for personalized therapies in infection-related respiratory symptoms
- **Focus:** ICU patients with respiratory failure during first two years of COVID-19 pandemic
- **Approach:** Integration of EHR data with molecular insights using machine learning
- **Goal:** Align patients with specific, data-driven therapies for improved outcomes

Key Components

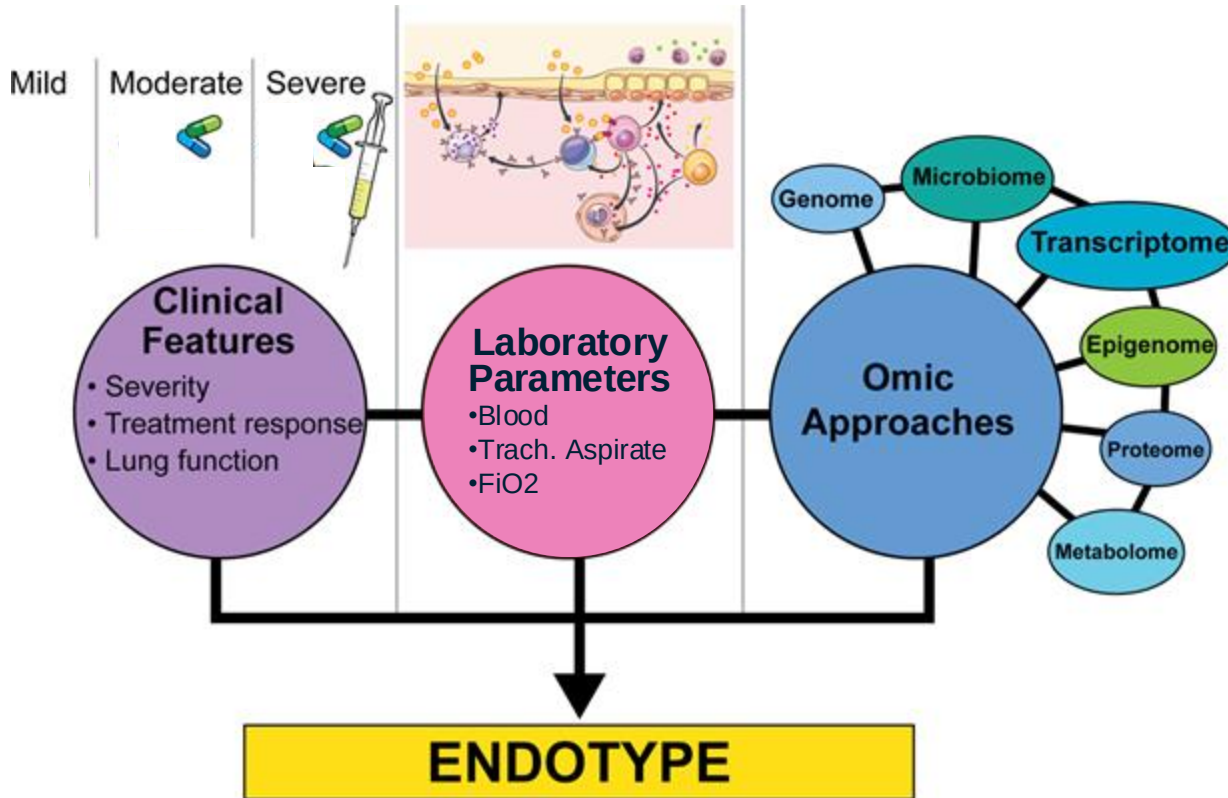
- Distinct molecular endotypes from molecular data from the **LUNG**
- ML methodologies for *real-time* analytical predictions
- Repeatable analytic solution for clinical trial success



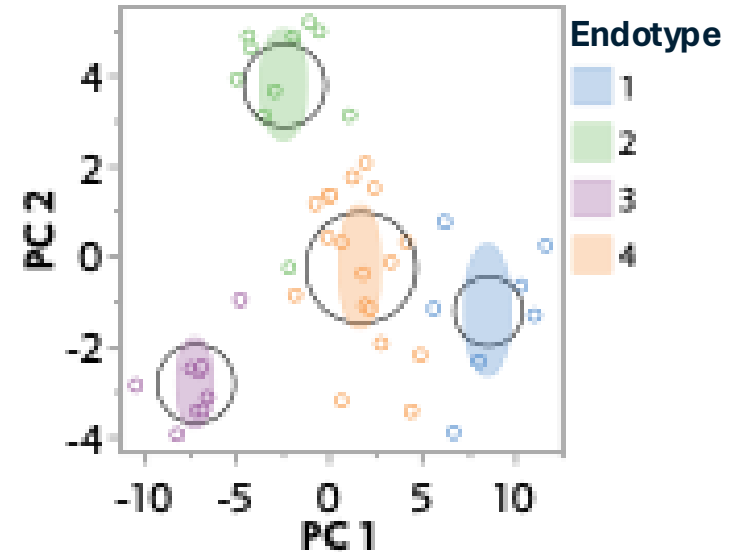
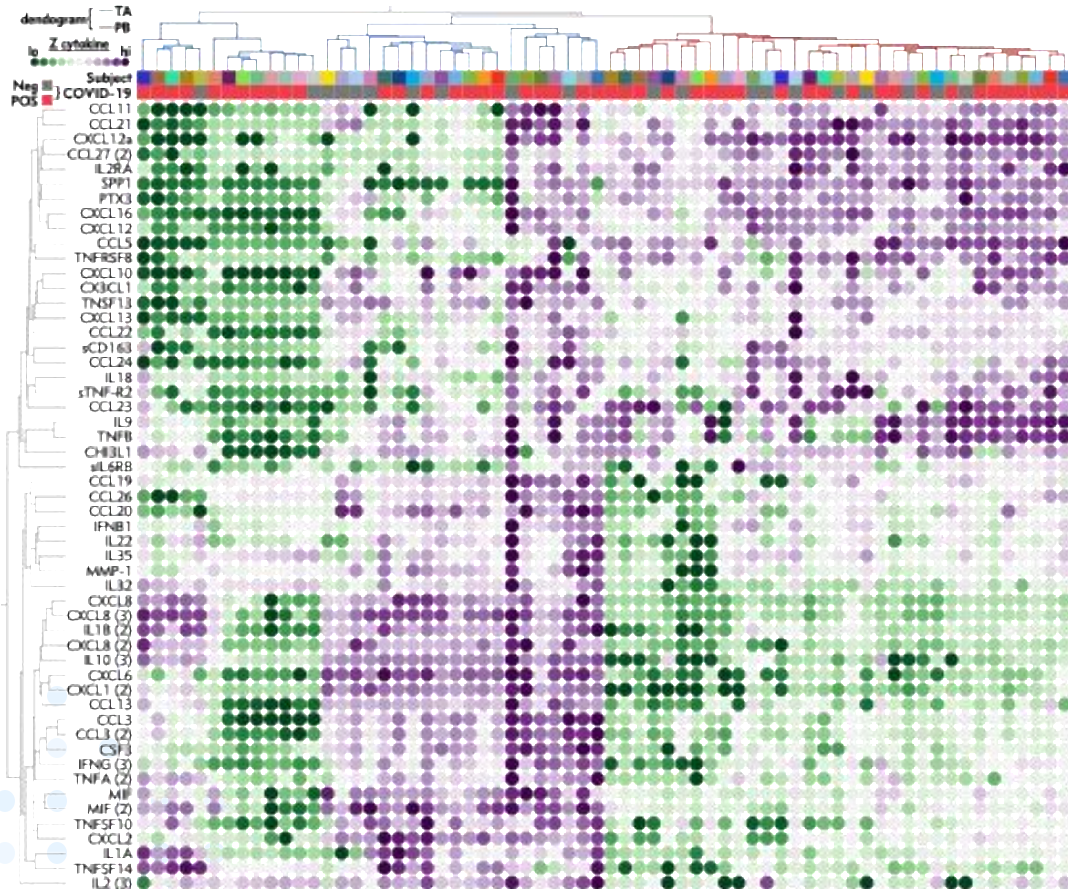
Clinical data source: Pathobiology of COVID-19



Patient & Molecular Endotyping



Compartment-aware Molecular Endotyping



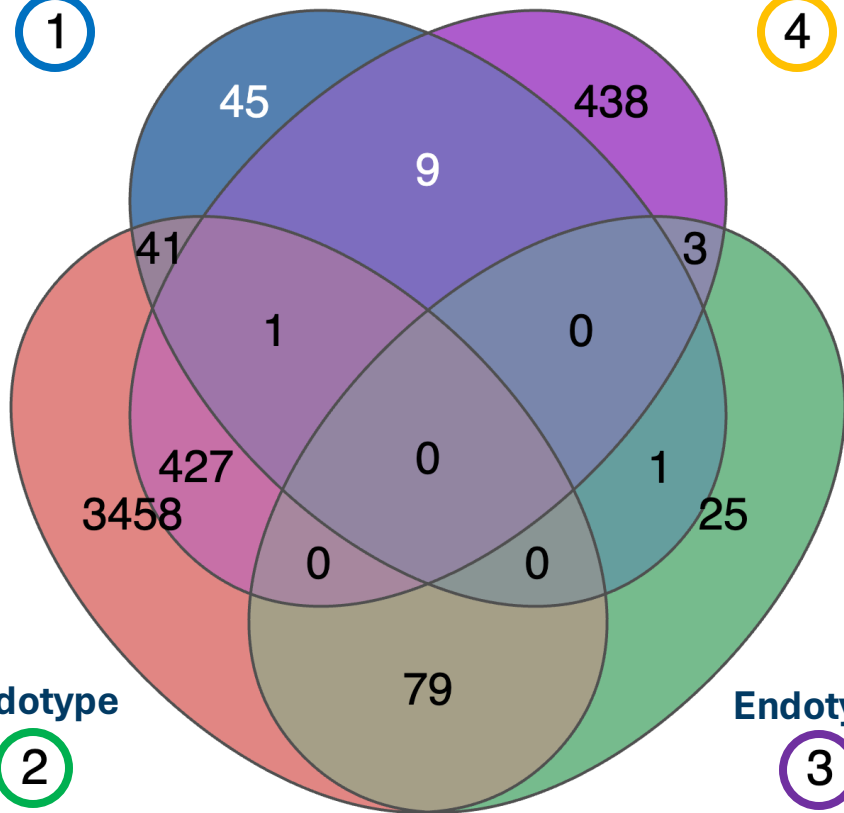
RNA-based Knowledge Graphs to Predict Drug Responsiveness

Endotype

①

Endotype

④



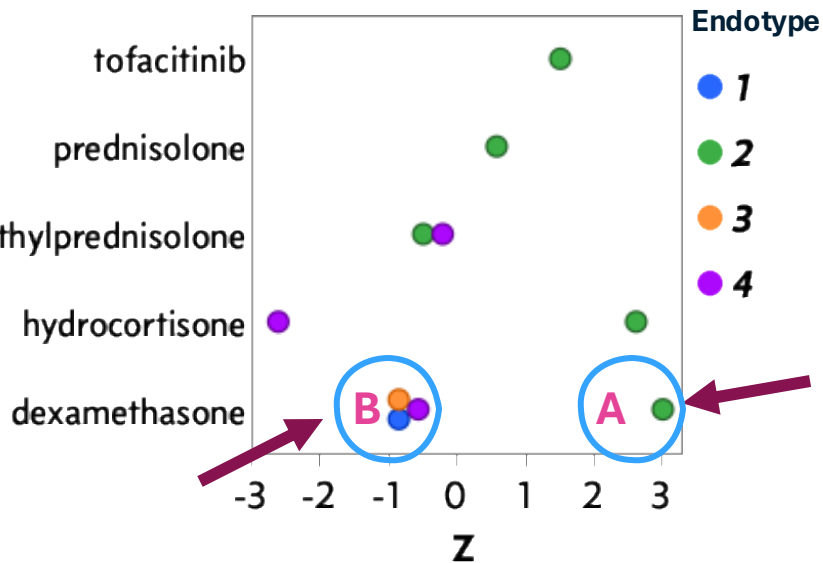
Endotype

②

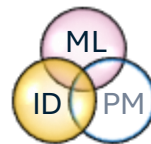
Endotype

③

Drug

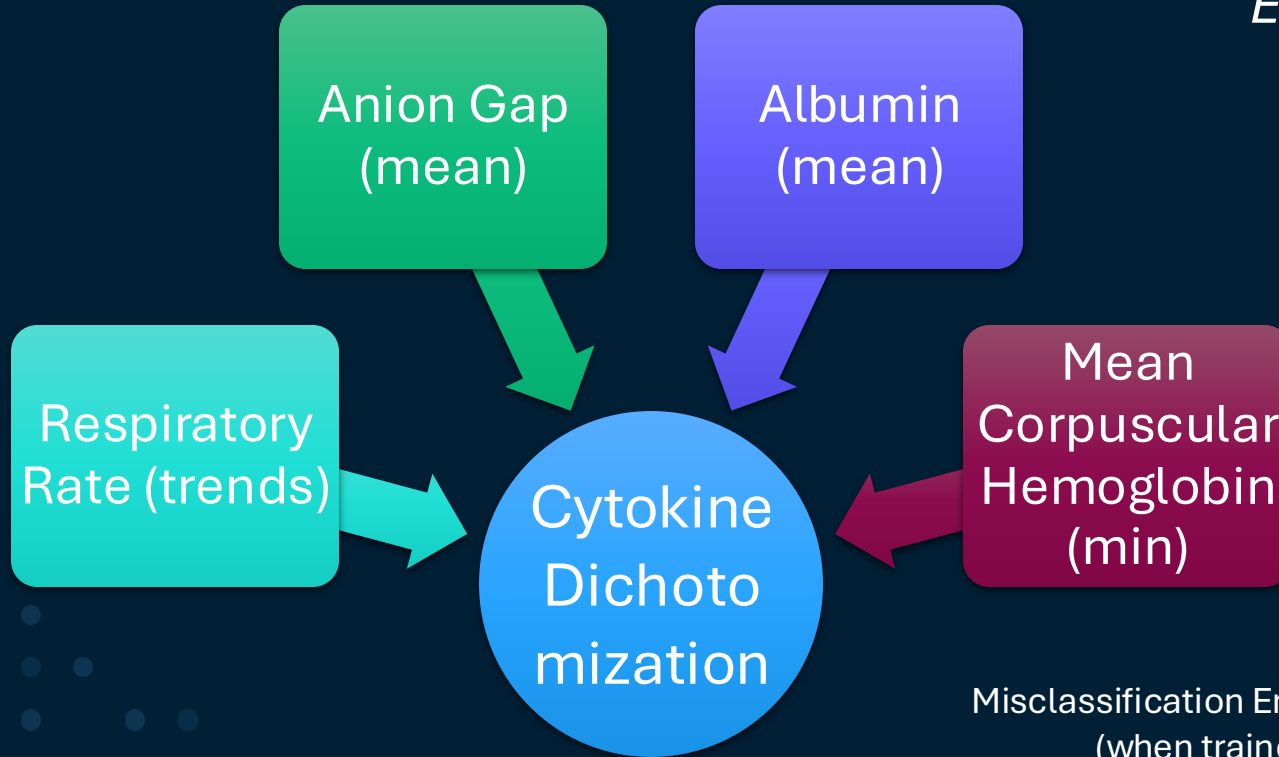


Dichotomize endotypes A and B



Clinically Informed Bayesian network

EHR43, Day -1 & 0, 1-fold



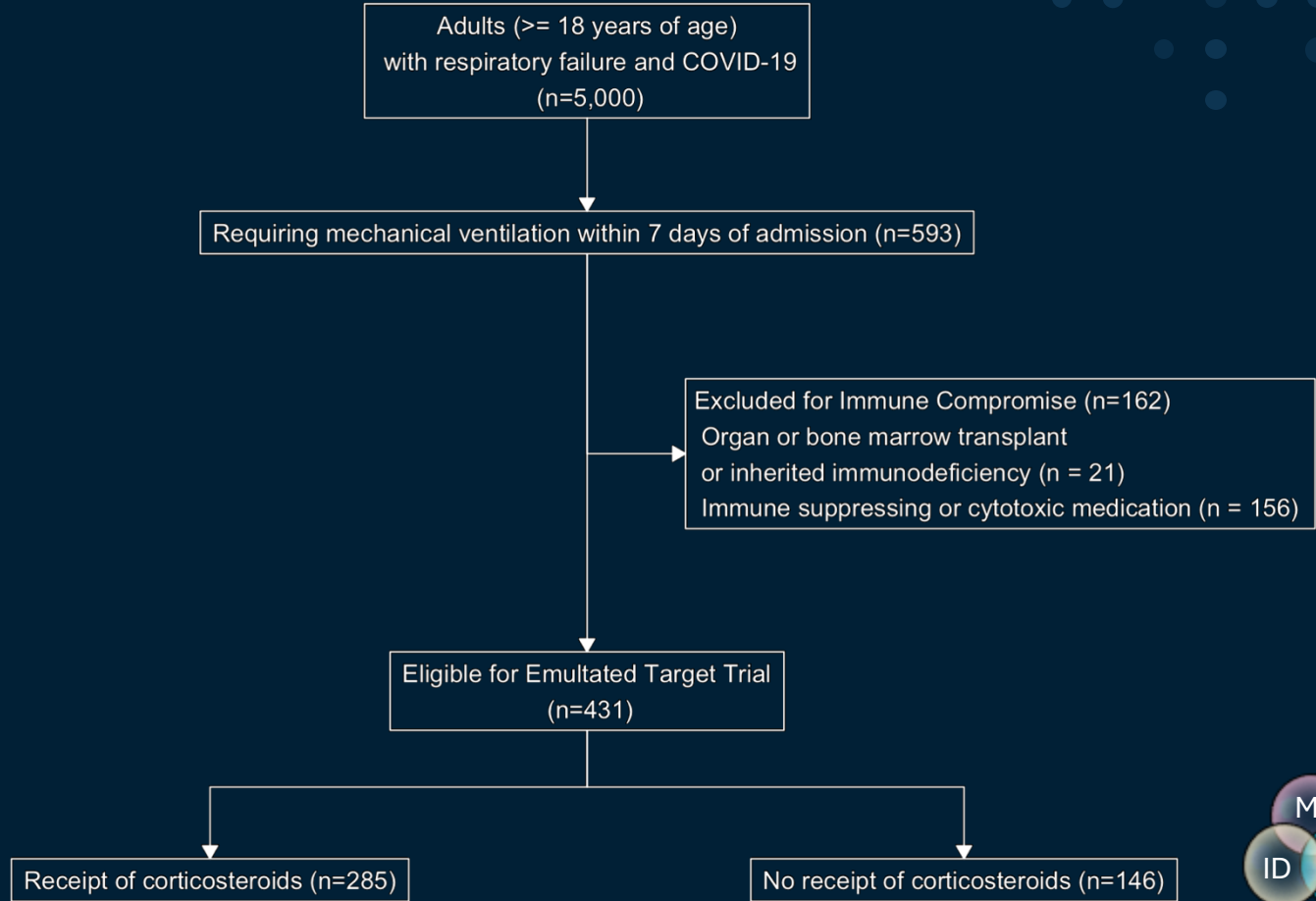
Misclassification Error Rate
(when trained on all
observations):

0.047



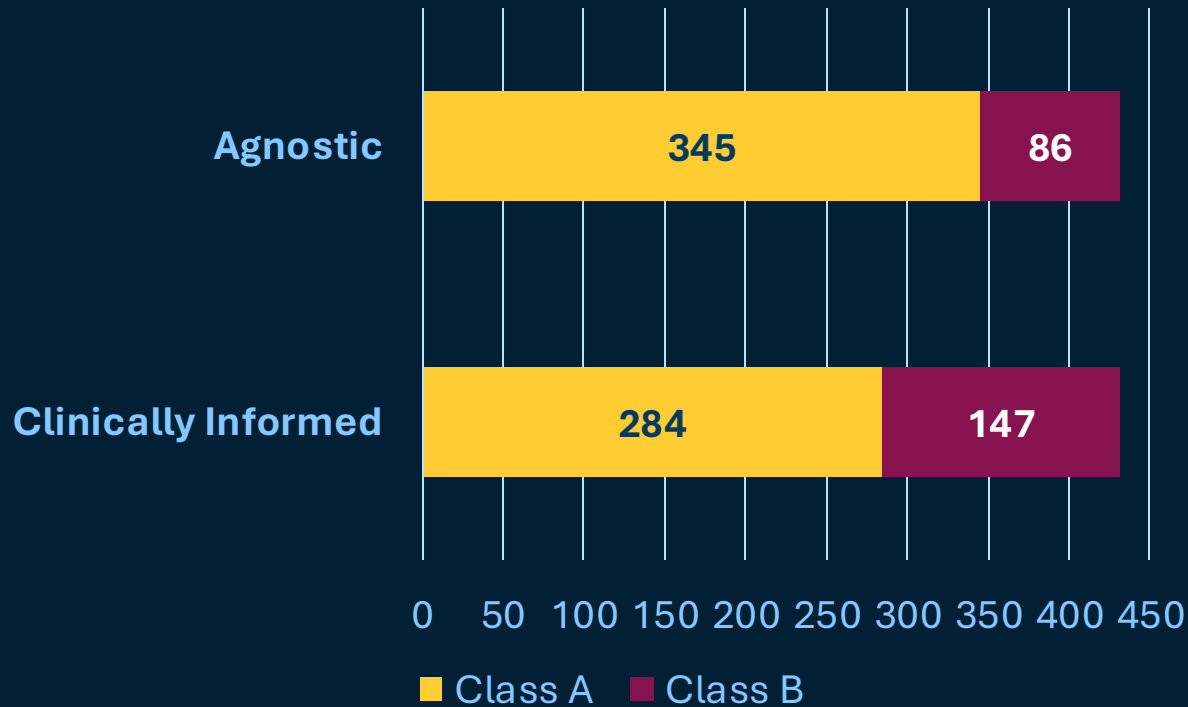
Ingestion of EHR data and Feature Engineering

- All UNC admissions from 1/1/2020 to 12/31/2022 with hypoxemia and positive SARS-CoV-2 PCR
- Processed through feature engineering pipeline



Bayesian network Scored $P(\text{class}) > 0.5$

Number of Patients per Predicted Class



Bayesian Networks Scored

P(classification) thresholds	CI	CA	Overlap	Concordance
P(class) > 0.5	431	431	431	352 (82%)
P(class) > 0.6	356	351	342	339 (99%)
P(class) > 0.7	355	344	336	336 (100%)
P(class) > 0.75	149	344	138	138 (100%)
P(class) > 0.8	43	344	37	37 (100%)

Primary Outcome

Untreated, the probability of death in the next 28 days is 0.472

Treated, the probability of death in the next 28 days is 0.434

28 patients were censored due on transfer to another hospital.

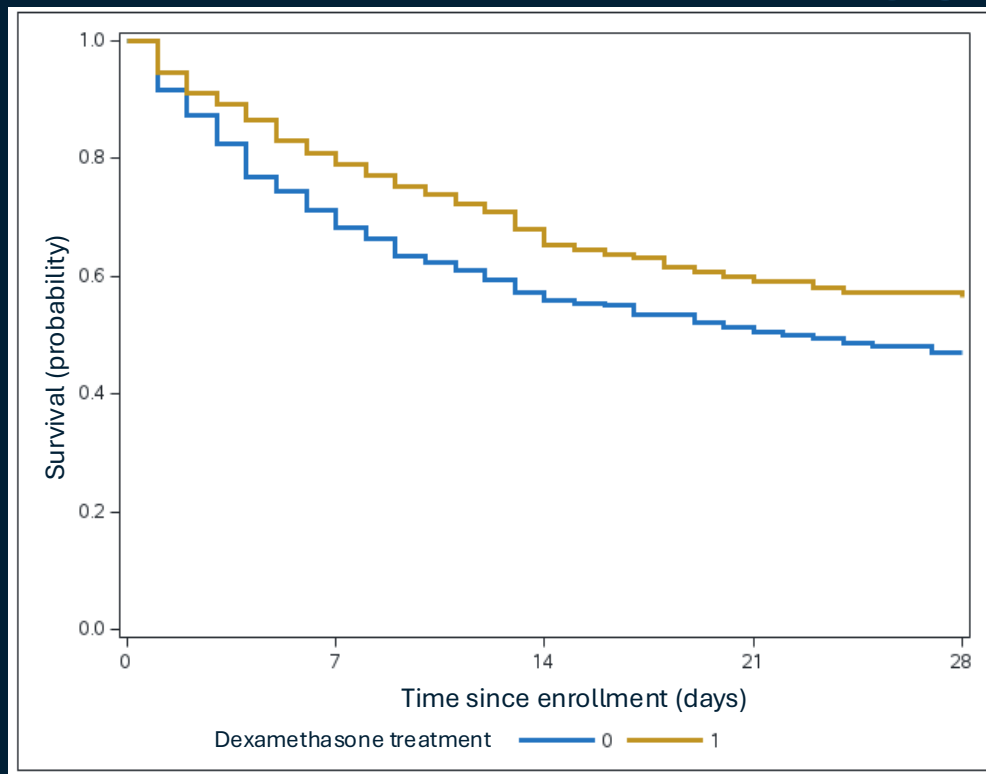
8 patients discharged to hospice treated as discharged deceased for analysis purposes.

Odds ratio = 0.86 ($p = 0.28$)

n = 401

28-day Mortality

Survival vs Time by dexamethasone treatment



Heterogeneity of Treatment Effect Analysis

Group A (n = 265)

- Treated, Pr = 0.446
- Untreated, Pr = 0.426

OR = 1.09 (p = 0.64)

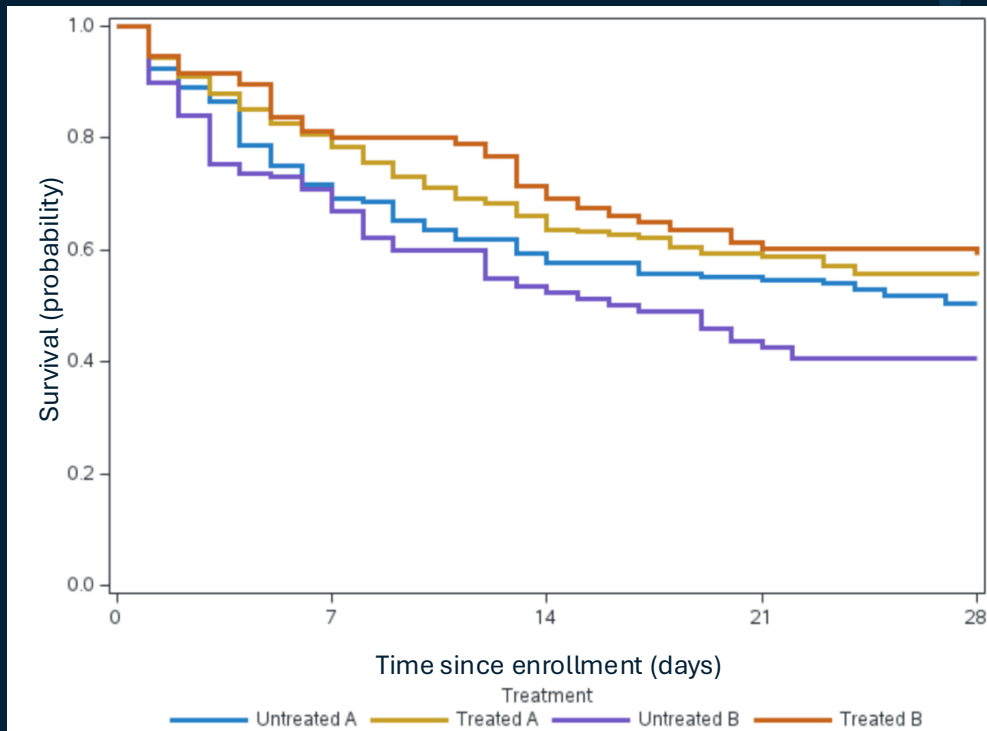
Group B (n = 136)

- Treated, p(mort) = 0.409
- Untreated, p(mort) = 0.552

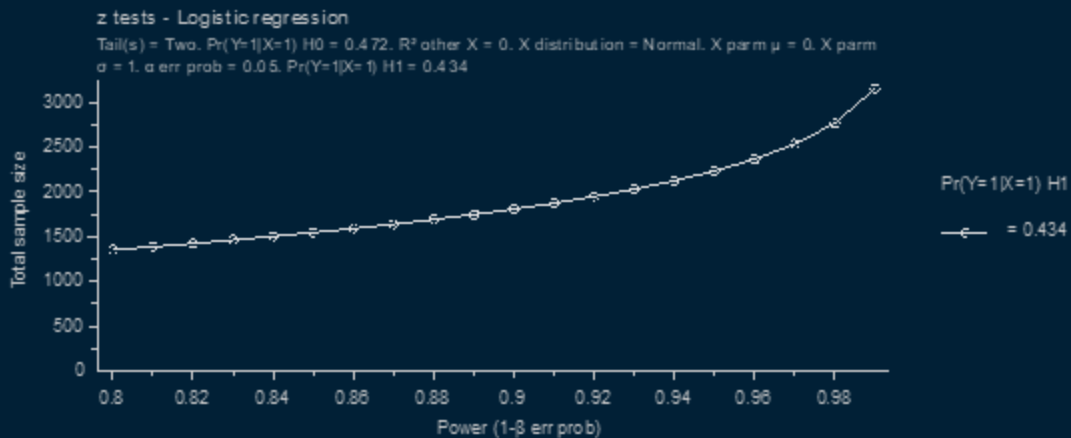
OR = 0.56 (p = 0.02)

Cytokine Dichotomization

401 patients, weighted - predicted by clinically informed Bayesian network



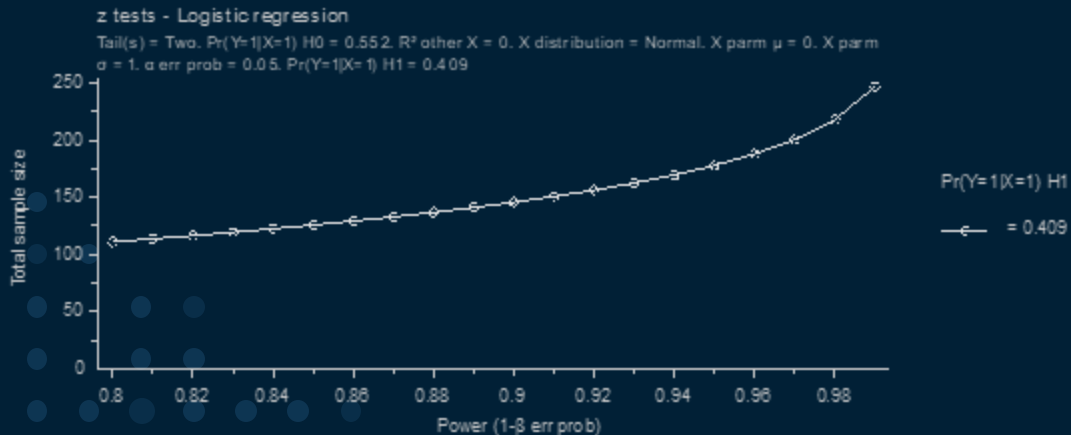
Clinical trial implications



Untargeted enrollment of mechanically ventilated patients.

Estimated sample size for 95% power = **2,236**

$\Delta = -92\%$



Targeted enrollment of mechanically ventilated patients with treatment responsive endotype.

Estimated sample size for 95% power = **179**

Conclusions

Machine Learning-Enabled Personalized Medicine for Rapidly Evolving Infectious Diseases

- A readily identifiable sub-population of COVID-19 ARDS defined by molecular phenotyping of tracheal aspirate samples appears to have a stronger 28-day mortality benefit to corticosteroids
 - **Molecular definition of syndromic disease may improve efficacy and targeting of therapies**
- Initial molecular endotyping identified dexamethasone as a key therapeutic with diverging upstream regulons which was reflected in clinical data
 - **Host directed therapeutic target identification is possible in the infectious disease space**
- Molecular analysis of Phase 1-equivalent sample sizes can drive predictions to enrich Phase 2/3 trial enrollment
 - **De-risking clinical trials may be possible by reducing cost and target enrollment**



Acknowledgments



UNC
SCHOOL OF MEDICINE





