Paper ML17

Machine Learning-Enabled Personalized Medicine for Rapidly Evolving Infectious Diseases

Jonathan Schisler, UNC-CH, Chapel Hill, USA
Matthew Wolfgang, UNC-CH, Chapel Hill, USA
Robert Hagan, UNC-CH, Chapel Hill, USA
Benjamin Sines, UNC-CH, Chapel Hill, USA
Scott McClain, SAS™, Cary, USA
Xi Jiang, SAS™, Cary, USA
Ella Pavlechko, SAS™, Cary, USA
Julia Florou-Moreno, SAS™, Cary, USA
Zhalen Hosseini, SAS™, Cary, USA
Vanunraj Valsaraj, SAS™, Cary, USA
Gabriel Risa, Rejuvenate Biomed, Heusden-Zolder, Belgium

ABSTRACT

Patients hospitalized with infection-related respiratory symptoms urgently require personalized therapies. Standardized clinical practices often result in varying effectiveness across diverse populations, whereas personalized medicine offers a tailored approach to address heterogeneous diseases. Despite extensive molecular and clinical data availability, a significant obstacle in precision medicine is the lack of real-time analytical tools to predict individualized treatment efficacy for specific cohorts. Our study focuses on patients admitted to intensive care units for respiratory failure treatment during the initial three years of the COVID-19 pandemic. Previously, we identified distinct molecular endotypes by analyzing omics and cytokine data in deep lung tissue and peripheral blood. In this presentation, we leveraged machine learning methodologies to integrate electronic health records (EHR) data with molecular insights, enhancing predictive capabilities. The value of this study lies in applying a repeatable analytic solution that aligns patients with specific, data-driven therapies, ultimately improving outcomes and enabling clinical trial success.

INTRODUCTION

The rapid evolution of infectious diseases, exemplified by the COVID-19 pandemic, has underscored the urgent need for personalized therapeutic strategies, particularly for patients with severe infection-related respiratory symptoms. In the first two years of the pandemic, intensive care unit (ICU) patients with respiratory failure due to SARS-CoV-2 exhibited heterogeneous responses to standard treatments, highlighting the limitations of one-size-fits-all approaches (Calfee, 2014). Molecular heterogeneity in acute respiratory distress syndrome (ARDS) associated with COVID-19 suggests that distinct endotypes—patient subgroups defined by molecular and clinical features—may drive differential outcomes (Seymour, 2019). Leveraging electronic health record (EHR) data and molecular insights, machine learning (ML) offers a promising avenue to identify these endotypes and align patients with data-driven therapies (Chen, 2021). This study integrates multi-omic profiling with ML to address the challenge of personalizing treatments for ICU patients with COVID-19-related respiratory failure, focusing on the therapeutic efficacy of dexamethasone as a case study.

METHODOLOGY

STUDY POPULATION AND DATA COLLECTION

This study utilized clinical and molecular data from ICU patients with COVID-19-related respiratory failure at UNC Health between April 14, 2020, and February 14, 2021 (**Figure 1**). Recruitment began shortly after North Carolina's stay-at-home order (March 30, 2020), representing the alpha wave of the SARS-CoV-2 pandemic. A total of 88 subjects were recruited to the study; deep lung samples and blood were used for cytokine and RNA sequencing analyses. For a subset of 39 subjects, multi-omics mass spectrometry was used to analyze deep lung samples. EHR data were processed from 5,000 adults (≥16 years) admitted between January 1, 2020, and December 31, 2022, with hypoxemia and a positive SARS-CoV-2 PCR. Patients requiring mechanical ventilation within 7 days of admission (n=593) were prioritized, with exclusions for immune compromise (n=162) due to organ/bone marrow transplants, inherited immunodeficiencies, or immunosuppressive medications, yielding a final cohort of 431 patients for emulated target trial analysis.

MOLECULAR PROFILING AND ENDOTYPING

Deep lung samples (tracheal aspirates) underwent multi-omic mass spectrometry (39 subjects) and cytokine/RNA sequencing (88 subjects) to identify molecular endotypes (**Figure 2A**). Compartment-aware endotyping distinguished lung-specific signatures from systemic responses (**Figure 2B**), integrating proteomic and transcriptomic data, a method validated in ARDS endotyping studies (Calfee, 2014). RNA-based knowledge graphs were constructed to predict drug responses, linking molecular pathways to potential clinical outcomes (**Figure 2C**) (Sweeny, 2016).

MACHINE LEARNING AND BAYESIAN NETWORK ANALYSIS

Feature engineering transformed EHR data into predictive variables, including demographics, comorbidities, and ventilation status. A clinically informed Bayesian network was developed to classify patients into molecular endotypes "A" and "B" (**Figure 2C**), scored with probability thresholds (P(class) >0.5 to >0.7) (**Figure 3**). Real-time ML predictions were validated against clinical outcomes, ensuring repeatability for clinical trial applications, a strategy supported by recent advances in critical care ML (Chen. 2021).

OUTCOME ASSESSMENT

The primary outcome was 28-day mortality, assessed via survival analysis comparing dexamethasone-treated versus untreated patients (**Figure 4A**). Secondary analyses explored treatment heterogeneity by endotype, using odds ratios (OR) and p-values to evaluate efficacy (**Figure 4B**).

RESULTS

MOLECULAR ENDOTYPES AND PATIENT CLASSIFICATION

RNA-based knowledge graphs identified dexamethasone as a key therapeutic, with upstream regulatory divergence between endotypes (**Figure 2C**). Bayesian network analysis classified 431 patients into two primary endotypes (Group A, n=284; Group B, n=147) based on molecular profiles from tracheal aspirates (**Figure 3A**). At P(class) >0.5, concordance reached 82% (352/431), improving to 100% at stricter thresholds (e.g., P(class) >0.75, 138/138) (**Figure 3B**).

TREATMENT OUTCOMES AND HETEROGENEITY

Overall, 28-day mortality probability was 0.472 in untreated patients versus 0.434 in dexamethasone-treated patients (OR=0.86, p=0.28, n=401), with 28 patients censored due to hospital transfer and 8 hospice discharges treated as deceased (**Figure 4A**). Group A showed no significant treatment effect (treated Pr=0.446, untreated Pr=0.425, OR=1.09, p=0.64), while Group B exhibited a significant mortality reduction with dexamethasone (treated Pr=0.409, untreated Pr=0.552, OR=0.58, p=0.02), indicating endotype-specific efficacy (**Figure 4B**) (Davenport, 2015).

DISCUSSION

CLINICAL IMPLICATIONS

These findings reveal a molecularly defined subpopulation of COVID-19 ARDS patients (Group B) with a more substantial mortality benefit from dexamethasone, aligning with the RECOVERY trial's broader evidence of corticosteroid efficacy (Horby, 2021). Integrating EHR and molecular data via ML enables rapid identification of treatment responders, potentially improving ICU outcomes. This approach validates molecular endotyping to enhance therapeutic precision, extending beyond corticosteroids to other host-directed therapies (**Figure 5**) (Seymour, 2019).

FUTURE WORK

Future studies should validate these endotypes in more extensive, multicenter cohorts and explore additional therapeutics predicted by RNA-based knowledge graphs. Incorporating real-time ML into clinical workflows could de-risk trials by enriching enrollment for responsive subgroups and reducing costs and sample sizes (Chen, 2021). Longitudinal analyses of evolving infectious diseases will further test the adaptability of this framework.

CONCLUSIONS

This study demonstrates that ML-enabled personalized medicine can identify therapeutically responsive subpopulations in rapidly evolving infectious diseases like COVID-19. By defining molecular endotypes linked to dexamethasone efficacy, we provide a scalable framework for improving treatment outcomes and trial design, with broader implications for precision medicine in infectious disease management.

REFERENCES

1. Calfee CS, Delucchi K, Parsons PE, et al. Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. *Lancet Respir Med*. 2014;2(8):611-620. PMID: 24853585.

2. Chen H, Zhou X, Li J, et al. Machine learning in critical care: state-of-the-art and a sepsis case study. *Biomed Eng Online*. 2021;20(1):101. PMID: 33789010.
3. Davenport EE, Antrobus RD, Lillie PJ, et al. Transcriptomic profiling facilitates classification of response to influenza challenge. *J Mol Med (Berl)*. 2015;93(1):105-114. PMID: 25213779.
4. Horby P, Lim WS, Emberson JR, et al. Dexamethasone in Hospitalized Patients with Covid-19 — Preliminary Report. *N Engl J Med*. 2021;384(8):693-704. PMID: 32678530.
5. Seymour CW, Kennedy JN, Wang S, et al. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. *JAMA*. 2019;321(20):2003-2017. PMID: 31104070.
6. Sweeney TE, Wong HR, Khatri P. Robust classification of bacterial and viral infections via integrated host gene expression diagnostics. *Sci Transl Med*. 2016;8(346):346ra91. PMID: 27384347.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Jonathan Schisler
University of North Carolina – Chapel Hill
111 Mason Farm Rd, MBRB Rm#3340C, Chapel Hill, NC 27599-7126
Phone: 919-843-8708
Email: schisler@unc.edu
Website: https://research.web.unc.edu/

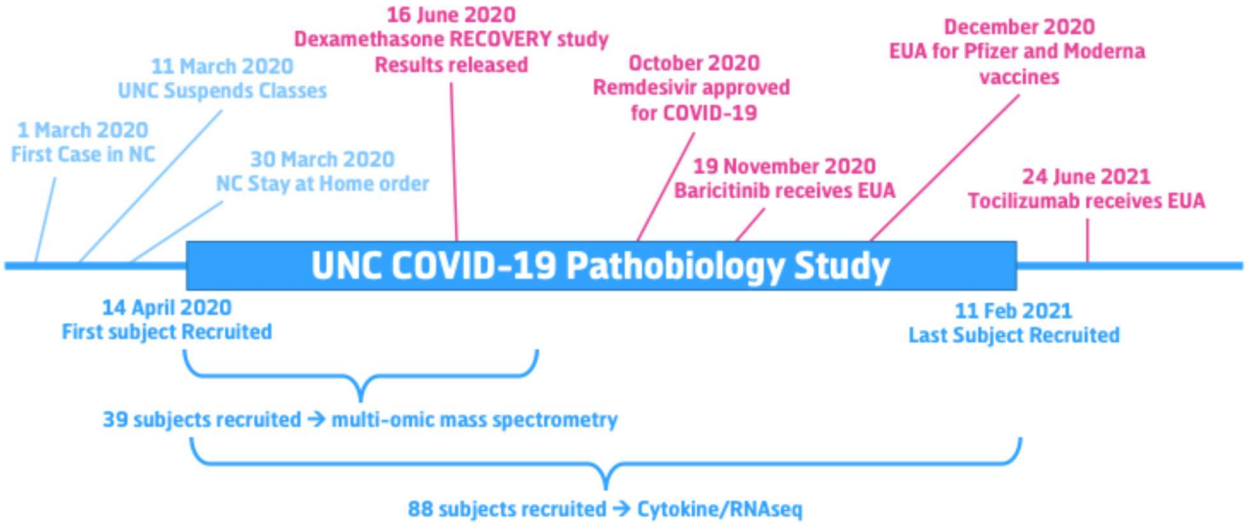


Figure 1: Study Design and Timeline. Overview of study timeline and cohort recruitment for integrating EHR and molecular data in COVID-19 respiratory failure patients

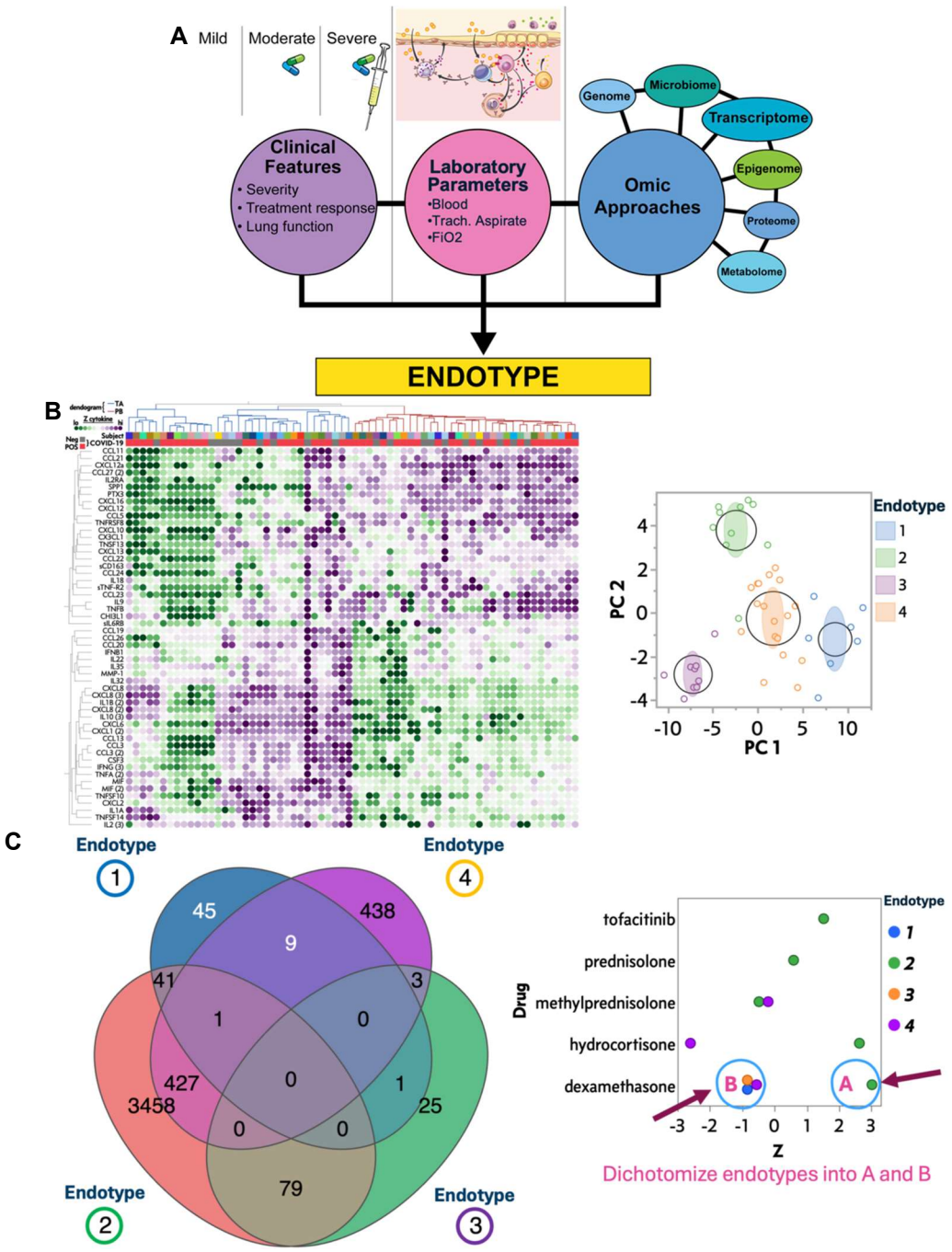
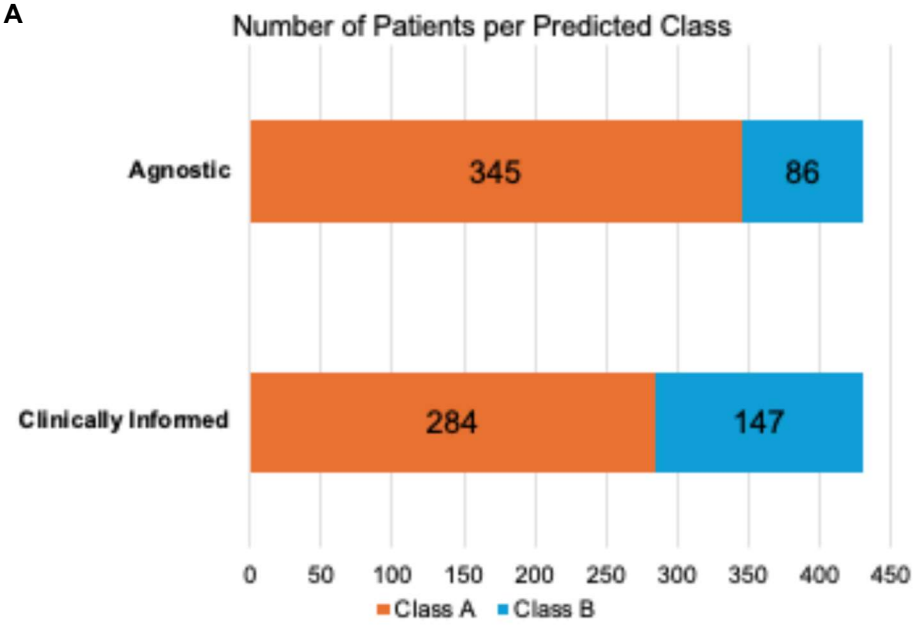


Figure 2. Workflow for identifying molecular endotypes and predicting drug responses in COVID-19 respiratory failure patients. *A, Data integration for endotype classification.* Clinical features (e.g., lung function, treatment response), laboratory parameters (e.g., blood tests), and omic approaches (e.g., metabolome, transcriptome, proteome) are integrated to define molecular endotypes, as illustrated by the interconnected nodes feeding into the "ENDOTYPE" output, adapted from doi.org/10.1016/j.jaci.2019.05.015. *B, Molecular endotype visualization.* A heat map of differentially expressed cytokines (left) and a scatter plot of principal component analysis (PCA, right) display lung-specific (red) and systemic (blue) molecular signatures, clustering patients into endotypes (1-4). Axes represent principal components 1 and 2, explaining 60% of the variance, with each point representing a patient sample. *C, Endotype-specific treatment prediction.* Venn diagrams show overlapping differentially expressed genes in tracheal aspirate samplers across endotypes, and a network knowledge graph illustrates treatment associations (e.g., dexamethasone, hydrocortisone) with molecular pathways, connecting differentially expressed genes to therapeutic responses. The four endotypes were dichotomized into “A” and “B” based on the predicted

A



B

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%)

Figure 3. Performance of Bayesian network in classifying COVID-19 respiratory failure patients into molecular endotypes. *A, Patient distribution by classification approach.* Bar graph displaying the number of patients classified as "A" (orange) or "B" (blue) using Agnostic (upper) or Clinically Informed (lower) at varying probability thresholds (P(class)). Bars represent patient counts. *B, Concordance across classification thresholds.* Summarizing the number of patients in Class A, Class B, Overlap, and Concordance at each P(class) threshold with concordance percentages (82%, 99%, 100%, 100%) reflecting model agreement.

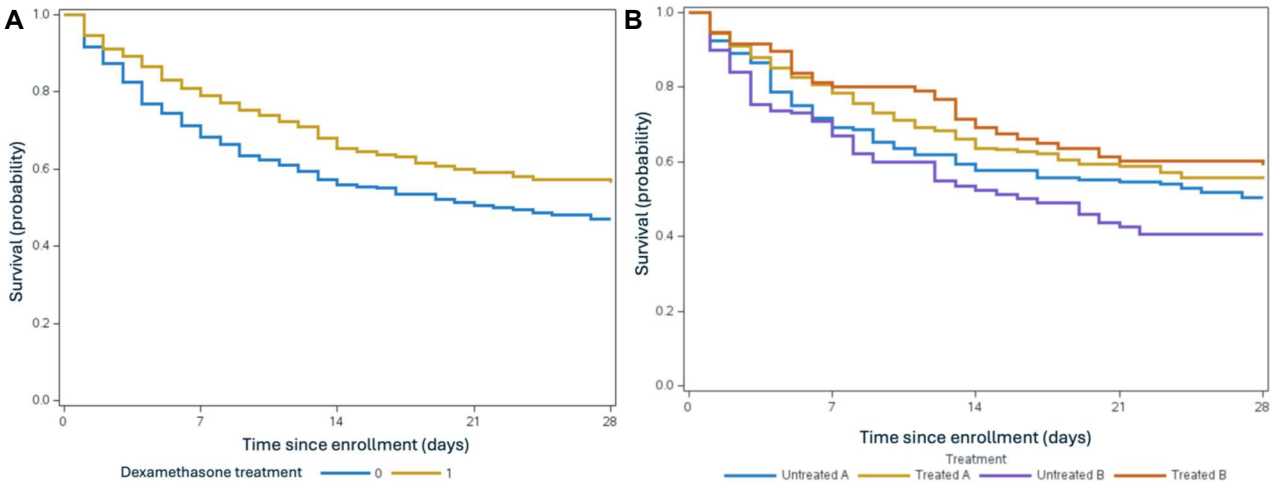


Figure 4. Kaplan-Meier survival curves illustrating 28-day mortality in COVID-19 respiratory failure patients by dexamethasone treatment and molecular endotype. *A, Overall survival by dexamethasone treatment.* Kaplan-Meier curve comparing survival probability over 28 days for untreated patients (blue line, 28-day mortality probability 0.472) and dexamethasone-treated patients (orange line, 28-day mortality probability 0.434), with a log-rank test p-value of 0.28 (n = 401). *B, Survival by endotype and treatment.* Kaplan-Meier curves stratified by molecular endotypes: Group A (n = 165: untreated BLUE, Pr = 0.426; treated GOLD, Pr = 0.446; OR = 1.09, p = 0.64) and Group B (n=136: untreated PURPLE, Pr = 0.552; treated ORANGE, Pr = 0.409; OR = 0.56, p = 0.02).

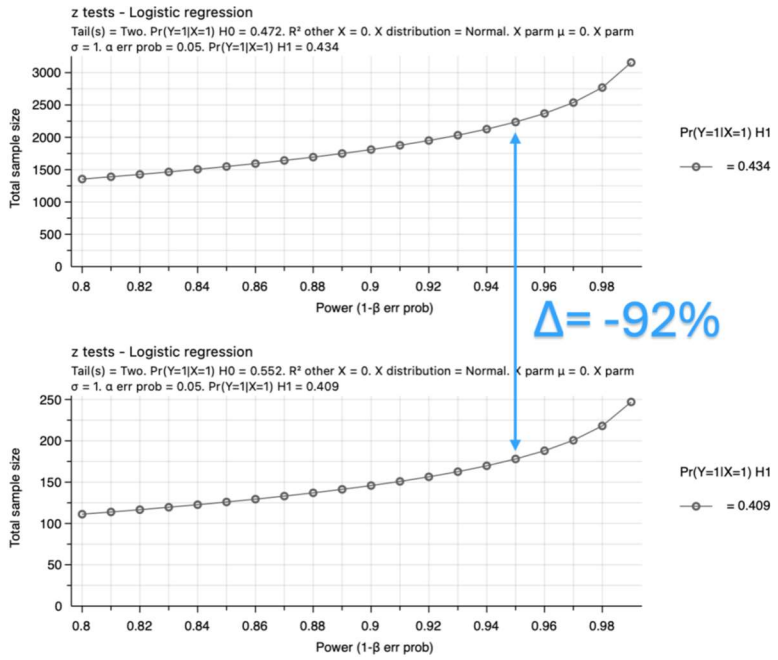


Figure 5. Impact of molecular endotyping on clinical trial efficiency for dexamethasone in COVID-19 respiratory failure patients. Power curves showing the relationship between sample size (x-axis) and statistical power (y-axis, 0–1.0) for detecting a treatment effect of dexamethasone. The upper plot is for an untargeted trial, and the lower plot is for the endotype-enriched trial design. The curve (black line) with a vertical line indicating required sample sizes for 95% power (dashed blue), marking the -92% reduction achieved by endotyping.