# Problem Statement

*Ted Liu, Parvati Jayakumar*

Ventilator-Associated Pneumonia (VAP) is a prevalent infection in critically ill patients, commonly linked to prolonged mechanical ventilation, increased hospital stays, and higher mortality rates. While certain pathogens associated with VAP have been identified, it remains unclear if they each exhibit unique biomarker profiles or distinctive pathophysiological patterns. Our project, in collaboration with Dr. Eric Morrell, a pulmonologist and critical care physician-scientist at the University of Washington Department of Pulmonary, Critical Care, and Sleep Medicine, seeks to explore these differences by (1) identifying if specific pathogens show unique biomarker signatures, (2) determining if subphenotypes exist within the VAP cohort, and (3) examining how these subphenotypes correlate with clinical outcomes, including ventilator-free days, severity of respiratory failure, and mortality.

Dr. Morrell has expressed challenges in processing clinical data into an analyzable format and implementing the rigorous statistical analyses required to answer these questions. Although he is experienced in pulmonary and critical care medicine, Dr. Morrell has limited training in computational and statistical methods, which are essential for addressing the aims of this project. The difficulty lies in ensuring that the project outcomes are both accurate and reliable, particularly since this research involves critically ill patients. Achieving a good level of accuracy and reliability is challenging due to the complexity of the data and the need for precise analysis to inform clinical decision-making effectively.

To meet these challenges, Dr. Morrell has requested assistance from our team with performing the necessary data engineering and statistical analysis to answer the questions identified in the project aims. Specifically, Dr. Morrell has asked for our help in developing an efficient, reproducible pipeline to generate datasets for analysis and performing rigorous statistical analysis of clinical data to address the project's key questions. To ensure accuracy and reliability, we will use validated statistical methods and advanced computational techniques, implementing rigorous data quality checks at every stage to eliminate errors and inconsistencies. Clear documentation and version control will be maintained throughout the process to ensure reproducibility and transparency. This collaborative effort will support Dr. Morrell's objectives, ensuring a robust and reliable investigation into pathogen-specific biomarkers, the potential for VAP subphenotypes, and their clinical impact.