# Research Plan

## Significance

### Clincial Impact

**Acute respiratory failure is a significant burden of disease.** Many hospitalized patients develop acute respiratory failure (SHINYstan Team, 2015), which is worrisome.

**more respiratory failure is a significant burden of disease.** Indeed ny hospitalized patients develop acute respiratory failure (SHINYstan Team, 2015), which is worrisome.

### Bayesian imputation

**Heterogeneous provider compliance and missing clinical data may limit implementation** of the prediction algorithm, the therapeutic interventions and the trial itself. Variables with strong predicitive power in our model may not be recorded in all patients or may be missing for the time window needed for prediction. To improve prediction for cases with incomplete data we can impute the missing data. Likelihood-based mixed effects models for incomplete data give valid estimates when the data are ignorably missing; that is, the parameters for the missing data process are distinct from those of the main model for the outcome, and the data are missing at random (MAR). This unlikely in our setting of clinical data in the EMR, because physicians will request test based on the patients comorbidities and current clinical conditions. Data will not be missing at random, but incompleteness will be associated with predictors and outcomes.

**Incomplete data can be imputed from auxilary data,** that is additional information available in the form of an auxiliary variable known to be correlated with the missing outcome of interest. For example, arterial blood gas oxygen saturation may be used to impute peripheral pulse oxymetry or oxygen therapy, if the latter are unavailable for the prediction time window, and vice versa.

first (Wang & Hall, 2010). second(Hall, Lipton, Katz, & Wang, 2014)

Joint modeling methods can help address bias caused by informative missing data in the estimation of the effect of]

Hall, C. B., Lipton, R. B., Katz, M. J., & Wang, C. (2014). Correcting bias caused by missing data in the estimate of the effect of apolipoprotein epsilon 4 on cognitive decline. *J Int Neuropsychol Soc*, 1–6. doi:10.1017/S1355617714000952

SHINYstan Team. (2015). SHINYstan: R package for interactive exploration of markov chain monte carlo output, version 0.1. Retrieved from https://github.com/jgabry/SHINYstan

Wang, C., & Hall, C. B. (2010). Correction of bias from non-random missing longitudinal data using auxiliary information. *Stat Med*, *29*(6), 671–679. doi:10.1002/sim.3821