Specific Aims

We propose to improve the prediction and prevention of respiratory failure and death in hospitalized patients by integrating complex Bayesian hierarchical modeling into data acquisition, patient triage and treatment implementation in electronic medical record based (EMR) surveillance.

Severe acute respiratory failure (ARF) requiring mechanical ventilation leads to increased mortality, increased cognitive and functional impairment. EMR surveillance can identify hospitalized patients at risk, days before their deteriorating conditions are typically recognized; earlier initiation of preventive interventions can reduce morbidity, mortality and expenses: My mentor Dr. Gong is leading a randomized multicenter trial to reduce mortality by triggering an individualized prevention checklist for patients identified as at risk.

Hierarchical models may perform better than classical models in large data sets} with spatial and temporal organization. We are particularly interested in fitting complex hierarchical Bayesian models to improve prediction, (1) by allowing model parameters to vary between patients, between medical floors, services or institutions and (2) by modeling variation in compliance and treatment effects during trial implementation. Patients seen by the same team, treated in the same setting or season will show similar clinical trajectories and responses. Especially in the subset with sparse or missing data, precision and accuracy of parameter estimates can be improved by pooling, when they are informed by data from all the other patients.

Heterogeneous provider compliance and missing clinical data may limit implementation of the prediction algorithm, the therapeutic interventions and the trial itself. I will advance Bayesian data imputation using auxiliary data with Dr. Hall. Seasonal effects and institutional learning may limit prediction accuracy. We will update our model continuously with new patient information, incorporate compliance and effectiveness of the interventions into the model and adjust for seasonal effects in one coherent model.

The integration of advanced statistical modeling with EMR surveillance to improve patient outcomes constitutes the unique innovation and power of my proposal. Implementation of Bayesian hierarchical modelling can be computationally challenging for Big Data. My co-mentor Dr. Gelman is leading the NSF-funded development of Stan, statistical software achieving faster convergence and parameter estimation based on novel Markov Chain Monte Carlo computer algorithms.

Under an exceptional and multidisciplinary combination of mentors, I will integrate complex hierarchical models into an ongoing "real time" EMR based multicenter trial and clinical decision algorithm, advance integrated data imputation and overcome current limitations in Bayesian computational implementation for very large multi-center EMR surveillance.

Our overall hypothesis is that complex hierarchical Bayesian modeling and data imputation will reduce morbidity and mortality from respiratory failure in hospitalized patients compared to the classical model.

Specific aims:

Aim 1:To improve early prediction of prolonged respiratory failure and death in hospitalized patients. We will implement a complex hierarchical Bayesian prediction algorithm, comparing it to the classical model. ***** SA 1a: To build a pragmatic EMR based hierarchical Bayesian model implemented in the ultra-fast statistical software Stan to predict a composite outcome [death or prolonged mechanical ventilation $>$ 48 hours] in our Montefiore Medical Center inpatients and compare it with the existing frequentist algorithm. ***** SA 1b: To further develop Bayesian data imputation algorithms of missing clinical data using auxiliary data, to identify the auxiliary measure properties, ceiling and floor effects and to test the imputations against manually verified data and published algorithms.

***** Aims 2: To integrate a complex Bayesian model into patient triage and treatment implementation. Our prediction algorithm will trigger individualized patient interventions in Dr. Gong's pragmatic multi-center trial. ***** SA 2a: To integrate patient triage and advance compliance into Dr. Gong's clinical trial and sustained quality improvement and to focus education efforts on the most effective components of the checklist intervention. ***** SA 2b: To update our model continuously with new incoming patients, to expand our model to include patients from other regional institutions and to incorporate provider compliance, seasonal effects and institutional learning into the model. (SHINYstan Team, 2015), which is worrisome.

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