Research Plan

A. Significance

Respiratory failure in hospitalized patients can be predicted and should be prevented.

Acute respiratory failure (ARF) requiring mechanical ventilation is common in hospitalized patients, consuming a disproportionate amount of health care resources in the USA[2]. Short term mechanical ventilation can be life saving, but prolonged mechanical ventilation often leads to multi-organ failure and death[2;3]. Most research focuses on *established* respiratory failure in the ICU, while detectable clinical signs and symptoms often herald the impending respiratory decompensation much earlier[4]. Dr. Gong co-developed the LIPS score to identify patients at high risk for Adult Respiratory Distress Syndrome in the emergency department[5], which proved equally able to discriminate the 587 patients in the cohort who progressed to severe ARF requiring > 48 hrs of mechanical ventilation. She also demonstrated that predictive scores deteriorate as early as 24-48 hours before ICU admission[1] [Figure 1]; but such ominous signs are either not recognized or not acted upon[6;7]. Early interventions and preventive measures(e.g antibiotic therapy, diuretics and head elevation) would be able to stop or reverse the clinical deterioration and/or prevent progression to multiple organ failure and prolonged mechanical ventilation or at least attenuate the subsequent clinical course[8;9;10;11].

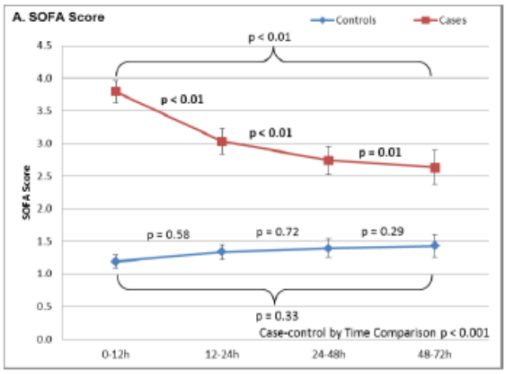

Figure 1: Deterioration of the Sequential Organ Failure Assessment score (SOFA) can be detected 24-48 hours before clinical deterioration leads to ICU admission; p-values reflect pair-wise comparisons between consecutive time intervals, adjusting for patient characteristics.[1].

Checklist intervention examples
Prevent respiratory insufficiency
Early goal directed therapy[12]
Adequate early antibiotics[13]
Decrease mechanical ventilation
Daily sedation break[14]
Spontaneous breathing trials[15]
Limit transfusion-related lung injury
Restrictive transfusion strategy[16]

Table 1: Examples of checklist interventions, references documenting effect.

A pragmatic clinical trial to predict and prevent mortality from respiratory failure in hospitalized patients. My mentor Dr. Gong is leading a NHLBI-funded multi-center cluster randomized pragmatic trial in two phases. (1) the first phase APPROVE aims to identify patients at risk by building classical logistic regression models based on electronic medical records (EMR) to Accurately Predict PROlonged Ventilation. (2) In the second phase PROOFCheck, identification of a patients at high risk triggers a decision support tool and bundled checklist interventions, proven to prevent organ failure in critically ill patients[12;13;14;15;16]. PROOFCheck is testing if the early implementation of a checklist of preventive measures (1), reduces severity of organ failure, mortality and duration of mechanical ventilation in patients at high risk identified by APRROVE.

Electronic medical records are an eminent example of richly structured and correlated Big Data. Exemplified by Dr. Gong's pragmatic trial, they hold enormous promise for outcomes research across a wide swath of clinical domains ranging from pediatrics to psychiatry, from maternal health to mortality from cancer[17;18;19;20;21;22]. However, large electronic medical data sets are not just bigger in that there are more instances of the same thing, (e.g. more patients would make data analysis only easier). Rather, there is more breadth to the data, and in the case of pragmatic trials, more heterogeneity, more subgroups, locations, or time granularity than is currently being modeled, more frequent and detailed measurements than can easily be incorporated into classical models. This currently limits the scientific hypotheses and clinical inferences, that can be explored and evaluated. In Dr. Gong's trial in particular, we desire more fine-grained predictions to individualize prevention.

We can individualize prevention by targeting patients at risk. Preventive measure, for example goal targeted resuscitation, decrease respiratory failure requiring mechanical ventilation, when they are initiated early[10]. However, an indiscriminate approach to prevention of respiratory failure in hospitalized patients will be ineffective, because only one in 30 hospitalized adults requires mechanical ventilation. Secondly, individualizing preventive and therapeutic measures specifically based on patient characteristics will be more efficient in preventing potentially irreversible end organ damage, while also leading to improved compliance by providers and cost effectiveness. So how can we improve and individualize prediction and prevention?

Hierarchical modeling is transformative for EMR-based prediction.

Observations and outcomes in EMRs and pragmatic trials will be nested hierarchically. For example, in APPROVE and PROOFCheck, repetitive oxygen saturation measurements will be similar in the same patients; the closer in time they are, the higher the correlation between repeated observations. Equally, patients seen by one and same hospitalist will tend to have similar outcomes, predicted by that physician's behavior and qualities. As an example, some physicians (or services) will follow a more liberal fluid management, others will emphasize early diureses; clearly this choice will summarily affect the respiratory failure risk of specifically those patients under this physician's care. Generally, physicians in large academic medical systems like ours are organized in services, which are integrated across wards, clustered in several hospitals. Consequently, the observations in our hospitalized patient cohort, their outcomes and their propensity to respond to treatments, all are hierarchically nested; this requires more than just fitting well-known models at larger scales.

Hierarchical models better exploit the fine-grained multilevel structures of electronic medical records and may therefore optimally predict acute respiratory failure leading to prolonged mechanical ventilation or death in our trial cohort. Fitting our predictive regression model, we would want the regression coefficients to vary by group (e.g. by service, by medical unit, by hospital), to realistically model the multifaceted correlations seen in actual clinical practice. The number of parameters to estimate grows very quickly and so do the potential interactions. Even with very large data sets, the sample size in each subgroup will shrink rapidly; estimates using least squares or maximum likelihood will become noisy and thus often become essentially useless. One solution lies in hierarchical modeling, where we estimate hyper-parameters and hyper-hyper-parameters (Figure 2), to represent how lower level parameters vary across different groupings[23].

Partial pooling is more efficient for prediction. Prediction based on partial pooling outperforms (a) the no-pooling and (b) the complete-pooling approaches, as can be shown mathematically[24] or via cross-validation[25]. Using the No-pooling approach, we estimate the model for each specific subset of interest separately. But this leads to far too many sub-classifications, thus too small samples in any given subgroup for useful inferences, if we fully explore the complexity and granularity, the richness of the EMR data,. Employing complete pooling or structural modeling constitutes the other extreme of the spectrum, but the implied hard constraints on the coefficients in different groups may lead to bias: we loose information, because we cannot learn from groups where we have more data. We choose the middle ground: Prediction using partial pooling or hierarchical modeling is especially effective for our richly organized EMR data, because the estimate of each individual parameter is simultaneously informed by data from all the other patients in our cohort, improving prediction in particular for subgroups with sparse data.[26]. Effron explained this apparent paradox well to non-statisticians in the Scientific American[27;28].

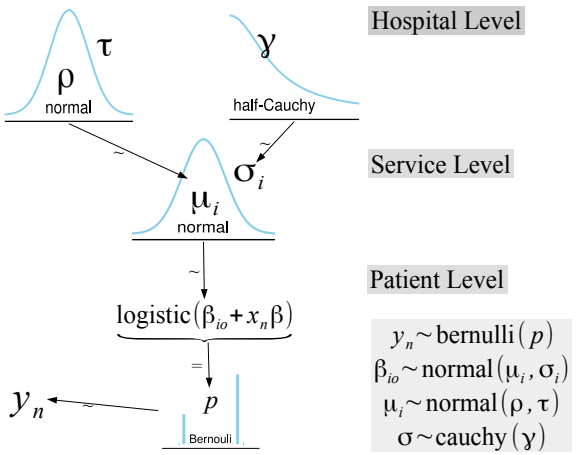

Figure 2: Distrogram to illustrated the hierarchical structure of patient trajectories. Outcome y_n for the n^{th} patient is a boolean indicator predicted by a logistic regression. We allow the patient (random) intercept β_{oi} to vary according to the i^{th} medical services the patient is under. The service level mean μ_i and within-service variance σ_i are modeled to vary by hospital. x_n is a vector of patient level predictors, β is a vector of regression coefficients.

Heterogeneous and incomplete clinical data may limit prediction and implementation. Variables with strong predictive power in our model may not be recorded in all patients or may be missing for the time window needed for prediction, limiting development of the prediction algorithm, implementation of the therapeutic interventions and the trial itself. Incomplete data are the hallmark of EMRs. In our data set we find for example that an arterial blood gas (ABG) to assess arterial oxygen tension is often unavailable for the prediction time window, because it was not requested by the physicians. To improve prediction for cases with incomplete data, we can impute the missing data using *multiple imputation*. Likelihood-based mixed effects models for incomplete data give valid estimates *if and only if* 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)[29]. However, this is an unreasonable assumption for EMRs; in our example, physicians will re-

quest ABGs based on the patients respiratory co-morbidity and clinical hypoxia symptoms. Data will not be MAR. Instead, incomplete data will be associated with predictors and outcomes; this could lead to biased imputations.

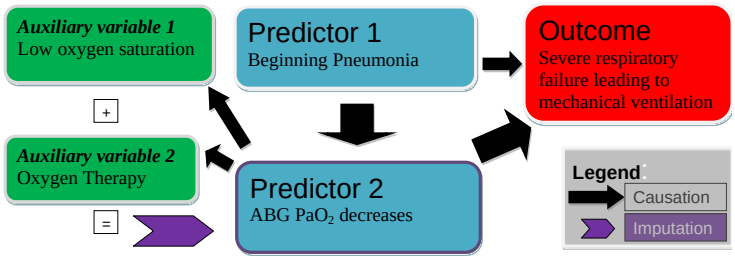

Figure 3: Incomplete data can hinder outcome prediction. We can impute incomplete data from auxilliary information. Pneumonia impairs oxygenation, causing respiratory failure, for example. If arterial blood gases (ABG) are missing, we can impute the arterial PaO_2 (oxygen tension) from oxygen therapy and/or peripheral oxygen saturation.[30].

Auxiliary data can be used to impute incomplete medical records. Auxiliary data are additional information available in the form of variables known to be correlated with the missing data of interest[31]. If the physician did not request an ABG, instead peripheral oxygen saturation and or oxygen therapy may be available and could be used to impute the arterial blood oxygen tension [Figure 3]. This approach avoids the perils associated with missing at random (MAR) assumptions, when fitting a non-ignorable missingness model[32]. Adding auxiliary variables not included in the main model for multiple imputation, in other words using additional information that is correlated with the missing outcome is an emerging approach to help correct bias[33;34;35], often relying on Bayesian methods[36;37]; joint hierarchical modeling, including auxiliary data to impute incomplete patient records, will improve the prediction model and facilitate the implementation of the prediction algorithm[30].

Seasonal effects and institutional learning can bias risk prediction and can thwart implementation or imperil the effectiveness of our efforts to mitigate the risks of severe respiratory failure in hospitalized patients. The composition of our hospital population, their co-morbidities and risk profiles change over time, altering which patient characteristics best predict severe adverse respiratory failure and mechanical ventilation. More importantly, during the implementation phase of previous preventive trials we noted that providers learn, changing their behavior as a result of trial participation. As trials progressed providers implemented previously underutilized interventions more frequently even before they were prompted. We term this effect institutional learning. On the other hand, the transition of junior and senior providers through their training and to other institutions and new personnel joining the staff, may led to lessons learned being forgotten again. Last but not least, respiratory disease is affected by seasonal and secular effects; influenza prevalence for example is seasonal and characterized by major and minor epidemics. Seasons and epidemics will affect the predictive power of any model and hence also alter the risk profile of our patients over time. Institutional culture and individual provider behavior change in response to trials and quality improvements interventions; patient populations change over time. Respiratory patients are plagued by seasonal deterioration. These temporal, seasonal and secular effects will alter the predictors of risk in our model and affect its implementation. We will include these and continuously update our model with new patient data to account for said changes in the risk profile. The integration EMR-triggered prediction and prevention with institutional learning, secular and seasonal effects as well as data imputations from auxiliary data within one coherent (Bayesian) model is certainly novel, but how can it be implemented in one coherent model?

B. Innovation

Bayesian hierarchical modeling is groundbreaking in EMR-based prediction, and particularly suited for joint hierarchical modeling. With their inherent flexibility and robustness[38;39], Bayesian hierarchical models may outperform classical models for EMR-based prediction owing to the integration of additional information through "partial pooling"[40] and the imputation of incomplete records from auxiliary data. Increases in computer power led to an expansion of applied Bayesian work[41;42], also in Big Data[43;44] and more recently in EMR-based prediction[45;46;47]. *However, we are unaware of any Bayesian hierarchical prediction model based on large EMRs.*

A brief introduction to Bayesian inference. Clinical decision-making is Bayesian[42]. Physicians continuously update their preliminary diagnosis with new information. Prior belief $P(A)$ in a diagnosis may be weakened by new laboratory information, leading to an updated diagnosis $P(A|B)$, based on the lab data[48;49].

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)} \quad (1)$$

According to Bayes' Theorem 1 prior information $P(A)$ is combined with new data, (known as the likelihood) to yield an *updated* estimate for the probability of a hypothesis $P(A)$, given the data $P(B)$ observed, called the posterior distribution $P(A|B)$[50]. Full Bayesian inference is based on priors. Statis-

ticians may object to prior choices (e.g. the cauchy distribution as prior for variance parameters[51] in Figure 2). Principled critique questions the subjectivity of priors. But are not *all* models (frequentist and Bayesian) based on subjective choices[52] (e.g. link function, correlation matrices) or (distributional) assumptions; (Bayesian or classical), these choices need to be reasonable and convincing to the intended audience and should be subjected to sensitivity analysis and model exploration. Combining frequentist with Bayesian thinking may advance Big Data science most[53;43].

The computational implementation of Bayesian hierarchical models. Bayesian inference for multi-layered models can often not be derived analytically; instead we calculate numerical approximations of the multi-dimensional integrals to obtain the posterior distributions of the parameters of interest. In technical terms, Markov Chain Monte Carlo (MCMC) based Bayesian inference methods sample from a posterior probability distribution after building a Markov chain. MCMC simulation replaces intractable analytical integration with empirical summaries of samples from the posterior distribution[54]: *Instead of analyzing the odds, we simulate throwing the dice repeatedly.*

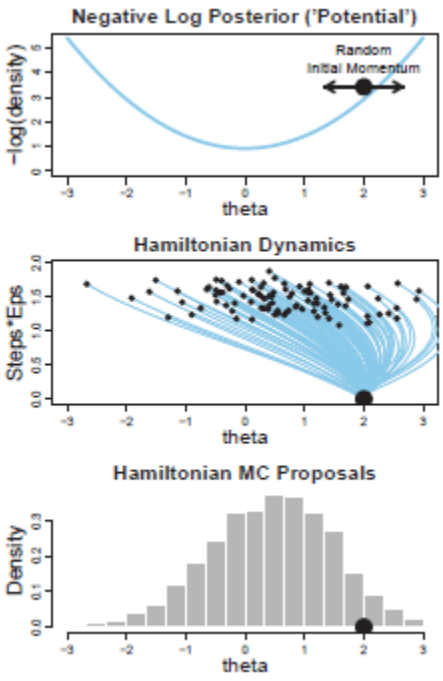

Figure 4: Hamiltonian MCMC uses momentum to optimize the next proposal. The current proposal's higher momentum (black dot) is indicated in the top panel. The middle panel illustrates how random samples are drawn to the mode of the posterior distribution (lower panel) improving convergence and effective sample size. Fig 14 in Kruschke[49].

Pushing the envelope of Bayesian EMR-based prediction, we will implement the Bayesian model in parallel in the ultra-fast probabilistic programming language software Stan developed by my co-mentor Dr. Gelman[55]. Stan's Hamiltonian Monte Carlo algorithms[55] and clever statistical formulation push the boundaries of computability[25]. For example non-centered parameterization allows sampling in the standardized normal space; this takes full advantage of the faster convergence and higher effective sample size of Stan to overcome computational limitations of Bayesian hierarchical models for Big Data[25]. Stan is based on Hamiltonian Monte Carlo (HMC)[25], a Markov chain Monte Carlo (MCMC) algorithm[56], which avoids the sensitivity to correlated parameters that plague many MCMC methods by introducing auxiliary momentum variables[57] as illustrated in Figure 4. HMC is dependent on tuning the reciprocal relationship of the crucial parameters step size and desired number of steps. Too low a step size wastes computing power, whereas a step size too large loses efficiency. Stan overcomes this with the No-U-Turn Sampler (NUTS), a recursive algorithm to automate HMC tuning[57].

Analyzing and advancing the practical clinical implementation of preventive interventions is decisive for outcome improvement and research. Imperfect provider compliance is a major concern also in our PROOFCheck trial, just as non-compliance is a major obstacle to the effective delivery of health care and improved outcomes in general[58]. The targeted interventions triggered by our EMR-prediction algorithm will only prevent respiratory failure if our physicians and nurses actually implement them. Improving fidelity of health care providers with evidence based interventions continues to be a challenge and is under-researched[59] and little is known on how to reproduce multi-faceted interventions (specially directed toward providers) to improve clinical outcomes[60]. As long as we do not understand what drives provider fidelity and patient compliance with the preventive measures proposed to our providers for their high risk patients[61], we ignore the best means to translate widely accepted interventions and new findings of outcomes research into practice[62]. We need to understand better what patient and/or provider characteristics hinder compliance with the triggered preventive checklist interventions to ensure care is in accordance with evidence-based best practices.

We use a pragmatic trial to investigate provider fidelity. Pragmatic trials like Dr. Gong's may result in more valid estimates of effectiveness for more realistic health care scenarios[63;64]; we will use her pragmatic trial data to investigate incomplete fidelity, heterogeneity and difficulty in clinical implementation. An relevant example for PROOFCheck is blood product management: transfusions increase the risk of acute severe respiratory failure with mechanical ventilation[65], but implementation of rational transfusion blood product management is still sketchy and very heterogeneous across the nation[66]. Weiss et al. demonstrated that direct prompting for best practices improves provider compliance in the ICU and outcomes such as duration of mechanical ventilation or

length of stay[67]. We hypothesize that fidelity will be associated with certain provider and patient characteristics; their investigation will allow more focused re-education efforts and adaptation of the checklist implementation.

Summary of the impact

Acute respiratory failure in hospitalized patients leading to prolonged mechanical ventilation with the inherent mortality and morbidity constitutes a serious health care challenge. We will tackle this by combining innovative approaches to data imputation with sophisticated hierarchical prediction models to form a near real-time EMR-based clinical decision tool with practical utility in critical care. We use the opportunity to investigate poor provider fidelity a serious and under-researched barrier to outcomes research and the implementation of evidenced-based care. *Our findings will have implications beyond our trial for any clinical research, indeed for the implementation of evidence-based-medicine at large.* Changes in reimbursement give providers a stake in patient outcomes and led to a keen interest in the prediction and prevention of adverse event in hospitalized patients. This project advances hierarchical Bayesian models to implement this paradigm shift in very large EMRs, triggering personalized interventions that deliver outcome improvements. This is novel and has not been attempted to our knowledge. But our impact goes beyond improving morbidity and mortality from respiratory disease in hospitalized patients through improved prediction and prevention, beyond investigating drivers of poor provider compliance. We will develop new methods to impute incomplete electronic medical records from auxiliary data and pioneer Bayesian hierarchical prediction models for large EMR data. Our proposal is unique and novel in its integration of cutting edge methods from clinical, statistical and computer science to fully realize the promise of Big Data in medicine.

C. Approach

My research project will be closely aligned with my mentor's NIH-funded pragmatic two phase trial. Aim 1 will utilize the processed data of APPROVE to improve the prediction model and Aim 2 will use the data from the implementation of PROOFCheck to investigate fidelity of the providers with the EMR-triggered interventions.

Aim 1: To improve incomplete data imputation and early prediction of acute respiratory failure.

Hypothesis: The integration of auxiliary data imputation and multi-level Bayesian modeling will improve prediction of severe respiratory failure in hospitalized patients compared to classical statistical approaches.

For specific aim 1a, we will build a pragmatic EMR-based hierarchical Bayesian model to predict a composite outcome [mechanical ventilation prolonged beyond 48 hours or death] in hospitalized adult and compare our Bayesian approach with the existing frequentist algorithm used by Dr. Gong in her pragmatic trial.

Population: We will include all adults patients, admitted to the Montefiore Medical Center during APPROVE, excluding only those who are chronically ventilated at home or who have Do not resuscitate orders at the time of hospital admission. APPROVE is Dr. Gong's prospective observational cohort study underway at Montefiore and the Mayo Clinic Rochester, described in detail under Significance; we will build our Bayesian hierarchical model based solely on Montefiore patients. We will divide the cohort into separate fitting and validation set.

Predictors: We will consider time-invariant and time-variant demographic and clinical data for inclusion as independent predictors in our model. Examples for demographics are gender, age, medical service or ward, examples for physiological and clinical predictors are heart rate, blood pressure or lab tests. Certain predictors will require summary aggregations and (logarithmic) transformations to induce variance stability.

Outcomes: Our primary outcome will be acute respiratory failure requiring mechanical ventilation longer than 48 hours, specified as positive for (a) mechanical ventilation lasting longer than 48 hours or (b) mechanical ventilation lasts less than 48 hours, but the patient died within 96 hours of the calculated score. Patients that are not on prolonged ventilation within 96 hours or discharged alive from the hospital will be considered negative.

Bayesian hierarchical modeling to reflect the nested structure of health care. We will build a Bayesian hierarchical multivariate logistic regression model of time-invariant and time-variant demographic, clinical and administrative variables. Our Bayesian hierarchical modeling will represent the multi-level nested structure of current health care, with levels for medical or surgical service the patient is under, the floor or ward where the patient is cared for, the institution the patient is admitted to. We may also consider other random effects for example for co-morbidity and other time-invariant patient specific descriptors. We illustrate this nested structure in a simplified logistic model with hierarchical levels for patient, service and hospital, analogous to Figure 2.

$$y_n \sim bernulli\Big(inv\_logit(\beta_0 + \beta_1 * PaO_2)\Big) \quad (2)$$

Patient level (2) On the left, we model at the patient level, the probability that the n^{th} patient will develop the dichotomous event y_n, (acute respiratory failure requiring prolonged mechanical ventilation), using say arterial oxygen tension PaO_2 as one predictor in a simple logistic regression model. However, patients are typically assigned to different services. Pulmonary service patients may have a lower baseline PaO_2, while surgical patients tend to have normal lung function.

$$\beta_{0i} \sim Normal(\mu_i, \sigma_i); \beta_{1i} \sim Normal(\kappa_i, \sigma_{\beta_1 i}) \quad (3)$$

Service level (3) On the left, we develop our hierarchical model further to allow random intercepts β_{0i} modeling that the average μ_i baseline arterial oxygen tension (PaO_2) may vary between a given medical $service_i$ and another say surgical $service_{i+1}$. Analogously, smaller changes in PaO_2 may be indicative of respiratory deterioration in a certain medical $service_i$, compared to a surgical $service_{i+1}$, where only a larger drop in arterial oxygen tension effectively predicts outcome. We may allow the regression coefficient for the slope β_{1i} to vary around different mean slopes κ_i at the service level i.

$$\mu_i \sim Normal(\rho, \tau); \sigma_{\beta_{1i}} \sim Cauchy(\gamma) \quad (4)$$

Hospital level (4) Even within one academic institution, some (city) hospitals may cater to a economically more disadvantages population, which is sicker on average. To reflect this, we may model the mean intercept μ_i for the services hierarchically at the hospital level. Analogously, patients may differ very much between services in one hospital, leading to a larger variance $\sigma_{\beta_{1i}}$, while another hospital may have a more homogeneous patient population leading to a more narrow distribution of regression parameters; we can model this *variation* σ_{β_1} of the mean slope μ_1 within services at a given hospital to capture the variability of PaO_2's predictive effect. We will compare our model to the classical frequentist prediction model currently build by Dr. Gong's statistical team.

Data Acquisition Data will be abstracted from a clinical data warehouse(see Environment and Resources). A multi-prong approach for capturing complete, longitudinal data in real-time, near real-time, or asynchronously from the EMR replica will be used. Montefiore Enterprise Clinical Research Management Systems will provide secured electronic data capture tools to streamline, quality control, normalize, and manage data collection and entry efforts. A fully de-identified, study specific database will be compiled for model development and validation.

$$p(y^{rep}|y) = \int p(y^{rep}|\theta)p(\theta|y)\,\mathrm{d}\theta \quad (5)$$

If our model is a good fit, then data generated by the model using the estimated parameters should have a distribution similar to the original data we observed. We illustrate this idea behind posterior predictive checking[68] in Figure 5, generated in our software package shinyStan[69]:

Model checking We will look at auto-correlation, trace-plots and calculate the Gelman and Rubin's MCMC Convergence Diagnostic $\hat{R}$ to evaluate the the convergence of our Markov chain Monte Carlo (MCMC) simulations using shinyStan, the interactive visual application to graphically explore hierarchical models, we developed[69]. Others reviewed shinyStan's installation and utility on YouTube. In evaluating our Bayesian model's predictive performance, exploratory graphical[70] and confirmatory formal posterior predictive assessment using discrepancies[71] will complement each other to compare the patient test set to simulated replications from our fitted hierarchical Bayesian model. As a simple example, for each draw of the estimated parameter θ from the posterior $p(\theta|y)$ we simulate data y^{rep} from the posterior predictive distribution $p(y^{rep}|y)$. Using the simulations of y^{rep} we can make various graphical displays comparing our observed data to the replications. As a more sophisticated approach we will graphically contrast the vector test statistics $T(y)$ versus replicated data $T(y^{rep})$ to detect a potential misfit of model to data[70;72] Analogously, we will use predictive validation to adjust for overfitting of our model and perform a sensitivity analysis of our priors on key model parameteters[25;68].

Model comparison We will assess the plausibility of our posited hierarchical model and its assumptions[68;71] and will compare it to the alternative classical model by Dr. Gong (based on a non-nested much simpler model[5]). As a simple approach, we will perform a nonparametric comparison of areas under the curve (AUC) of the correlated receiver operating characteristics (ROC) curves[73] to assess their respective predictive performance[74;47], in other words to investigate if the hierarchical modeling improves prediction of acute severe respiratory failure over the simpler classical model used by Dr. Gong. We will compare the models based on a different test set from the same population, to avoid biasing the comparison in favor of our more complicated, (possibly overfitted) hierarchical model. Cross-validation is widely used to compare statistical models for estimating out-of-sample prediction

error[75]. However, in our case, we operate on the limits of computability and repeatedly fitting our Bayesian hierarchical model to leave-one-out samples, could be computationally too expensive[76]. Besides, for multi-level data, leaving partitioning the data for cross-validation should probably consider the hierarchical structure itself; indeed, cross-validation may not always be a sensitive instrument for model comparison[77]. We will also explore the predictive information of our hierarchical model using posterior predicitve simulations and realized discrepancies[76;68;71]. In parallel sampling[78], we will compare our Bayesian Model to the classical simpler algorithms using the minimum χ^2 discrepancy, essentially equivalent to the classical goodness-of-fit test statistic[71]. For additional validation, we will train our model with patient data from other participating institutions (Mayo Clinic Rochester and Florida) and test if our model outperforms the classical prediction models even in other ecological settings (say Mayo Clinic Rochester) or if the classical model is based on data from all institutions.

Technical approach Boolean combinations of data matching and natural language processing of the prediction algorithms will be used to scan a real time copy of the hospital's clinical and administrative data including demographic, monitoring, pharmacy, laboratory, and physician notes for risk factors and physiological abnormality. The rule engine (implemented in Java) will send out the alert to providers.

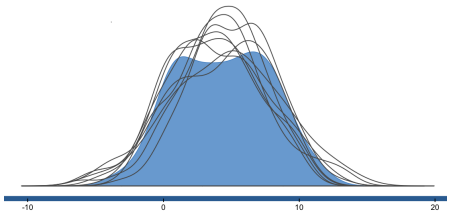

Figure 5: Exploratory posterior predictive check with our software shinyStan[69]: The most basic exploratory graphic[70] is simply a comparison of the entire data set (here the distribution of observed outcomes shaded in blue) to a reference distribution, (distributions of simulated outcomes shown as thin lines). An approximate match suggest a reasonable fit.

For specific aim 1b, we develop new Bayesian data imputation algorithms for missing clinical data using auxiliary data and we identify auxiliary measure properties (ceiling, floor and threshold effects). Missing data are a characteristic limitation of large electronic medical records and may bias our prediction model[17]. Electronically medical records measurements not updated 24 hours earlier than the selected start time will be considered missing; as an illustrative example, we formulated a simplistic model illustrated in [Figure 3]. We predict acute respiratory failure, the dichotomous compound outcome Y in Equation 6; we integrate this with a model for latent arterial oxygen tension Ω in a logistic regression model in Equation 7, contingent on having PaO_2 from an arterial blood gas (ABG) or not. Nota bene, ABGs will certainly not be missing at random, but contingent on the PaO_2 value and respiratory outcome Y.

$$Y \sim Binom(\mu, n); \mu = inv\_logit(\beta_0 + \beta_1 * \Omega) \quad (6)$$

$$\Omega = I(observed = true) * PaO_2 + I(observed = false) * \delta \quad (7)$$

$$\delta \sim Normal(\theta, \tau); \theta = \gamma_0 + \gamma_1 * O_2Sat + \gamma_2 * O_2Therapy \quad (8)$$

If the PaO_2 from the ABG is observed, we will use it to predict the outcome Y. If no ABG was obtained, we impute the latent oxygen tension Ω with a regression model from the auxiliary data O_2 Saturation and O_2 therapy in Equation 8. We will identify the auxiliary measure properties, ceiling and floor and potential threshold effects effects, test the imputations against manually verified data and published algorithms and compare them to simple and multiple imputation strategies in Dr. Gong's trial[79;80].

Aim 2: To model temporality (institutional learning, seasons) and investigate provider compliance.
To focus education efforts and improve implementation of preventive or therapeutic measures, we will investigate predictors of provider behavior. To most closely reflect the realistic situation of actual academic and community medical delivery settings, we need to take temporal and seasonal changes into account.

For specific Aim 2a, we will investigate provider compliance with the individual components of the checklist. During the second phase (PROOFCheck) of Dr. Gong's pragmatic trial, providers of a patient identified as high risk by the frequentist prediction algorithm will be prompted electronically to implement concrete preventive and corrective measures from a list of widely accepted interventions. During roll-out, providers receive targeted education on prevention and best practice. During PROOFCheck, an interactive notification algorithm will suggest to the physicians patient specific interventions from the checklist to the clinicians via an electronic clinical interface.

Prediction of adverse events is useful only if followed by effective preventive action. We will use data from PROOFCheck, the second phase of Dr. Gong's pragmatic trial to analyze provider compliance (fidelity) with the proposed interventions. We will investigate which provider and patient characteristics predict compliance with which components of the intervention checklist to identify drivers of poor provider fidelity. Results will inform our compliance retraining for PROOFCheck in which I will actively participate during my second year.

Population: Hospitalized adults identified by the APPROVE algorithm as high risk and intubated patients will be included in Dr. Gong's PROOFCheck. PROOFCheck will limit recruitment to wards found to have higher prevalence of severe adverse respiratory events during the first phase of Dr. Gong's trial (APPROVE). Patients chronically ventilated at home or who have DNR, will be excluded. I will include all data from all participating centers. We anticipate enrollment of 12,000 patients over 4 years.

Outcomes,exposures and predictors: My primary outcome will be provider compliance, a dichotomous event, defined as positive if the provider ordered the prompted preventive intervention. In order to measure and demonstrate compliance with the checklist, near real time (same day) transaction logs evidence for compliance will be recorded electronically. We will consider time-invariant and time-variant provider and patient demographic and clinical data. Two hypothetical examples of provider and patient demographics as predictors fidelity: (1) junior residents may be less comfortable with stricter blood transfusion triggers compared to seasoned physician assistants; (2) providers fidelity with evidence based treatment recommendations may be contingent on patient gender, say for heart failure[82], likely an important predictor of respiratory failure in APPROVE. Time-variant patient characteristics (e.g. lab values) could determine provider fidelity; for example borderline blood hemoglobin concentration may influence compliance with PROOFCheck blood transfusion recommendations.

Study design and model building This is a prospective observational cohort study to investigate sustained provider fidelity with EMR-triggered preventive interventions in PROOFCheck, Dr. Gong's pragmatic multicenter trial. We will build a Bayesian hierarchical multivariate logistic regression model of time-invariant and time-variant demographic, clinical and administrative variables, with levels for service, ward and institution, analogously to aim 1; the hierarchical structure is to reflect certain biases and attitudes ingrained in certain specialties or hospitals, which may lead to different associations between provider and patient characteristics and fidelity with treatment alerts, e.g. service-specific reluctance to use triggers to minimize blood cell transfusion[83].

For specific Aim 2b, to reflect changing risk profiles over time, we will adjust our Bayesian model to update continuously with new incoming patients and adapt our model to include temporal effects, like institutional learning, seasonal or endemic phenomena. Seasonal changes could for example be modeled by adding another level above the hospital level to our patient-service-hospital hierarchy illustrated in Figure 2. This would allow the hospital level mean to vary over time to reflect increases in the severity and prevalence of chronic obstructive pulmonary disease in the winter or to smooth over differences in annual flu prevalence.

Preliminary work and feasibility of our research plan

<table>
<tr><td colspan="2">Pre K-award period</td><td rowspan="2">2016
K01 starts</td><td>Q1</td><td>Q2</td><td>Q3</td><td>Q4</td><td rowspan="2">2017</td><td>Q1</td><td>Q2</td><td>Q2</td><td>Q4</td><td rowspan="2">2018</td><td>Q1</td><td>Q2</td><td>Q3</td><td>Q4</td></tr>
<tr><td colspan="2" rowspan="2">IRB approval, extract and select variables, implement e-triggers</td><td colspan="4">Aims 1a and 1b →</td><td>Aim 1→</td><td colspan="3">Aims 2a</td><td colspan="4">Aim 2b</td></tr>
<tr><td colspan="4">Joint (incomplete data and prediction) model</td><td colspan="4">Investigate provider compliance to inform training</td><td colspan="4">Integrate temporal effects like seasons and provider learning</td></tr>
<tr><td colspan="2">APPROVE</td><td rowspan="2">U1</td><td colspan="5">PROOFCheck</td><td colspan="6">PROOFCheck</td><td colspan="4">PROOFCheck --→ →</td></tr>
<tr><td>Fit</td><td>Validate</td><td>Train</td><td colspan="2">Randomize</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td colspan="2">Compliance</td><td colspan="2">RCT completed</td><td>Follow up</td></tr>
</table>

Timeline, detailing quarterly progress through the K01 training period: My research plan is well aligned with APRROVE and PROOFCheck, my mentor's trial. Time consuming preliminary work (IRB approval, computerized data collection and cleaning, aggregation and standardization, identification of important predictors of respiratory failure) is already well under way. Cluster randomization [for hospital 1-6] begins soon after my K01 starts. Aim 1 will have considerable overlap into the second year, when concurrent fidelity analysis will inform compliance retraining for PROOFCheck. Rich data will sustain my final integration of temporal effects in the model.

Successful execution of my research plan is facilitated by its integration in my mentor's trial, illustrated above. We have already abstracted data from 68,000 patients from Monte and Mayo for derivation and validation for APPROVE. The computability of our Bayesian model hinges on its effective computational implementation. My co-mentor, Dr. Gelman is personally invested in the realization of cutting-edge Bayesian models through our allied R01 research project. Several standalone components of my research proposal will lead to high impact publications: missing data imputation using auxiliary data is novel, as is the analysis of poor provider compliance.

My research is well aligned with NIH funding opportunities, institutional priorities and emerging paradigms

Together with my mentors Drs. Gong, Gelman and Hall, we are working on an related R01 application to further develop Bayesian computational algorithms, using Dr. Gong's trial as use case. Culminating a PhD, my K01 training will give me the competitive edge to lead similar multi-disciplinary NIH applications as early stage principle investigator. I am particularly interested to extend our Bayesian tools to the "Perioperative Surgical Home"[84].