Specific Aims

Acute respiratory failure (ARF) requiring mechanical ventilation is common in hospitalized patients; prolonged mechanical ventilation often leads to multi-organ failure. We will model electronic medical record (EMR) and pragmatic clinical trial data to predict and to prevent acute severe respiratory failure in hospitalized patients.

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 two phase pragmatic clinical trial: APPROVE, phase 1, develops a classical algorithm to identify patients at risk; PROOFCheck, phase 2, aims to improve outcomes by triggering a prevention checklist targeting those patients the APPROVE algorithm identifies.

Hierarchical modeling may be transformative for EMR-based prediction and prevention and exploit the nested hierarchical granularity typical for EMRs. We propose to fit a more sophisticated hierarchical prediction algorithm than currently developed in APPROVE; we propose (a) to allow model parameters to vary between patients, medical floors, services or institutions and (b) to model temporal effects, e.g. seasonal effects, shifting population characteristics or heterogeneous provider behavior which might otherwise limit prediction accuracy. Incomplete clinical data limit prediction algorithms, but are characteristic for EMRs. I will develop new data imputation algorithms using auxiliary data, a novel approach to overcome issues with missing at random assumptions.

Hierarchical models improve prediction over classical approaches owing to additional information , gained from (a) modeling the rich spatial and temporal organization of EMRs more realistically, (b) imputing incomplete data from auxiliary data and (c) partial pooling. Patients treated by the same team, in similar settings will show similar clinical trajectories and responses. Partial pooling will improve precision and accuracy by informing parameter estimates with data from all other patients, using information from different but related subsets, especially in subgroups with sparse data. The near real-time *integration* of auxiliary data imputation and hierarchical modeling with partial pooling into one coherent EMR-surveillance model is groundbreaking.

Novel algorithms push the envelope of computability for Bayesian prediction models. We choose Bayesian inference, novel for EMR prediction, for its flexibility in hierarchical modeling. Computational implementation can be challenging. My co-mentor Dr. Gelman is leading the NSF-funded development of the probabilistic programming language Stan. His novel algorithm achieves much faster model convergence and parameter estimation. My second co-mentor Dr Hall is also a seasoned Bayesian statistician. He will supervise me for clever statistical formulation or transformation to further push the boundaries of computability for large EMRs. My exceptional and multidisciplinary team of mentors is lead by Dr. Gong with her clinical angle on Big Data science. Together, we will integrate innovative approaches to data imputation with advanced hierarchical prediction models to form a near real-time EMR-based clinical decision tool with practical utility in critical care. The integration of pioneering statistical modeling with pragmatic clinical EMR-surveillance constitutes our unique innovation.

Hypothesis: *Compared to classical approaches, joint hierarchical Bayesian models improve prediction, prevention and compliance analysis in a pragmatic trial to prevent respiratory failure in hospitalized patients.*

Specific aims

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

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 inpatients.
SA 1b: To further develop Bayesian data imputation algorithms of missing clinical data using auxiliary data, to identify auxiliary measure properties (ceiling, floor and threshold effects). To integrate imputation and prediction in one coherent hierarchical Bayesian model and to assess the predictive performance compared to the classical algorithm by Dr. Gong by the area under the curve (AUC) of their receiver operating characteristics (ROC).

Aim 2: To model temporality (institutional learning, seasons) and investigate provider compliance.

SA 2a: To investigate patient and provider characteristics as drivers of poor provider compliance in PROOFCheck, Dr. Gong's pragmatic trial, to inform the ongoing PROOFCheck trial implementation and to focus our retraining.
SA 2b: To update our model continuously with new incoming patients to reflect their changing risk profile and to model institutional learning and temporal effects like seasons and endemics.

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 [Figure 1] [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].

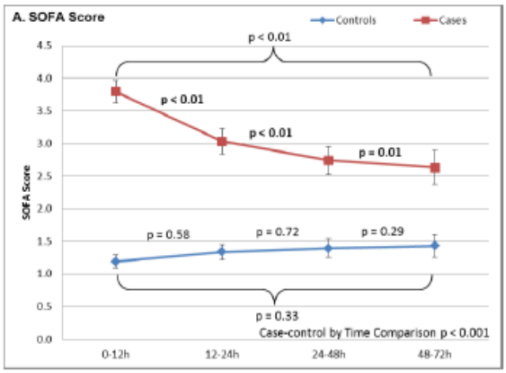

Figure 1: Deterioration of the Sequential Organ Failure Assessment score (SOFA) can be detected 24-48 hours before clinical deterioration leads to ICU admission; p-values reflect pairwise comparisons between consecutive time intervals, adjusting for patient characteristics. [1].

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]. CITE CLINICAL TRIALS.GOV. *Hypothesis:* Early implementation of a checklist of preventive measures (1), reduces severity of organ failure, mortality and duration of mechanical ventilationin 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" approach, as can be shown mathematically or via cross-validation [24]. Using (a) 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 (b) "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. [25].

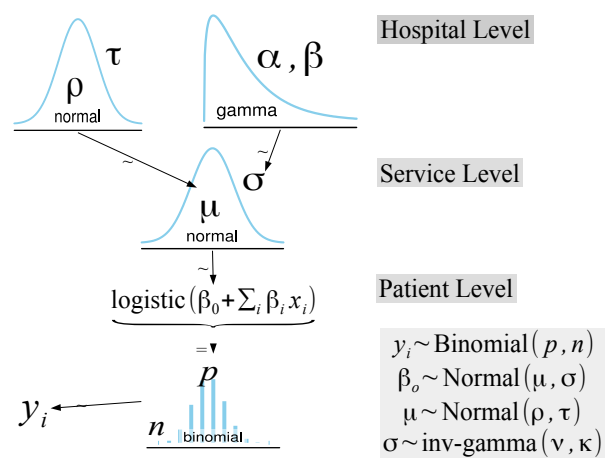

Figure 2: Distrogram to illustrated the hierarchical structure of patient trajectories. Patient level regression coefficient β_0 varies across medical services, service level mean μ and within-service variance σ vary by hospital.

We hypothesize that hierarchical modeling may better identify hospitalized patients at risk for acute respiratory failure leading to prolonged mechanical ventilation or death than the APPROVE prediction algorithm.

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) [26]. However, this is an unreasonable assumption for EMRs; in our example, physicians will request 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.

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 [27]. 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].

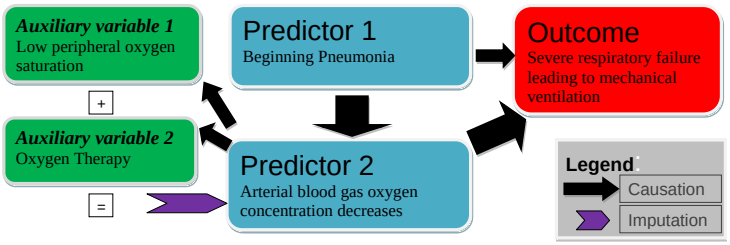

Figure 3: Incomplete data can hinder outcome prediction, but we can impute incomplete data from auxilliary information. For example, pneumonia (causing to low oxygen tension), may cause respiratory failure. If arterial blood gas results are missing, we can impute the oxygen tension from oxygen therapy and/or peripheral oxygen saturation. [28].

This approach avoids the perils associated with missing at random (MAR) assumptions, when fitting a non-ignorable missingness model [29]. 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 [30, 31, 32], often relying on Bayesian methods for the multiple imputations approach [33, 34]; joint hierarchical modeling, including the use of auxiliary data to impute incomplete patient records will improve the prediction model and facilitate smoother implementation of the algorithm into the clinical trial [28].

We hypothesize that joint hierarchical modeling, incorporating the use of auxiliary data to impute incomplete patient records, will improve our prediction model compared to classical prediction with multiple imputation.

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 therefore include institutional learning, seasonal effects 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?

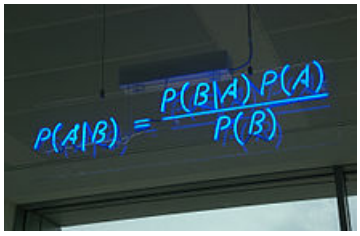

Figure 4: Bayes Theorem [35] The posterior probability $P(A|B)$ is the prior probability $P(A)$ updated with the likelihood $P(A)$

B. Innovation

Bayesian hierarchical modeling is groundbreaking in EMR-based prediction, and particularly suited for joint hierarchical modeling. With their inherent flexibility and robustness [36, 37], Bayesian hierarchical models may outperform classical models for EMR-based prediction owing to the integration of additional information through "partial pooling" [38] and the imputation of incomplete records from auxiliary data. Massive increases in computer power and better Markov Chain Monte Carlo algorithms led to an expansion of applied Bayesian work

expanded rapidly [39, 40], also in Big Data [41] and more recently in EMR-based prediction [42, 43, 44]. *However, we are unaware of any Bayesian hierarchical prediction model based on large EMRs.*

A brief introduction to Bayesian inference and its computational implementation. According to Bayes' Theorem [35], shown in Figure 4, prior information $P(A)$ is combined with new data $P(B)$, (known as the likelihood) to yield an *updated* estimate for the probability of a hypothesis $P(A)$ given the data $P(B)$, called the posterior distribution $P(A|B)$ [45]. The Bayesian approach is analogous to clinical decision-making [40]. Physicians continuously update their preliminary diagnosis as new information comes in. Prior belief $P(A)$ in a diagnosis may be weakened by new laboratory information $P(B)$, leading to an updated diagnosis given the lab data $P(A|B)$ [46]. Bayesian inference for multi-layered (hierarchical) 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 [47]: *Instead of analyzing the odds, we simulate throwing the dice repeatedly.* But MCMC simulations can be slow to converge for large EMRs.

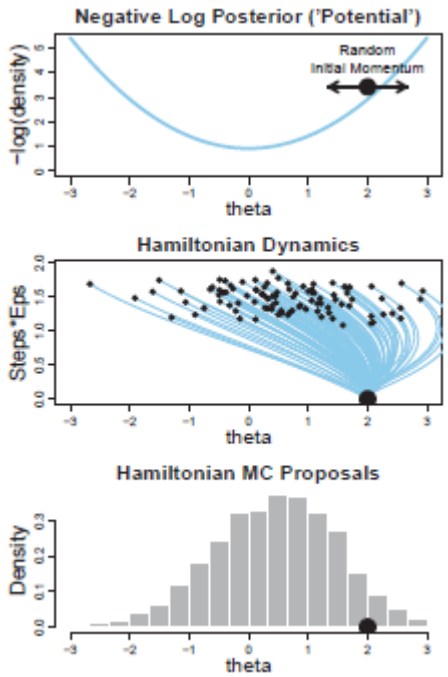

Figure 5: Hamiltonian MCMC uses momentum to optimize the next proposal. The higher momentum of the current position (black dot) is indicated in the top panel. The middle panel illustrates how random samples are drawn to the mode of the posterior distribution (lower panel) leading to faster convergence and higher effective sample size. Fig 14.1 in Kruschke [45].

Pushing the envelope of Bayesian inference for 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 [48]. Stan's Hamiltonian Monte Carlo algorithms [48] and clever statistical formulation push the boundaries of computability [24]. 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 [24]. Stan is based on Hamiltonian Monte Carlo (HMC) [24], a Marcov chain Monte Carlo (MCMC) algorithm [49], which avoids the sensitivity to correlated parameters that plague many MCMC methods by introducing auxiliary momentum variables [50] as illustrated in Figure 5. HMC is dependent on tuning the reciprocal relationship of the crucial parameters step size and desired number of steps. If the latter is too large, efficiency is low, if to small we see undesirable random walk behavior. To overcome this, Dr. Gelman's team developed and implemented the No-U-Turn Sampler (NUTS), a recursive algorithm to automate and optimize tuning the HMC [50].

Analyzing and advancing implementation is decisive for outcome improvement and research. Imperfect fidelity (poor 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 [51]. 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 [52] and little is known on how to reproduce multi-faceted interventions (specially directed toward providers) to improve clinical outcomes [53]. 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 [54], we ignore the best means to translate widely accepted interventions and new findings of outcomes research into practice [55]. 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 accepted evidence based best practices.

We use a pragmatic trial to investigate provider fidelity. Pragmatic trials like Dr. Gong's may result in valid more estimates of effectiveness for more realistic health care scenarios [56, 57]; we will use her pragmatic trial data to investigate incomplete fidelity, heterogeneity and difficulty in clinical implementation. An example with problematic fidelity relevant for PROOFCheck is blood product management: transfusions increase the risk of acute severe respiratory failure with mechanical ventilation [58], but implementation of rational transfusion blood product management is still sketchy and very heterogeneous across the nation [61]. 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 [62]. 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 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 perioperative 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 the study period, excluding only those who are chronically ventilated at home or who have Do not resuscitate orders at the time of hospital admission (Table 1: Inpatient population at Montefiore Medical Center.) As part of APPROVE, the pragmatic trial by Dr. Gong described in detail under Significance, two 3-month, prospective, observational cohort studies are underway at several Albert Einstein College of medicine and Mayo Clinic Rochester and Florida sites; we will build our Bayesian hierarchical model based solely on Einstein patients: Patients from the first three months will serve as the fitting cohort, while the 2nd cohort will serve as the validation and test set.

Predictors: Many independent variables are candidates for potential inclusion into our Bayesian hierarchical model. From the multicenter LIPS study, clinical factors most closely associated with prolonged mechanical ventilation have already been identified [5]. We will consider these and additional time-invariant and time-variant demographic and clinical data. Examples for demographics are gender, age, medical service or ward, examples for physiological and clinical predictors are heart rate, blood pressure or lab tests, respectively. Certain predictors will require summary aggregations and (logarithmic) transformations to induce variance stability.

Outcomes: Our primary dichotomous outcome will be acute respiratory failure requiring mechanical ventilation longer than 48 hours. Outcomes are 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 will also consider other random effects for example for co-morbidity and other time-invariant patient specific descriptors. We illustrate this nested structure in a simple logistic model with hierarchical levels for patient, service and hospital.

$$Y \sim Binom(\alpha, n); \alpha = inv_log(\beta_0 + \beta_1 * PaO_2) \quad (1)$$

Patient level (1) On the left, we model at the patient level, the probability $alpha$ that a patient will develop the dichotomous event Y, acute respiratory failure requiring mechanical ventilation, using arterial oxygen tension PaO_2 as a predictor in a simple logistic regression model. However, patients are typically assigned to hospital services. Pulmonary service patients may have a lower baseline PaO_2, while surgical patients tend to have normal lung function.

$$\beta_0 \sim Normal(\gamma_0, \tau_{\beta_0}); \beta_1 \sim Normal(\gamma_1, \tau_{\beta_1}) \quad (2)$$

Service level (2) On the left we develop our hierarchical model to allow different intercepts β_0 representing the average PaO_2 of patients in the various medical and surgical services estimating different service level mean intercepts γ_0. Under a medical service, smaller changes in PaO_2 may be indicative of respiratory deterioration compared to a surgical service, where larger drop in arterial oxygen tension predicts outcome. We may allow the regression coefficient for the slope β_1 to vary around different mean slopes γ_1 at the service level.

$$\gamma_0 \sim Normal(\delta_{tau_0}, \zeta); \tau_{\beta_1} \sim Normal(\delta_{tau_1}, \kappa) \quad (3)$$

Hospital level (3) Some hospitals may cater to an economically disadvantages population, which is sicker on average. To reflect this, we may model the mean intercept γ_0 for the services hierarchically at the hospital level. Patients differ very much between services in some hospitals, while others may have a more homogeneous patient distribution; we can model this *variation* τ_{β_1} of the mean slope γ_1 within services at a given hospital to capture the variability of PaO_2's predictive effect.

In APPROVE, the compared frequentist prediction model will be build by Dr. Gong's statistical team employing the published LIPS score [5] to identify hospitalized patients at risk for prolonged mechanical ventilation and death. The score at the selected start time will be used to determine the best cutoff score to identify the patients with the highest risk of developing prolonged mechanical ventilation.

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. When possible, and where collection of additional complementary information is warranted, we will use the Retrieve Form for Data Capture IS THIS THE CORRECT REFERENCE?[63],an IHE30 IS THIS THE CORRECT REFERENCE? [64] standard for gathering new data within a user's current application environment (EMR in this case) to meet the requirements of an external system. Secured electronic data capture tools provided by the Montefiore Enterprise Clinical Research Management System will be used to streamline, quality control, normalize, and manage data collection and data entry efforts. A fully de-identified, study specific database of all study variables will be compiled for model development and validation. When subsequently patient data from additional regional medical centers are incorporated.

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 [65]. Others reviewed shinyStan's installation and utility on YouTube. In evaluating our Bayesian model's predictive performance, exploratory graphical [66] and confirmatory formal posterior predictive assessment using discrepancies [67] 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:

$$p(y^{rep}|y) = \int p(y^{rep}|\theta)p(\theta|y)\,\mathrm{d}\theta \quad (4)$$

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 6, generated in our software package shinyStan [65]: 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 [66, 69] 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 [24, 68].

Model comparison We will assess the plausibility of our posited hierarchical model and its assumptions [68, 67] 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 [70] to assess their respective predictive performance [71, 44], 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 [72]. 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 [73]. 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 [74]. We will also explore the predictive information of our hierarchical model using posterior predicitve simulations and realized discrepancies [73, 68, 67]. In parallel sampling [75], 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 [67]. 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.

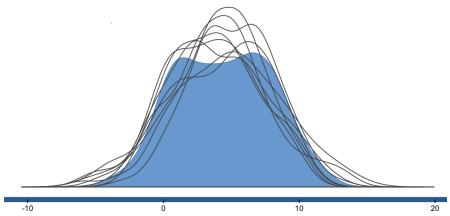

Figure 6: Exploratory posterior predictive check with our software shinyStan [65]: The most basic exploratory graphic [66] is simply a comparison of the entire data set (here the distribution of observed outcomes shaded in blue) to a reference distribution, (distributions of simulated outcomes shown as thin lines). An approximate match suggest a reasonable fit.

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.

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 combine the prediction of the dichotomous compound outcome Y, (defined as acute respiratory failure leading to mechanical ventilation and death) using latent arterial oxygen tension Ω in a logistic regression model in Equation 5 and 6:

$$Y \sim Binom(\mu, n) \quad (5)$$

$$\mu = inv_log(\beta_0 + \beta_1 * \Omega) \quad (6)$$

We may have the 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. 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:

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

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

Our imputation approach exploits the temporal relationship between variables in the longitudinal electronic medical records [19]. 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 the simple and multiple imputation strategies planned for Dr. Gong's pragmatic trial [76, 77]. We will perform posterior cross validation checks to investigate the appropriateness of our assumptions and incomplete data model [78].

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 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. Hospital beds and annual numbers of ventilated patients for each hospital are presented in Table 2.

Hospital	Beds	Vent[a]
Montefiore	620	815
Weiler	396	xx
Wakefield	369	xx
Mayo Clinic	xxx	440
BIDMC	646	2,204
Total	xxx	xx

Table 2: PROOFCheck hospital population ([a]patients ventilated)

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 [79], 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 medical or surgical specialties or hospital wards, which may lead to different associations between provider and patient characteristics and provider fidelity with treatment prompts, for example service-specific reluctance to use triggers to minimize blood cell transfusion [80].

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 learn-

ing, seasonal or endemic phenomena. Seasonal changes could for example be modeled by adding another level above the hospital level modeled in Equation 3 to our patient-service-hospital hierarchy illustrated in Figure 2:

Equation 9 would thus 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.

$$\delta_{tau_0} \sim Normal(\phi, \psi) \quad (9)$$

Limitations 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 colspan="4">Aim 1→ 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="4">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 broad research plan is facilitated by its integration in APPROVE and PROOFCheck, my mentor's ongoing trial, as illustrated in the Timeline above. 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; developing new missing data imputation using auxiliary data is novel, as is the analysis and improvement 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 aligned R01 application to the BD2K initiative in response to FAO PA-14-156 to further develop Bayesian computational algorithms for the software Stan, using Dr. Gong's trial as use case. This K01 training and the PhD will give me the competitive edge to lead similar multi-disciplinary NIH applications as early stage principle investigator. Entering the field of Bayesian EMR-based prediction at its dawn with such a rigorous training will give me the opportunity to establish myself as a leading investigator and makes me a desirable collaborator and key player in my institution. I am particularly interested to extend our Bayesian tools to the "Perioperative Surgical Home" [81], where I published before [82] incorporating record matching across our regional academic databases [83].

References

[1] S. Yu, S. Leung, M. Heo, G. J. Soto, R. T. Shah, S. Gunda, and M. N. Gong. Comparison of risk prediction scoring systems for ward patients: a retrospective nested case-control study. *Crit Care*, 18(3):R132, 2014. PMID: 24970344.

[2] H. Wunsch, W. T. Linde-Zwirble, D. C. Angus, M. E. Hartman, E. B. Milbrandt, and J. M. Kahn. The epidemiology of mechanical ventilation use in the United States. *Crit Care Med*, 38(10):1947–1953, Oct 2010. PMID: 20639743.

[3] V. M. Ranieri, F. Giunta, P. M. Suter, and A. S. Slutsky. Mechanical ventilation as a mediator of multisystem organ failure in acute respiratory distress syndrome. *JAMA*, 284(1):43–44, Jul 2000. PMID: 10872010.

[4] J. M. Rohde, A. J. Odden, C. Bonham, L. Kuhn, P. N. Malani, L. M. Chen, S. A. Flanders, and T. J. Iwashyna. The epidemiology of acute organ system dysfunction from severe sepsis outside of the intensive care unit. *J Hosp Med*, 8(5):243–247, May 2013. PMID: 23401431.

[5] M. S. Herridge, A. M. Cheung, C. M. Tansey, A. Matte-Martyn, N. Diaz-Granados, F. Al-Saidi, A. B. Cooper, C. B. Guest, C. D. Mazer, S. Mehta, T. E. Stewart, A. Barr, D. Cook, A. S. Slutsky, and C. C. C. T. G. . One-year outcomes in survivors of the acute respiratory distress syndrome. *N Engl J Med*, 348(8):683–693, Feb 2003. PMID: 12594312.

[6] K. M. Hillman, P. J. Bristow, T. Chey, K. Daffurn, T. Jacques, S. L. Norman, G. F. Bishop, and G. Simmons. Duration of life-threatening antecedents prior to intensive care admission. *Intensive Care Med*, 28(11):1629–1634, Nov 2002. PMID: 12415452.

[7] P. McQuillan, S. Pilkington, A. Allan, B. Taylor, A. Short, G. Morgan, M. Nielsen, D. Barrett, G. Smith, and C. H. Collins. Confidential inquiry into quality of care before admission to intensive care. *BMJ*, 316(7148):1853–1858, Jun 1998. PMID: 9632403.

[8] N. Naeem and H. Montenegro. Beyond the intensive care unit: a review of interventions aimed at anticipating and preventing in-hospital cardiopulmonary arrest. *Resuscitation*, 67(1):13–23, Oct 2005. PMID: 16150531.

[9] E. Rivers, B. Nguyen, S. Havstad, J. Ressler, A. Muzzin, B. Knoblich, E. Peterson, M. Tomlanovich, and E. G.-D. T. C. G. . Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med*, 345(19):1368–1377, Nov 2001. PMID: 11794169.

[10] E. P. Rivers. Early goal-directed therapy in severe sepsis and septic shock: converting science to reality. *Chest*, 129(2):217–218, Feb 2006. PMID: 16478830.

[11] I. A. Mitchell, H. McKay, C. Van Leuvan, R. Berry, C. McCutcheon, B. Avard, N. Slater, T. Neeman, and P. Lamberth. A prospective controlled trial of the effect of a multi-faceted intervention on early recognition and intervention in deteriorating hospital patients. *Resuscitation*, 81(6):658–666, Jun 2010. PMID: 20378235.

[12] M. M. Levy, A. Artigas, G. S. Phillips, A. Rhodes, R. Beale, T. Osborn, J.-L. Vincent, S. Townsend, S. Lemeshow, and R. P. Dellinger. Outcomes of the Surviving Sepsis Campaign in intensive care units in the USA and Europe: a prospective cohort study. *Lancet Infect Dis*, 12(12):919–924, Dec 2012. PMID: 23103175.

[13] W. S. Lim, S. V. Baudouin, R. C. George, A. T. Hill, C. Jamieson, I. Le Jeune, J. T. Macfarlane, R. C. Read, H. J. Roberts, M. L. Levy, M. Wani, M. A. Woodhead, and P. G. C. o. t. B. T. S. S. o. C. C. . BTS guidelines for the management of community acquired pneumonia in adults: update 2009. *Thorax*, 64 Suppl 3:iii1–ii55, Oct 2009. PMID: 19783532.

[14] J. Barr, G. L. Fraser, K. Puntillo, E. W. Ely, C. Gélinas, J. F. Dasta, J. E. Davidson, J. W. Devlin, J. P. Kress, A. M. Joffe, D. B. Coursin, D. L. Herr, A. Tung, B. R. H. Robinson, D. K. Fontaine, M. A. Ramsay, R. R.

Riker, C. N. Sessler, B. Pun, Y. Skrobik, R. Jaeschke, and A. C. o. C. C. M. . Clinical practice guidelines for the management of pain, agitation, and delirium in adult patients in the intensive care unit. *Crit Care Med*, 41(1):263–306, Jan 2013. PMID: 23269131.

[15] T. D. Girard, J. P. Kress, B. D. Fuchs, J. W. W. Thomason, W. D. Schweickert, B. T. Pun, D. B. Taichman, J. G. Dunn, A. S. Pohlman, P. A. Kinniry, J. C. Jackson, A. E. Canonico, R. W. Light, A. K. Shintani, J. L. Thompson, S. M. Gordon, J. B. Hall, R. S. Dittus, G. R. Bernard, and E. W. Ely. Efficacy and safety of a paired sedation and ventilator weaning protocol for mechanically ventilated patients in intensive care (Awakening and Breathing Controlled trial): a randomised controlled trial. *Lancet*, 371(9607):126–134, Jan 2008. PMID: 18191684.

[16] P. C. Hébert, G. Wells, M. A. Blajchman, J. Marshall, C. Martin, G. Pagliarello, M. Tweeddale, I. Schweitzer, and E. Yetisir. A multicenter, randomized, controlled clinical trial of transfusion requirements in critical care. Transfusion Requirements in Critical Care Investigators, Canadian Critical Care Trials Group. *N Engl J Med*, 340(6):409–417, Feb 1999. PMID: 9971864.

[17] B. B. Dean, J. Lam, J. L. Natoli, Q. Butler, D. Aguilar, and R. J. Nordyke. Review: use of electronic medical records for health outcomes research: a literature review. *Med Care Res Rev*, 66(6):611–638, Dec 2009. PMID: 19279318.

[18] R. Amarasingham, B. J. Moore, Y. P. Tabak, M. H. Drazner, C. A. Clark, S. Zhang, W. G. Reed, T. S. Swanson, Y. Ma, and E. A. Halm. An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data. *Med Care*, 48(11):981–988, Nov 2010. PMID: 20940649.

[19] C. A. Welch, I. Petersen, J. W. Bartlett, I. R. White, L. Marston, R. W. Morris, I. Nazareth, K. Walters, and J. Carpenter. Evaluation of two-fold fully conditional specification multiple imputation for longitudinal electronic health record data. *Stat Med*, 33(21):3725–3737, Sep 2014. PMID: 24782349.

[20] L. Smeeth, S. L. Thomas, A. J. Hall, R. Hubbard, P. Farrington, and P. Vallance. Risk of myocardial infarction and stroke after acute infection or vaccination. *N Engl J Med*, 351(25):2611–2618, Dec 2004. PMID: 15602021.

[21] S. Davé, I. Petersen, L. Sherr, and I. Nazareth. Incidence of maternal and paternal depression in primary care: a cohort study using a primary care database. *Arch Pediatr Adolesc Med*, 164(11):1038–1044, Nov 2010. PMID: 20819960.

[22] S.-L. Man, I. Petersen, M. Thompson, and I. Nazareth. Antiepileptic drugs during pregnancy in primary care: a UK population based study. *PLoS One*, 7(12):e52339, 2012. PMID: 23272239.

[23] J. Bafumi and A. Gelman. Fitting Multilevel Models When Predictors and Group Effects Correlate. SSRN Scholarly Paper ID 1010095, Social Science Research Network, Rochester, NY, September 2007.

[24] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian data analysis*, volume 2. Taylor & Francis, 2014.

[25] A. Gelman. Multilevel (Hierarchical) Modeling: What It Can and Cannot Do. *Technometrics*, 48(3):432–435, August 2006.

[26] D. B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

[27] M. J. Daniels, C. Wang, and B. H. Marcus. Fully Bayesian inference under ignorable missingness in the presence of auxiliary covariates. *Biometrics*, 70(1):62–72, Mar 2014. PMID: 24571539.

[28] C. B. Hall, R. B. Lipton, M. J. Katz, and C. Wang. Correcting Bias Caused by Missing Data in the Estimate of the Effect of Apolipoprotein epsilon 4 on Cognitive Decline. *J Int Neuropsychol Soc*, pages 1–6, Nov 2014. PMID: 25389642.

[29] C. Wang and C. B. Hall. Correction of bias from non-random missing longitudinal data using auxiliary information. *Stat Med*, 29(6):671–679, Mar 2010. PMID: 20029935.

[30] X.-L. Meng. Multiple-Imputation Inferences with Uncongenial Sources of Input. *Statist. Sci.*, 9(4):538–558, 11 1994.

[31] L. M. Collins, J. L. Schafer, and C. M. Kam. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychol Methods*, 6(4):330–351, Dec 2001. PMID: 11778676.

[32] D. B. Rubin. Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 91(434):473–489, 1996.

[33] M. J. Daniels and J. W. Hogan. *Missing data in longitudinal studies: Strategies for Bayesian modeling and sensitivity analysis*. CRC Press, 2008.

[34] J. L. Schafer. *Analysis of incomplete multivariate data*. CRC press, 1997.

[35] M. Bayes and M. Price. An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, frs communicated by mr. price, in a letter to john canton, amfrs. *Philosophical Transactions (1683-1775)*, pages 370–418, 1763.

[36] J. B. Carlin. Meta-analysis for 2 x 2 tables: a Bayesian approach. *Stat Med*, 11(2):141–158, Jan 1992. PMID: 1349763.

[37] A. J. Sutton, N. J. Cooper, K. R. Abrams, and A. Ades. *Evidence synthesis for decision making in healthcare*, volume 132. John Wiley & Sons, 2012.

[38] A. Gelman. *Red state, blue state, rich state, poor state: why Americans vote the way they do*. Princeton University Press, 2009.

[39] D. Ashby. Bayesian statistics in medicine: a 25 year review. *Stat Med*, 25(21):3589–3631, Nov 2006. PMID: 16947924.

[40] D. J. Spiegelhalter, J. P. Myles, D. R. Jones, and K. R. Abrams. Bayesian methods in health technology assessment: a review. *Health Technol Assess*, 4(38):1–130, 2000. PMID: 11134920.

[41] C. Yoo, L. Ramirez, and J. Liuzzi. Big data analysis using modern statistical and machine learning methods in medicine. *Int Neurourol J*, 18(2):50–57, Jun 2014. PMID: 24987556.

[42] B. E. Himes, Y. Dai, I. S. Kohane, S. T. Weiss, and M. F. Ramoni. Prediction of chronic obstructive pulmonary disease (COPD) in asthma patients using electronic medical records. *J Am Med Inform Assoc*, 16(3):371–379, 2009. PMID: 19261943.

[43] O.-P. Ryynänen, E. J. Soini, A. Lindqvist, M. Kilpeläinen, and T. Laitinen. Bayesian predictors of very poor health related quality of life and mortality in patients with COPD. *BMC Med Inform Decis Mak*, 13:34, 2013. PMID: 23496851.

[44] J. Wu, J. Roy, and W. F. Stewart. Prediction modeling using EHR data: challenges, strategies, and a comparison of machine learning approaches. *Med Care*, 48(6 Suppl):S106–S113, Jun 2010. PMID: 20473190.

[45] J. Kruschke. *Doing Bayesian data analysis: A tutorial introduction with R*. Academic Press, 2014.

[46] J. K. Kruschke. Bayesian estimation supersedes the t test. *J Exp Psychol Gen*, 142(2):573–603, May 2013. PMID: 22774788.

[47] K. Abrams and B. Sansó. Approximate Bayesian inference for random effects meta-analysis. *Stat Med*, 17(2):201–218, Jan 1998. PMID: 9483729.

[48] Stan Development Team. Stan: A C++ Library for Probability and Sampling, Version 2.5.0, 2014.

[49] N. Metropolis and S. Ulam. The monte carlo method. *Journal of the American statistical association*, 44(247):335–341, 1949.

[50] M. D. Homan and A. Gelman. The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *The Journal of Machine Learning Research*, 15(1):1593–1623, 2014.

[51] M. M. Duncan, P. A. McIntosh, C. D. Stayton, and C. B. Hall. Individualized performance feedback to increase prenatal domestic violence screening. *Matern Child Health J*, 10(5):443–449, Sep 2006. PMID: 16710766.

[52] D. A. Davis, M. A. Thomson, A. D. Oxman, and R. B. Haynes. Changing physician performance. A systematic review of the effect of continuing medical education strategies. *JAMA*, 274(9):700–705, Sep 1995. PMID: 7650822.

[53] M. Campbell, R. Fitzpatrick, A. Haines, A. L. Kinmonth, P. Sandercock, D. Spiegelhalter, and P. Tyrer. Framework for design and evaluation of complex interventions to improve health. *BMJ*, 321(7262):694–696, Sep 2000. PMID: 10987780.

[54] B. S. Mittman. Creating the evidence base for quality improvement collaboratives. *Ann Intern Med*, 140(11):897–901, Jun 2004. PMID: 15172904.

[55] R. E. Glasgow and K. M. Emmons. How can we increase translation of research into practice? Types of evidence needed. *Annu Rev Public Health*, 28:413–433, 2007. PMID: 17150029.

[56] P. Selby, G. Brosky, P. I. Oh, V. Raymond, and S. Ranger. How pragmatic or explanatory is the randomized, controlled trial? The application and enhancement of the PRECIS tool to the evaluation of a smoking cessation trial. *BMC Med Res Methodol*, 12:101, 2012. PMID: 22824225.

[57] G. Tosh, K. Soares-Weiser, and C. E. Adams. Pragmatic vs explanatory trials: the pragmascope tool to help measure differences in protocols of mental health randomized controlled trials. *Dialogues Clin Neurosci*, 13(2):209–215, 2011. PMID: 21842618.

[58] H. E. Kenz and P. Van der Linden. Transfusion-related acute lung injury. *Eur J Anaesthesiol*, 31(7):345–350, Jul 2014. PMID: 24892308.

[59] M. Ranucci, U. Di Dedda, S. Castelvecchio, L. Menicanti, A. Frigiola, G. Pelissero, S. , and C. O. R. S. Group. Impact of preoperative anemia on outcome in adult cardiac surgery: a propensity-matched analysis. *Ann Thorac Surg*, 94(4):1134–1141, Oct 2012. PMID: 22698773.

[60] Practice guidelines for perioperative blood management: an updated report by the american society of anesthesiologists task force on perioperative blood management. *Anesthesiology*, 122(2):241–275, Feb 2015. PMID: 25545654.

[61] D. S. Likosky, D. C. FitzGerald, R. C. Groom, D. K. Jones, R. A. Baker, K. G. Shann, C. D. Mazer, B. D. Spiess, and S. C. Body. Effect of the perioperative blood transfusion and blood conservation in cardiac surgery clinical practice guidelines of the Society of Thoracic Surgeons and the Society of Cardiovascular Anesthesiologists upon clinical practices. *Anesth Analg*, 111(2):316–323, Aug 2010. PMID: 20488928.

[62] C. H. Weiss, F. Moazed, C. A. McEvoy, B. D. Singer, I. Szleifer, L. A. N. Amaral, M. Kwasny, C. M. Watts, S. D. Persell, D. W. Baker, J. I. Sznajder, and R. G. Wunderink. Prompting physicians to address a daily checklist and process of care and clinical outcomes: a single-site study. *Am J Respir Crit Care Med*, 184(6):680–686, Sep 2011. PMID: 21616996.

[63] H.-B. Rothenhäusler, B. Grieser, G. Nollert, B. Reichart, G. Schelling, and H.-P. Kapfhammer. Psychiatric and psychosocial outcome of cardiac surgery with cardiopulmonary bypass: a prospective 12-month follow-up study. *Gen Hosp Psychiatry*, 27(1):18–28, 2005. PMID: 15694215.

[64] M. Rotte, C. Baerecke, G. Pottag, S. Klose, E. Kanneberg, H.-J. Heinze, and H. Lehnert. Insulin affects the neuronal response in the medial temporal lobe in humans. *Neuroendocrinology*, 81(1):49–55, 2005. PMID: 15809512.

[65] shinyStan Team. shinyStan: R Package for Interactive Exploration of Markov Chain Monte Carlo Output, Version 0.1, 2015.

[66] A. Gelman. Exploratory data analysis for complex models. *Journal of Computational and Graphical Statistics*, 13(4), 2004.

[67] A. Gelman, X.-L. Meng, and H. Stern. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica sinica*, 6(4):733–760, 1996.

[68] A. Gelman, Y. Goegebeur, F. Tuerlinckx, and I. Van Mechelen. Diagnostic Checks for Discrete Data Regression Models Using Posterior Predictive Simulations. *Applied Statistics*, pages 247–268, 2000.

[69] A. Buja, D. Cook, and D. Swayne. Inference for data visualization. In *Talk given at Joint Statistical Meetings*, 1999.

[70] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44(3):837–845, Sep 1988. PMID: 3203132.

[71] P. J. Newcombe, B. H. Reck, J. Sun, G. T. Platek, C. Verzilli, A. K. Kader, S.-T. Kim, F.-C. Hsu, Z. Zhang, S. L. Zheng, V. E. Mooser, L. D. Condreay, C. F. Spraggs, J. C. Whittaker, R. S. Rittmaster, and J. Xu. A comparison of Bayesian and frequentist approaches to incorporating external information for the prediction of prostate cancer risk. *Genet Epidemiol*, 36(1):71–83, Jan 2012. PMID: 22890972.

[72] A. Vehtari and J. Lampinen. Bayesian model assessment and comparison using cross-validation predictive densities. *Neural Comput*, 14(10):2439–2468, Oct 2002. PMID: 12396570.

[73] A. Gelman, J. Hwang, and A. Vehtari. Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6):997–1016, 2014.

[74] W. Wang and A. Gelman. Difficulty of selecting among multilevel models using predictive accuracy, 2014.

[75] P. Congdon. Bayesian predictive model comparison via parallel sampling. *Computational statistics & data analysis*, 48(4):735–753, 2005.

[76] J. L. Huntington and A. Dueck. Handling missing data. *Curr Probl Cancer*, 29(6):317–325, 2005. PMID: 16311133.

[77] J. A. Sloan and A. Dueck. Issues for statisticians in conducting analyses and translating results for quality of life end points in clinical trials. *J Biopharm Stat*, 14(1):73–96, Feb 2004. PMID: 15027501.

[78] A. Gelman, G. King, and C. Liu. Not asked and not answered: Multiple imputation for multiple surveys. *Journal of the American Statistical Association*, 93(443):846–857, 1998.

[79] J. L. Cook, K. L. Grady, M. Colvin, S. M. Joseph, M. A. Brisco, M. N. Walsh, and f. t. g. A. D. W. G. . Sex Differences in the Care of Patients With Advanced Heart Failure. *Circ Cardiovasc Qual Outcomes*, Feb 2015. PMID: 25714825.

[80] L. T. Goodnough, J. H. Levy, and M. F. Murphy. Concepts of blood transfusion in adults. *Lancet*, 381(9880):1845–1854, May 2013. PMID: 23706801.

[81] T. R. Vetter, A. M. Boudreaux, K. A. Jones, J. M. Hunter, Jr, and J.-F. Pittet. The perioperative surgical home: how anesthesiology can collaboratively achieve and leverage the triple aim in health care. *Anesth Analg*, 118(5):1131–1136, May 2014. PMID: 24781579.

[82] M. H. Andreae and D. A. Andreae. Regional anaesthesia to prevent chronic pain after surgery: a Cochrane systematic review and meta-analysis. *Br J Anaesth*, 111(5):711–720, Nov 2013. PMID: 23811426.

[83] R. Kaushal, G. Hripcsak, D. D. Ascheim, T. Bloom, T. R. Campion, Jr, A. L. Caplan, B. P. Currie, T. Check, E. L. Deland, M. N. Gourevitch, R. Hart, C. R. Horowitz, I. Kastenbaum, A. A. Levin, A. F. H. Low, P. Meissner, P. Mirhaji, H. A. Pincus, C. Scaglione, D. Shelley, J. N. Tobin, and N. Y. C.-C. D. R. N. . Changing the research landscape: the New York City Clinical Data Research Network. *J Am Med Inform Assoc*, 21(4):587–590, 2014. PMID: 24821739.