Specific Aims

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

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

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

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

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

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

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

Specific aims:

Aim 1: To improve early prediction of prolonged respiratory failure and death in hospitalized patients.

We will implement a complex hierarchical Bayesian prediction algorithm, comparing it to the classical model.

SA 1a: To build a pragmatic EMR based hierarchical Bayesian model implemented in the ultra-fast statistical software Stan to predict a composite outcome [death or prolonged mechanical ventilation > 48 hours] in our Montefiore Medical Center inpatients and compare it with the existing frequentist algorithm.

SA 1b: To further develop Bayesian data imputation algorithms of missing clinical data using auxiliary data, to identify the auxiliary measure properties, ceiling and floor effects and to test the imputations against manually verified data and published algorithms.

Aim 2: To integrate a complex Bayesian model into patient triage and treatment implementation.

Our prediction algorithm will trigger individualized patient interventions in Dr. Gong's pragmatic multi-center trial.

SA 2a: To integrate patient triage and advance compliance into Dr. Gong's clinical trial and sustained quality improvement and to focus education efforts on the most effective components of the checklist intervention.

SA 2b: To update our model continuously with new incoming patients, to expand to other regional institutions and to incorporate provider compliance, seasonal effects and institutional learning into the model.

Aim 1: Predict and impute Aim 2a: Investigate provider compliance and effectiveness of the individual components of the intervention checklist Aim 2b: Model institutional learning and seasonal effects, incorporate new patients and other institutions Exploratory Aim: Implement and sustain, offer to other institutions

Research Plan

A. Significance

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

Acute respiratory failure requiring mechanical ventilation is common in hospitalized patients and consumes a disproportionate amount of health care resources in the United States [2]. Short term mechanical ventilation can be live saving, but prolonged mechanical ventilation often leads to multi-organ failure and death in a third of patients [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 the development of Adult Respiratory Distress Syndrome (ARDS) 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 (e.g appropriate antibiotic therapy, diuretics and chest physiotherapy) and preventive measures (e.g. 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].

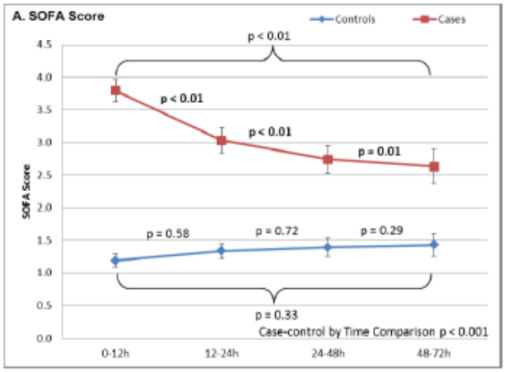

Figure 1: Deterioration of the Sequential Organ Failure Assessment score (SOFA) can be detected 24 to 48 hours before clinical deterioration leads to ICU admission [1].

A pragmatic trials 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 (PROOFCheck). The trial aims to identify patients at risk by building classical logistic regression models based on electronic medical records (EMR) to Accurately Predict PROlonged Ventilation (APPROVE). Identification of a patients at high risk triggers a decision support tool and bundled checklist of processes of care known to be beneficial in critically ill patients to PRevent Organ Failure (PROOFCheck). CITE CLINICAL TRIALS.GOV. The hypothesis is that in patients at high risk for prolonged respiratory failure or death, the early implementation of PROOFCheck, ideally already on the floor before ICU admission, reduces severity of organ failure, mortality and duration of mechanical ventilation. Pragmatic trials like this may result in more realistic and valid estimates of effectiveness and improved health resource allocation [12, 13].

Electronic medical records are an eminent example of richly structured and correlated Big Data and hold enormous promise for outcomes research [14, 15], exemplified by Dr. Gong's pragmatic trial. 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 pragmatic trial in particular, we desire more fine-grained information on the predictions to individualize prevention.

We can individualize prevention 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?

Observations and outcomes in our pragmatic PROOFcheck trial will be nested hierarchically. For example, repeated oxygen saturation measurements will be similar in the same patients, the more so, the closer they are in time. Patients seen by one and same hospitalist in our system 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 service or physician's care. Generally, physicians in academic medical centers like ours are organized in services, which are integrated across wards, clustered in hospitals. Concretely, the observations in our hospitalized patient cohort at Montefiore Medical Center, 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; richer models can exploit this fine-grained multilevel structures for optimal prediction of acute respiratory failure in our trial cohort. Finally, differences in regional health care environments predict patient, provider and institutional behavior and consequently outcomes, therefore risk patterns in one institution or even one region may not translate well to others; adjustments are surely necessary to translate our algorithms to different ecological settings.

Bayesian hierarchical modeling is transformative for EMR based prediction.

Bayesian methods are old, but novel in EMR surveillance. Reverend Thomas Bayes (Portrait in Fig.1)[16] formulated the Bayes theorem already in 1763 as an alternative statistical model. With a well developed theory, Bayesian methods are only novel in so far as they were rarely used in medical research until computers and Markov Chain Monte Carlo algorithms became widely available in the 1990s, leading to an expansion of applied Bayesian work [17, 18], more recently also in Big Data [19]. Past constraints in computational implementation have largely been overcome, except maybe for very large complex data sets like EMRs. We are not aware of any Bayesian hierarchical prediction model based on large EMRs to date. However, now ultra-fast Hamiltonian Monte Carlo algorithms [20] and clever statistical formulation or transformation allow to push the boundaries of computability, also for very large EMR data sets [21].

Figure 2: Thomas Bayes formulated the Bayes Theorem, published in a posthumous paper in 1763. [16].

A too brief introduction to Bayesian inference. In Bayesian inference, prior information is combined with new data to yield an updated posterior estimate for the probability of a hypothesis. The Bayesian approach is hence analogous to clinical decision-making. Physicians continuously update their preliminary diagnosis as new information comes in. Prior belief in a diagnosis may be weakened by new laboratory information, leading to an alternative disease hypothesis; or new lab information may reaffirm the initial diagnosis. [22]. Bayesian methods are particularly suited for flexible hierarchical modeling [23, 24].

Hierarchical modeling exploits the richly organized heterogeneity of electronic medical records. With their inherent flexibility and robustness, Bayesian hierarchical models may outperform classical models for prediction in large EMR data sets with spatial and temporal organization [25]. Above we described our multilevel EMR, consisting of repeated visits by patients with different medical plans, ages and medical conditions, treated by different services integrated in different wards and hospitals. Fitting our predictive regression model, we would want the regression coefficients to vary by group (by service, by medical unit, by hospital), to realistically model the complex correlations seen in actual clinical practice: The number of parameters to estimate grows very quickly and so do the potential interactions. Reciprocally, 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, to represent how lower level parameters vary across different groupings [26].

Prediction based on "partial pooling" outperforms the no-pooling and complete-pooling approach. Employing hierarchical modeling for our prediction model will be more efficient, as can be shown mathematically or via cross-validation [21]. "No pooling" would be an alternative approach to estimate the model for each specific subset of interest separately. But addressing and exploring the complexity and granularity, the richness of the EMR data would lead to far too many sub-classifications, thus too small samples in any given subgroup

for useful inferences. "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 [27].

We hypothesize that Bayesian hierarchical modeling may better identify hospitalized patients at risk for acute respiratory failure and prolonged mechanical ventilation than classical prediction algorithms.

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. To improve prediction for cases with incomplete data, we can impute the missing data. Informative loss by non-ignorable incomplete data may bias risk prediction or may hamper the implementation of the prediction algorithm. 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) [28]. However, this is an unreasonable assumption for our electronic medical records, for example because physicians will request test based on the patients co-morbidity and current clinical conditions. Data will not be missing at random, instead incomplete data will be associated with predictors and outcomes.

Auxiliary data can be used to impute incomplete medical records. Auxiliary date are additional information available in the form of variables known to be correlated with the missing data of interest [30]. In our data set we find that arterial blood gas oxygen tension is often unavailable for the prediction time window, because it was not requested by the physicians. In this case peripheral oxygen saturation and or or oxygen therapy may be used to impute the peripheral 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 [31]. 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 [32, 33, 34], often relying on Bayesian methods for the multiple imputations approach [35, 36]; joint modeling and multiple imputations could both be used also to impute incomplete medical records [37]. 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 [29]. Moreover, auxiliary data imputation for incomplete electronic medical records is underdeveloped; methodologically, their development is an innovative hallmark of this proposal.

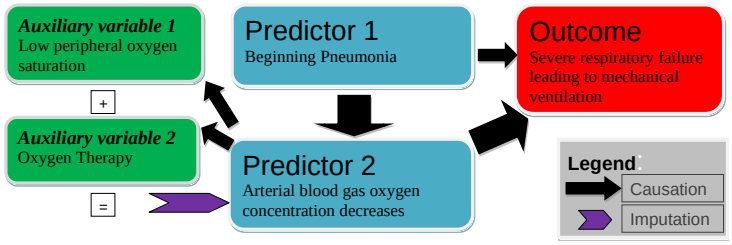

Figure 3: Incomplete data can hinder outcome prediction, but we can impute incomplete data from auxilliary information. For example, pneumonia (causing to low oxygen tension), may cause respiratory failure. If arterial blood gas results are missing, we can impute the oxygen tension from oxygen therapy and/or peripheral oxygen saturation. [29].

Seasonal effects, provider compliance 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. Non-compliance is a major obstacle to the effective delivery of health care and improved outcomes [38]. The personalized interventions triggered by our EMR-prediction algorithm will only prevent respiratory failure if our physicians and nurses actually implement them. Improving compliance of health care providers with evidence based interventions continues to be a challenge and is under-researched [39]. We need to understand better what patient and/or provider characteristics predict compliance with the preventive checklist interventions we suggest to ensure care is in accordance with accepted evidence based best practices.

Like for any institution, the composition of the Montefiore Medical Center 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, we noted during the implementation phase of previous preventive trials, that providers learn, changing their behavior as a result of trial participation. As trails progressed providers implemented previously underutilized interventions more frequently even before they were prompted. The transition of junior and senior providers through their training and to other institutions and new personnel joining the staff, may on the other hand lead to these improvements being lost. We term these effects institutional learning. 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 alter the risk profile 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 during the implementation of our trial to account for said changes in the risk profile. The integration of provider compliance, secular and seasonal effects in a EMR-triggered prevention within one coherent (Bayesian) model is novel.

B. Innovation

Focus on prevention of critical adverse outcomes in hospitalized patients. 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. It makes sense to focus early intervention on patients at high risk for poor outcomes. This project advances hierarchical Bayesian models to implement this paradigm shift in very large electronic medical records, triggering personalized interventions that drive outcome improvements.

Improve imputation of incomplete electronic medical records. Incomplete patient data, typical for actual clinical records can hinder the development and execution of our prediction algorithms. We will further Bayesian methods to impute incomplete missing data from additional auxiliary data to overcome this limitation. Beyond improving prediction and patient outcomes in our clinical trial, the Bayesian methods we propose to develop can be employed to impute incomplete electronic medical records in other settings.

Integrate advances in critical care with cutting edge computational statistics. To often advances in statistical modeling and clinical science are worlds apart. We want to address a critical problem of scalability in Big Data inference, but are equally motivated by our practical use case, improving our pragmatic clinical trial. The strength of our proposal, therefore, is the integration of disparate disciplines, critical care and computational statistics.

Summary of the impact

We tackle a serious health care challenge by integrating advanced hierarchical modeling into a pragmatic clinical trial. Beyond improving morbidity and mortality from respiratory disease in hospitalized patients through improved prediction and prevention, we will develop new methods to impute incomplete electronic medical records from auxillary data and scale Bayesian hierachical models to use in 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 critical care.

C. Approach

Hypothesis: Our overall hypothesis is that the integration of complex hierarchical Bayesian modeling and data imputation into a pragmatic clinical trial will reduce morbidity and mortality from respiratory failure in hospitalized patients compared to the classical statistical approach.

My research project will be closely aligned with my mentor's NIH-funded pragmatic trial in which Dr. Gong employs the published LIPS score [5] (based on a classical frequentist model) to identify hospitalized patients at risk for prolonged mechanical ventilation and death and initiates targeted interventions from a bundle of widely accepted care practices to prevent such adverse respiratory events.

Aim 1: To improve early prediction of prolonged respiratory failure and death in hospitalized patients.

For specific aim 1a, we will build a pragmatic EMR-based hierarchical Bayesian model implemented in the ultra-fast statistical software Stan 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.

Table 1: Table describing the population?

City	N[a]	%Silly
San Diego	289	41%
Seattle	262	32%
Galveston	261	15%
St Louis	269	7%
New York	271	4%
Baltimore	231	2%
Total	1,586	21%

[a]Inpatient population at Montefiore Medical Center.

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.

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.

Study Design: Two 3-month, prospective, observational cohort studies are underway at Montefiore Medical Center. Patients from the first three months will serve as the fitting cohort, while the 2nd cohort will serve as the validation and test set.

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?[40],an IHE30 IS THIS THE CORRECT REFERENCE? [41] 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, site identifier variables will be obfuscated to blind investigators.

Statistical Model: 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 consider random effects for co-morbidity and other time-invariant patient specific descriptors.

Score development and computational implementation 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. The initial cutoff score used will be when 25 percent of the patients in the fitted sample with high risk are actually positive. The simple effective prediction point in time for a positive patient is the earliest contiguous

time before the event where the score indicates a high risk. We will implement our hierarchical Bayesian model in the ultra-fast statistical software Stan, a probabilistic programming language for specifying models in terms of probability distributions [20].

Posterior predictive checking, predictive validations and model comparison In evaluating its predictive performance, we will perform posterior predictive assessment using discrepancies [42] to compare the patient test set to simulated replications from our fitted hierarchical Bayesian model, predictive validation to adjust for overfitting of our model and a sensitivity analysis of our priors on key model parameteters [21, 43]. We will compare our Bayesian Model to the frequentist algorithms using the minimum χ^2 discrepancy, essentially equivalent to the classical goodness-of-fit test statistic [42].

Exploratory analysis As an exploratory analysis, we will investigate if a score deterioration over time (even less than the cut off for high risk,) improves prediction, in particular if it reduces false negative rates, in which case we shall incorporate it into our prediction algorithm.

Missing data Missing data are characteristic limitation of large electronic medical records and may bias our prediction model [14]. Electronically medical records measurements not updated 24 hours earlier than the selected start time will be considered missing; details on handling and imputing missing data are provided under specific aim 1b, below.

For specific aim 1b, we will advance Bayesian data imputation algorithms of missing clinical data using auxiliary data, to identify the auxiliary measure properties, ceiling and floor 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 [44, 45]. We will exploit the temporal relationship between variables in the longitudinal electronic medical records [46]. We will perform posterior cross validation checks to investigate the appropriateness of our assumptions and incomplete data model [47].

Aim 2: Include temporal changes and provider behavior and treatment effectivness in our Bayesian model

To most closely reflect the realistic situation of actual real time academic and community medical delivery settings, we need to take temporal and seasonal changes changes into account and consider provider behaviro and the differential effectiveness of the interventions.

Aim 2a: Investigate provider compliance and effectiveness of the individual components of the checklist.

Prediction of adverse events is useful only if followed by effective preventive action: As part of Dr. Gong's randomized trial, providers of a patient identified as high risk by our Bayesian hierarchical model will be prompted to implement concrete preventive and corrective measures. We will investigate a) what predicts provider compliance with the prevention checklist and b) which components of the intervention checklist are most effective.

For specific Aim 2a, we will collect patient and provider characterizes and record provider compliance with EMR-triggered alerts in real time to investigate predictors of provider behavior in generalized linear models. We will also analyze predictors of the individual effectiveness of the various components of the intervention checklist.

For specific Aim 2b, to reflect changing risk profiles over time, we will update our Bayesian model with new patients continuously and tune our model to include temporal effects, like the above changes in patient population, institutional learning learning, seasons or epidemics. we will further develop our Bayesian model to continuously update the prediction model with new incoming patients. , expand it to include other regional institutions and seasonal effects and institutional learning into the model.

Population: Hospitalized adults identified as high risk for developing severe acute respiratory failure with prolonged mechanical ventilation and intubated patients will be included. Patients chronically ventilated at home or who have Do not resuscitate orders, will be excluded. We will limit the population to patients from those wards where we find severe adverse respiratory events to be most prevalent and worrisome. (Table 1: Inpatient population at Montefiore Medical Center.)

Intervention: During the PROOFcheck intervention period, providers receive targeted education on prevention and best practice. As during the PROOFcheck intervention period, during our Bayesian implementation, when a patient at high risk is identified, his or her physician is notified by pager and/or electronic clinical interface; an interactive notification algorithm will suggest patient specific interventions from the checklist to the clinicians.

Control The control condition will be the PROOFcheck intervention period. We will collect provider characteristics to investigate them as predictors of provider compliance.

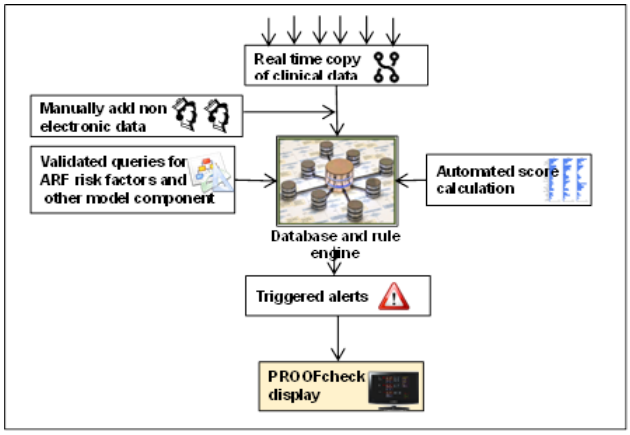

Figure 4: Implementation algorithm. A rule engine scans the database, comprising a real-time copy of the electronic medical record complemented with manually entered data. An alert is send out to the provider, if the prediction score exceeds the threshold.

Outcomes: The primary outcome will be provider compliance. Secondary outcomes include mortality, multi-organ-failure as indicated by the daily SOFA score for the 7 days after patient identification, 28 day ventilator free days, ICU and hospital length of stay, discharge to home, and survival in 6 and 12 months after hospital discharge. These outcomes were selected to be important to patients, clinically relevant and easy to determine from routinely collected data.

Study Design: This is a prospective observational cohort study to investigate provider compliance and effectiveness of sustained EMR-surveillance subsequently to PROOFcheck.

Institutional review and data safety and monitoring board (DSMB): the IRB approved both the observational cohort study to develop the algorithm and the subsequent pragmatic trial. A DSMB is already in place to monitor for patient safety and adverse effects on clinicians from the alerts and decision support tools. We will use these as templates for subsequent IRB applications.

Implementation and individualization of preventive interventions using EMR-context During PROOFcheck, Clinicians already receive instructions on how to assess and intervene for their high risk patients. Obviously, the proposed interventions will only prevent respiratory failure if our physicians and nurses implement them. We will individualize the suggested measures to make them more patient specific and measure compliance. The patient's record will be scanned to provide context. For example, if the patient's mental status is undocumented, the clinician will be prompted with pertinent questions and preventive measure are proposed in accordance to those responses. In order to measure and demonstrate compliance with the checklist, near real time (same day) transaction logs and evidence for compliance will be recorded through electronically.

Weiss et alt 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 [48].

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. Figure 4 illustrates the implementation algorithm. The rule engine (implemented in Java) will send out the alert to providers. We will integrate Bayesian patient triage into the infrastructure of Dr. Gong's pragmatic trial, if we can show that our Bayesian prediction algorithm is superior to the existing algorithm. Alternatively, we will investigate provider compliance continuing with the classical prediction algorithm. The compliance information will be reviewed for various items during the course of study so that sites could be targeted for re-education.

Exploratory Aim: Sustained implementation of our Bayesian algorithm and extension to other regional institutions.

We will continue to trigger individualized patient interventions to the providers in the Einstein system after the PROOFcheck trial, if the EMR triggered intervention checklist proves to be effective. We will integrated our Bayesian model into the PROOFcheck infrastructure to sustain the outcome improvements, if our Bayesian imputation and prediction outperforms the classical algorithms. We will package and standardize our prediction model

to tests its predictive power and offer its implementation at other regional institutions, for example our partners in the New York Clinical Research Data Network.

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] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

[18] 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.

[19] 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.

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

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

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

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

[24] 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.

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

[26] 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.

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

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

[29] 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.

[30] 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.

[31] 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.

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

[33] 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.

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

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

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

[37] G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs. *Longitudinal data analysis*. CRC Press, 2008.

[38] 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.

[39] 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.

[40] 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.

[41] 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.

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

[43] 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.

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

[45] 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.

[46] 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.

[47] 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.

[48] 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.