

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.

Research Plan

A. Significance

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

Acute respiratory failure requiring mechanical ventilation is common and consumes a disproportionate amount of health care resources in the United States [1]. Short term mechanical ventilation can be life saving, but prolonged mechanical ventilation often leads to multi-organ failure and death in a third of patients [1, 2]. Most research focuses on established respiratory failure in the ICU, while detectable clinical signs and symptoms often herald the impending respiratory decompensation [3]. These serious abnormalities often occur 8-48 hours before ICU admission, but are either not recognized or not acted upon [4, 5]. Early interventions (e.g. appropriate antibiotic therapy, diuretics and chest physiotherapy) and preventive measures (e.g. head elevation) may be able to stop or reverse the clinical deterioration before severe acute respiratory failure requires mechanical ventilation [6, 7, 8, 9].

A pragmatic trials to predict and prevent mortality from respiratory failure in hospitalized patients.

My mentor Dr. Gong is leading a randomized multi-center trial to reduce mortality from acute respiratory failure requiring mechanical ventilation. The trial aims to identify patients at risk by building classical logistic regression models based on electronic medical records (EMR). Identification of a patients at high risk triggers a prevention intervention CITE CLINICAL TRIALS.GOV. The hypothesis is that implementation of a prevention checklist reduces mortality.

Electronic medical records are an eminent example of richly structured and correlated Big Data, and hold enormous promise for outcomes research [10]. EMR have more useful data than can be analyzed in a scientifically meaningful way by existing statistical inference tools. This currently limits the scientific hypotheses and clinical inferences, that can be explored and evaluated. Large electronic medical data sets are not just bigger in that there are more instances of the same thing, (this would make data analysis only easier). Rather, there is more breadth to the data: more subgroups, locations, or time granularity than is currently being modeled, more frequent and detailed measurements than can easily be incorporated into standard models, more information on the population units being measured, and more fine-grained information on the predictions desired.

Large electronic medical records are nested hierarchically. Clinical observations are nested within patients, e.g. repeated glucose measurements will be similar in the same patients. Patients seen by the same provider will have similar outcomes predicted by provider behavior and qualities. Providers are integrated in institutions. Institutions are nested geographically in counties and regions. Health care environments predict patient and provider behavior and outcomes. Patients seen by the same team, treated in the same setting will have similar propensity to respond to interventions. Large electronic medical records require more than just fitting well-known models at larger scales; they require richer models to exploit fine-grained multilevel structures and to map to predictive questions of interest.

Bayesian hierarchical modeling is transformative for complex Big Data.

Bayesian methods are old and new in health sciences and Big Data. Based on the theory outlined in a posthumous 1763 paper by Reverend Thomas Bayes (Portrait in Fig.1)[11], the Bayesian approach is based on an alternative statistical model and is particularly suited for hierarchical modeling [12, 13]. Old and well established, 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 [14, 15], also in Big Data [16]. Past constraints in computational implementation have largely been overcome by modern Hamiltonian Monte Carlo Algorithms, except for very large complex data sets [17].



Figure 1: Reverend Thomas Bayes is credited with formulating the Bayes Theorem, published in 1763 in a posthumous paper [11].

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 test results come 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. A particular challenge of the proposed research project is to communicate the meaning and value of this underused statistical approach, the Bayesian model, to our fellow clinical researchers [18].

Hierarchical modeling exploits the richly organized heterogeneity of electronic medical records. With their inherent flexibility and robustness, Bayesian models may outperform classical models in large data sets with spatial and temporal organization [19]. Consider our multilevel electronic medical records data set consisting of repeated visits by patients with different ages and medical conditions in different services integrated in different hospitals in different states with different medical plans. Fitting the 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. Regardless, we will want to estimate various hyper-parameters and hyper-hyper-parameters, to represent how lower level parameters vary across different groupings [20].

"Partial pooling" outperforms the no-pooling and complete-pooling approach. Hierarchical modeling is more efficient, as can be shown mathematically or via cross-validation [17]. "No pooling" is one approach to estimate the model for each specific subset of interest separately. Addressing and exploring the complexity and granularity, the richness of the data may lead to far too many sub-classifications, thus too small samples in any given subgroup for useful inferences. "Complete pooling" or structural modeling is another approach, but the implied hard constraints on the coefficients in different groups may lead to bias, in particular for groups with sparse data; we lose information, because we cannot learn from groups where we have more data. In hierarchical modeling, the estimate of each individual parameter is simultaneously informed by data from all the other units; this is what makes "partial pooling" or hierarchical modeling especially effective [21].

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 and thus improve mortality in Dr. Gong's pragmatic clinical trial.

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 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) [22]. 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 data are additional information available in the form of variables known to be correlated with the missing data of interest. For example, arterial blood gas oxygen saturation may be used to impute peripheral pulse oxymetry or oxygen therapy, if the latter are unavailable for the prediction time window, and vice versa. This approach avoids the perils associated with missing at random (MAR) assumptions, when fitting a non-ignorable missingness model [23]. 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 [24, 25, 26], often relying on Bayesian methods for the multiple imputations approach [27, 28]; joint modeling and multiple imputations could

both be used also to impute incomplete medical records [29]. 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 [30]. Moreover, auxiliary data imputation for incomplete electronic medical records is underdeveloped; methodologically, their development is an innovative hallmark of this proposal.

Seasonal effects, provider compliance and institutional learning can bias risk prediction. Non-compliance is a major obstacle to the effective delivery of health care and improved outcomes [31]. The personalized intervention triggered by our EMR-prediction algorithm will only prevent respiratory failure if our physicians and nurses implement them. Improving compliance of health care providers with evidence based interventions continues to be a challenge and is under-researched [32]. Institutional behavior changes in response to trials and quality improvements interventions; patient populations can change over time. Respiratory patients are plagued by seasonal deterioration, which could lead to bias in our model. These seasonal and secular effects will alter the predictors of risk in our model. We will therefore include 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 pragmatic trial with one coherent 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 critical care with 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 hierarchical 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 complex hierarchical Bayesian modeling and data imputation will reduce morbidity and mortality from respiratory failure in hospitalized patients compared to the classical model.

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 [8]. However, an indiscriminate universal approach to prevention of respiratory failure in hospitalized patients will be ineffective,

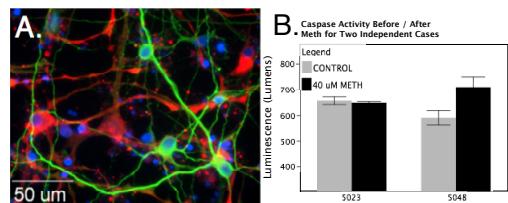


Figure 2: Change in SOFA over time before

because only one in 30 hospitalized adults requires mechanical ventilation. Individualized preventive interventions targeting patients at high risk will be more efficient in preventing potentially irreversible end organ damage, while also possibly leading to improved compliance by providers and cost effectiveness. My mentor Dr. Gong, demonstrated that predictive scores can signal clinical deterioration as early as 24-48 hours before ICU transfer or rapid response team interventions (Fig 1) [33]. 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 [34], which proved equally able to discriminate the 587 patients in the cohort who progressed to severe ARF requiring > 48 hrs of mechanical ventilation. My research project will be closely aligned with her NIH-funded pragmatic trial using the LIPS score (based on a classical frequentist model) to identify hospitalized patients at risk for prolonged mechanical ventilation and death and initiate targeted interventions from a bundle of already 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 it with the existing frequentist algorithm used by Dr. Gong in her pragmatic trial.

Population: Adult patients admitted to the Montefiore Medical Center during the study period will be included; we will exclude patients 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 [34]. 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 (logarithmic) transformations to induce variance stability and summary aggregations.

Outcomes: The dichotomous outcome of interest 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?[35],an IHE30 IS THIS THE CORRECT REFERENCE? [36] 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,

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%
<i>Total</i>	1,586	21%

^aInpatient population at Montefiore Medical Center.

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 include 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 [37].

Posterior predictive checking, predictive validations and model comparison In evaluating its predictive performance, we will perform posterior predictive checking to compare the 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 parameters [17, 38]. WE WILL COMPARE OUR BAYESIAN MODEL TO THE FREQUENTIST USING ...

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 [10]. 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 [39, 40]

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.

References

References

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] I. A. Mitchell, H. McKay, C. Van Leuven, 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.
- [10] 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.
- [11] 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.
- [12] J. B. Carlin. Meta-analysis for 2 x 2 tables: a Bayesian approach. *Stat Med*, 11(2):141–158, Jan 1992. PMID: 1349763.
- [13] 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.
- [14] D. Ashby. Bayesian statistics in medicine: a 25 year review. *Stat Med*, 25(21):3589–3631, Nov 2006. PMID: 16947924.
- [15] 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.
- [16] C. Yoo, L. Ramirez, and J. Liuzzi. Big data analysis using modern statistical and machine learning methods in medicine. *Int Neurorol J*, 18(2):50–57, Jun 2014. PMID: 24987556.

- [17] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian data analysis*, volume 2. Taylor & Francis, 2014.
- [18] J. K. Kruschke. Bayesian estimation supersedes the t test. *J Exp Psychol Gen*, 142(2):573–603, May 2013. PMID: 22774788.
- [19] A. Gelman. *Red state, blue state, rich state, poor state: why Americans vote the way they do*. Princeton University Press, 2009.
- [20] 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.
- [21] A. Gelman. Multilevel (Hierarchical) Modeling: What It Can and Cannot Do. *Technometrics*, 48(3):432–435, August 2006.
- [22] D. B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.
- [23] 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.
- [24] X.-L. Meng. Multiple-Imputation Inferences with Uncongenial Sources of Input. *Statist. Sci.*, 9(4):538–558, 11 1994.
- [25] 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.
- [26] D. B. Rubin. Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 91(434):473–489, 1996.
- [27] M. J. Daniels and J. W. Hogan. *Missing data in longitudinal studies: Strategies for Bayesian modeling and sensitivity analysis*. CRC Press, 2008.
- [28] J. L. Schafer. *Analysis of incomplete multivariate data*. CRC press, 1997.
- [29] G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs. *Longitudinal data analysis*. CRC Press, 2008.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] 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.

- [35] 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.
- [36] 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.
- [37] Stan Development Team. Stan: A C++ Library for Probability and Sampling, Version 2.5.0, 2014.
- [38] A. Gelman, Y. Goeghebeur, F. Tuerlinckx, and I. Van Mechelen. Diagnostic Checks for Discrete Data Regression Models Using Posterior Predictive Simulations. *Applied Statistics*, pages 247–268, 2000.
- [39] J. L. Huntington and A. Dueck. Handling missing data. *Curr Probl Cancer*, 29(6):317–325, 2005. PMID: 16311133.
- [40] 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.