SUPPLEMENTAL MATERIAL

"Assessing Hospital Performance Following Percutaneous Coronary Intervention Using Big Data"

By Jacob V Spertus; Sharon-Lise T Normand; Robert Wolf; Matt Cioffi; Ann Lovett; Sherri Rose

This appendix supplies details regarding (1) our methodology to impute data, (2) details of the three methods we employed for effect estimation, (3) bootstrapping for confidence intervals, (4) sensitivity analysis for multiple patient admissions, and (5) the causal assumptions underlying our approaches.

Missing Data and Imputation Procedure

Our final cohort was obtained after merging with CHIA billing data. Many records in the Mass-DAC registry did not merge and were dropped. These records did not merge because some hospitals recorded key fields needed for matching differently in the registry and billing data. The severity of this issue varied considerably across hospitals and resulted in the loss of 25% of the registry cohort. Despite this problem, we used the billing data because it enriched our variable set and mortality did not vary significantly between merged and unmerged records. Nevertheless, this issue emphasizes the importance of working with hospitals to ensure high-quality data and diagnosing the extent to which dropping records based on missingness could bias results. A few variables in the Mass-DAC registry also had missing cells, especially the stenosis percentage fields. These variables were checked by coders as "not available" in the NCDR data collection tool, indicating that the variables were truly not measured. This could occur if patients transferred in from another hospital and did not undergo a second diagnostic catheterization to assess stenosis at the time of recording.

Imputation of missing values for variables in the Mass-DAC registry was implemented with a SAS callable user interface to the IVEWare software package. IVEWare employs a multivariate sequential regression approach, iteratively computing multiple regressions for each imputed variable using observed data from the other covariates in the imputation model[1]. Continuous, binary, or categorical variables are imputed using linear, logistic, Poisson, or multinomial models, respectively. Some variables in our data were conditional on the presence of other variables, meaning they were not measured if another variable took on a certain value. These were not imputed via regression, but rather set to a default value. For example, if a patient did not have heart failure, they would not receive a NYHA classification,

and we would not impute this. Instead a value of 0 was assigned and treated as its own class in the analysis process. One complete imputed set was obtained and utilized for the study analyses without adjustment of variance to account for the imputation.

In general, missing data may be imputed if they are "missing at random," wherein the presence or absence of a missing cell does not depend on its value given observed variables. For a complete discussion of missing data, see Gelman and Hill chapter 25.[2]

Approaches to Effect Estimation

Regression Only

Our data consisted of N = 9325 admissions at 24 hospitals. Denote y_{ij} as the binary mortality outcome for the *i*th patient at hospital *j*, the matrix of confounders by *X*, and the set of regression parameters by $\theta = \{a_1, a_2, \ldots, a_{24}, \beta\}$ where a_j is the hospital-specific intercept for the j[th] hospital and β is a vector of coefficients associated with the confounders. Furthermore, we use the indicator $I_{ij} = 1$ if patient *i* was treated at hospital *j* and 0 otherwise. We used standard logistic regression to estimate θ (**Supplemental Figure 1**) when using the clinical set of 11 confounders. For the full confounder set, β was a vector of 225 coefficients and we estimated θ using penalized maximum likelihood as described below.

To determine the adjusted mortality at the j^{th} hospital, we set the hospital intercept of every patient in the state equal to a_j and then computed $E(y_{ij})$ by $\text{logit}^{-1}(a_j + \beta^T X_i)$, which represents the counterfactual estimate of 30-day mortality for the i^{th} patient that would have been observed had the i^{th} patient undergone his/her PCI at hospital *j*. We then average these estimates over all patients to obtain $E(y_j) = \frac{1}{N}\sum_{i=1}^{N} E(y_{ij})$, the overall adjusted mortality at hospital *j*.

Penalized Maximum Likelihood

Traditionally, estimation of outcome or treatment regressions involves the application of substantive (e.g., domain) knowledge to preselect a workable subset of confounders from which to form estimates.[3] However, preselection of features in a causal inference problem can undermine the opportunities offered by large data sets, where the information contained in existing confounders can boost predictive accuracy or lead to new insights if properly accounted for by investigators. The penalized maximum likelihood method permits estimation in this setting by restricting the size of the regression coefficients and includes techniques such as ridge regression, the lasso, and their generalization, the elastic net.[4]

In this paper, we utilized elastic nets, a procedure that selects confounders to minimize a penalized maximum likelihood equation. Under the elastic net penalty, we maximize the penalized log-likelihood with respect to θ:

$$\sum_{i=1}^{N}\left\{y_i(I_{i1}a_1 \dots I_{i24}a_{24} + \beta^T X_i) - \log\left(1 + e^{I_{i1}a_1 \dots I_{i24}a_{24} + \beta^T X_i}\right)\right\} - \lambda \sum_{p=1}^{225}\left[(1 - \alpha)\,\beta_p^2 + \alpha\left|\beta_p\right|\right]$$

The elastic net penalty is the second term in this equation while the first term is simply the ordinary log-likelihood for binomial logistic regression. The strength of the penalty is determined by the value of the tuning parameter, λ. When $\lambda = 0$, the ordinary maximum likelihood results are recovered. We chose an α value of 0.80, yielding a mix of the lasso penalty ($\left|\beta_p\right|$) allowing for variable selection, and the ridge penalty (β_p^2), used to reduce multicollinearity and encourage the coefficients of correlated confounders to be shrunk towards each other.[5] For the mortality regressions, we estimated a penalized logistic regression with the penalty size λ varying from e^{-10} to e^{2} and assessed the 5-fold cross-validation deviance at each value of λ (**Supplemental Figure 2**)[6]. We then selected the coefficients from the regression with the λ value that yielded the lowest cross-validated deviance, which we present in the main paper (**Figure 4**). For theoretical and practical reasons, we penalized only the confounder coefficients and did not penalize the hospital specific intercepts.[4]

The multinomial (24 hospitals) propensity score regression with full confounders was estimated in an identical fashion using the multinomial deviance, yielding $(24 - 1) * 225 = 5175$ coefficients (a

set for each hospital), most of which were identically zero. Note that only 23 probabilities require estimation for the 24 hospitals because the final hospital's probability for patient i is given by $1 - \sum_{j=1}^{23} P(I_{ij} = 1)$. Because of the large number of hospital probabilities requiring estimation, estimation of the multinomial regression can have convergence issues, even with a small number of confounders. For example, even with our parsimonious set of 11 clinical confounders, convergence was challenging, requiring the use of a ridge penalty to estimate the regression parameters. We again chose the coefficients corresponding to the λ minimizing cross-validated deviance (**Supplemental Figure 2**). Because we used the ridge penalty, none of these coefficients was identically zero and the full set of 11 clinically chosen confounders remained in the regression. We implemented our penalized regression in the R package *glmnet.*[6]

Augmented Inverse Probability Weighting (A-IPW)

Augmented inverse probability weighting (A-IPW) is an advance incorporating the commonly used propensity score. The method combines a propensity score regression with outcome predictions, in our case generated from a logistic regression as described above. The double-robustness of this approach means that it yields unbiased estimates of the hospital effect if *either* the propensity score or outcome regression is consistently estimated. That is, if either regression accounts for confounders in a way that reflects their true relationship to hospital and outcomes.[7-9] Using doubly-robust approaches is appealing because it is rare that the true structure of the regression is known, and the doubly-robust nature gives the researcher two opportunities to get it right.[7,10]

To implement A-IPW, we estimated a multinomial regression that provides 24 estimated probabilities for each patient – each probability reflects the likelihood that a patient undergoes PCI at the hospital given the patient's observed confounders. We used elastic net penalized regression for the full confounder set and ridge penalized regression for the clinically informed confounder set. To stabilize extreme weights for each hospital we divided the propensity scores for that hospital by the proportion of

total patients who underwent PCI at the hospital. Next, we estimated the respective logistic regressions linking 30-day mortality with the full confounders (via elastic net penalized regression) and again with the parsimonious confounder sets (via standard logistic regression). The propensity scores from the hospital regression and the mortality estimates from the outcome regression were combined using the A-IPW equation as follows:

$$E(y_j) = \frac{1}{N}\sum_{i=1}^{N} E(y_{ij}) + \frac{I_{ij}}{P(I_{ij} = 1 \mid X_i)}\{y_i - E(y_{ij})\}$$

Thus if patient *i* wasn't treated at hospital *j*, then that patient's contribution to the A-IPW estimate is simply their expected (counterfactual) outcome as predicted from regression. However, if the patient was treated at hospital *j*, A-IPW includes the additional weighting from the propensity score, $P(I_{ij} = 1 \,|X_i)$.

Targeted Maximum Likelihood Estimation

Targeted maximum likelihood estimation (TMLE) is a general framework that extends parametric maximum likelihood estimation for semiparametric and nonparametric models. It focuses on estimating parameters that are features of a probability distribution while making minimal assumptions. The TMLE algorithm for the parameters targeted in this paper consists of updating an initial mortality regression with information from the propensity score in an effort to reduce bias for the parameter of interest. This second step can therefore be thought of as a bias-reduction step and is detailed below. The TMLE is a double robust, efficient, consistent, well-defined loss-based substitution estimator of the parameter of interest.[11]After generating initial estimates for $E(Y_{ij})$ in the same fashion as for the regression only approach, denoted Q_{ij}^0, we update the estimates with information from the multinomial propensity score regression to obtain a revised estimate, Q_{ij}^1. This is accomplished by regressing the outcome y_i on a new covariate generated from the propensity scores with offset Q_{ij}^0 to obtain a coefficient ϵ, and averaging the fluctuated outcomes:

$$\text{logit}\left(Q_{ij}^{1}\right) = \text{logit}(Q_{ij}^{0}) + \epsilon\left[\frac{I_{ij}}{P\left(I_{ij} = 1 \mid X_i\right)}\right]$$

$$E\left(y_j\right) = \frac{1}{N}\sum_{i=1}^{N} Q_{ij}^{1}$$

The algorithm is easy to implement in R using the *tmle* package.[12]

Hospital-Specific Excess Mortality Risk

The excess mortality risk at a hospital was determined using the estimates $E(y_j)$ and their average across hospitals. Specifically, we determined the difference in the adjusted mortality at that hospital, and a statewide adjusted mortality calculated by taking the average adjusted mortality at each hospital: $E\left(y_j\right) - E(y) = E\left(y_j\right) - \frac{1}{24}\sum_{j=1}^{24} E(y_j)$. A high excess risk indicates a hospital that is performing poorly compared with the rest of the state, while negative excess risk indicates a better than average hospital.

Bootstrapping

We next determined standard errors and confidence intervals to account for the uncertainty associated with the mortality estimates. We accomplished this through the use of bootstrapping, a robust statistical technique for constructing confidence intervals. We resampled hospitals *with replacement* B times, using all patients in the sampled hospitals, [13] estimating the parameters of interest on each bootstrapped set, and using the distribution of the recalculated estimates to construct standard errors and confidence intervals. To match the exact confidence intervals more precisely, we perform a transformation on the percentiles of the bootstrap samples to give bias-corrected intervals.[14]

Multiple Patient Admissions: Sensitivity Analysis

We allowed for multiple admissions from single patients to contribute to our data as long as the admissions were more than 30-days apart. Treating admissions from repeat patients as independent is

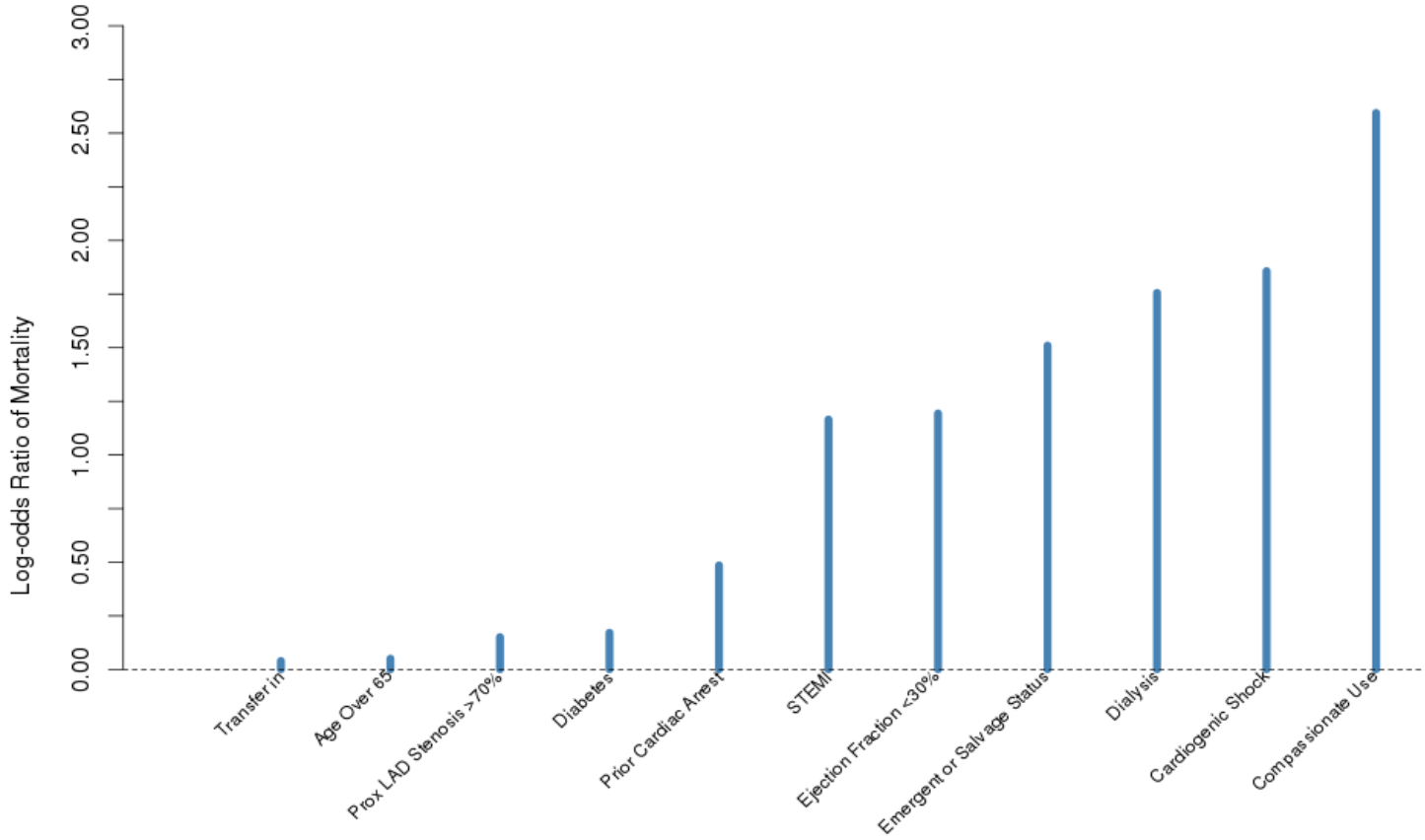
typical in profiling.[15] In our data 352 patients had more than one admission, contributing an additional 373 admissions to our data. Only 7 of these patients died within 30-days after their last procedure. We eliminated these additional admissions by randomly sampling a single admission for each of the 373 patients. Minor shifts in the patient distributions lead to a few small changes in outlier classification for borderline hospitals (**Supplemental Figure 3**). For example, hospital I moves from a non-outlier to an outlier in TMLE with full confounders. However, substantive inferences like kappa statistics are not significantly affected. Special techniques may be required to handle correlation introduced by multiple admissions if they make up a substantial portion of the data.

Causal Assumptions

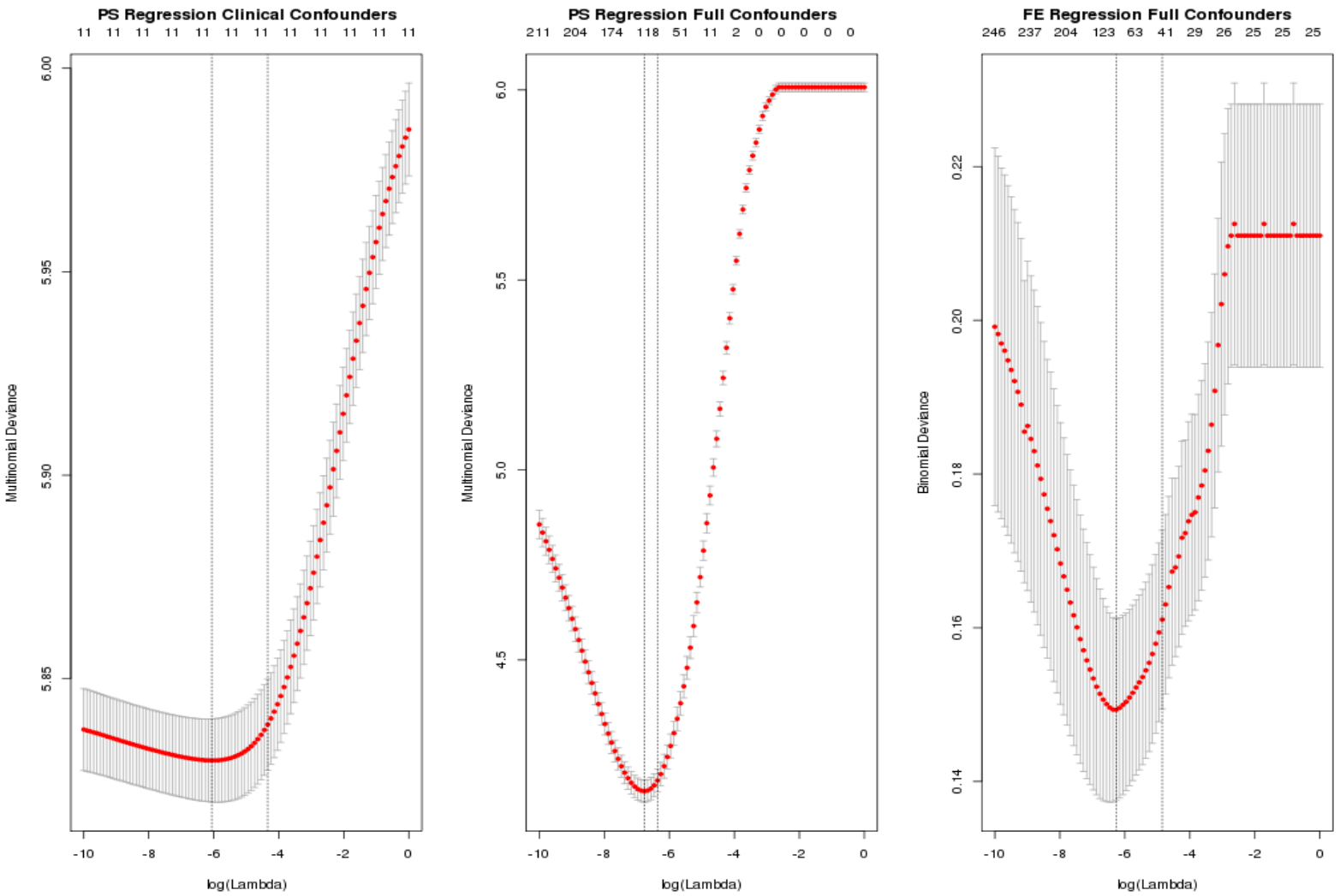
Additional untestable assumptions are required to enrich the interpretation of hospital effects as causal. Key causal assumptions include the stable unit treatment assumption (SUTVA) and no unmeasured confounding.[16] Under SUTVA, we assume both that the hospital assigned to a subject does not interfere with the potential outcomes for other subjects and that there is no variation within hospitals. We have reason to doubt the validity of the SUTVA assumption because care may vary significantly within hospitals, for instance because different doctors perform operations at the same center. SUTVA is untestable, but such realities compel us to be cautious when drawing causal conclusions from our analysis. No unmeasured confounding, a causal assumption frequently violated in practice, assumes that the potential outcomes are independent of the hospital conditioning on the adjustment set of confounders. The inclusion of many potential confounders from rich data sets, as in this paper, can bolster the validity of this assumption. Even when causal assumptions may not be justified, the target parameters defined in this paper are still interesting statistical parameters with interpretations as average effects adjusted for measured confounding.

References

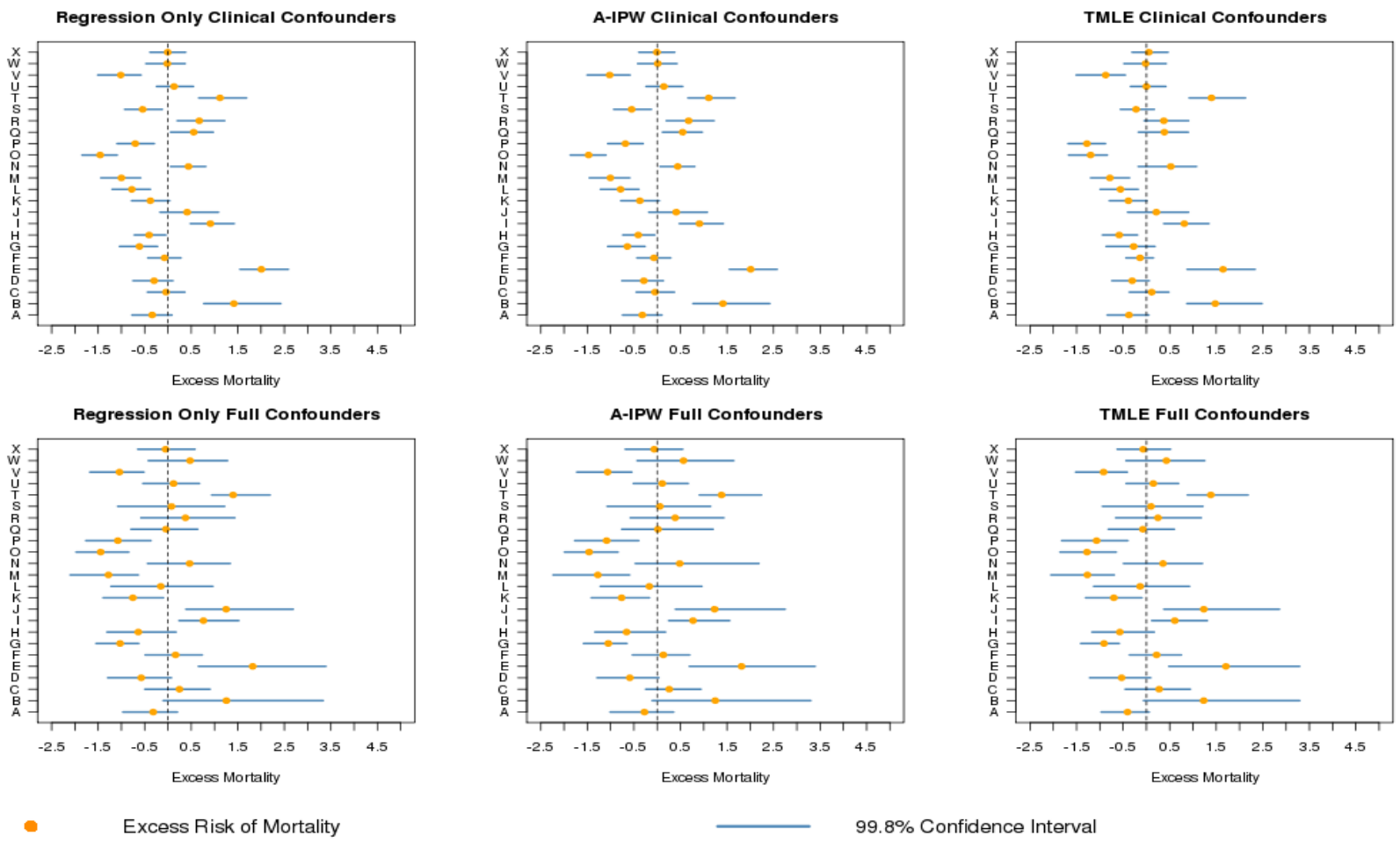
1. Raghunathan TE, Solenberger P, van Hoewyk J. IVEware: imputation and variance estimation software user guide. Available at: http://citeseerx.ist.psu.edu/showciting?cid=1929984 accessed March 1st, 2016.
2. Gelman A, Hill J. Data analysis using regression and multilevel/hierarchical models. New York: Cambridge University Press, 2007.
3. Brookhart A, Schneeweiss S, Rothman K, Glynn R, Avorn J, Stürmer T. Variable selection for propensity score models. Practice of Epidemiology. 2006;163:1149-1156.
4. Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society. 2005; 67: 301-320
5. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning second edition. New York: Springer, 2009.
6. Friedman J, Hastie T, Simon N, Tibshirani R. Package 'glmnet.' Available at https://cran.r-project.org/web/packages/glmnet/glmnet.pdf. Accessed February 19, 2016.
7. van der Laan MJ, Robins JM. Unified Methods for Censored Longitudinal Data and Causality. New York: Springer Science+Business Media, 2003.
8. Farrell MH. Robust inference on average treatment effects with possibly more covariates than observations. Journal of Econometrics. 2015; 189:1-23.
9. Varewyck M, Goetghebeur E, Eriksson M, Vansteelandt S. On shrinkage and model extrapolation in the evaluation of clinical center performance. Biostatistics. 2014; 15(4):651-664.
10. Lunceford J, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. Statistics in Medicine. 2004; 23: 2937-2960.
11. van der Laan MJ, Rose S. Targeted learning: causal inference for observational and experimental data. New York: Springer, 2011.
12. Gruber S, van der Laan M. Package 'tmle.' Available at https://cran.r-project.org/web/packages/tmle/tmle.pdf. Accessed February 19, 2016.
13. Field CA, Welsh AH. Bootstrapping clustered data. Journal of the Royal Statistical Society: Series B. 2007; 69: 369-390.
14. Efron B. Better bootstrap confidence intervals. Journal of the American Statistical Association. 1984; 82: 171-185.
15. Massachusetts Data Analysis Center. Cardiac Study-Annual Reports. Available at http://www.massdac.org/index.php/reports/cardiac-study-annual/. Accessed May 23, 2016.
16. Rubin DB. Causal inference using potential outcomes: design, modeling, decisions. Journal of the American Statistical Association. 2005; 100(469):322-331



Supplemental Figure 1: Estimated coefficients for the logistic mortality regression model with clinically selected confounders. Confounders are indexed on the x-axis, sorted by size. The y-axis denotes the log-odds ratios associated with each confounder. A higher log-odds ratio indicates increased chance of mortality when that confounder is present.



Supplemental Figure 2: Five-fold cross validation deviance plots. The log of the lambda values appears on the x-axis. Moving from left to right, a larger lambda value selects a more parsimonious model. The number of non-zero coefficients at a given lambda appears above the plot. The lambda giving the lowest cross-validated deviance is denoted by the dotted line on the left hand side. The dotted line to the right shows the lambda value with the largest deviance within one standard error of the smallest, which gives a more parsimonious model.



Supplemental Figure 3: Estimates of hospital effects with random sampling of a single admission from patients with multiple. Orange dots are point estimates of excess risk; blue lines are 99.8% confidence intervals (Boneferonni adjusted for multiple testing).