Targeted maximum likelihood estimation and hospital readmissions

Background: Hoping to improving quality of hospital care, the US and other jurisdictions financially penalize hospitals with poor (confounder-adjusted) 30-day readmission rates. Although hospital administrative data are information-rich, confounder adjustment tends to be crude. Non-parametric machine learning techniques can take advantage of these rich data to predict readmission, but cannot isolate the independent effect of hospitals on readmission risk.

Research Design: To estimate the effect of care at different hospitals on 30-day readmission risk, we used targeted maximum likelihood estimation (TMLE), which allowed us to use a non-parametric machine learning technique (random forest) to take advantage of the rich confounder data. We used an 11-year cohort of 65-year-old patients from 20 hospitals in Montreal, Canada, and developed three models to estimate the marginal readmission risk at each of the hospitals after hospitalization for heart failure, acute myocardial infarction (AMI), and pneumonia. We controlled for hundreds of confounders including outpatient drug prescriptions, medical procedures, and diagnoses. We compared the TMLE-estimated risk to a logistic regression model that only crudely adjusted for confounding.

Results: Within each hospital, crude readmission risk varied widely across the twenty hospitals for AMI 2,525 / 15,746 (16%), heart failure 5,520 / 24,847 (22%), and pneumonia 3,183 / 20,421 (16%). When crudely controlling for confounding, and compared to an arbitrary reference hospital, the odds ratio ranged from 0.95-1.02 for AMI, 0.92-1.04 for heart failure, and 0.96-1.04 for pneumonia. When we applied TMLE, the odds ratio ranged from 0.57-2.30 for AMI, 0.50-1.85 for heart failure, and 0.47-1.55 for pneumonia.

Conclusion: Our results suggest that crudely adjusted hospital readmission rates will underestimate the differences between hospitals. Using TMLE, we could take advantage of rich confounder data, and reveal wide differences in quality of care between hospitals.

# Introduction

In the US and other jurisdictions, administrators have sought to improve quality of care by financially penalizing hospitals with poor readmission rates.1 To avoid penalizing high quality of care at hospitals that admit sicker (more likely to be readmitted) patients, readmission rates are adjusted for patient-level confounding.2 Although hospital readmission rates are typically only adjusted for a few well-known confounders such as age, sex, previous readmissions, and summarized comorbidity scores,3 healthcare administrative data are often information-rich, including drug prescriptions, diagnoses, and medical procedures.

Some epidemiologists have argued for the application of machine learning techniques to handle newly available information-rich data sources.4 Non-parametric machine learning techniques can accurately discriminate patient readmission risk using hundreds of variables in a computationally efficient way.5 Non-parametric models also allow us to avoid specifying a functional form, making it easier to detect complex relationships like multi-way interactions. However, most machine learning techniques were developed for prediction rather than inference; we cannot use them alone to isolate (target) effects measures of specific variables, such as care at a particular hospital, on readmission risk.

The targeted maximum likelihood estimator (TMLE), is a doubly-robust technique that uses propensity scores to estimate target parameters of interest, and allows the incorporation of machine learning techniques.6 In TMLE, two (possibly non-parametric) two models are developed: one to estimate the probability of exposure (the propensity score), and another model to estimate the probability of the outcome. These two probabilities are combined in a parametric model with only the parameter of interest. In this way, the discriminative power of non-parametric models can be used to extract estimates of parameters of interest.

Although some studies have used the rich confounder data in combination with machine learning techniques to predict hospital readmissions7,8, no study to our knowledge has used these data to draw causal inference on the effect of quality of care on readmissions. In this study, we sought to estimate the independent effect of hospital care on the 30-day readmission for twenty Montreal hospitals, within three different admission diagnoses (pneumonia, heart failure, and acute myocardial infarction). We used a non-parametric machine learning technique, (random forest9), with TMLE to take advantage of the rich confounder data and minimize bias in our estimate of readmission risk.

# Methods

## Study Design

We used a cohort extracted from a Canadian provincial (Quebec) administrative database of hospitalizations, obtained from the *Régie de l’assurance maladie du Québec* (RAMQ). We enrolled patients into this cohort on the month that two conditions were satisfied: 1) they had at least one diagnosis of a respiratory illness (the exact list of respiratory International Classification of Diseases, 9th Revision [ICD-9] codes is given in Table [icd\_codes] the Appendix) between January 1st, 1996 and March 31, 2006 (the study period), while living in the 2006 census metropolitan area of Montreal, and 2) were at least 65 years of age. We used this cohort because it contains the majority of 65-year-old patients who were hospitalized in the region during the study period.

From among this cohort, we selected hospital discharges for those who had accrued at least one continuous year in the cohort preceding the day of admission. We restricted our data to only the discharges from the twenty hospitals with the most discharges of patients 65 years of age or older within the study period; the twenty hospitals accounted for 75% of all such discharges. We only selected hospital discharges which resulted from hospital stays of at least one day. Therefore, the earliest possible hospital discharge was January 2, 1997.

From among the identified hospital discharges, we selected only those with one of three high-volume admission diagnoses with high rates of hospital readmissions: pneumonia, acute myocardial infarction (AMI), and heart failure. We identified each of the admission diagnoses using ICD-9 codes; for pneumonia we used codes ranging from 480-487, for heart failure we used all 428 codes, and for AMI we used all 410 codes. The following methods were applied individually to all three disease subsets.

## Hospital readmissions

The unit of analysis was the hospital discharge; a person could be discharged multiple times. A hospital readmission was defined as an emergency hospital admission to any Quebec hospital in the 30 days following a discharge. A person who died or had a non-emergency readmission in the 30 days following discharge was considered not readmitted.

## Confounders and Risk Factors

For each hospital discharge, we collected plausible confounders that measured states at the time of, or prior to, admission. We used the demographic characteristics (age at time of admission (years), sex, birth year-month), the number of previous readmissions (within the preceding year), the admission diagnosis (as measured by the specific ICD-9 code). We also included the day of week of discharge, which has been previously shown to have an association with readmissions10, and the month of discharge, because we hypothesized that readmission risk would vary by season in Montreal.

Additionally, for each discharge, we collected the Quebec hospital diagnoses, Quebec hospital procedures, and drugs dispensed outside of the hospital but inside Quebec, in the year preceding the admission. The hospital procedures were recorded in the Canadian Classification of Diagnostic, Therapeutic, and Surgical Procedures (CCP) system. Hospital diagnostic codes were coded using the ICD-9 system. Finally, drugs which were prescribed and dispensed outside the hospital, and were being taken on the day of admission were also recorded for each patient in the *code commune* system, which categorizes drugs based on the chemical compound. To ease computation, before fitting any model, we removed any diagnosis, procedure or drug that occurred less than 30 times among all discharges. We chose 30 because it appeared to be a natural breakpoint; if the number of variables included is a function of the threshold, then the first derivative of dropped at 30 for all three disease categories.

We believed that residential location would strongly affect the probability of admission to the hospital nearest that census tract. We included it in our models because we also expected it to crudely approximate a (expected) confounder: socio-economic status. We used the residential postal code at the time of admission to assign each patient in the cohort to a census tract, as defined by the 2006 Canadian census. (Census tracts contain between 2,500 and 8,000 people, and, at the time of their creation, are demarcated so as to maximize homogeneity of socioeconomic characteristics.)11

## Statistical Analyses

For each discharge , we sought to estimate the effect of each of the twenty hospitals on 30-day readmission (), accounting for the vector of confounders (). To estimate this risk, we used targeted maximum likelihood estimation, which consisted of several steps. We first estimated of a model of the propensity score (using random forest described below). Next, we estimated of a model of readmission risk based on the confounders and the variables for each of the hospitals . We then calculated (sometimes referred to as the clever covariate) described in equation [logistic\_clever\_covariate]

(where is the indicator function which evaluates to 1 when its argument is true, and 0 otherwise), and solved for all in fluctuation function described in equation [logistic\_fluctuation\_function].

We solved for all twenty by regressing the 30-day readmission outcome (with a logit link function) onto (with no intercept) offset by the inverse logit of the initial estimate of readmission risk . Finally, for each discharge, we computed the estimated risk of 30-day readmission for all twenty counterfactual conditions (the risk of readmission for every discharge as if they had attended different hospital) using Equation [Q\_star\_logistic].

For each hospital, we then calculated the mean readmission risk () and associated odds ratio.

To estimate both models and , we used a random forest, a non-parametric model based on decision trees.9 Decision trees use the independent variables () to repeatedly split data into partitions that are as homogeneous as possible with respect to the outcome of interest (specifically measured with the Gini coefficient12). Random forest improves decision trees by using bootstrap aggregation (bagging); multiple decision trees are grown on bootstrap replicates (sampled with replacement) to avoid overfitting. Additionally, within each tree, only a sample of the covariates is used (in our case we used a square root of the number of variables included in the mode, rounded down).

For both models and , we arbitrarily chose to grow 1200 trees, and then measured the accuracy as a function of the number of trees to ensure that growing further trees would be unlikely to improve accuracy. Because the model was used solely to estimate the *probability* of admission to to specific hospitals (and not to predict exactly which hospital was attended), when calculating the Gini coefficient to build the trees we configured the model to favor calibration over discrimination: we weighted each of the twenty predicted hospitals by the inverse of the proportion of discharges at that hospital. When measuring the accuracy for each discharge, we only used trees for which the discharge was “out-of-bag”, that is, we only used trees for which the bootstrap sample did not include the discharge.

To describe importance of the covariates in both models and , for each variable, we measured the decrease in the Gini coefficient for each partition in which the variable was used, in every tree. A low Gini (i.e. higher decrease in Gini) means that a particular predictor variable plays a greater role in partitioning the data into the defined classes. We plotted the densities of variables with four different classes (census tract, procedure, diagnosis and drug) at different levels of Gini decreases.

Random forest classifies each item by majority vote: each tree in the forest assigns each discharge to a specific class. Although the vote proportion is between zero and one, it is not calibrated well as a probability. To calibrate the vote proportion, we used Platt scaling13 (logistic regression of the outcome () on to the vote proportion).

When the probability of exposure is very low, that discharge would receive a large weight in estimating . For any below some fixed value , we set to . We recomputed our analyses at 31 different values of , ranging from to , decreasing the exponent at intervals of 0.1.

Finally, we compared our results of our analysis with a logistic regression for 30-day readmission. In this model, we included only the age, sex, number of previous admissions, and the Charlson comorbidity score (Elixhauser version)14, along with indicator variables representing the hospitals themselves.

## Software

The data were cleaned and prepared for statistical analysis using the Postgres relational database (version 9.2.6). We implemented our models using the R statistical package (version 3.1.1),15 using the “bigrf” package (version 0.1.11) to grow the random forests.16 We plotted our figures using the “ggplot2” package (version 1.0.0) .17 All the code to develop used to process our data, fit our models, and typeset this article is [available for download at Github](https://github.com/nograpes/tmle_readmissions).

# Results

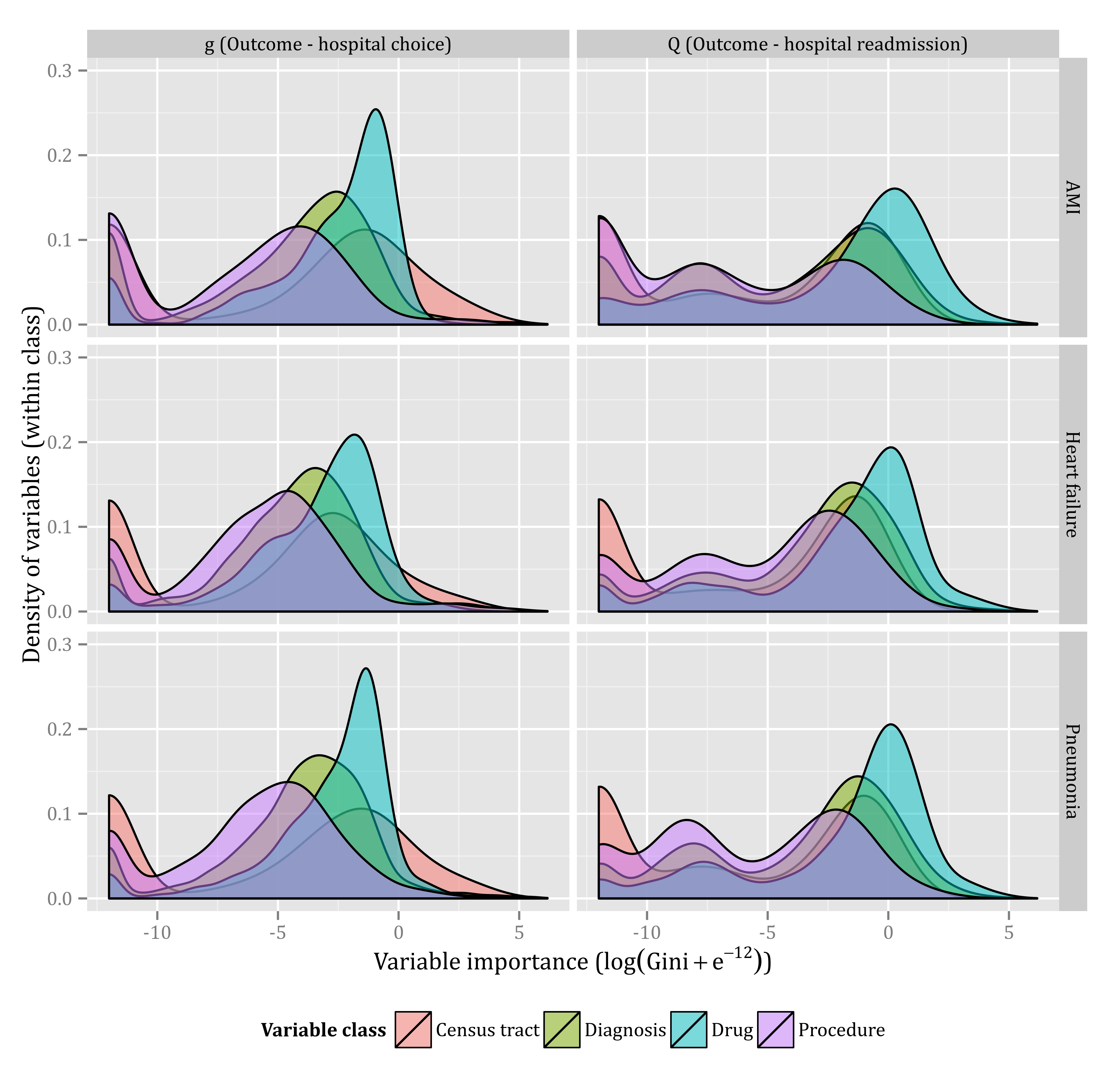
Over the course of January 2, 1996 to March 31, 2006, 482,064 people were entered into our cohort. Among these, 16,521 were ever admitted for pneumonia, 13,884 were ever admitted for AMI, and 15,822 were ever admitted for heart failure. People ever admitted for pneumonia had a mean (median) 1.2 (1) pneumonia admissions, heart failure patients had a mean (median) 1.6 (1) heart failure admissions, and AMI patients had a mean (median) 1.1 (1) AMI admissions. In total, we analyzed 20,421 pneumonia discharges, 15,746 AMI discharges, and 24,847 heart failure discharges.

The accuracy of the random forest (for both models and ) did not appear to improve significantly beyond 125 trees (see Figure [fig:error\_rate\_for\_g\_and\_Q] in the appendix). In Figure [fig:variable\_importance\_by\_model\_and\_class] we plot the importance of variables (as measured by the Gini coefficient) in the random forest models for four variable classes, for all disease subsets for both the and model. Although census tracts were found to be important in prediction of hospital choice, the other three variable classes were had a high density of important variables as well. The prescription drugs in particular had a high proportion of important variables, and generally the lowest proportion of unimportant variables. For the model, the variable density appeared bimodal within variable importance for all four variable classes. Additionally, the pre-admission drug prescriptions appeared to be strongly important in predicting readmission for all pneumonia, heart failure, and AMI admissions.

The predicted probability of admission to any particular hospital () was less than 5% in 88% of cases (across all disease subsets and hospitals). We set (the lower bound of when used to fit the values for ), to two different values, and . Across all disease subsets and hospitals, 39% of discharge/hospital combinations and a less than , and 4% had a less than . Figure [fig:dist\_g\_zoomed] (in the Appendix) describes the histogram of when it is below 0.05 for each disease/hospital combination separately.

The unadjusted proportion of patients readmitted in 30 days varied across hospitals for each disease subset (Tables 1-3). The linear correlation between the proportion of deaths during hospital stay and the proportion readmitted was (0.19, , ) among AMI, heart failure, and pneumonia admissions respectively. Using a model that adjusts for a few well-known confounders, for AMI, heart failure, and pneumonia respectively, one, three, and five hospitals had significantly different odds than the reference hospital. Notably, the significant odds ratios are all relatively small, with point estimates ranging from 0.92 - 1.04. In contrast, in the TMLE models, at both values of , for all admission diagnoses, nearly all of the hospitals had significantly different odds than the reference hospital.

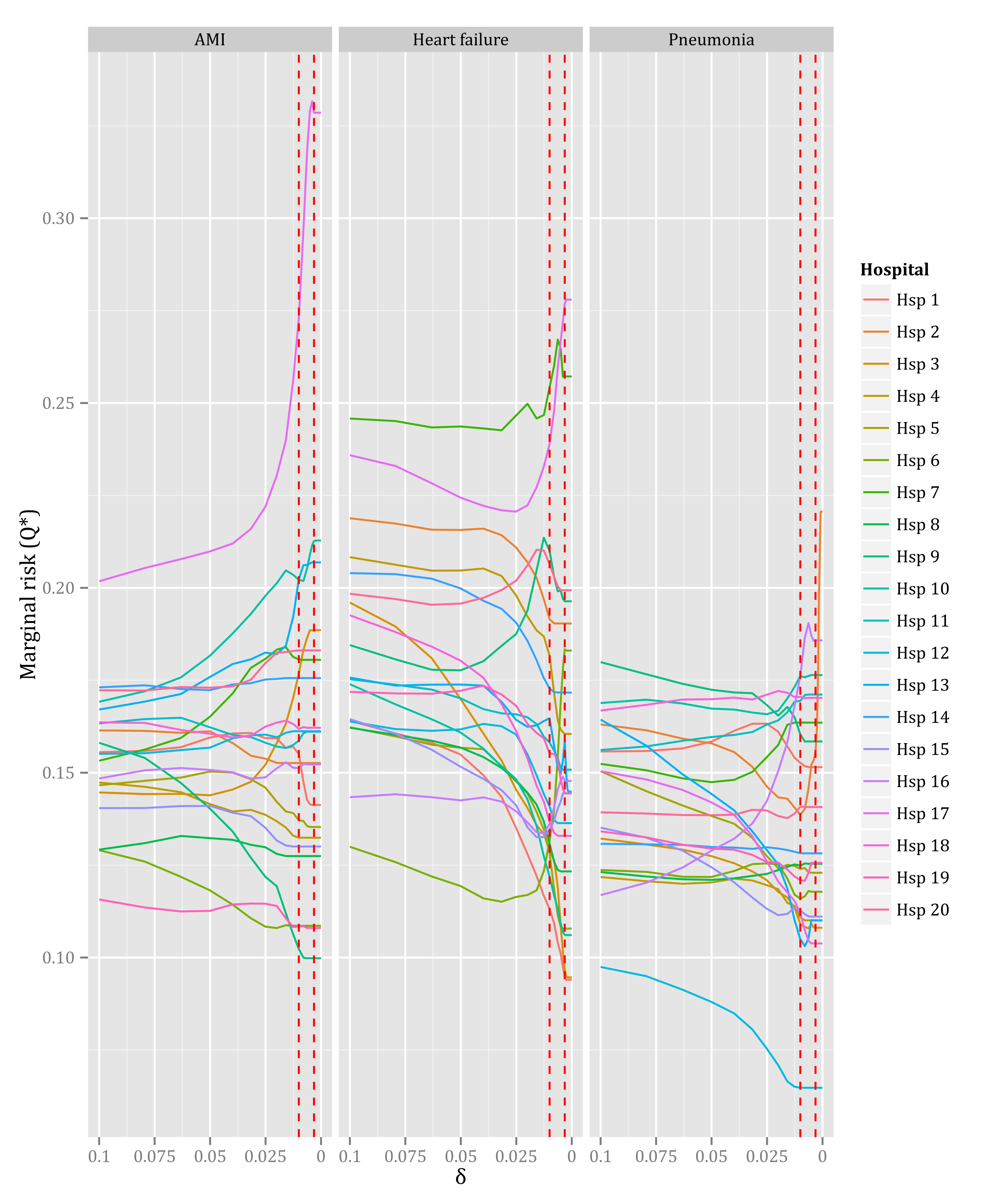
In some hospitals and disease subsets, the parameter , (the lower bound on the probability of exposure ) had a considerable effect on the marginal risk and the associated odds ratios. For example, for AMI (shown in Table 1), the marginal risk for hospital 17 increases by six percent when decreases from to . In Figure [fig:effect\_of\_gbound], we display the marginal risk for each of the twenty hospitals and disease subsets as a function of the parameter . For many hospitals, the effect was quite strong; for pneumonia admissions, hospital 16 went from having the second-lowest marginal risk when to having the highest marginal risk when .



[fig:variable\_importance\_by\_model\_and\_class]

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Admitted | Died (%) | Discharged | Readmitted (%) | Odds ratio (95% CI) | Marginal risk | Odds ratio (95% CI) | Marginal risk | Odds ratio (95% CI) | Marginal risk |
| 763 | 112 (15) | 651 | 105 (16) | 0.98 (0.95-1.01) | 0.16 | 0.86 (0.83-0.89) | 0.16 | 0.77 (0.74-0.81) | 0.14 |
| 1557 | 148 (10) | 1409 | 191 (14) | 0.97 (0.95-1.00) | 0.15 | 0.85 (0.82-0.87) | 0.15 | 0.85 (0.82-0.87) | 0.15 |
| 606 | 83 (14) | 523 | 84 (16) | 0.98 (0.95-1.02) | 0.16 | 1.01 (0.97-1.05) | 0.18 | 1.09 (1.04-1.14) | 0.19 |
| 1022 | 125 (12) | 897 | 136 (15) | 0.97 (0.94-1.00) | 0.15 | 0.72 (0.69-0.74) | 0.13 | 0.72 (0.69-0.74) | 0.13 |
| 729 | 150 (21) | 579 | 98 (17) | 0.98 (0.95-1.02) | 0.16 | 0.75 (0.72-0.77) | 0.14 | 0.73 (0.71-0.76) | 0.14 |
| 826 | 119 (14) | 707 | 106 (15) | 0.98 (0.94-1.01) | 0.15 | 0.57 (0.54-0.60) | 0.11 | 0.57 (0.54-0.60) | 0.11 |
| 1491 | 241 (16) | 1250 | 216 (17) | 0.99 (0.96-1.01) | 0.16 | 1.04 (1.01-1.06) | 0.18 | 1.03 (1.01-1.06) | 0.18 |
| 1270 | 198 (16) | 1072 | 138 (13) | 0.95 (0.92-0.98) | 0.13 | 0.69 (0.67-0.71) | 0.13 | 0.69 (0.67-0.71) | 0.13 |
| 780 | 152 (19) | 628 | 130 (21) | 1.01 (0.97-1.05) | 0.19 | 0.54 (0.51-0.56) | 0.1 | 0.52 (0.50-0.54) | 0.1 |
| 778 | 124 (16) | 654 | 123 (19) | 1.01 (0.97-1.05) | 0.19 | 1.19 (1.15-1.23) | 0.2 | 1.27 (1.22-1.31) | 0.21 |
| 705 | 125 (18) | 580 | 97 (17) | 0.99 (0.96-1.03) | 0.17 | 0.89 (0.85-0.92) | 0.16 | 0.90 (0.86-0.94) | 0.16 |
| 1284 | 266 (21) | 1018 | 166 (16) | 0.99 (0.96-1.02) | 0.16 | 0.90 (0.88-0.93) | 0.16 | 0.90 (0.88-0.93) | 0.16 |
| 739 | 86 (12) | 653 | 110 (17) | 0.99 (0.95-1.02) | 0.16 | 1.19 (1.16-1.23) | 0.2 | 1.22 (1.18-1.27) | 0.21 |
| 1307 | 184 (14) | 1123 | 210 (19) | (Reference) | 0.18 | (Reference) | 0.18 | (Reference) | 0.18 |
| 1152 | 168 (15) | 984 | 129 (13) | 0.97 (0.95-1.01) | 0.15 | 0.70 (0.68-0.73) | 0.13 | 0.70 (0.68-0.73) | 0.13 |
| 408 | 70 (17) | 338 | 43 (13) | 0.97 (0.93-1.01) | 0.15 | 0.84 (0.80-0.88) | 0.15 | 0.84 (0.80-0.89) | 0.15 |
| 807 | 123 (15) | 684 | 134 (20) | 1.02 (0.99-1.06) | 0.2 | 1.76 (1.72-1.81) | 0.27 | 2.30 (2.23-2.37) | 0.33 |
| 894 | 144 (16) | 750 | 116 (15) | 0.98 (0.95-1.01) | 0.16 | 0.91 (0.87-0.94) | 0.16 | 0.91 (0.87-0.95) | 0.16 |
| 499 | 94 (19) | 405 | 50 (12) | 0.95 (0.91-0.99) | 0.13 | 0.57 (0.53-0.61) | 0.11 | 0.57 (0.53-0.61) | 0.11 |
| 1025 | 184 (18) | 841 | 143 (17) | 0.99 (0.96-1.02) | 0.17 | 1.05 (1.02-1.09) | 0.18 | 1.05 (1.02-1.09) | 0.18 |

The proportion of those who were readmitted within 30 days is caluculated using the number discharged alive as the denominator. The confidence intervals for the odds ratios for the parameters in the logistic regression model were calculated using the profile likelihood method. The marginal risk for the odds ratios was calculated by using the regression model to calculate the mean predicted probability of readmission for every admission, except individually fixing the hospital attended to one hospital. The parameter represents the lower bound on the probability of exposure to that hospital (); we display odds ratios and marginal risks for two versions of the TMLE model with varying levels of .



[fig:effect\_of\_gbound]

# Discussion

Using targeted maximum likelihood estimation (TMLE) to adjust precisely for measured confounders, we found that the differences in marginal risk of 30-day hospital readmission in twenty Montreal hospitals were much stronger than a model that only crudely adjusted for confounding readmission risk suggested. Additionally, our study revealed some practical positivity violations for some hospitals, suggesting that the relative readmission risk may not always be estimable from observed data.

Our study has several strengths. By using a doubly-robust estimation technique, and by accurately adjusting for thousands of plausible confounders, we minimized the bias in our estimates of the effect of hospital care on readmissions. Our work suggests that the difference the bias reduction was not trivial; in assessing the effect of hospitals on readmission, it has a substantive effect as well. Also, since we did not have to restrict our cohort to a single healthcare insurance network, we had a large cohort of patients from all socioeconomic classes. Because we had complete access to all hospital visits in the province, we could accurately measure which patients were readmitted.

Other hospital readmission studies have applied statistical and machine learning algorithms to readmission data to develop predictive models7,19, including one using the data used in this study.8 Most studies, including our own, found relatively poor accuracy. No study to our knowledge has used machine learning algorithms to draw causal inference target parameters. Predictive models of hospital readmissions may not be very accurate, but our study demonstrates that machine learning techniques can improve our ability to draw inference on target parameters.

Some authors3,20 believe that by using readmission rates as a quality metric, we assume that readmissions are preventable. Hoping to develop a quality metric that compares preventable readmissions, some researchers have attempted to identify which individual readmissions were preventable. Some studies have clinicians classify individual readmissions as preventable21–23, despite evidence that clinicians cannot reliably measure preventability.24 Other studies use pairs of admission/readmission diagnosis codes that identify “potentially preventable” readmissions.25 However, the proportion of those actually preventable among the “potentially” preventable differs among hospitals26, meaning that potentially preventable readmissions are not an adequate proxy for preventable readmissions.20

But to estimate the effect of an exposure (like hospital care) on an outcome (like readmission), we do not need to identify exactly which individuals would not have had the outcome if they were not exposed.27 Some readmissions are unpreventable: no matter where they were treated, they would be readmitted. If patients were randomized among different hospitals, the number of unpreventable readmissions would be (asymptotically) the same among all hospitals, and any difference in readmission rates would be the “preventable proportion”. Since the patients were not randomized to each hospital, we attempted to recreate that situation by controlling for confounding. Assuming that we have adequately controlled for confounding, we have estimated the independent effect of each hospital on readmission risk, without identifying whether *individual* readmissions were preventable. In this study, practical positivity violations occur when large subgroups of the hospitalized patients are rarely admitted to specific hospitals. Practical positivity violations can bias our estimates of the parameter of interest, because our risk estimates are heavily dependent on the few admitted patients from certain subgroups, and on the precision of our estimate of the probability of their attendance. For some hospitals, our estimates for marginal risk were sensitive to the parameter , which set a lower bound on the probability of exposure , suggesting practical positivity violations. We believe that the discovery of practical positivity violations is an important finding: observational data may not provide us with enough information to meaningfully compare certain hospitals. To avoid adjusting for a variable on the causal pathway between hospital care and readmission, we did not adjust for hospital length of stay.28 We also excluded all diagnoses and procedures that occurred during the hospital admission, because these covariates were also effectively mediators between hospital care and readmission.

The major competing risk for 30-day hospital readmission is death, but others include moving outside the study area, or admission to a hospital for a non-emergency reason. In our analysis, we did not account for these competing risks. If patients died within 30 days of discharge more often at one hospital than another, we could have biased our estimate of readmission risk. Similarly, if patients died during the hospitalization more often at some hospitals than others, it could have created a selection bias (left censorship) in which hospitals with better care were discharging sicker (but still living) patients, who would be more likely to be readmitted. Also, there is no special significance of 30 days in readmission, except for the fact that it is (recently) widely used as a cutoff. In future work, we plan to account for both left censorship and competing risks in a model that estimates the effect of hospital care on time-to-readmission.

Differing *admission* practices can strongly affect the rates of readmission.29 In some cases, like a major trauma, admission is certain, but in most cases, there is some variation in practice of how patients are admitted. Since most (89% in this cohort) patients who are readmitted within 30 days are readmitted to the same hospital that they were discharged from, a hospital that is more likely to admit patients will have a higher readmission rate. In future work, we plan to study the effect of the probability of admission on readmission.

Entry to our cohort was dependent on having one diagnosis of a respiratory illness in an inpatient or outpatient setting. Respiratory illness was defined rather broadly, including extremely common diagnoses such as “cough”. We expect that the the majority of 65-year-old patients who would be hospitalized would have at least one respiratory illness diagnosis in an outpatient setting. We cannot, however, exclude the possibility that parameter estimates were affected by selection bias with respect to the full population of 65-year-old patients.

The effect of hospital care on readmissions is confounded by a vast spectrum of health-related states of the admitted patients. In the absence of a clear theoretical basis of the structure of that confounding, we can 1) identify relatively few, well-understood and measurable confounders to include in our model, or 2) forgo any theoretical understanding of the structure of confounding, and attempt to identify the broadest measurable set of even faintly plausible confounders. The first option has some advantages: in a situation where data collection is expensive, it may not be plausible to measure thousands of variables. Additionally, by reducing the confounders to a well-understood few, the model gains credibility because it can be shown that the confounders are having the expected effect. Non-parametric techniques such as random forest don’t allow us to look (easily) at the individual effects of the confounders, and even in a parametric model it would be difficult to analyze thousands of variables. We summarized the densities of the effects a few classes of variables in Figure [fig:variableimportancebymodelandclass], but this still does not allow the variable-by-variable analysis typical in an epidemiologic study. Also, by including many confounders we also risk *inducing* bias, such as the M-bias30,31. However, the recent availability of large scale healthcare administrative data has put us into the situation where the cost of data collection is relatively low. By using machine learning techniques like random forest, we also automatically fit multi-way interactions that we would be unlikely to explore in a model fit “by hand”. Finally, because the structure of the confounding is unclear, we cannot assess if M-bias is present, and some research suggests that the scale of M-bias may be small when compared to traditional confounding.32 We argue that in this situation, where we have a large data set, thousands of measurable confounders, and little understanding of the structure of confounding, the second option is more appropriate.

Despite a relatively standardized data collection process, some hospitals may have idiosyncratic code usage patterns, leading to differing specificity and sensitivity of some diagnostic and procedural codes. This possible differential misclassification could have biased our estimate of the parameters of interest.

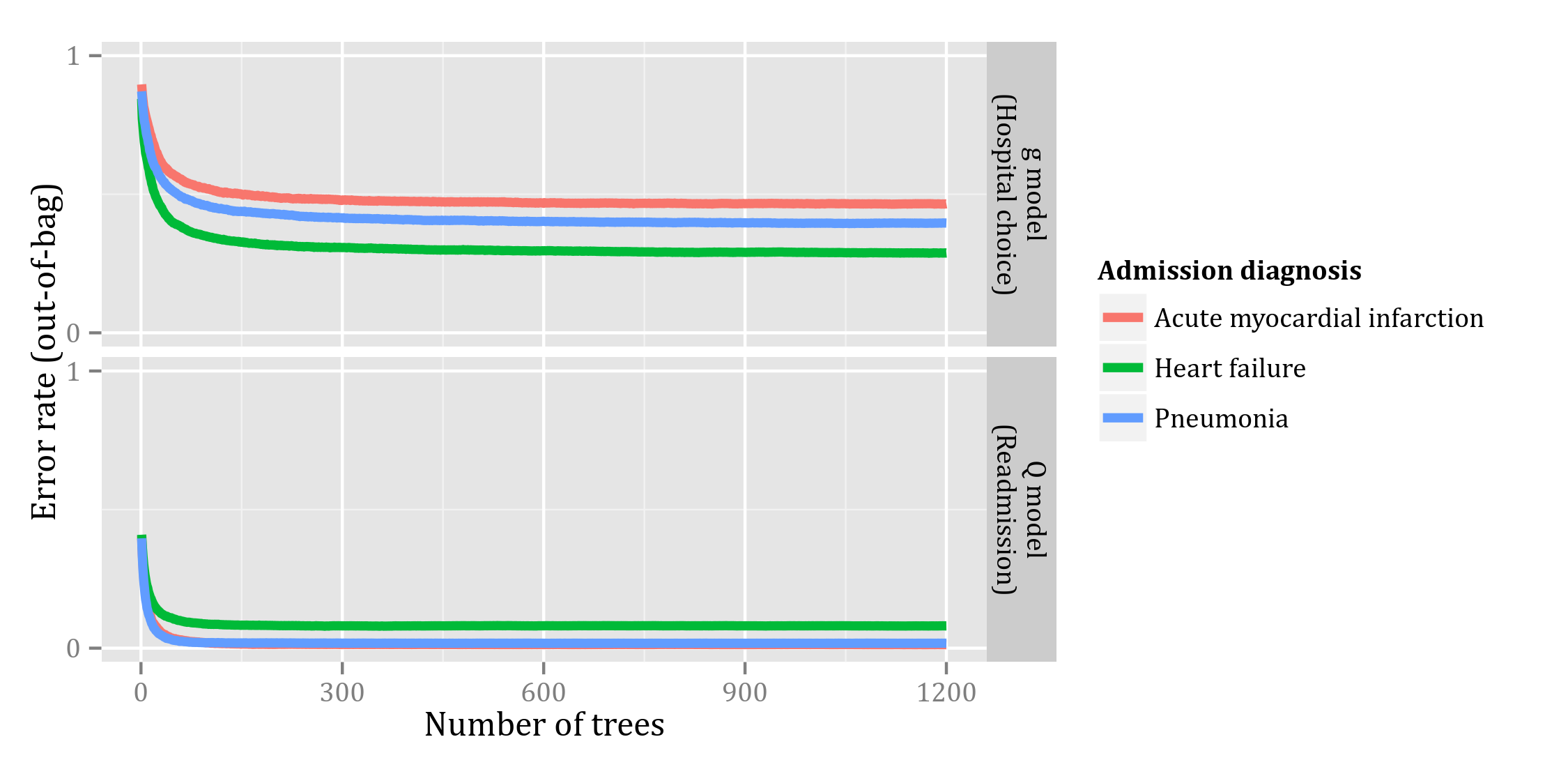
The unit of analysis in this study was the discharge, but each discharge was “clustered” within a patient. The expected within-cluster homogeneity could have biased our estimates of variance, and our parameter estimates. However, because the number of clusters (unique patients) was relatively high when compared to the sample size (the number of discharges), we do not expect that our parameter or variance estimates to be biased very strongly.

Beside random forest, we could have used many other machine learning techniques on these data, many of which we explored in other work.8 Also, some ensemble machine learning techniques, (in particular SuperLearner33 which is commonly used with TMLE), are available, that combine any number of other machine learning techniques. We found that in these data, ensemble learning techniques were too computationally expensive. We selected random forest because of its relative simplicity, and because our variables were nearly all binary, for which decision trees are particularly suitable.

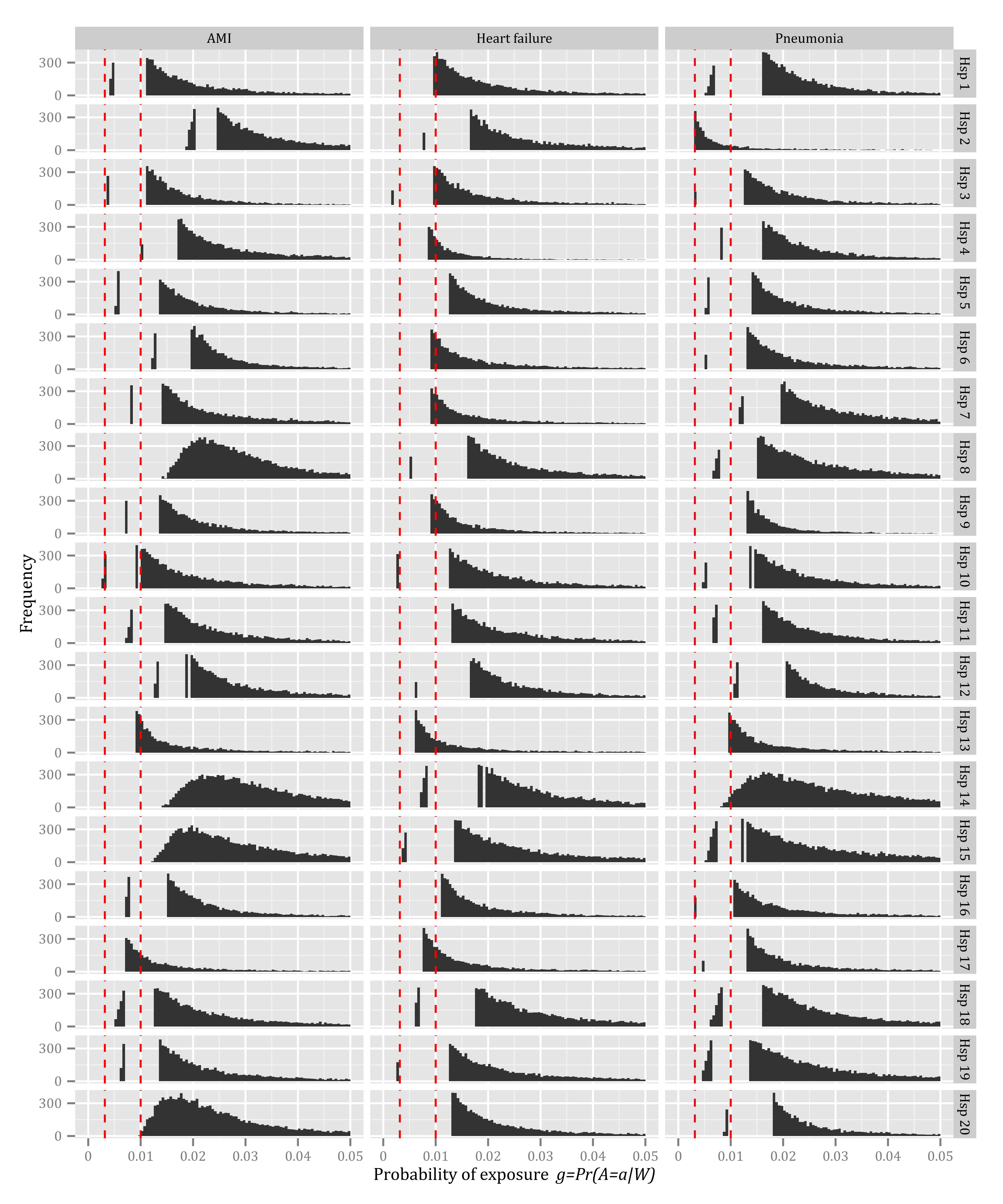
Calibration of the random forest vote proportions in the model strongly affected estimates of our parameter of interest in the update step. Other articles using non-parametric techniques typically combined them with other models in an ensemble learner (like SuperLearner). The final step in (many) ensemble learners is to combine all the probability estimates in parametric model, which would effectively calibrate the probabilities. In our study, a single, non-parametric technique was used, so an additional, separate calibration step was necessary to convert the ranking scores into a probability estimate.

Hospital readmissions can be a relatively crude proxy for quality of care, but they can still provide valuable insight. In a seminal research article on quality of care measures, Donabedian writes: “But how precise do estimates of quality have to be? At least the better methods have been adequate for the administrative and social policy purposes that have brought them into being. The search for perfection should not blind one to the fact that present techniques of evaluating quality, crude as they are, have revealed a range of quality from outstanding to deplorable.”34 Our work suggests that, when finely adjusted for confounding, hospital readmissions reveal wide differences in hospital quality of care.

# Appendix



[fig:errorrateforgandQ]



[fig:distgzoomed]

ll&100&Primary tuberculous infection, primary tuberculous complex101&Tuberculous pleurisy in primary progressive tuberculosis108&Other primary progressive tuberculosis109&Primary tuberculous infection, unspecified110&Tuberculosis of lung, infiltrative111&Tuberculosis of lung, nodular112&Tuberculosis of lung with cavitation113&Tuberculosis of bronchus114&Tuberculous fibrosis of lung115&Tuberculous bronchiectasis116&Tuberculous pneumonia (any form)117&Tuberculous pneumothorax118&Other pulmonary tuberculosis119&Pulmonary tuberculosis, unspecified203&Plague primary pneumonic204&Plague secondary pneumonic205&Plague pneumonic, unspecified219&Tularaemia221&Pulmonary anthrax249&Glanders259&Melioidosis320&Faucial diphtheria321&Nasopharyngeal diphtheria322&Anterior nasal diphtheria323&Laryngeal diphtheria329&Diphtheria, unspecified330&Whooping cough, bordetella pertussis (B.pertussis)331&Whooping cough, bordetella parapertussis (B.parapertussis)338&Whooping cough, other specified organism339&Whooping cough, unspecified organism340&Streptococcal sore throat529&Chickenpox&Chickenpox (varicella), uncomplicated551&Measles, postmeasles pneumonia739&Ornithosis741&Specific diseases due to coxsackie virus, epidemic pleurodynia790&Adenovirus793&Rhinovirus798&Other viral infection799&Viremia, unspecified&Unspecified viral infection830&Rickettsioses, Q-fever1124&Candidiasis, of lung1149&Coccidioidomycosis1150&Infection by histoplasma capsulatum1151&Infection by histoplasma duboisii1159&Histoplasmosis, unspecified1309&Toxoplasmosis1363&Pneumocystosis3820&Suppurative and unspecified otitis media, acute suppurative otitis media3824&Unspecified suppurative otitis media3829&Unspecified otitis media4609&Acute nasopharyngitis (common cold), acute nasopharyngitis (common cold)4618&Acute sinusitis, other4619&Acute sinusitis, unspecified4629&Acute pharyngitis, acute pharyngitis4639&Acute tonsillitis, acute tonsillitis4640&Acute laryngitis4641&Acute tracheitis4642&Acute laryngotracheitis4643&Acute epiglottitis4644&Acute laryngitis and tracheitis, croup4650&Acute laryngopharyngitis4658&Other multiple sites4659&Acute upper respiratory infections, unspecified site&URTI, unspecified4660&Acute bronchitis4661&Acute bronchiolitis4789&Other and unspecified diseases of upper respiratory tract4800&Viral pneumonia, pneumonia due to adenovirus4801&Viral pneumonia, pneumonia due to respiratory syncytial virus4802&Viral pneumonia, pneumonia due to parainfluenza virus4808&Viral pneumonia, pneumonia due to other virus, not elsewhere classified4809&Viral pneumonia, viral pneumonia, unspecified4819&Pneumococcal pneumonia4820&Other bacterial pneumonia, pneumonia due to klebsiella pneumoniae4821&Other bacterial pneumonia, pneumonia due to pseudomonas4822&Pneumonia due to haemophilus influenzae (h.influenzae)4823&Other bacterial pneumonia, pneumonia due to streptococcus4824&Other bacterial pneumonia, pneumonia due to staphylococcus4828&Other bacterial pneumonia, pneumonia due to other specified bacteria4829&Other bacterial pneumonia, bacterial pneumonia, unspecified4839&Pneumonia due to other specified organism4841&Cytomegalic inclusion disease4843&Pneumonia in infectious diseases classified elsewhere, whooping cough4845&Pneumonia in infectious diseases classified elsewhere, anthrax4846&Pneumonia in infectious diseases classified elsewhere, aspergillosis4847&Pneumonia in other systemic mycoses4848&Pneumonia in other infectious diseases4859&Bronchopneumonia, organism unspecified4869&Pneumonia, organism unspecified&Pneumonia, unspecified4870&Influenza, with pneumonia4871&Influenza (flu) NOS&Influenza, with other respiratory manifestations4878&Influenza, with other manifestations4909&Bronchitis, not specified as acute or chronic4910&Simple chronic bronchitis4911&Mucopurulent chronic bronchitis4918&Other chronic bronchitis4919&Chronic bronchitis, unspecified5070&Pneumonitis due to solids and liquids, due to inhalation of food or vomit5071&Due to inhalation of oils and essences5078&Pneumonitis due to solids and liquids, other5110&Pleurisy, without mention of effusion or current tuberculosis5111&With effusion, with mention of a bacterial cause other than tuberculosis5118&Pleurisy, other specified forms of effusion, except tuberculosis5119&Pleurisy, unspecified pleural effusion5130&Abscess of lung5131&Abscess of mediastinum5180&Other diseases of lung, pulmonary collapse5184&Other diseases of lung, acute oedema of lung, unspecified5188&Other diseases of lung, other diseases of lung, not elsewhere classified5192&Other diseases of respiratory system, mediastinitis7806&Chills&General symptoms, pyrexia of unknown origin&General symptoms: fever, not otherwise specified&Hyperthermia7841&Symptoms involving head and neck, throat pain7860&Dyspnoea and respiratory abnormalities&Shortness of breath7861&Symptoms involving respiratory system and other chest symptoms, stridor7862&Symptoms involving respiratory system and other chest symptoms, cough7865&Symptoms involving respiratory system and other chest symptoms, chest pain&Pleurodynia7953&Nonspecific positive culture findingsV018&Other communicable diseases

1. Institute of Medicine (U.S.). Division of Health Care Services, Lohr KN, Institute of Medicine (U.S.). Committee to Design a Strategy for Quality Review and Assurance in Medicare, United States. Health Care Financing Administration. *Medicare : A Strategy for Quality Assurance*. Washington, D.C.: National Academy Press; 1990.

2. US Government. Public Law 111 - 148 - Patient Protection and Affordable Care Act. 2012. Available at: <http://www.gpo.gov/fdsys/pkg/PLAW-111publ148/content-detail.html>.

3. Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: a systematic review. *JAMA: The Journal of the American Medical Association*. 2011;306(15):1688-1698. doi:[10.1001/jama.2011.1515](http://dx.doi.org/10.1001/jama.2011.1515).

4. Glymour MM, Osypuk TL, Rehkopf DH. Invited Commentary: Off-Roading With Social Epidemiology–Exploration, Causation, Translation. *American Journal of Epidemiology*. 2013;178(6):858-863. doi:[10.1093/aje/kwt145](http://dx.doi.org/10.1093/aje/kwt145).

5. Friedman JH, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*. 2010;33(1):1-22. Available at: <http://www.jstatsoft.org/v33/i01>.

6. Laan MJ van der, Rose S. *Targeted Learning: Causal Inference for Observational Data*. New York, NY: Springer New York; 2011. Available at: <http://link.springer.com/10.1007/978-1-4419-9782-1>. Accessed July 19, 2014.

7. He D, Mathews SC, Kalloo AN, Hutfless S. Mining high-dimensional administrative claims data to predict early hospital readmissions. *Journal of the American Medical Informatics Association*. 2014;21(2):272-279. doi:[10.1136/amiajnl-2013-002151](http://dx.doi.org/10.1136/amiajnl-2013-002151).

8. Hosseinzadeh A, Izadi M, Verma A, Precup D, Buckeridge D. Assessing the Predictability of Hospital Readmission Using Machine Learning. 2013. Available at: <http://www.aaai.org/ocs/index.php/IAAI/IAAI13/paper/view/6475/6431>.

9. Breiman L. Random Forests. *Machine Learning*. 2001;45(1):5-32. doi:[10.1023/A:1010933404324](http://dx.doi.org/10.1023/A:1010933404324).

10. Walraven C van, Bell CM. Risk of death or readmission among people discharged from hospital on Fridays. *CMAJ: Canadian Medical Association journal = journal de l’Association medicale canadienne*. 2002;166(13):1672-1673.

11. Statistics Canada. 2006 Census Dictionary. 2007. Available at: <http://www12.statcan.ca/english/census06/reference/dictionary/index.cfm>. Accessed August 15, 2014.

12. Gini C. *Variabilità E Mutabilità: Contributo Allo Studio Delle Distribuzioni E Delle Relazioni Statistiche.* Bologna: C. Cuppini; 1912.

13. Platt JC. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. In: *Advances in Large Margin Classifiers*. MIT Press; 1999:61-74.

14. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity Measures for Use with Administrative Data: *Medical Care*. 1998;36(1):8-27. doi:[10.1097/00005650-199801000-00004](http://dx.doi.org/10.1097/00005650-199801000-00004).

15. Team RC. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2014. Available at: <http://www.R-project.org/>.

16. Lim A, Breiman L, Cutler A. *Bigrf: Big Random Forests: Classification and Regression Forests for Large Data Sets*.; 2014. Available at: <https://github.com/aloysius-lim/bigrf>.

17. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. Springer New York; 2009. Available at: <http://had.co.nz/ggplot2/book>.

18. Silverman BW. *Density Estimation for Statistics and Data Analysis*. CRC press; 1986.

19. Walraven C van, Dhalla IA, Bell C, et al. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. *Canadian Medical Association Journal*. 2010;182(6):551-557. doi:[10.1503/cmaj.091117](http://dx.doi.org/10.1503/cmaj.091117).

20. Clarke A. Are readmissions avoidable? *BMJ: British Medical Journal*. 1990;301(6761):1136.

21. Witherington EMA, Pirzada OM, Avery AJ. Communication gaps and readmissions to hospital for patients aged 75 years and older: observational study. *Quality and Safety in Health Care*. 2008;17(1):71-75. doi:[10.1136/qshc.2006.020842](http://dx.doi.org/10.1136/qshc.2006.020842).

22. Stanley A, Graham N, Parrish A. A review of internal medicine re-admissions in a peri-urban South African hospital. *South African Medical Journal = Suid-Afrikaanse Tydskrif Vir Geneeskunde*. 2008;98(4):291-294.

23. Ruiz B, García M, Aguirre U, Aguirre C. Factors predicting hospital readmissions related to adverse drug reactions. *European Journal of Clinical Pharmacology*. 2008;64(7):715-722. doi:[10.1007/s00228-008-0473-y](http://dx.doi.org/10.1007/s00228-008-0473-y).

24. Walraven C van, Jennings A, Taljaard M, et al. Incidence of potentially avoidable urgent readmissions and their relation to all-cause urgent readmissions. *Canadian Medical Association Journal*. 2011;183(14):E1067-E1072. doi:[10.1503/cmaj.110400](http://dx.doi.org/10.1503/cmaj.110400).

25. Halfon P, Eggli Y, Prêtre-Rohrbach I, Meylan D, Marazzi A, Burnand B. Validation of the potentially avoidable hospital readmission rate as a routine indicator of the quality of hospital care. *Medical care*. 2006;44(11):972-981.

26. Walraven C van, Bennett C, Jennings A, Austin PC, Forster AJ. Proportion of hospital readmissions deemed avoidable: a systematic review. *Canadian Medical Association Journal*. 2011;183(7):E391-E402. doi:[10.1503/cmaj.101860](http://dx.doi.org/10.1503/cmaj.101860).

27. Hernán MA, Robins JM. *Causal Inference (Draft - May 14, 2014 )*.; 2014. Available at: <http://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/>. Accessed August 5, 2014.

28. Ash AS, Fienberg SE, Louis TA, Normand S-LT, Stukel TA, Utts J. Statistical issues in assessing hospital performance: Commission by the Committee of Presidents of Statistical Societies (white paper). 2012. Available at: <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Statistical-Issues-in-Assessing-Hospital-Performance.pdf>. Accessed June 4, 2012.

29. Epstein AM, Jha AK, Orav EJ. The Relationship between Hospital Admission Rates and Rehospitalizations. *New England Journal of Medicine*. 2011;365(24):2287-2295. doi:[10.1056/NEJMsa1101942](http://dx.doi.org/10.1056/NEJMsa1101942).

30. Pearl J. Causal diagrams for empirical research. *Biometrika*. 1995;82(4):669-688. doi:[10.1093/biomet/82.4.669](http://dx.doi.org/10.1093/biomet/82.4.669).

31. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology (Cambridge, Mass)*. 2003;14(3):300-306.

32. Liu W, Brookhart MA, Schneeweiss S, Mi X, Setoguchi S. Implications of M Bias in Epidemiologic Studies: A Simulation Study. *American Journal of Epidemiology*. 2012;176(10):938-948. doi:[10.1093/aje/kws165](http://dx.doi.org/10.1093/aje/kws165).

33. Laan MJ van der, Polley EC, Hubbard AE. Super Learner. *Statistical Applications in Genetics and Molecular Biology*. 2007;6(1). doi:[10.2202/1544-6115.1309](http://dx.doi.org/10.2202/1544-6115.1309).

34. Donabedian A. Evaluating the quality of medical care. *The Milbank memorial fund quarterly*. 1966:166-206.