Abstract

Background: Hospitals in the US and other jurisdictions are being financially penalized based on estimates of their effect on 30-day readmission risk, after adjustment for confounders. Although hospital administrative data is information-rich, confounder adjustment in these models tends to be crude, risking residual confounding. 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 marginal 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 marginal readmission risk at each of the hospitals after hospitalization for heart failure, acute myocardial infarction (AMI), and pneumonia. To control for confounding, we modeled the effect of hundreds of types of pre-admission outpatient drug prescriptions, medical procedures, and diagnoses on readmission risk.

Results: Within 30 days of discharge, there were 5,520 / 24,847 (22%) heart failure readmissions, 3,183 / 20,421 (16%) pneumonia readmissions, and 2,525 / 15,746 (16%) AMI readmissions. Within each hospital, there was a wide variation in crude readmission risk across the twenty hospitals for pneumonia 3,183 / 20,421 (16%), heart failure 5,520 / 24,847 (22%), and AMI 2,525 / 15,746 (16%). However, after control for confounding, the marginal risk for readmission within all hospitals was nearly the same (within each admission category).

Conclusion: Although crude 30-day readmission risk estimates show variation between different hospitals, no difference was found in marginal risks after control for confounding.

1 Introduction

In the early eighties, hospital administrators in the US sought to reduce hospitalization costs by changing the reimbursement system. Instead of paying hospitals per day of hospitalization, hospitals were paid a fixed rate for the type of hospitalization and the procedures performed. Following implementation of this law, the length of stay at hospitals dropped dramatically, although evidence exists that is was already in decline.

Some worry that the new system created a perverse incentive to discharge patients early, and admit them again at a later date. To ensure proper quality of care in the hospitals while keeping costs controlled, administrators have sought to establish useful quality of care metrics. Hospital readmissions have been identified as a simple metric that can establish a baseline of care; if an abnormally high number of patients from a certain hospital are quickly readmitted, it could indicate poor quality of care.

Since hospitals admit patients with varying risk of readmission, it is important to accurately estimate the effect of hospital treatment on readmission independent of patient-level confounders. Without effectively controlling for confounding, we risk unfairly penalizing hospitals that treat sicker (more likely to be readmitted) patients. Fortunately, hospital and outpatient administrative data is information-rich. Drug prescriptions, diagnoses, and medical procedures can provide important information on how the effect of hospital care on readmission risk is confounded by patient health.

However, in most statistical models of readmission risk, the hospital administrative data is simplified to a few well-known confounders (age, sex, previous readmissions), and sometimes a summary "comorbidity score". If each drug, diagnosis and procedure was modeled with a separate covariate, the model would be very computationally expensive to fit. Such a model would also be very unwieldy to develop; analyzing how inclusion or exclusion of variables affects the model would be impossible to do effectively with hundreds of covariates.

By summarizing confounders into crude risk scores, we risk "residual confounding" leading to biased effect estimates. Furthermore, to compare hospitals, we are only interested in estimating one parameter, the independent effect of hospitals on readmissions. We are only interested in the other variables insofar as they confound the effect of hospitals on readmissions, estimating the individual effect of each of these variables is unnecessary, and statistically wasteful.

Non-parametric machine learning techniques are available that will let us accurately discriminate patient readmission risk using hundreds of variables in a computationally efficient way. Furthermore, because they are non-parametric, we avoid having to specify a functional form, and can find complex "interactions" between variables. On the other hand, these non-parametric techniques don't allow us to isolate (target) the effect 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 from non-parametric (or parametric) models. To use TMLE, a model is developed to estimate the probability of exposure (the propensity score), and also fit another model to estimate the probability of outcome. These two probabilities are combined in a parametric model with only the parameter of interest, inversely weighted by the probability of exposure, and offset by the probability of the outcome. In this way, the discriminative power of non-parametric models can be used to extract estimates of parameters of interest.

In this study, we sought to estimate the marginal risk of readmission within 30 days of discharge for three different admission diagnoses (pneumonia, heart failure, and acute myocardial infarction) for each of twenty Montreal hospitals. We used a ten-year cohort of 65 year old people, for which we had outpatient and inpatient diagnoses and procedures, and outpatient drugs prescriptions. We used a non-parametric machine learning technique, (random forest), with TMLE to take advantage of the rich confounder data and provide less biased estimates of readmission risk at each hospital.

2 Methods

2.1 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 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 represents the majority of 65-year olds 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 time of admission. We restricted our data to only the discharges from the twenty hospitals with the most discharges of patients 65 years 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, the three initial conditions selected by the Centers for Medicare and Medicaid Services (CMS) to implement the Hospital Readmissions Reduction Program mandated by the Affordable Care Act. 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.

2.1.1 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. For the purposes of the survival analysis, "time zero" was the day of discharge; a person was considered right-censored if, before an emergency readmission, they died, had a non-emergency readmission, they changed residence to a location outside of Quebec, or the study ended.

2.2 Confounders and Risk Factors

For each hospital discharge, we colllected variables that measured states at the time of admission, or events that occurred prior to the hospital admission, and which may confound the relationship between hospital care and readmission. 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 readmissions, and the month of discharge, because we hypothesized that readmission risk would vary by seasons 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 f of the threshold, then the first derivative of f 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.)

3 Statistical Analyses

To esimate the marginal risk of 30-day readmission for each of the twenty hospitals (A_i), we used targeted maximum likelihood estimation. The process consisted of three steps 1) estimation of a propensity score $g(W)$, 2) an initial estimate of a model of readmission risk based on all confounders W_i and the variables for each of the hospitals $Q(A, W)$, 3) calculation $h(A, W)$ (sometimes referred to as the clever covariate) described in Equation 1

$$h(A, W) = \frac{I(A = a)}{g(A = a|W)} \tag{1}$$

where I is the indicator function which evaluates to 1 when its argument is true, and 0 otherwise, and 4) a final step in which we fit twenty models (Q_i^*), in which we regress the 30-day readmission outcome Y_i onto $h(A_i, W)$ offset by the initial estimate of readmission risk $Q(A_i, W_i)$ as described in Equation 2.

$$Q^* = expit(logit(Q) + \epsilon \times 1/g) \tag{2}$$

To estimate both models g and Q, we used a random forest, a non-parametric model based on decision trees[1]. Decision trees

use the independent variables to repeatedly split data into partitions that are as homogenous as possible with respect to the outcome of interest (specifically measured with the Gini coefficient[2]). Complex interactions and non-linearities are naturally modeled within this framework. 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, rounded down, included in the model).

For both Q and g, 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 significantly 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), we configured the model to favour 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 g and Q, 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 traditionally classifies each item by majority vote; we converted this into a probability by taking the proportion of votes for each hospital (using only out-of-bag trees for each discharge).

To avoid convergence problems for this one parameter model, we fit the parameter using a quasi-Newton method simultaneously discovered by Broyden[3], Fletcher[4], Goldfarb[5] and Shanno[6] (BFGS).

3.1 Time-to-event analysis

We also estimated the effect of the twenty hospitals on the time-to-readmission. In general, we followed the method for (non-collaborative) TMLE for survival analysis outlined in Stitleman, with some deviation.

The steps for TMLE in survival analysis are: 1) Estimate g the probability of exposure (attending each hospital) 2) Estimate g_2 the probability of right censorship 3) Estimate $S0$ the survival function for each person for each level of exposure. 4) Combine g, g_2 and S_0 into the clever covariate h(t,A,W)

$$h(t, A, W) = \frac{I(A = a)}{g(A = a|W) \prod_{i=1}^{t} 1 - g_2(i|A, W)} \frac{S_0(t_k|A, W)}{S_0(t|A, W)} \tag{3}$$

which is then fit in a single-parameter (no intercept) model against a binary, time-varying outcome, offset by the conditional failure probability (as derived from S_0).

The estimate g, we reused the random forest model that we developed for the 30-day radmission risk analysis. In addition, we needed to develop models that estimated the survival function of both readmission (model Q), and right-censoring (model g_2). We estimated both survival functions in two steps, we first fit a Cox proportional hazards model, and then used that model to estimate a baseline survival function using the technique described by Fleming and Harrington**fleming_nonparametric_1984** . Stitleman described a technique in which a binary outcome would be fit to every unit of person-time. Although this model would be more flexible, it would be computationally very onerous because not only did we have thousands of covariates and tens of thousands of people in each disease cohort, they each were in the cohort for hundreds of days, leading to an information matrix with hundreds of millions of elements.

For both Q and g_2, we fit a regularized Cox proportional hazards model, using cyclical coordinate descent to estimate the parameters efficiently in our sparse but large information matrix. We penalized the likelihood by the ℓ_2 norm, also known as the

lasso penalty. The scale of the penalty was determined by the parameter λ. We optimized the selection of the penalty scale for the best partial likelihood using a nested a 10-fold cross-validation. Within each fold we assessed 100 λ-values spaced evenly between max(λ)×10-4 and max(λ), where max(λ) was the smallest λ-value that would result in a model with no non-zero coefficients.

For each hospital, for each person-day, we computed the "clever covariate", which is a function of the readmission survival function, the censoring survival function, and the probability of hospital exposure. Using logistic regression (without an intercept term), we regressed the binary readmission status offset by the inverse logit of the conditional failure probability, at every unit of person-time, on to the clever covariate, yielding a single parameter epsilon. We then iterated the procedure: we repeated the same logistic regression, except that we offset the binary readmission status by the inverse logit of the conditional failure probability plus the epsilon in the previous iteration. We iterated until the epsilon was less than 1^-10.

Using the sum of the epsilons for all iterations for each hospital, we could then compute an "updated" survival function for each person for each hospital, which allowed us to estimate the expected time-to-readmission for each person as if they were cared for at each separate hospital. We compared the distribution of marginal survival times, as well as the marginal mean and median survival times.

3.2 Comparison to assessment with comorbidity scores

Finally, we compared our results of our analysis with 1) a logistic regression for 30-day readmission and 2) a Cox proportional hazards model for the time-to-readmission. In these models, we included only the age, sex, number of previous admissions, and the Charlson comoribidity score (Elixhauser version), along with indicator variables representing the hospitals themselves.

3.3 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 3.1.0)[7]. We implemented the random forest using the "bigrf" package (version 0.1.11)[8]. We implemented the GLM fitting using coordinate descent using the "glmnet" package (version 1.9.8)[9]. We plotted our figures using the "ggplot2" package (version 1.0.0)[10]. All the code to develop used to fit clean our data, fit our models, and layout this paper is available for download at Github.

4 Results

Over the course of January 2, 1996 to March 31, 2006, 482 064 people were entered into our cohort.

Among these, x were ever admitted for pneumonia, y were ever admitted for AMI, and z were ever admitted for heart failure.

People ever admitted for pneumonia had a mean of x and median x2 pneumonia admissions , ever heart failure patients had a mean h1 and median h2 heart failure admissions, and ever AMI patients had a mean a1 and median a2 heart failure admissions.

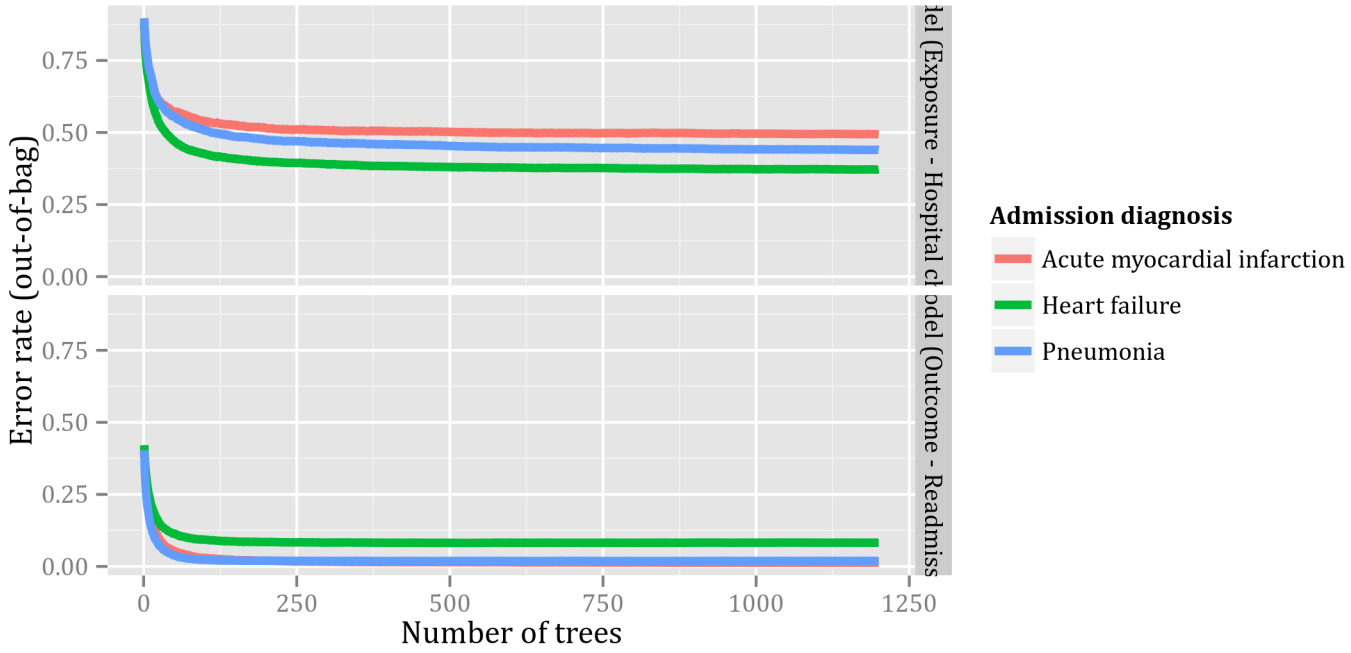

Figure 1: Error rate for both random forest models of hospital choice (g) and readmission (Q) as a function of the number of trees grown. For each admission, only out-of-bag trees were used to predict the given outcome.

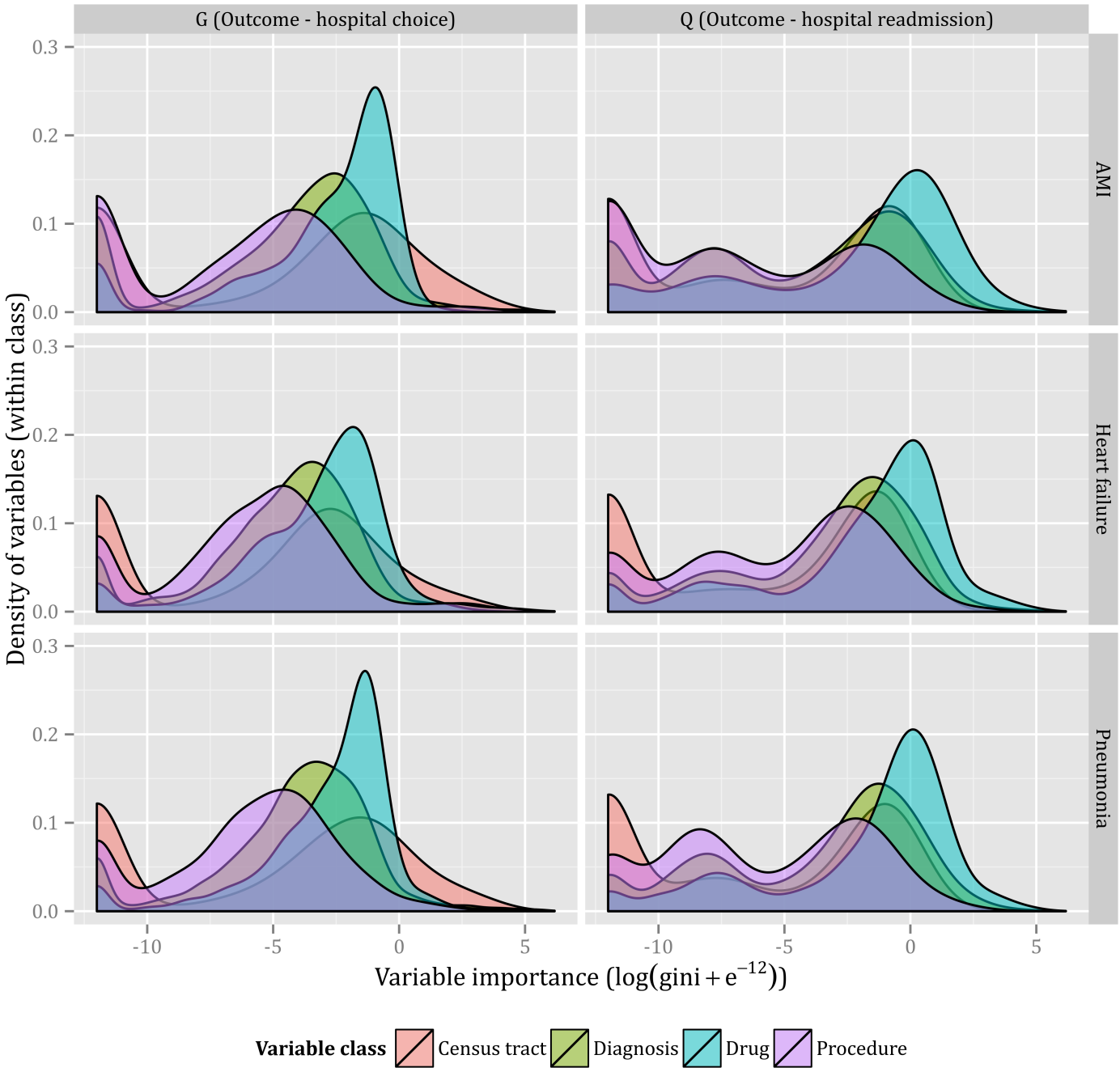

Figure 2: Variable importance by model and variable class. A descriptive sentence.

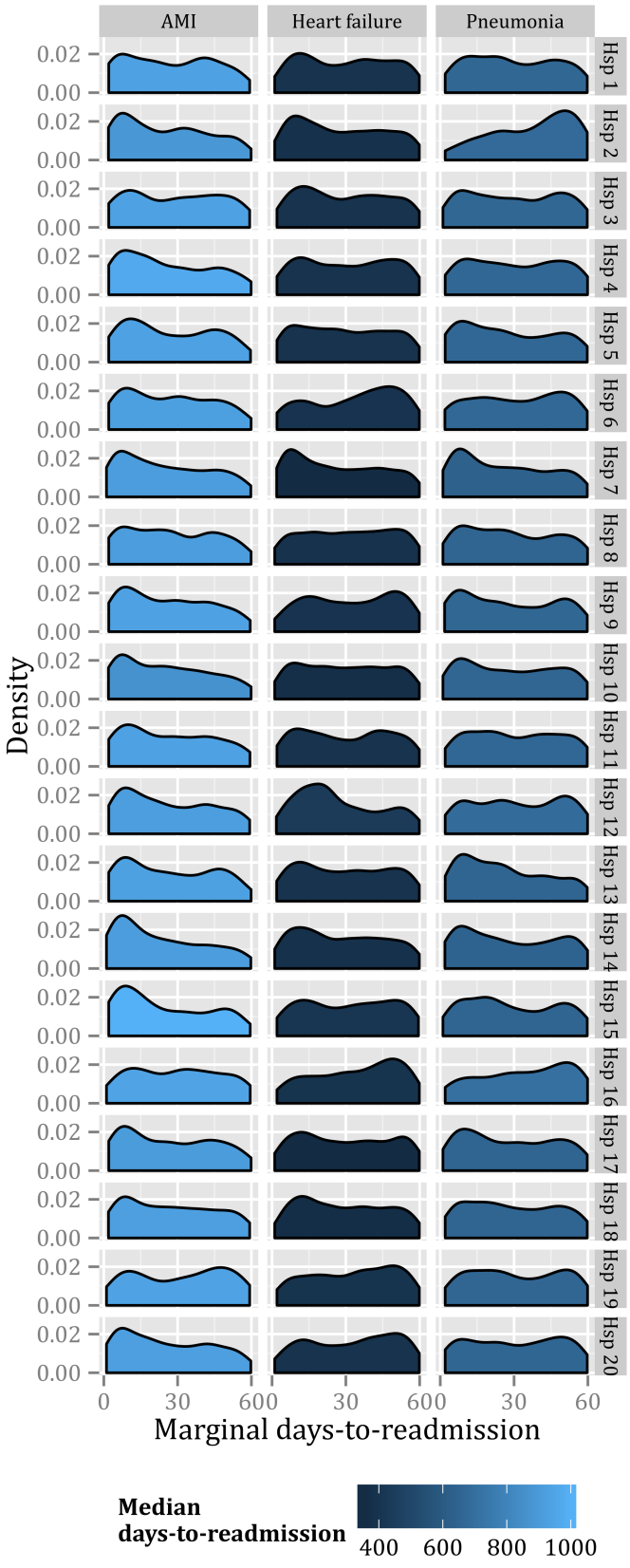

Figure 3: Marginal days-to-readmission.

Hospital							30-day readmission		Time-to-readmission	
	Admitted	Died	(%)	Discharged	Readmitted	(%)	Odds Ratio (95% CI)	Marginal risk	Hazard Ratio (95% CI)	Marginal mean TTE (median)
1	763	112	0.15	651	105	0.16	0.98 (0.95-1.01)	0.16	0.92 (0.81-1.04)	916 (933)
2	1557	148	0.10	1409	191	0.14	0.97 (0.95-1.00)	0.16	0.89 (0.81-0.99)	877 (888)
3	606	83	0.14	523	84	0.16	0.98 (0.95-1.02)	0.16	1.02 (0.90-1.16)	923 (939)
4	1022	125	0.12	897	136	0.15	0.97 (0.94-1.00)	0.16	0.76 (0.68-0.85)	950 (973)
5	729	150	0.21	579	98	0.17	0.98 (0.95-1.02)	0.16	0.94 (0.83-1.06)	920 (936)
6	826	119	0.14	707	106	0.15	0.98 (0.94-1.01)	0.16	0.81 (0.72-0.92)	913 (928)
7	1491	241	0.16	1250	216	0.17	0.99 (0.96-1.01)	0.16	0.91 (0.82-1.01)	893 (916)
8	1270	198	0.16	1072	138	0.13	0.95 (0.92-0.98)	0.16	0.88 (0.79-0.98)	900 (918)
9	780	152	0.19	628	130	0.21	1.01 (0.97-1.05)	0.16	0.92 (0.81-1.03)	920 (939)
10	778	124	0.16	654	123	0.19	1.01 (0.97-1.05)	0.16	1.09 (0.97-1.23)	855 (865)
11	705	125	0.18	580	97	0.17	0.99 (0.96-1.03)	0.16	0.97 (0.85-1.10)	919 (934)
12	1284	266	0.21	1018	166	0.16	0.99 (0.96-1.02)	0.16	1.04 (0.94-1.16)	902 (920)
13	739	86	0.12	653	110	0.17	0.99 (0.95-1.02)	0.16	0.93 (0.82-1.05)	916 (932)
14	1307	184	0.14	1123	210	0.19	(Reference)	0.16	(Reference)	899 (918)
15	1152	168	0.15	984	129	0.13	0.97 (0.95-1.01)	0.16	0.78 (0.70-0.87)	979 (1004)
16	408	70	0.17	338	43	0.13	0.97 (0.93-1.01)	0.16	0.78 (0.66-0.92)	936 (950)
17	807	123	0.15	684	134	0.20	1.02 (0.99-1.06)	0.16	1.07 (0.95-1.20)	898 (912)
18	894	144	0.16	750	116	0.15	0.98 (0.95-1.01)	0.16	0.92 (0.82-1.04)	911 (927)
19	499	94	0.19	405	50	0.12	0.95 (0.91-0.99)	0.16	0.84 (0.73-0.98)	927 (941)
20	1025	184	0.18	841	143	0.17	0.99 (0.96-1.02)	0.16	0.88 (0.78-0.98)	908 (923)

Table 1: Acute myocardial infarction (AMI)

Hospital							30-day readmission		Time-to-readmission	
	Admitted	Died	(%)	Discharged	Readmitted	(%)	Odds Ratio (95% CI)	Marginal risk	Hazard Ratio (95% CI)	Marginal mean TTE (median)
1	1229	141	0.11	1088	248	0.23	1.00 (0.97-1.03)	0.22	0.98 (0.90-1.07)	416 (387)
2	2071	166	0.08	1905	441	0.23	1.02 (0.99-1.05)	0.22	0.96 (0.89-1.04)	404 (375)
3	1243	134	0.11	1109	285	0.26	1.03 (1.00-1.07)	0.22	1.08 (0.99-1.17)	415 (386)
4	1076	122	0.11	954	214	0.22	1.01 (0.97-1.04)	0.22	0.87 (0.79-0.95)	416 (386)
5	1550	181	0.12	1369	288	0.21	0.99 (0.96-1.02)	0.22	0.94 (0.87-1.02)	412 (383)
6	827	107	0.13	720	128	0.18	0.97 (0.94-1.00)	0.22	0.87 (0.79-0.97)	419 (388)
7	2917	386	0.13	2531	666	0.26	1.04 (1.02-1.07)	0.22	1.05 (0.98-1.13)	372 (344)
8	1456	197	0.14	1259	232	0.18	0.97 (0.94-1.00)	0.22	0.91 (0.84-0.99)	412 (382)
9	881	111	0.13	770	157	0.20	0.98 (0.95-1.02)	0.22	0.88 (0.80-0.97)	418 (388)
10	1410	149	0.11	1261	311	0.25	1.01 (0.99-1.05)	0.22	1.05 (0.97-1.14)	391 (361)
11	1297	153	0.12	1144	258	0.23	1.01 (0.98-1.04)	0.22	1.00 (0.92-1.08)	414 (384)
12	1323	162	0.12	1161	192	0.17	0.92 (0.89-0.95)	0.22	0.75 (0.68-0.82)	459 (432)
13	1231	102	0.08	1129	262	0.23	1.00 (0.97-1.03)	0.22	0.90 (0.82-0.98)	415 (386)
14	2110	234	0.11	1876	424	0.23	(Reference)	0.22	(Reference)	405 (376)
15	1389	190	0.14	1199	203	0.17	0.97 (0.94-1.00)	0.22	0.89 (0.81-0.97)	431 (401)
16	681	94	0.14	587	111	0.19	0.98 (0.94-1.01)	0.22	0.95 (0.85-1.05)	421 (392)
17	1438	139	0.10	1299	328	0.25	1.04 (1.01-1.07)	0.22	1.09 (1.01-1.18)	367 (338)
18	1984	212	0.11	1772	438	0.25	1.03 (1.00-1.06)	0.22	1.12 (1.04-1.20)	386 (357)
19	932	99	0.11	833	163	0.20	0.98 (0.95-1.01)	0.22	0.91 (0.83-1.00)	418 (389)
20	1048	167	0.16	881	171	0.19	0.99 (0.96-1.02)	0.22	0.93 (0.85-1.02)	416 (386)

Table 2: Heart failure

Hospital							30-day readmission		Time-to-readmission	
	Admitted	Died	(%)	Discharged	Readmitted	(%)	Odds Ratio (95% CI)	Marginal risk	Hazard Ratio (95% CI)	Marginal mean TTE (median)
1	1184	176	0.15	1008	159	0.16	1.00 (0.98-1.03)	0.16	1.01 (0.92-1.11)	661 (657)
2	199	11	0.06	188	31	0.16	1.02 (0.97-1.08)	0.16	0.89 (0.75-1.07)	679 (674)
3	1085	132	0.12	953	160	0.17	1.01 (0.98-1.04)	0.16	1.02 (0.93-1.12)	663 (657)
4	863	91	0.11	772	113	0.15	1.00 (0.97-1.03)	0.16	0.89 (0.80-0.99)	665 (659)
5	923	147	0.16	776	143	0.18	1.04 (1.01-1.07)	0.16	1.00 (0.90-1.11)	665 (659)
6	788	136	0.17	652	89	0.14	1.00 (0.96-1.03)	0.16	0.97 (0.87-1.08)	668 (662)
7	2194	228	0.10	1966	328	0.17	1.03 (1.00-1.05)	0.16	1.00 (0.92-1.08)	637 (629)
8	1485	243	0.16	1242	173	0.14	0.99 (0.97-1.02)	0.16	0.92 (0.84-1.00)	655 (648)
9	990	166	0.17	824	158	0.19	1.04 (1.01-1.08)	0.16	0.95 (0.86-1.05)	664 (657)
10	1214	139	0.11	1075	181	0.17	1.01 (0.99-1.04)	0.16	1.04 (0.95-1.14)	659 (654)
11	892	147	0.16	745	119	0.16	1.02 (0.98-1.05)	0.16	0.93 (0.84-1.03)	666 (661)
12	1102	185	0.17	917	91	0.10	0.96 (0.93-0.98)	0.16	0.84 (0.76-0.92)	685 (679)
13	1914	204	0.11	1710	325	0.19	1.03 (1.00-1.05)	0.16	0.97 (0.90-1.05)	649 (647)
14	1980	278	0.14	1702	263	0.15	(Reference)	0.16	(Reference)	645 (639)
15	1365	179	0.13	1186	163	0.14	1.00 (0.97-1.03)	0.16	0.95 (0.87-1.04)	653 (646)
16	541	77	0.14	464	46	0.10	0.96 (0.93-1.00)	0.16	0.80 (0.70-0.91)	707 (704)
17	1338	163	0.12	1175	193	0.16	1.02 (0.99-1.05)	0.16	1.06 (0.97-1.16)	655 (649)
18	1356	168	0.12	1188	200	0.17	1.02 (0.99-1.04)	0.16	1.03 (0.95-1.13)	656 (650)
19	1020	123	0.12	897	122	0.14	0.99 (0.96-1.02)	0.16	0.90 (0.82-1.00)	662 (655)
20	1152	171	0.15	981	126	0.13	0.98 (0.95-1.01)	0.16	0.82 (0.75-0.91)	658 (651)

Table 3: Pneumonia

5 Discussion

In this study, we estimated marginal risk of 30-day readmission adjusted for patient readmission risk in twenty Montreal hospitals, and found that the marginal risk differences were not as strong as the crudely estimated readmission risk suggested.

In this study, we estimated the marginal risk of 30-day readmission and the mean time-to-readmission with machine learning techniques (random forest) and doubly-robust parameter estimation (TMLE) in twenty hospitals in Montreal.

Several readmission risk models have been criticized on the basis that they failed to distinguish preventable readmissions from readmissions due to chance alone. Despite evidence that clincians cannot reliably distinguish preventable and non-preventable readmissions, pairs of diagnosis codes have been developed that are "potentially preventable". For example, following a hospitalization for surgery, a readmission for infection would be considered preventable, but a readmission for trauma would not.

We did not atempt to measure whether *individual* readmissions were preventable, we instead opted to place readmissions within a counterfactual framework, and only identify the difference in readmission risk between hospitals, after controlling for differences in case-mix.

The counterfactual model clarifies the study question: What is the difference in the proportion with an emergency readmission within 30 days if all patients had attended Hospital A vs Hospital B? The notion of a "preventable" readmission is implicit; if a patient would have been readmitted if treated at another hospital, then the readmission was preventable. Since we cannot directly observe the desired proportions, we estimate it by attempting to recreate exchangeability (control for confounding) among the populations that visited different hospitals.

Following other hospital readmission work, we did not include hospital length of stay as a risk factor for readmission. Although we have strong reason to believe that the length of stay may affect readmission risk (and varies between hospitals), it acts as a mediator between hospital care and readmission; if we included it in our models, we risk biasing our estimate of the hospital's effect on readmission. However, unlike many other hospital readmission studies, 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. For example, hospital-acquired infections are one source of preventable readmissions; if we controlled for diagnosis codes indicating hospital-acquired infections, we would attenuate the effect that the hospital care was having on readmissions.

The major competing risk for hospital readmission is death, but others include moving outside the study area, or admission to a hospital for a non-emergency reason. In our 30-day readmission risk analysis, these effects were essentially ignored, but in the TMLE survival analysis we modeled the probability of right censorship.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 censorhsip) in which hospitals with better care were discharging sicker (but still living) patients, who would be more likely to be readmitted. This effect could have biased both of our results.

There is no special significance of 30 days in readmission, except for the fact that it is (recently) widely used as a cutoff. Interestingly, the distribution of expected survival times in the first sixty days appears to be bimodal (although some hospitals did not display this property), but the "dip" between the modes occurs a little earlier, roughly around day 20. These two modes suggest sharp changes in the conditional failure probability, which may reflect differing mechanisms of hospital's effect on readmission.

Differing *admission* practices can strongly affect the rates of readmission. Since most patients (x% in this cohort) return to the same hospital after readmission, if a hospital is more likely to admit patients than others, it will have a higher readmission rate. In some cases, like a major trauma, admission is certain, and in other cases, like a mild infection, admission is very improbable, but in most cases, there is some uncertainty in how patients are admitted. 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.

Strictly speaking, however, we cannot generalize our results to the non-respiratory illness population.

Non-parametric techniques such as random forest don't allow us to look (easily) at the individual effects of the confounders. This is somewhat disconcerting; we would like to be able to identify if the confounders that we expect to have an effect on the outcome are having the expected effect. We summarized the densities of the effects a few classes of variables, but this still does not allow the fine-grained analysis typical in an epidemiologic study. Even in a semiparameteric model, such as the Cox model we used for the survival analysis where we could identify hazard ratios directly from the model, it is difficult to analyze hundreds of fitted parameters in a meaningful way.

Hospital readmissions are confounded by a vast spectrum of health-related states; in the absence of a clear theoretical basis of the structure of that confounding, we have a few options. We can 1) identify relatively few, well-understood and measureable confounders to include in our model, or 2) forgo any theoretical understanding of the structure of confounding, and attempt to identify the broadest measureable set of even faintly plausible confounders.

In the absence of a clear theoretical basis of the structure of the confounding, we have to choose among two broad appraoches. We can 1) ignore the complexity and include confounders that we understand well and can measure well or 2) forgo any theoretical understanding of the structure of confounding, and attempt to identify the broadest measureable set of confounders.

Indeed, the first option has many advantages. In a situation where data collection is expensive, you want to choose the most efficient set of confounders. Additionally, by reducing the confounders to a well-understood few, your model gains credibility because you can show that the confounders are having the expected effect.

In the second option we lose the ability assess the "face validity" of our model, and we also risk inducing bias (such as M-bias).

However, the recent availability of large scale healthcare administrative data has put us into the situation where the second choice is at least possible; our data are no longer too expensive to collect. Additionally, because we have good reason to believe that readmission is confounded by a vast spectrum of health states, and because the structure of that confounding is poorly understood, we might as well use what we have.

We argue that in this situation, where we have a large dataset, thousands of measureable confounders, and little understanding of the structure of confounding, the second option is more appropriate.

Our work suggests that the difference between the approaches is not only a practical one; in assessing the effect of hospitals on readmission, it has a substantive effect as well.

We argue that if a large number of covariates are to be used, we lose the ability to identify individual effects, so we might as well take advantage of non-parametric machine learning techniques.

Calibration of the random forest vote proportions in the Q model strongly affected estimates of our parameter of interest in the Q* update step. Other articles using non-parametric techniques typically combined them with other models in an ensemble learner (in particular SuperLearner has been often used with TMLE). The final step in the ensemble learner is to combine all the probability estimates in parametric model, which would effectively calibrate the probabilities. When a single, non-parametric technique is used, an additional, seperate calibration step is necessary to convert the ranking scores into a probability estimate.

References

[1] Leo Breiman. ``Random Forests''. English. In: *Machine Learning* 45.1 (2001), pp. 5–32. ISSN: 0885-6125. DOI: 10.1023/A:1010933404324. URL: http://dx.doi.org/10.1023/A:1010933404324.

[2] Corrado Gini. *Variabilità e Mutabilità: Contributo allo studio delle distribuzioni e delle relazioni statistiche.* ita. Bologna: C. Cuppini, 1912.

[3] C. G. Broyden. ``The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations''. en. In: *IMA Journal of Applied Mathematics* 6.1 (1970), pp. 76–90. ISSN: 0272-4960, 1464-3634. DOI: 10.1093/imamat/6.1.76. URL: http://imamat.oxfordjournals.org/cgi/doi/10.1093/imamat/6.1.76 (visited on 06/02/2014).

[4] R. Fletcher. ``A new approach to variable metric algorithms''. en. In: *The Computer Journal* 13.3 (Mar. 1970), pp. 317–322. ISSN: 0010-4620, 1460-2067. DOI: 10.1093/comjnl/13.3.317. URL: http://comjnl.oupjournals.org/cgi/doi/10.1093/comjnl/13.3.317 (visited on 06/02/2014).

[5] Donald Goldfarb. ``A family of variable-metric methods derived by variational means''. en. In: *Mathematics of Computation* 24.109 (Jan. 1970), pp. 23–23. ISSN: 0025-5718. DOI: 10.1090/S0025-5718-1970-0258249-6. URL: http://www.ams.org/jourcgi/jour-getitem?pii=S0025-5718-1970-0258249-6 (visited on 06/02/2014).

[6] D. F. Shanno. ``Conditioning of quasi-Newton methods for function minimization''. en. In: *Mathematics of Computation* 24.111 (Sept. 1970), pp. 647–647. ISSN: 0025-5718. DOI: 10.1090/S0025-5718-1970-0274029-X. URL: http://www.ams.org/jourcgi/jour-getitem?pii=S0025-5718-1970-0274029-X (visited on 06/02/2014).

[7] R. Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2014. URL: http://www.R-project.org/.

[8] Aloysius Lim, Leo Breiman, and Adele Cutler. *bigrf: Big Random Forests: Classification and Regression Forests for Large Data Sets*. R package version 0.1-9. 2014. URL: https://github.com/aloysius-lim/bigrf.

[9] Jerome H. Friedman, Trevor Hastie, and Rob Tibshirani. ``Regularization Paths for Generalized Linear Models via Coordinate Descent''. In: *Journal of Statistical Software* 33.1 (2010), pp. 1–22. ISSN: 1548-7660. URL: http://www.jstatsoft.org/v33/i01.

[10] Hadley Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009. ISBN: 978-0-387-98140-6. URL: http://had.co.nz/ggplot2/book.