1   Introduction

It is particularly important to accurately estimate the independent effect of hospital treatment on readmission because it is being used to financially penalize hospitals.

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.

2   Methods

2.1   Data

2.1.1   Cohort selection

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 year in the cohort at 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.

2.1.2   Disease types

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.

2.1.3   Hospital readmissions

The unit of analysis in all models was the hospital discharge; a single patient 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.

2.2   Confounders

For each hospital discharge, we colllected variables that measured 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 study period), 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 hospital diagnoses, hospital procedures, and drugs dispensed outside of the hospital recorded in the year preceding the admissionin in Quebec. The procedures performed 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 records chemical compound being taken. 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.

2.3 Descriptive analysis

2.4 Models

2.4.1 Model G - probability of exposure

We developed a multinomial model g that predicted the probability of exposure (admission to a particular one of twenty hospitals).

2.4.2 Model Q - probability of outcome

2.4.3 Variable importance

For each tree, we calculated the out-of-bag decrease in Gini impurity when comparing the split nodes of the tree to the root. The Gini impurity at any node is, for all out-of-bag discharges.

For each hospital, the proportion of patients that chose this hospital p_h, multiplied by the probability of guessing that this patient would choose this hospital based on the distribution of patients $1 - p_h$. Summed over all hospitals, this is the Gini impurity. Gini impurity is 1 when all patients chose the same hospital.

To assess the importance of the variables in predicting which hospital a patient will choose, for each tree, we calculate the decrease in Gini impurity between the root of the tree and the leaves. If the patients within the leaves of the tree made relatively homogeneous hospital choices, this variable predicts hospital choice well, and the Gini impurity will decrease.

To calculate the Gini variable importance criterion, we summed the Gini impurity decrease over all trees.

2.4.4 Model Q* - updated targeted maximum likelihood estimation

2.5 Software

We implemented the random forest using the "bigrf" [1] package . We implemented the GLM fitting using coordinate descent using the "glmnet" package [2]. We plotted our figures using the "ggplot2" package [3].
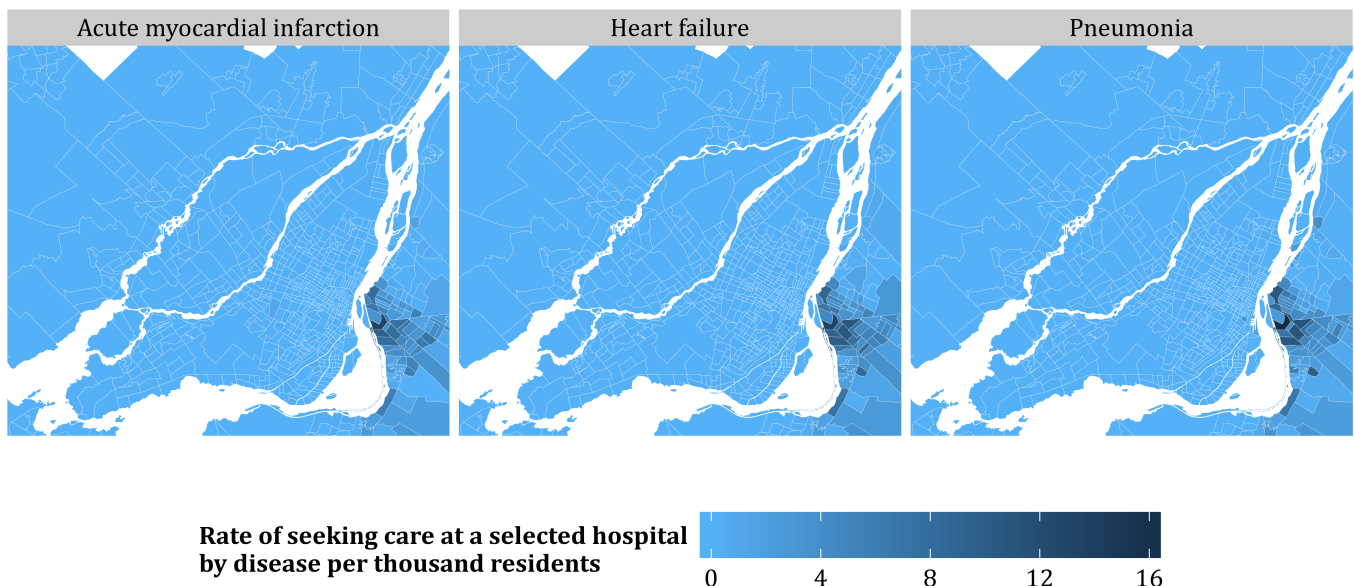
3 Results

Figure 1: Rates of seeking care at a selected hospital by admission diagnosis. A descriptive sentence
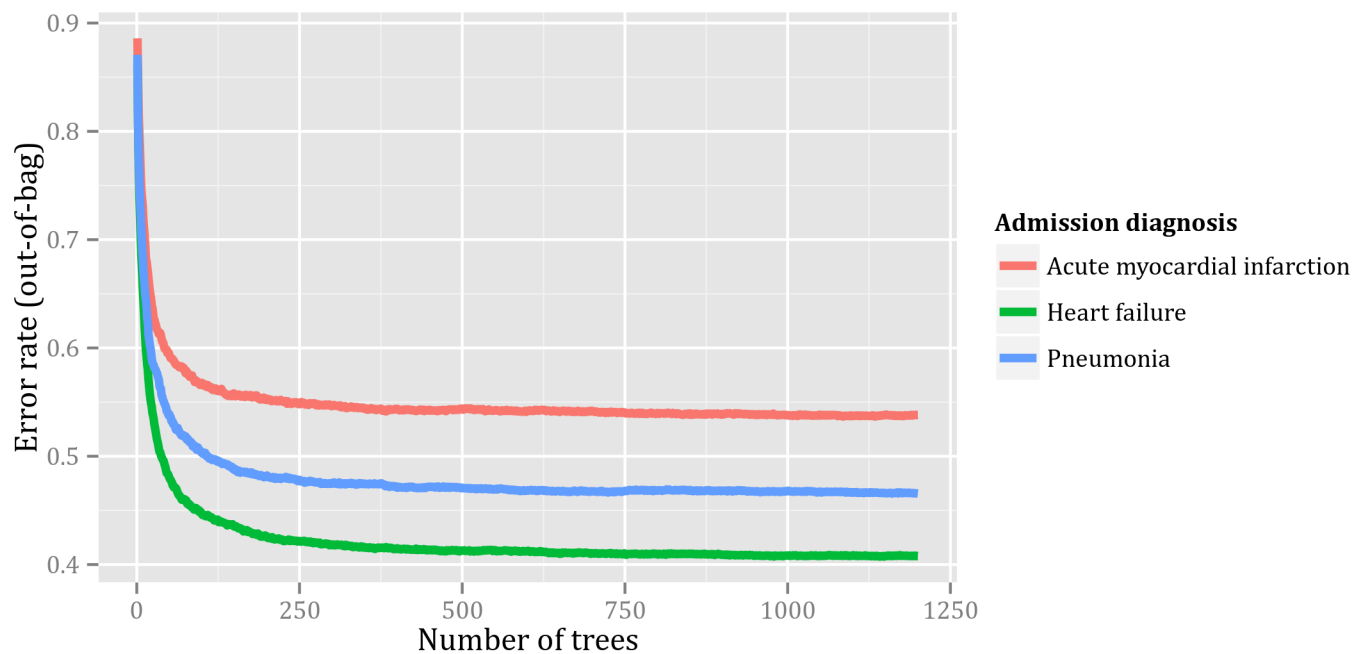
Figure 2: Error rate for random forest model of hospital choice. A descriptive sentence.
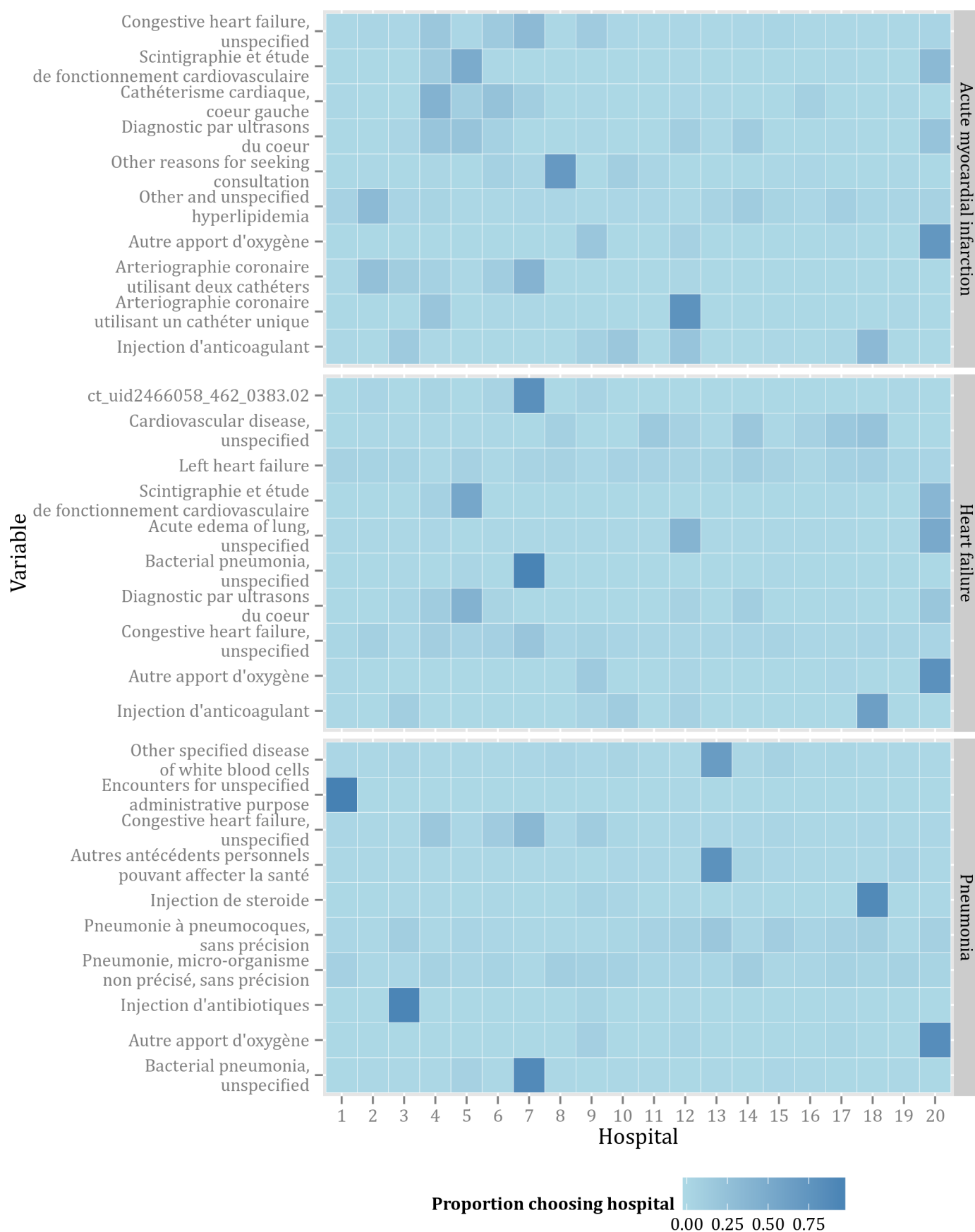
Figure 3: 10 most important variables for the G model, by disease. A descriptive sentence.
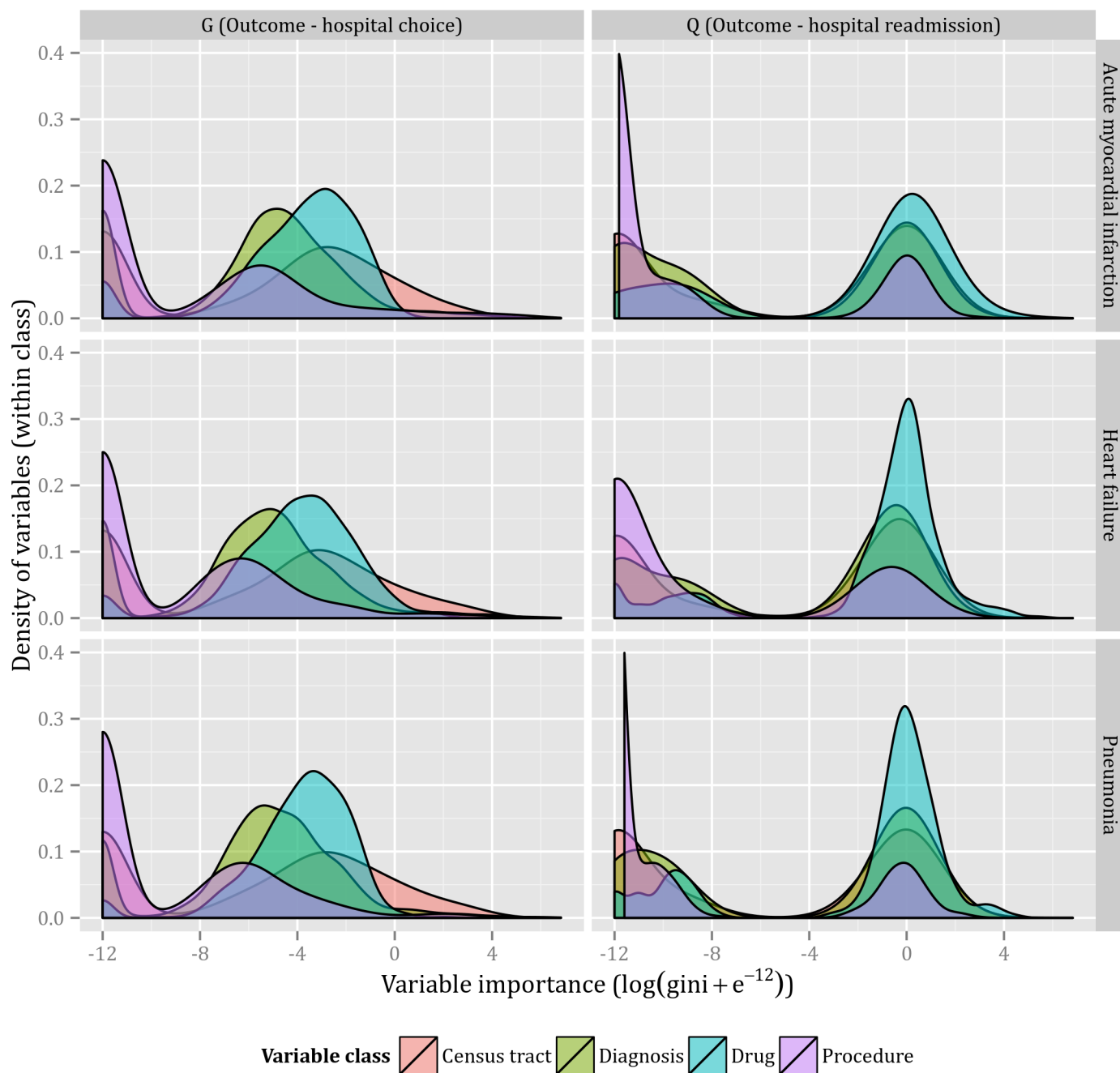
Figure 4: Variable importance by model and variable class. A descriptive sentence.

4 Discussion

References

[1] 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.

[2] 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.

[3] 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.