Dartmouth Health

Analytics Institute / Value Reporting & Analytics
DARTMOUTH HEALTH

Readmission Models for Congestive Heart Failure Patients

July 08, 2022

Table of Content

Issue/context

- Heart failure readmission – institutional priority
  - Percent of 30-day readmissions are on the rise
    - DH CHF patients' readmission performance fell below national average in 30 years
  - DH is incurring a readmission penalty
    - penalty of about $[redacted] MHMH & Cheshire

Research questions

Primary:

- Determine strong predictors of 30-day readmission for pts with index admission of CHF
- Build a predictive model to predict readmissions – add on!

Other:

- Examine readmission performance by specialty, admit facility, & discharge disposition

Data

- Source: eDH/Clarity
- Covers index admissions with discharge dates from January, 2016 – March, 2022 (and readmissions through April, 2022)
- Index Admission
  - Any inpatient admission at MHMH, Cheshire, APD or New London
  - a final primary billing Dx of CHF
  - Excludes IP admissions with discharge dispositions of deceased/died or left against medical advice
- Readmission
  - All cause 30-day readmission
- Data courtesy: Peter R. Lang (Healthcare Data Analyst - Sr, Performance Measurement Group)

Variables

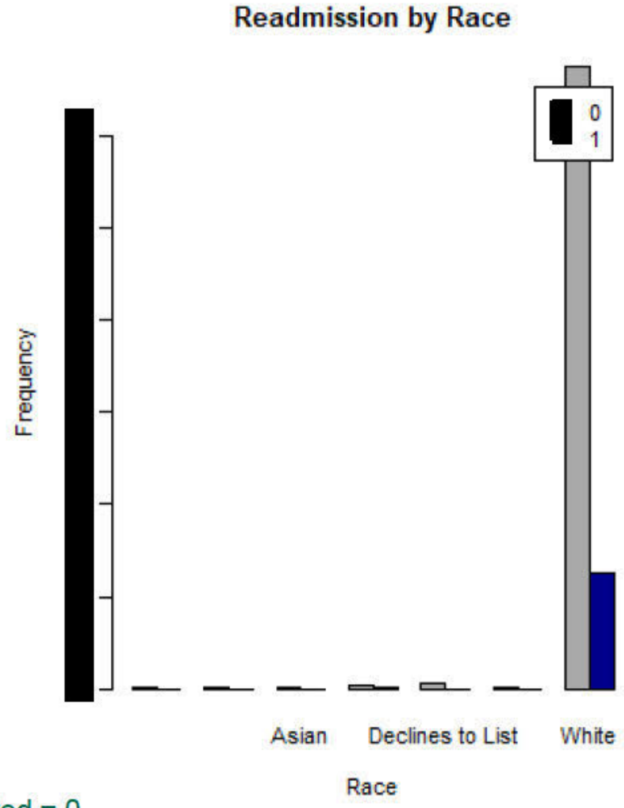
- Modelling focused on these paired down list of actionable variables derived from the initial analysis
- Variables grouped by categories
- A lot of binary and categorical variables
- Recoded age data

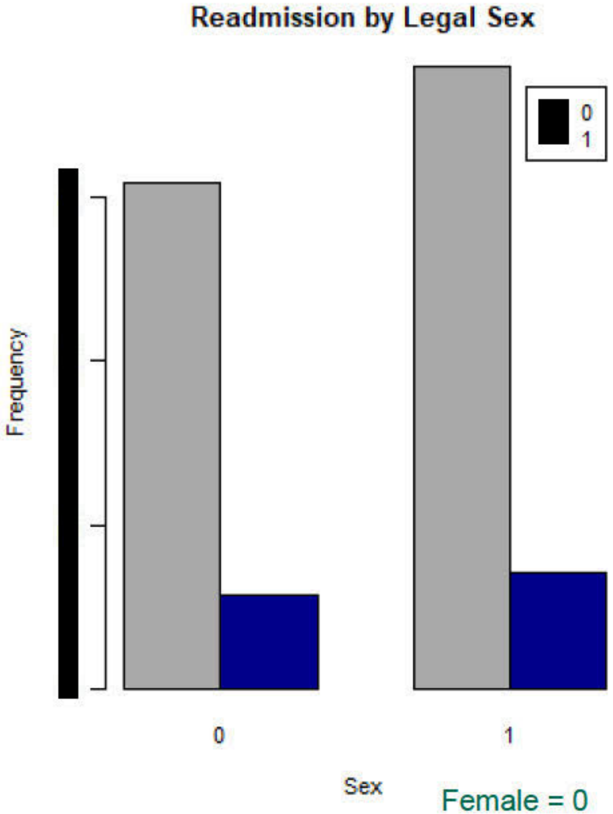
Categories	Variables	Values
Social Determinants of Health		
	Age	0-14, 15-44, 45-64, >64
	Sex	Male/Female = 1/0
	Race	American Indian/Alaska Native, Asian, Black or African American, Declines to List Unknown/Ur
Location of Care		
	Admit Facility	MHMH, Cheshire, ADP, New London
	Admit Specialty	Hospital Medicine, Cardiology, Other
	Discharge Disposition	AMA, Deceased/Died, Home, Home with VNA, Other, Other Facility, SNF/Rehab
Comorbidities		
	Obstructive Sleep Apnea	Yes/No = 1/0
	Atrial Fibrillation	Yes/No = 1/0
	Chronic Kidney Disease	Yes/No = 1/0
	Dementia	Yes/No = 1/0
	CVA	Yes/No = 1/0
Change in Biometrics		
	Systolic Blood pressure	Mean
	Creatinine	Mean
	Hematocrit	Mean
	proBNP	Mean
	Weight	Mean
Medications on Discharge		
	Ace Inhibitor	Yes/No = 1/0
	Based Beta Blocker	Yes/No = 1/0
	Aldosterone Antagonist	Yes/No = 1/0
	Diuretic	Yes/No = 1/0
	Angiotensin Receptor Blocker	Yes/No = 1/0
	SGLT-2 Inhibitor	Yes/No = 1/0
Process of Care		
	Completed Post Hospital Phone Call	Yes/No = 1/0
	Completed PCP Visit within 2 Weeks of discharge	Yes/No = 1/0
	Cardiology Visit within 4 Weeks of discharge	Yes/No = 1/0
	MyDH Active	Yes/No = 1/0
	Nutrition Consult	Yes/No = 1/0
	Cardiac Rehab Referral	Yes/No = 1/0

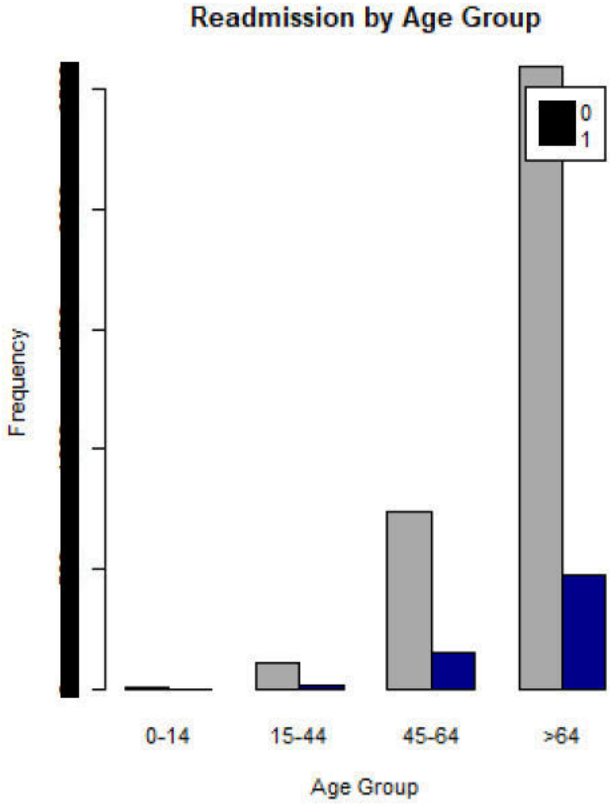
Exploratory Data Analyses

Social Determinants of Health

- Readmission varies by race, sex and age group
- Race highly skewed towards ‘White’ given the geography
- Lower occurrences of readmitted pts compared to non-readmitted pts





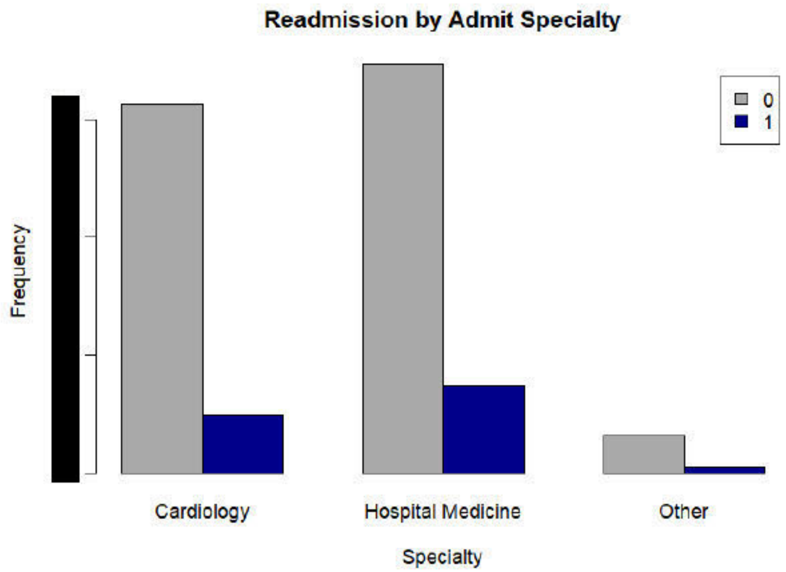


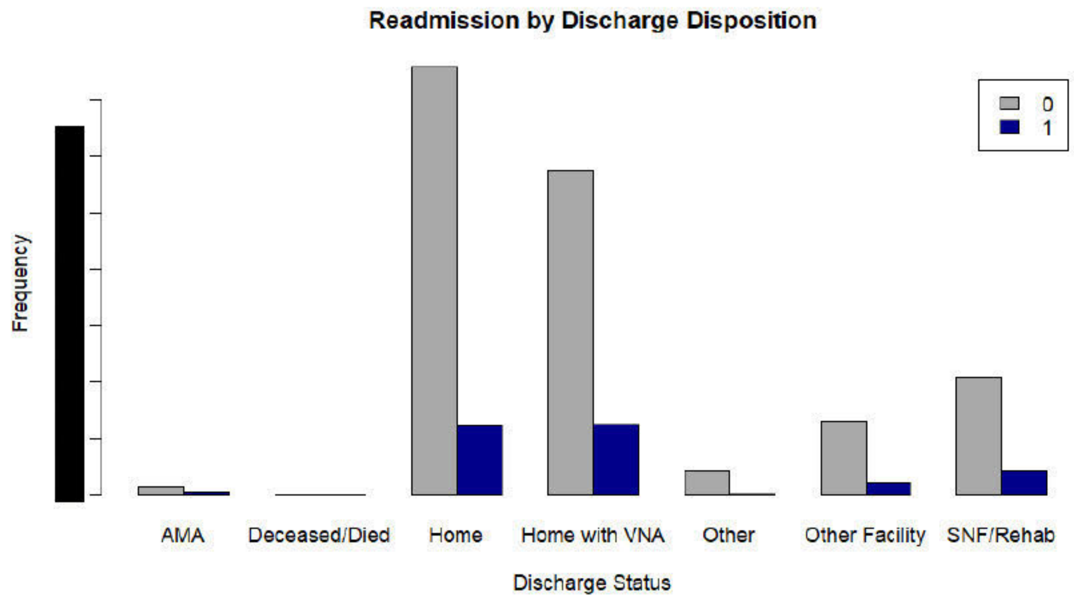
Not readmitted = 0
Readmitted = 1

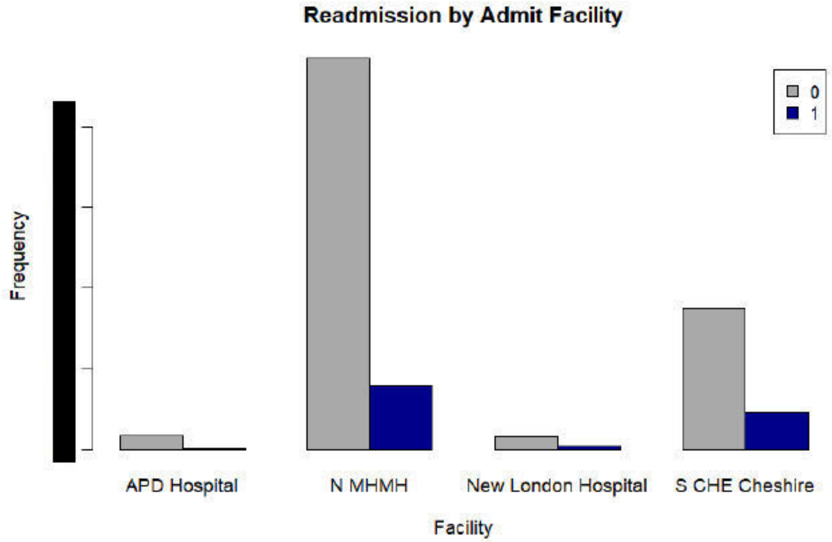
Female = 0
Male = 1

Location of Care

- Readmission varies by admit specialty, facility & discharge disposition
- Lower occurrences of readmitted pts compared to non-readmitted pts



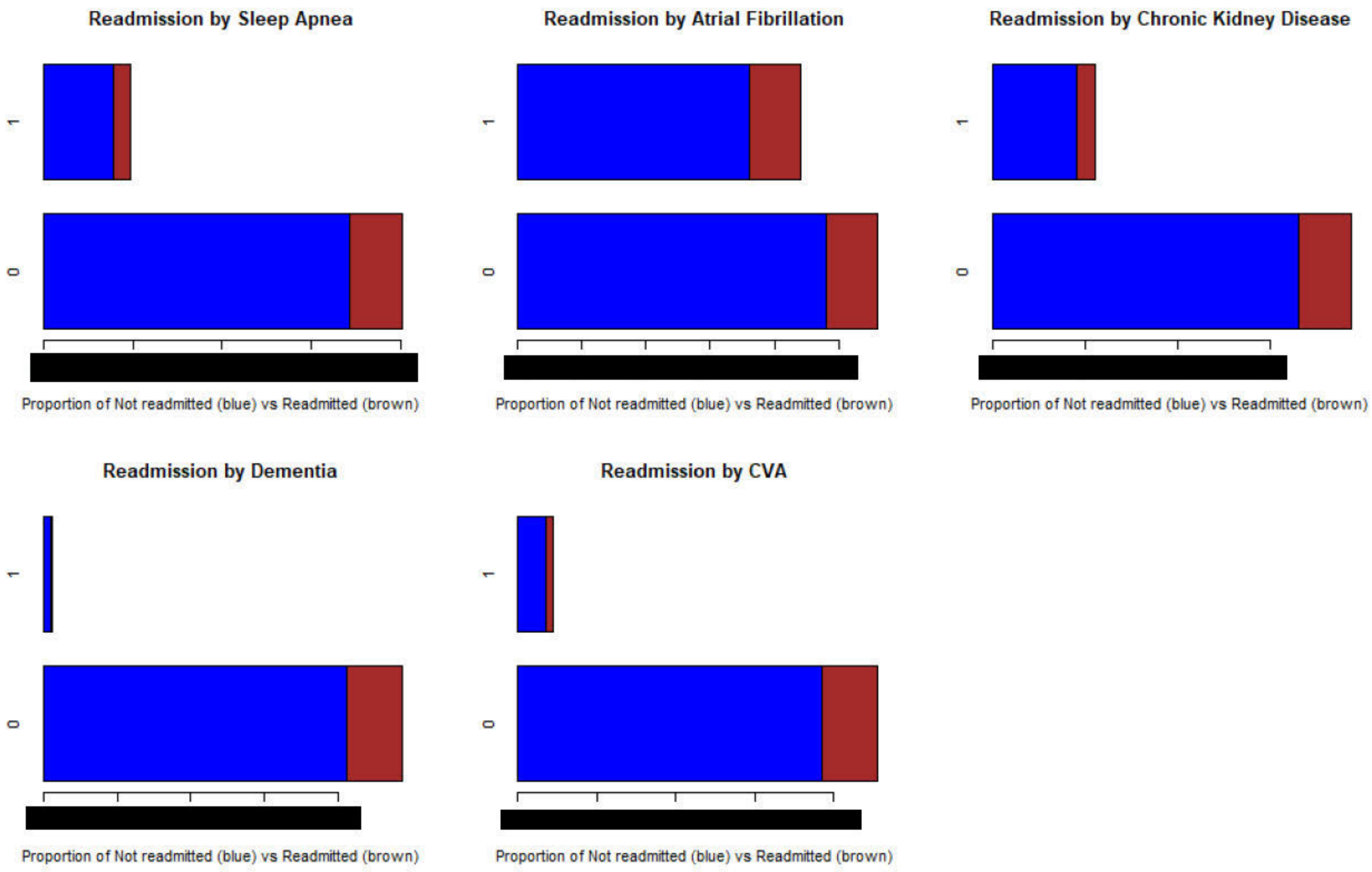




Not readmitted = 0
Readmitted = 1

Comorbidities

- From these plots, we see that there are differences in the proportions for each group and for each of the 5 predicting variables related to pre-existing conditions.

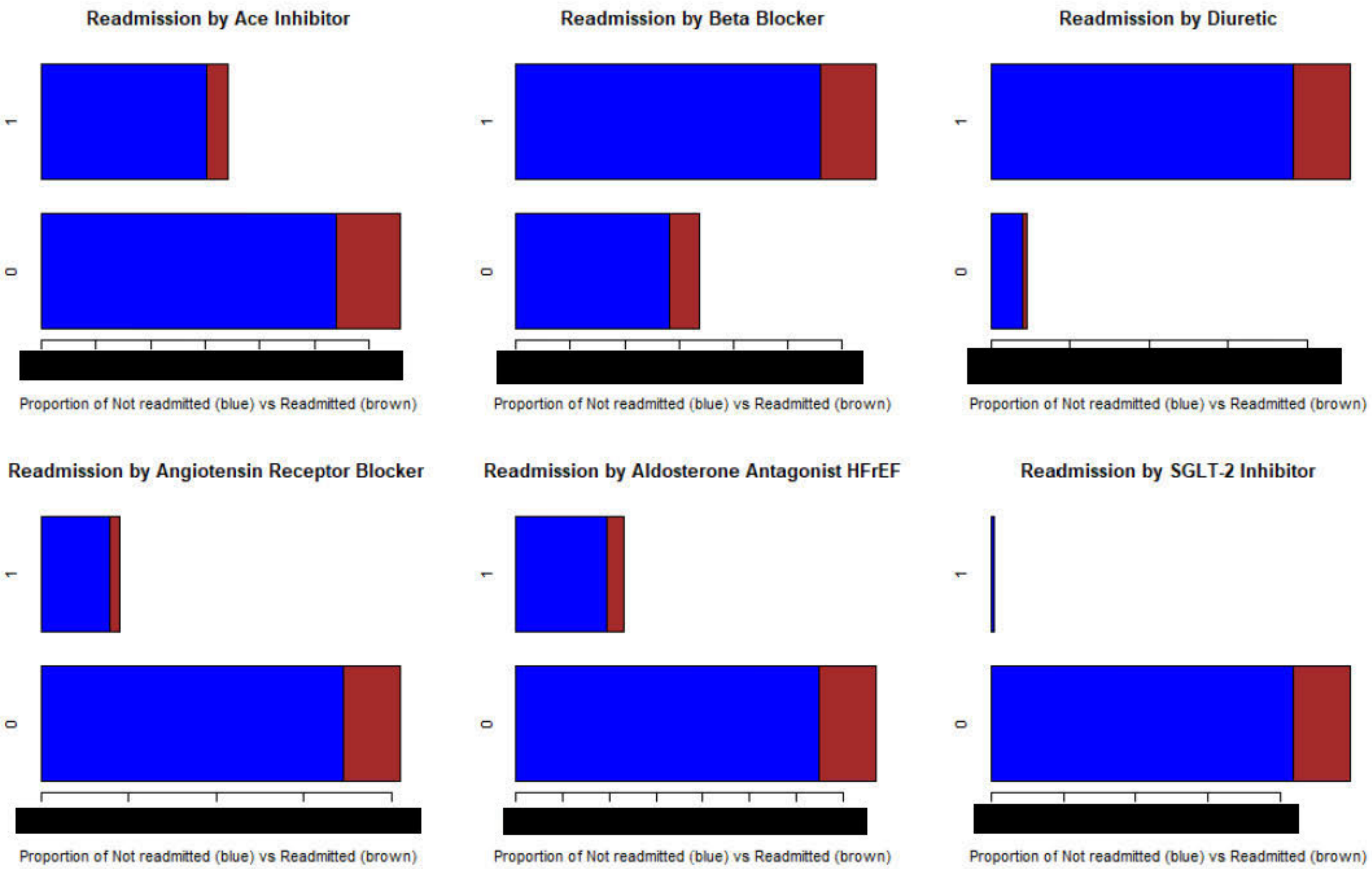


Condition present = 1
Condition not present = 0

Proportions add up to 100% or 1.0

Medications at Discharge

- From these plots, we see that there are differences in the proportions for each group and for each of the 6 predicting variables related to medications prescribed at discharge.

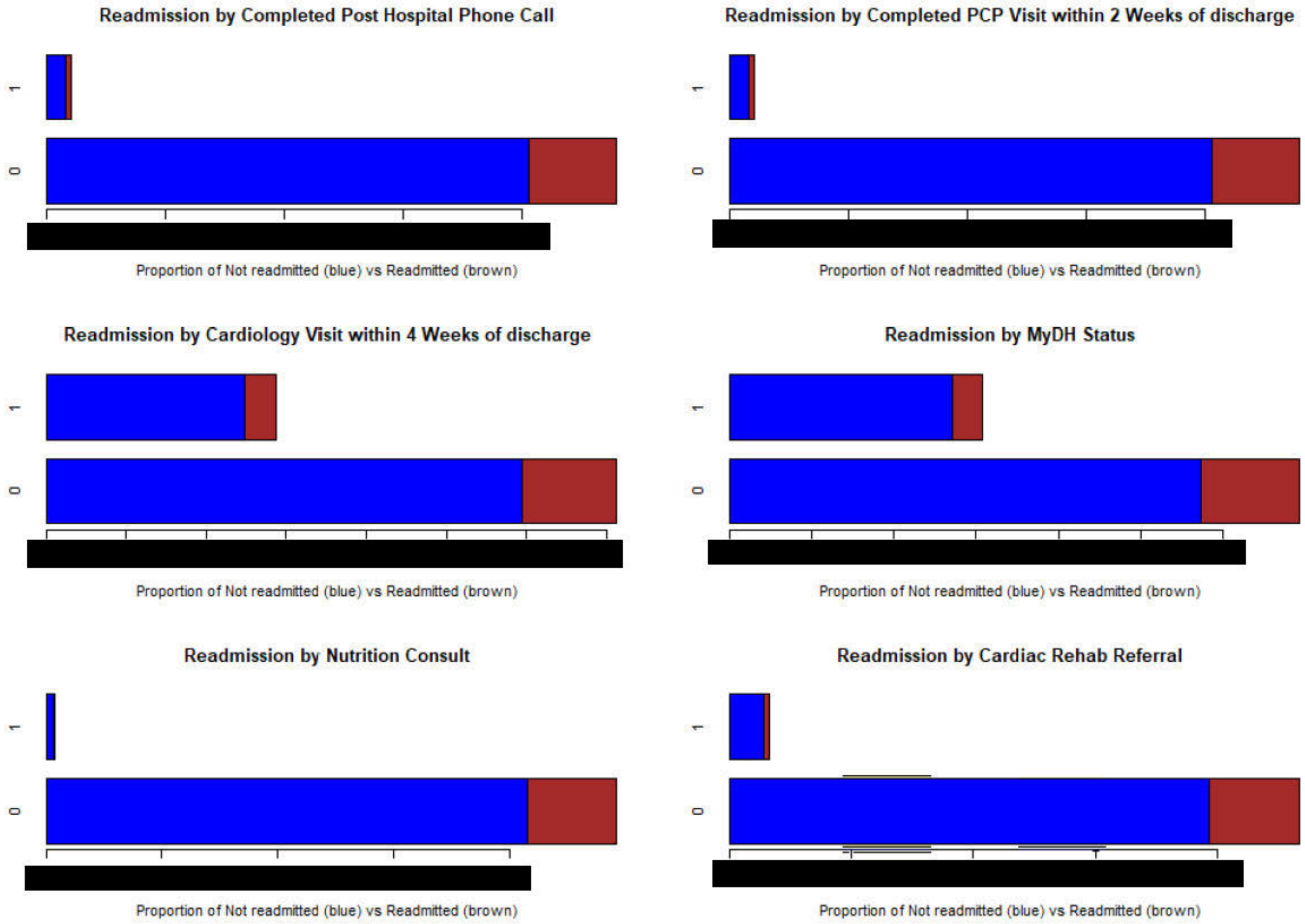


Medications at discharge = 1
Medications not at discharge = 0

Proportions add up to 100% or 1.0

Process of Care

- From these plots, we see that there are differences in the proportions for each group and for each of the 6 predicting variables related to different aspects of inpatient & post-discharge care processes.

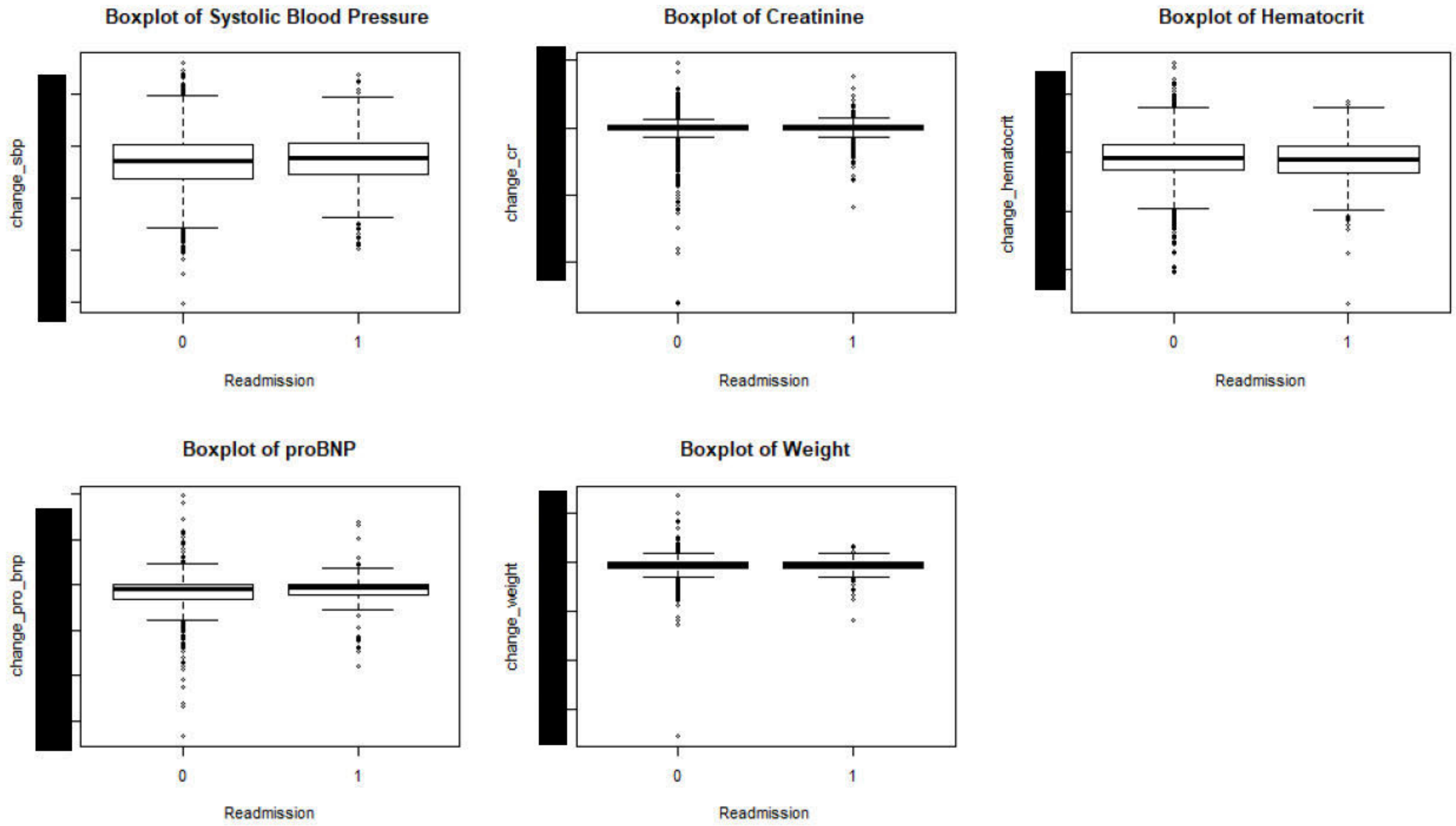


Care occurred = 1
Care did not occur = 0

Proportions add up to 100% or 1.0

Changes in Biometrics

- Boxplots on the continuous variables
- Mean and median between readmitted and not-readmitted are similar – variability does not seem to be an issue
- Within individual group, heavy tail exists (might or might not be outliers)



Not readmitted = 0
Readmitted = 1

Data Cleaning

Missing Values

- Biometrics data with most missing values
- Imputed continuous missing values using mean
- Deleted rows (<11) with missing categorical values

Variable	Percent Missing
change_pro_bnp	[redacted]
change_hematocrit	[redacted]
change_cr	[redacted]
change_weight	[redacted]
change_sbp	[redacted]
legal_sex	[redacted]
age	[redacted]
age_group	[redacted]
prim_enc_csn_id	[redacted]
obs_sleep_apnea_flag	[redacted]
a_fib_flag	[redacted]
ckd_flag	[redacted]
dementia_flag	[redacted]
cva_flag	[redacted]
ace_inhibitor_at_disch	[redacted]
beta_blocker_at_disch	[redacted]
diuretic_at_disch	[redacted]
arb_at_disch	[redacted]
aldo_at_disch	[redacted]
slgt_at_disch	[redacted]
post_hosp_call	[redacted]
pcp_within2wks	[redacted]
cardiology_within4wks	[redacted]
my_dh_active	[redacted]
nutrition_consult	[redacted]
card_rehab_consult	[redacted]
hospital_service_cat	[redacted]
disharge_disp_cat	[redacted]
loc_name	[redacted]
readmission_counter	[redacted]
race	[redacted]

Dataset Split

- Split dataset into:

  Train (80%) -> for fitting models

  Test (20%) -> for predictions

Modelling

Models

- Logistic Regression:
  - Used for non-continuous binary or categorical outcome data (e.g., readmitted or not, yes/no, 0/1)
  - Output based on probability (between 0 &1) and log odds (between –ve infinity to +ve infinity)

- Variable selection models: these models drop or add vars based on AIC & BIC scores (penalizes model complexity - measures variance and bias tradeoff)
  - Forward Stepwise Regression using BIC:
    - Starts with model with no variable to full model (logistic model in this case) and adds vars gradually based on Bayesian Information Criterion (BIC) scores
  - Backward Stepwise Regression using AIC
    - Starts with model with full model (logistic model in this case) and removes vars accordingly based on Akaike Information Criterion (AIC) scores

Models and Selected Variables

Categories	Variables	Logistic Regression	Forward Stepwise Regression using BIC	Backward Stepwise Regression using AIC
Social Determinants of Health	Age			
	Sex			
	Race			
Location of Care	**Admit Facility**	X (Cheshire)		X (Cheshire)
	Admit Specialty			
	Discharge Disposition	X (all categories except AMA)		X (all categories except AMA)
Comorbidities	**Obstructive Sleep Apnea**	X		X
	Atrial Fibrillation	X	X	X
	Chronic Kidney Disease			
	Dementia			
	CVA			X
Change in Biometrics	Systolic Blood pressure			
	Creatinine			
	Hematocrit	X		X
	proBNP			
	Weight			
Medications on Discharge	**Ace Inhibitor**	X	X	X
	Based Beta Blocker			
	Aldosterone Antagonist			
	Diuretic			
	Angiotensin Receptor Blocker	X		X
	SGLT-2 Inhibitor			
Process of Care	**Completed Post Hospital Phone Call**	X	X	X
	Completed PCP Visit within 2 Weeks of discharge			
	Cardiology Visit within 4 Weeks of discharge			
	MyDH Active	X	X	X
	Nutrition Consult			
	Cardiac Rehab Referral			

Models and Selected Variables

- All three models selected almost same variables as significant:

  Discharge Disposition (Home, Home with VNA, Other, Other Facility, SNF/Rehab)
  Admit Facility (Cheshire)
  Obstructive Sleep Apnea
  Atrial Fibrillation
  CVA
  Ace Inhibitor
  Angiotensin Receptor Blocker
  Completed Post Hospital Phone Call
  MyDH Active
  Hematocrit
  Systolic Blood Pressure

Evaluation of Models

Logistic model is highly significant (p values ~ 0)
Based on AIC scores, Backward Stepwise Regression is preferred
Based on BIC scores, Forward Stepwise Regression is preferred

Models	Purpose	Criterion	AIC	BIC	Significant	P-value
Logistic Regression	variable selection & prediction	deviance test/ p-value	[redacted]			
Forward Stepwise Regression using BIC	variable selection & prediction	BIC				
Backward Stepwise Regression using AIC	variable selection & prediction	AIC				

Focus on Logistic Regression Model

Logistic Model Results

- Full model results – variables with * are significant!

```
Call:
glm(formula = readmission_counter ~ ., family = binomial, data = dataTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.2096  -0.6211  -0.5114  -0.3893   2.8864

Coefficients:
                                        Estimate Std. Error z value Pr(>|z|)
[redacted]
(Intercept)
legal_sex1
obs_sleep_apnea_flag1
a_fib_flag1
ckd_flag1
dementia_flag1
cva_flag1
ace_inhibitor_at_disch1
beta_blocker_at_disch1
diuretic_at_disch1
arb_at_disch1
aldo_at_disch1
slgt_at_disch1
post_hosp_call1
pcp_within2wks1
cardiology_within4wks1
my_dh_active1
nutrition_consult1
card_rehab_consult1
hospital_service_catHospital Medicine
hospital_service_catOther
disharge_disp_catHome
disharge_disp_catHome with VNA
disharge_disp_catOther
disharge_disp_catOther Facility
disharge_disp_catSNF/Rehab
loc_nameN MHMH
loc_nameNew London Hospital
loc_nameS CHE Cheshire
raceAsian
raceBlack or African American
raceDeclines to List
raceUnknown/Unavailable
raceWhite
age_group15-44
age_group45-64
age_group>64
change_sbp
change_cr
change_pro_bnp
change_hematocrit
change_weight
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2825.5  on 3272  degrees of freedom
Residual deviance: 2689.3  on 3231  degrees of freedom
```

Logistic Model Results

- Odds & probability calculations for significant variables only – interpretations in the next slide!

Variables	odds	probability	Compared to base case
a_fib_flag1	[redacted]	[redacted]	higher
ace_inhibitor_at_disch1	[redacted]	[redacted]	lower
post_hosp_call1	[redacted]	[redacted]	higher
cardiology_within4wks1	[redacted]	[redacted]	lower
my_dh_active1	[redacted]	[redacted]	lower
disharge_disp_catHome	[redacted]	[redacted]	lower
disharge_disp_catHome with VNA	[redacted]	[redacted]	lower
disharge_disp_catOther	[redacted]	[redacted]	lower
disharge_disp_catOther Facility	[redacted]	[redacted]	lower
disharge_disp_catSNF/Rehab	[redacted]	[redacted]	lower
loc_nameS CHE Cheshire	[redacted]	[redacted]	higher
change_sbp	[redacted]	[redacted]	higher
change_hematocrit	[redacted]	[redacted]	lower
hospital_service_catHospitalMedicine (not-significant)	[redacted]	[redacted]	lower
hospital_service_catOther (not-significant)	[redacted]	[redacted]	lower

Logistic Model Interpretations

Interpretation of each variable coefficient is based on holding all other predictors constant:

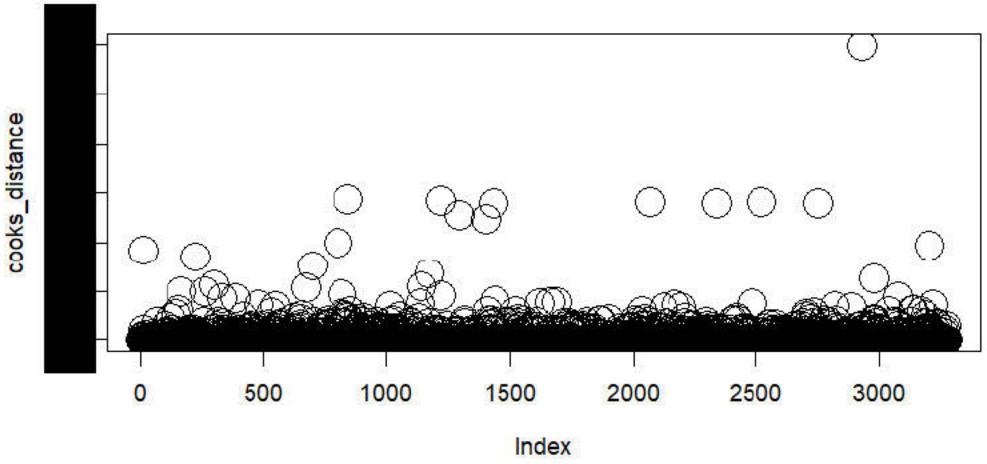
- The patient's probability of getting readmitted increases by:
  [redacted] if had a comorbidity of atrial fibrillation
  [redacted] if a patient had a post-hospital discharge call within 2 days of the index admission discharge
  [redacted] if the admitting location is Cheshire compared to MHMH

- The patient's probability of getting readmitted lowers by:
  [redacted] with ace inhibitors at discharge
  [redacted] if had cardiology visit within 4 weeks of discharge
  [redacted] if had myDH active
  [redacted], [redacted] if discharged to Home, Home with VNA, Other, Other Facility, & SNF/Rehab respectively compared to the base case of AMA
  [redacted] if the admitting specialty is Hospital Medicine compared to the base case of Cardiology (non-significant)
  [redacted] if the admitting specialty is Other compared to the base case of Cardiology (non-significant)

- With 1 unit increase in systolic blood pressure, the probability of getting readmitted increases by [redacted]
- With 1 unit increase in hematocrit, the probability of getting readmitted decreases by [redacted]

Outliers

- No outliers based on cook's distance using two thresholds of greater than 1 and 4/n
- The dispersion parameter is less than [redacted], over dispersion is not a concern for this model




```
> model_log$deviance/model_log$df.res
[1] [redacted]
```

Collinearity

- None of our variables has a Variance Inflation Factors (VIF) over the threshold of 10, so multicollinearity is not a problem among predictive variables!

```
> vif_log_model <- vif(model_log)
> vif_log_model
                          GVIF Df GVIF^(1/(2*Df))
legal_sex
obs_sleep_apnea_flag
a_fib_flag
ckd_flag
dementia_flag
cva_flag
ace_inhibitor_at_disch
beta_blocker_at_disch
diuretic_at_disch
arb_at_disch
aldo_at_disch
slgt_at_disch
post_hosp_call
pcp_within2wks
cardiology_within4wks
my_dh_active
nutrition_consult
card_rehab_consult
hospital_service_cat
disharge_disp_cat
loc_name
race
age_group
change_sbp
change_cr
change_pro_bnp
change_hematocrit
change_weight
>
```

[redacted]

Predictions

Evaluation of predictive models

- Prediction accuracy is pretty high for all three models - [redacted]
- Based on non-existent or very low sensitivity, models' ability to predict readmitted pts are not good (could be because of class-bias - readmitted [redacted] and non-readmitted [redacted]).
- Based on high specificity, models' ability to predict non-readmitted pts are pretty high.
- Given the context of the problem, sensitivity should be prioritized. Readmission is an issue, so not predicting a patient as likely to be readmitted when a patient actually might be readmitted could cause a lot of dangerous and expensive health complications.

Models	Accuracy	Sensitivity	Specificity
Logistic Regression	[redacted]	[redacted]	[redacted]
Forward Stepwise Regression using BIC	[redacted]	[redacted]	[redacted]
Backward Stepwise Regression using AIC	[redacted]	[redacted]	[redacted]

Based on using 0.5 as cutoff for predicting Y=1

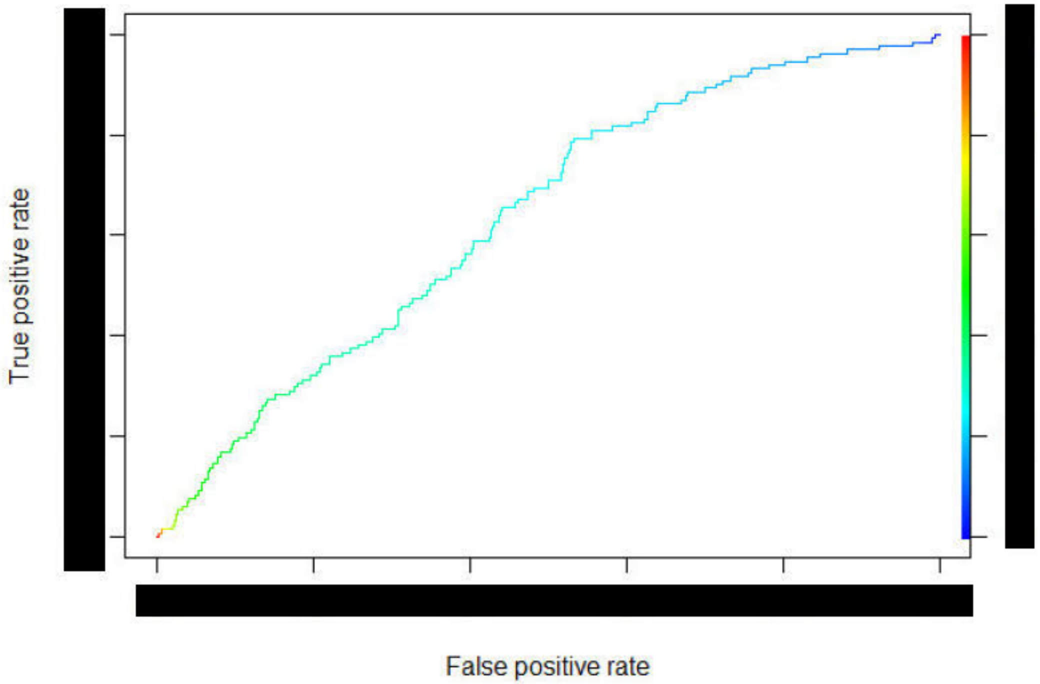
Logistic model prediction

	pred_outcome_log_0.5	
Readmission	No	Yes
No	[redacted]	
Yes		

Using 0.5 as threshold for predicting readmission, about [redacted] of predictions match with the actual observed.

ROC curve and Area Under the Curve value provides an aggregate measure of performance across all possible classification thresholds. About [redacted] of the predictions are correct.

```
> auc.perf@y.values
[[1]]
[1] [redacted]
```



Prediction model can be improved!

Challenges & Future Considerations

- Impute missing values differently
  - Use predictive models such as linear regression, predictive mean matching for continuous values and logistic model for categorical variables
- Apply machine learning approaches such as regularized regressions (e.g., lasso & elastic net) or feature selection methods (e.g., Principal Component Analysis (PCA)) on all available variables for robust analysis (as opposed to just paired down variables)
- Address class bias present in the dataset. 80% is not-readmitted and only 20% is readmitted.
- Improve prediction models
- Use a more reliable weight variable – standing weight
- Perform residual analysis for checking goodness of fit for this model - although it is possible for a model to have explanatory/predictive power while still being a poor fit!

  - Transform continuous variables to create normal distribution if needed

Dartmouth Health

Analytics Institute / Value Reporting & Analytics
DARTMOUTH HEALTH

Thank you.

For questions, contact Sukriti Raut [redacted]