American Journal of Public Health and Epidemiology
Volume.8, Number 4;July-August, 2022;
ISSN: 2644-0032 | Impact Factor: 6.16
https://zapjournals.com/Journals/index.php/ajphe
Published By: Zendo Academic Publishing

CARDIAC PATIENT READMISSIONS: AN IN-DEPTH INVESTIGATION OF INFLUENTIAL VARIABLES

Benjamin David Mitchell, Sarah Elizabeth Harrison, James William Anderson

Article Info

Keywords: American healthcare, access, cost-effectiveness, quality, Hospital Readmissions Reduction Program, healthcare legislation.

Abstract

American healthcare is at a critical juncture, facing the challenge of balancing accessibility, cost-effectiveness, and quality. The complex interplay of rising healthcare expenses, evolving legislative frameworks, shifting demographics, and varying definitions of quality necessitates innovative solutions to maximize the value proposition in the U.S. healthcare system. One significant strategy aimed at achieving this balance is the Hospital Readmissions Reduction Program (HRRP), an integral component of the 2010 Patient Protection and Affordable Care Act (PPACA). The HRRP penalizes hospitals with higher-than-expected readmission rates by deducting up to 3% of their total Medicare payments. Hospital readmissions, as identified by the Centers for Medicare and Medicaid Services, not only incur substantial financial costs but are also amendable. Over 2,000 hospitals are subject to these penalties, resulting in a collective loss of approximately $280 million in Medicare funds annually. Beyond the financial aspect, this issue has broader societal ramifications, including the economic externalities of lost wages and productivity, as well as the potential for enhanced patient care.

1.1 Introduction and Motivation

American health care consumers struggle to find balance between access to health care services, the cost of health care, and the quality of the health care. Due to rising costs, new legislation, changing population health care needs, and discrepancies in defining quality, maximizing the value proposition within American health care is complicated. One approach has been the Hospital Readmissions Reduction Program (HRRP), part of the 2010 Patient Protection and Affordable Care Act (PPACA),that penalizes hospitals with higher-than-expected readmission rates, up to 3% of their total Medicare payments (http://go.cms.gov/1L93Lh4). Hospital readmissions, according to Centers for Medicare and Medicaid, are common, costly, and most importantly correctable. More than 2,000 hospitals are penalized for readmission and therefore forfeit about $280 million in

Medicare funds annually. It is not just cost of unnecessary hospitalizations but a social impact (externalities) in lost wages and production as well as improvement in care.

This study aims to identify the factors that can predict characteristics of cardiac patient readmissions. Specifically, weaimed to achieve two goals: 1) Identify significant risk factors that can predict cardiac patient readmissions and 2) Develop a Risk Score for Readmission and validate it.

Identification of risk factors that can predict, with varying levels of certainty, whether a cardiac patient will be readmitted following discharge from the hospital can advance the practice of medicine. More importantly, quality of care can be improved with providers mindful of characteristics in their patient population which can increase the risk of readmissions this type of risk scoring tool is customized to patient specific data, resulting in a more targeted approach. Consequently, generalization of this score is possible across various regions.

1.2 Literature Review

The U.S. healthcare system historically spends far more per capita on health care than the rest of the world.

Data from The World Bank shows the U.S. spends approximately $9,146 for health per capita. Only Norway and Switzerland spend more (The World Bank, 2015). The Patient Protection and Affordable Care Act (PPACA) is the first attempt to reduce cost while improving access, through insurance, and thus improving quality. Success in these areas remains to be proven. In an attempt to juggle these varied attributes, numerous population health management (PM) initiatives are being developed in an attempt to provide systemic solutions (Snowdon, 2014)

Hospital readmissions are found in 20% of Medicare beneficiaries costing $19 billion annually. According to Ban off, Milner, Rimar, Greer and Canavan (2016), Heart Failure (HF) is the most common of cardiac-related readmissions; alone it accounts for $1 billion. Readmission risk assessment can be used to help target the delivery of these resource-intensive interventions to the patients at greatest risk. Past studies have identified the benefit of interventions to reduce admissions. Unfortunately, an effective readmission remains elusive; transitional care interventions may reduce readmissions among chronically ill. (Falvey, Burke, Malone, Ridgeway, McManus, & StevensLapsley, 2016).

Readmission scores are not unique but often lack specificity to the disease at hand. Van Walraven and associates (2010) developed the LACE index which relies on four variables; LOS (“L”), acuity of the admission (e.g., emergency admission) (“A”), Charlson Comorbidity Index score (“C”), and the number of previous emergency department visits in the past 6 months (“E”). The LACE index has been validated using a mix of medical and surgical patients. Wang and colleagues (2014) test the LACE index with patients with HF and find that the index does not predict unplanned readmission within 30 days reliably. Similarly, in a study of general medical patients in the United Kingdom, the LACE index shows fair predictive value for 30-day readmission with a C-statistic of 0.55 (Cotter, Bhalla, Wallis, & Biram, 2012), and in medical patients in Canada, the index identifies50% the patients readmitted within 30 days of discharge but does not identify the other half (Gruneir et al., 2011). Choudhry and coauthors (2013) have been developing allcause hospital readmission risk-prediction models to identify adult patients at high risk for 30-day readmission upon admission and discharge. Unfortunately, the evidence to date points to a score that may generally identify readmission risk but fail in identifying specific to the cardiovascular population.

Within cardiovascular medicine, several attempts have been made to categorize readmission risk. A systematic review of statistical models to predict a HF patient’s risk of readmission by Ross and colleagues (2008) reveal substantial inconsistencies in patient characteristics that are predictive of readmission in this population. Most models rely on retrospective administrative data; however, a few relied on real-time administrative data. Some of the models incorporate primary data collection, an effort that often creates limits practical application. Banoff et.

al. (2016) advance an automated algorithm within the EMR which transitioned the model to a usable tool in the clinical setting called the HOSPITAL score. The HOSPITAL score includes seven variables: hemoglobin at discharge, discharge from oncology, sodium level at discharge, having a procedure, type of admission, number of admissions in past year, and LOS. The HOSPITAL score has fair discriminatory power for prediction of 30-day readmission in medical patients (Donze, Aujesky, Williams, & Schnipper, 2013). However, the HOSPITAL score does not include information from nursing assessments in their estimation of risk for 30-day readmission. Both also lack data on the patient's condition throughout the hospitalization.

Valid risk adjustment methods are required for calculation of risk-standardized readmission rates, which are used for hospital comparison, public reporting, and reimbursement determinations. Models that are designed for these purposes will have good predictive ability; be deployable in large populations; use reliable data that can be easily obtained; and use variables that are clinically related to and validated in the populations in which use is intended. This can be very easily automated and is very convenient for physicians charged with making discharge decisions. This risk score formula provides an opportunity to identify high risk patients at the time of discharge. It may help them enroll in preventive programs, if the doctor recommends such action and increase reimbursement potential from third party payers. This risk score can be used as a proactive measure to screen out high risk patients. This formula groups cardiac patients in to three groups: Low risk; medium risk and high risk based on ranges of the risk score. The risk score is based t-test results and on Chi-square values from Logistic regression results. The statistical results are based large sample and hence reliable.

2. Methodology

Statistical results have been obtained from statistical models (Logistic regression, T-tests, and Discriminant Analysis). Several interventions that involve multiple components (e.g., patient needs assessment, medication reconciliation, patient education, arranging timely outpatient appointments, and providing telephone follow-up) have successfully reduced readmission rates for patients discharged to home. To help Sanford Health System direct resources and services to patients with greater likelihood of readmission, a number of risk stratification methods are available. Outcomes can better define the role of home-based services, information technology, mental health care, caregiver support, community partnerships, and new transitional care personnel (Kripalani, Theobald, Anctil, and Vasilevskis, 2014). Logistic regression and T-test results identify (statistically) significant variables that can predict cardiac readmissions. De-identification of Protected Health Information is in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. The data set is de-identified and approved by the Sanford Privacy Board and in addition is also IRB approved by University of South Dakota. Sanford has multiple hospitals that admit for cardiovascular events. There are four regional hospitals and almost 40 critical access hospitals. The researchers identified records of 19,263 cardiac patients out of which 2,687 have been readmitted once or more. Each patient, to be part of the sample, has had at least one cardiac event. A random sample of 33,642 non-cardiac patient records has been selected for further analysis in the future. A univariate test (T-test for mean differences) and Cohen's D are used for feature selection and Logistic Regression and Discriminant Analysis models are developed. These two are multivariate models.

3. Data Description

The sample from a large integrated health system in the Midwest is used in this study which includes data from 19,263 cardiac patients. Of these 2,687 cardiac patients are readmitted (READ) and 16,576 (Not READ) are not readmitted. The 2,687 readmitted cardiac patients are coded as 1 and the other 16,576 are coded as 0.

Table 1: Variable Description

Variable	Description
Problem List	Diagnoses listed
CardEv	# of Cardiac Events
CSLOS	Length of stay for each cardiac event
HSDRx	# of prescriptions during time period
Gender	Male or Female
PatAge	Patient Age in years
BMI	Most recent Body Mass Index
A1C	A1C
L Diast	Most recent Diastolic blood pressure
L Syst	Most Systolic blood pressure
No shows	# of appointments that the patient missed
Race	Patient race
Marital status	Married or Single
Alcohol Use	Self-reported alcohol use
Patient Location	Patient Location (SF/Fargo or not)
Discharge Location	Home or SNF

The researchers have16 variables on these 19,263 patients. A t-test is performed for mean differences between these two groups – READ and Not READ. Variables with insignificant t-scores are dropped, since they do not contribute to the group separation. More specifically, the research team drop race (t = 0.395), marital status (t = 0441), systolic pressure (t = 0.64), alcohol use (t = 1.924) and patient location (t= 0.386). Researchers also drop discharge location, A1C, and smokeless tobacco use due to too many missing variables or low Cohen's D.

The remaining 8variables have significant t-scores and/or high Cohen's D and are used in the multivariate models (Logistic regression and Discriminant). They are: Cardiac Events, Problem List, Patient Age, BMI, Diastolic Pressure, No Shows, HSD Rx, and Gender. Many of these 8 explanatory variables have missing values and if any one of these explanatory variables are missing, that record (patient) is dropped from analysis.

The final sample to be used in various analyses consists of 6,064 patients. This is divided into two samples – 4,869 patients in the training sample and 1,195 patients in the validation sample. Out of 4,869 patients in the training sample, 1,669 are READ patients and 3,200 are Not READ patients. Out of 1,195 patients in the validation sample, 417 are READ cardiac patients and 778 are Not READ cardiac patients.

4.1 Descriptive Statistics

A summary of descriptive statistics is in Table 2. For readmitted cardiac patients (READ) and not readmitted cardiac patients (Not READ), this table reports the mean, the standard deviation, and T-statistics for several explanatory variables of interest. Mean values indicate that the READ patients have higher A1C, number of prescriptions, longer problem list, longer hospital stay, more cardiac events and are older than the control group. However, READ cardiac patients have lower mean values for body mass index and diastolic blood pressure. T-tests for mean difference indicate that cardiac events, number of no shows, number of prescriptions, length of stay and problem list are significantly different between the two groups at the 1 percent level. T-test results also indicate that patient age, body mass index, and diastolic pressure are significantly different between the two groups. Effect size as measured by Cohen's D is significant (large) for the following variables: cardiac events, no shows, number of prescriptions, length of stay, and problem list.

Table 2: Descriptive Statistics

Variables	Group Code	N	Mean	Std. Deviation	T-statistic	Cohen's D
PatAge	1	2687	70.54	13.25	2.52[b]	0.048
	0	16575	69.82	16.41		
BMI	1	2613	29.76	6.86	-2.94 [a]	-0.059
	0	15103	30.20	7.94		
CardEv	1	2687	2.16	1.37	35.78[a]	0.911
	0	16575	1.20	0.59		
CSLOS	1	2686	5.45	5.75	7.74 [a]	0.182
	0	6775	4.48	4.86		
A1C	1	1818	6.68	1.59	2.42[b]	0.066
	0	8172	6.58	1.45		
LSystolic	1	2687	124.11	20.47	0.64	0.013
	0	16571	123.84	19.67		
LDiastolic	1	2687	68.99	13.10	-3.67 [a]	-0.053
	0	16571	69.98	12.92		
NoShows	1	2187	7.90	12.55	10.74 [a]	0.292
	0	10729	4.91	7.27		
HSDRx	1	2631	365.17	396.43	14.99 [a]	0.336
	0	15399	242.72	328.20		
ProbList	1	2687	6.04	3.77	22.49 [a]	0.494
	0	16575	4.31	3.19		

Group code: 1 = Readmitted (READ) CP
0 = Not READ CP [a] two-tailed significance at < 0.01 level [b] two-tailed significance at < 0.05 level
PatAge = Patient Age; CardEv = Cardiac Event; CSLOS = Cardiac surgery length of stay
HSDRx = # of prescriptions over 3 years; ProbList = Diagnoses listed
Cardiac Events include: ASA, Arrhythmia, CVD, Angina, AmbulatoryCardiacMonitoring ICD9 89.50, RhythmEKG ICD9,ElectrographicTelemetry ICD9 89.54, Dyspnea, & Renal disease.

4.2: Correlation Analysis **Table 3: Pearson Correlation Coefficients**

	BMI	CE	LOS	LD	NS	HSD	GEND	PL
BMI	1.000							
CE	.002	1.000						
LOS	.007	.077	1.000					
LD	.017	-.020	-.012	1.000				
NS	.008	.071	.018	.041	1.000			
HSD	.008	.053	.010	-.131	.232	1.000		
GEND	.039	-.050	-.038	-.008	.009	-.001	1.000	
PL	.060	.401	.119	-.004	.022	.027	-.029	1.000

CE = Cardiac Event; LOS = length of stay; LD =Last diastolic; NS = No Shows; GEND = gender; HSD = # of prescriptions over 3 years; PL = Diagnoses listed

There are a number of strong correlations among the variables and the Pearson correlation coefficients are reported in Table 3. No Shows variable is positively correlated with Cardiac events, HSD_Rx, and to diastolic pressure. Cardiac events are strongly correlated with Problem list indicating patients with multiple health problems have more cardiac events. There is a positive relationship between gender and BMI indicating higher BMI for males. There is a positive association between length of stay and cardiac events, indicating a higher risk for longer stays. Diastolic blood pressure is negatively correlated with HSD_Rx. Even though some of these relationships among independent variables are significant at conventional levels, none of the correlations are greater than 0.401. Only one correlation (out of 45) is greater than 0.4 and it is at 0.401 between Cardiac events and Problem List. Judge, Griffiths, Hill and Lee (1985), suggest that multicollinearity problems arise only when the correlations among independent variables are higher than 0.8. Hence, the degree of collinearity present among independent variables appears to be too small to invalidate estimation results. The VIF values are also computed and all eight of them are less than 1.209 and this also indicates that multicollinearity is not an issue. Only if a VIF value exceeds 10, multicollinearity is a concern.

5.1Multivariate Model – Logistic Regression

Using the independent variables in a multivariate context, however, allows one to examine their relative explanatory power and can lead to better predictions since the information which is contained in the cross-correlations among variables is utilized. A primary objective of many multivariate statistical techniques is to classify entries correctly into mutually exclusive groups. Discriminant analysis and logistic regression are examples of such multivariate models. In this study, the following logistic regression (LOGIT) model is proposed:

$\Pr(Y=1|X) = F(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_k)$

The dependent variable Y is a dichotomous (0, 1) variable representing the two groups, cardiac patients readmitted (Y=1) and cardiac patients not readmitted (Y=0) firms. The independent variables X_1 , X_2 , ... X_K include: Gender, BMI, Problem list, Cardiac events, Length of stay, Last diastolic, No shows, and HSD Rx.

It is assumed that no exact linear dependencies exist among X's across k, and that the relationship between Y's and X's are non-linear or logistic (i.e., $P(Y=1|X) = \exp(\Sigma\beta_K X_K) / [1 + \exp(\Sigma\beta_K X_K)]$.)

The null hypotheses would be: $H_0 : \beta_k = 0$, where k = 1,….k

LOGIT results appear in Table 4. Of the eight explanatory variables, six are statistically significant and those six are discussed here.

Table4: Logistic Regression Analysis Results

$$(Y=1|X) = \square_0 + \square_1\,Gender_i + \square_2\,BMI_i + \square_3\,CardEv_i + \square_4\,CSLOS_i + \square_5\,LDiast_i + \square_6\,NoSows_i + \square_7\,HSDRx_i + \square_8\,ProbList_i$$

VARIABLE	MODEL I COEFFICIENT (CHI-SQUARE)	MODEL II COEFFICIENT (CHI-SQUARE)
INTER	-3.450 (135.14) [a]	-3.283 (217.14) [a]
Gender	0.13 (0.03)	0.413 (41.53) [a]
BMI	-0.015 (6.81) [a]	-0.011 (7.08) [a]
CardEv	1.991 (776.36) [a]	1.155 (812.07) [a]
CSLOS	-0.058 (11.50) [a]	---------- ----------
LDiast	-0.005 (2.80)	-0.005 (4.04) [b]
NoShows	0.022 (26.66) [a]	0.001 (50.00) [a]
HSDRx	0.000 (28.41) [a]	0.001 (51.89) [a]
ProbList	0.063 (24.25) [a]	0.019 (4.16) [b]
N	**4,869**	**9,299**

[a] two-tailed significance at < 0.01 level [b] two-tailed significance at < 0.05 level

DEPENDENT VARIABLE: 1 = READCP 0 = NOTREADCP

NAGELKERKE R SQUARE = 0.432; 0.259

MODEL LOG LIKELIHOOD = 4435.71; 7159.02 % CORRECTLY CLASSIFIED = 81.7; 84.0

H_2 (null) suggests that there is no statistically significant difference in BMI between READ and Not READ groups.

The coefficient estimate for the cash turnover ratio is -0.015 and is statistically significant at the 0.01 level. This suggests BMI values are different between the two groups. Interestingly, average BMI scores are slightly higher for READ group. Both groups are somewhat obese. H_3 (null) suggests that there is no statistically significant difference in the Cardiac Events measure between READ and Not READ groups. The coefficient estimate for Cardiac Events variable is 1.991 and is highly statistically significant at the 0.0001 level. This suggests that Cardiac Events measure is significantly different between the two groups. READ patients had, on average, much higher cardiac events than the control group.

H_4 (null) suggests that there is no statistically significant difference in length of stay between READ and Not READ
groups. The coefficient estimate for Cardiac Events variable is -0.058 and is highly statistically significant at the 0.001 level. This suggests that length of stay is significantly different between the two groups. READ patients have had, on average, longer stay than the control group.

H_6 (null) suggests that there is no statistically significant difference in the No Shows measure between READ and
Not READ groups. The coefficient estimate for the No Shows variable is 0.022 and is highly statistically significant at the
0.0001 level. This suggests that No Shows measure is significantly different between the two groups. READ patients missed, on average, more appointments than the control group. This is along the expected lines.

H_7 (null) suggests that there is no statistically significant difference in the HSD Rx measure between READ and Not READ groups. The coefficient estimate for HSD Rx variable is 0.0001 and is highly statistically significant at the 0.0001 level. This suggests that number of prescriptions measure is significantly different between the two groups. READ patients had, on average, many more prescriptions than the control group.

H_8 (null) suggests that there is no statistically significant difference in the problem list measure between READ and Not READ groups. The coefficient estimate for the problem list variable is 0.063 and is highly statistically significant at the 0.0001 level. This suggests that the problem list measure is significantly different between the two groups. READ patients had, on average, many more health problems listed than the control group. This is called comorbid illnesses in the literature and has been significant in predicting readmission in prior literature.

5.2 Classification by Multivariate Models

O'Leary (1987) recommends validating decision support systems (alternatively Risk Score algorithms) against other statistical models, if tests against human experts are very expensive. First, the multiple discriminant analysis model is employed in this study as a content validation tool to evaluate the Risk Score. The purpose of discriminant analysis (DA) is to find the linear combination of risk factors that best discriminates between groups that are partitioned. DA is often applied to problems where the dependent variable is dichotomous. DA classifies entries into mutually exclusive groups by maximizing the inter-group to intra-group variance-covariance from a set of predictor variables. Conventional statistical methods such as DA and Logistic regression (Logit) attempt to arrive at group separation by simultaneously considering all attributes.

The discriminant analysis results are described in table 5. The canonical correlation for the discriminant function is 0.535, and the Chi-square statistic is 1639.51 suggesting significance at p=0.0001 level. The relation between hospital readmissions and the risk factors that are contained in the model appears to be strong. Wilk's lambda, a measure of residual discrimination, is 0.724 and suggests that other factors outside the model may also influence readmissions. However, to an extent, it is not critical to include every variable that might be significant for the purpose of our study. This is because, adding every variable to the model will complicate the model, make it less parsimonious, and impractical to use.

TABLE 5: DISCRIMINANT ANALYSIS RESULTS

PANEL A: DA: TRAINING SAMPLE RESULTS

CLASSIFICATION MATRIX **PREDICTED GROUP** TOTAL

ACTUAL GROUP % CORRECT NOTREADCP READCP

NOTREADCP 88.9% 2845 355 3200

READCP 64.9% 586 1083 1669

TYPE I ERROR : 11.1% **
TYPE II ERROR : 35.1%

PANEL B: DA: HOLDOUT SAMPLE RESULTS

CLASSIFICATION MATRIX **PREDICTED GROUP** TOTAL
ACTUAL GROUP % CORRECT NOTREADCP READCP
NOTREADCP 90.6% 705 73 778
READCP 66.9% 138 279 417
TYPE I ERROR : 9.4% **
TYPE II ERROR : 33.1%

** - Type I (II) error is defined as the percentage of NOTREADCP(READCP) patients that were classified as READCP (NOTREADCP) patients.

Table 4 gives the classification matrix obtained from DA. Panel A of Table 4 gives the training sample results indicating that the 8-variable discriminant analysis model classifies 64.9 percent of the readmitted patients and 88.9 percent of the not-readmitted patients correctly.

Panel B of Table 4 shows that when the discriminant model is employed to analyze the holdout sample, 66.9 percent of the readmitted and 90.6 percent of the not-readmitted cases are grouped correctly.

Researchers also performed a logistic regression (Logit) analysis using the same data sets for the training and the holdout samples and the logit results are reported in Table 6. The training sample results indicate that the 8-variable logit model classifies 65.7 percent of the readmitted patients and 90.1 percent of the not-readmitted patients correctly (see panel A of Table 5). When the logit model is applied to analyze the holdout sample, 66.4 percent of the readmitted patients and 91.6 percent of the not-readmitted patients are grouped correctly (see panel B of Table 5). When you analyze the Type II errors, (33.1% and 33.6% for the two models in the Holdout sample), it shows that both models perform equally well, but there is certainly room for improvement.

TABLE 6: LOGIT ANALYSIS RESULTS

PANEL A: LOGIT: TRAINING SAMPLE RESULTS

CLASSIFICATION MATRIX **PREDICTED GROUP** TOTAL
ACTUAL GROUP % CORRECT NOTREADCP READCP
NOTREADCP 90.1% 2883 317 3200
READCP 65.7% 573 1096 1669
TYPE I ERROR : 9.9% **
TYPE II ERROR : 34.3%

PANEL B: LOGIT: HOLDOUT SAMPLE RESULTS

CLASSIFICATION MATRIX **PREDICTED GROUP** TOTAL
ACTUAL GROUP % CORRECT NOTREADCP READCP
NOTREADCP 91.6% 713 65 778
READCP 66.4% 140 277 417
TYPE I ERROR : 8.4% **
TYPE II ERROR : 33.6%

** - Type I (II) error is defined as the percentage of NOREADCP(READCP) patients that were classified as READCP (NOREADCP) patients.

6. Preliminary Risk Score for Cardiac Patient Readmission

A preliminary Risk Score for Cardiac Patient Readmission is given in **Appendix A**. This risk score is based on t-statistics and Chi-square statistics (from logistic regression). D. The dataset used in this study is unique in the population it covers, specifically within a region including various rural and mid-size urban centers, distinct from larger urban hospitals with a high volume of patients in close proximity. Our risk score formula has the ability, unlike previously published scores, to predict potential cardiac patient readmission at the time of discharge (not after). As such, it has the potential for automation and uptake in clinic flow. Input factors are readily available for patients and their readmission risk score can be easily calculated. The formula groups cardiac patients in to three groups: Low risk; medium risk and high risk based on ranges of the risk score. The risk score is based t-test results and on Chi-square values from Logistic regression results.

7. Conclusion

The current study is novel in its approach to predicting and ultimately providing a prescriptive direction to healthcare decision-makers in examining cardiac readmissions with the introduction of a new algorithm. Existing tools/models for predicting 30-day readmission and LOS have been limited. As previously highlighted, current industry standards such as the LACE index are non-specific and rely solely upon clinical drivers with prediction rates hovering around 50% (Cotter, Bhalla, Wallis, & Biram, 2012). Our findings not only provide increased ability to both predict those at-risk for readmission but also in our prediction of identifying those not at-risk adding to the literature in this area.

This is particularly timely in the transition from fee for service to value-based healthcare in which there is an increasing responsibility to begin identifying and working to mitigate the determinants, whether clinical, biological, or social, contributing to health outcomes including readmissions. As such, this research has the potential to help reduce overall hospital readmission rates and allow hospitals to utilize their resources more efficiently to enhance interventions for high-risk patients by contributing to: 1) successful identification of significant risk factors that can predict cardiac patient readmissions; 2) development of a risk score for readmission; and 3) early identification of the at-risk population and introducing preventive healthcare measures (exercise, education, therapy etc.) to reduce hospital readmission rates prior to and following discharge. In alignment with current healthcare transition, this work assists not only in providing a direction towards cost reduction but importantly makes strides towards increased quality of life for cardiac patients through data-driven preventative efforts for at-risk identified populations.

As shown in this study, readmission is not only affected by clinical indicators, but socioeconomic factors of patients as well. However, programs like HRPP hold healthcare providers accountable, making it necessary for the healthcare industry to leverage existing medical and social data to identify patients at risk and develop necessary interventions. Data-driven methodologies such as the ones performed in this algorithm development are necessary to provide a framework for balancing resource utilization towards such patients with the risk of reduced payments. Our algorithm uniquely collates predictors across both clinical and social determinants of health, specifically contributing to the social drivers crucial to readmissions as patients are discharged into their social environments. In our findings, two drivers viewed as proxies of social environments, are unique within the cardiac readmission research literature. Drivers of patient no shows or missed medical appointments, include both patient behavior characteristics but also potential social determinants impacting means to attend appointments (i.e., lack of transportation, lack of social support able to provide transportation, poverty, distance to hospital).

Similarly, number of prescriptions provides a proxy for comorbid illnesses, drug interactions, and age which are crucial to patient engagement with their healthcare. As such, a risk score revealing these important socioeconomic drivers, along with clinical determinants, are poised to provide healthcare systems with actionable insights on how to intervene with patient populations shown to be at risk for readmission.

This risk score formula, unique within in the published literature, provides an opportunity to identify high risk patients at the time of discharge. The analyses utilized not only allows us to score a patient at various levels of risk for readmission (predictive), it also provides information on the individual drivers for each score (prescriptive). This statistical approach will allow providers to compare and find the best fit on an individual level. On a population level of analytics this tool allows a look across the whole healthcare enterprise. In addition, once validated will allow for prescriptive analytics. This novel statistical approach proves to be a unique tool and a smart collaboration within in the healthcare space.

Limitations to the current study include a lack of data in potentially important drivers of readmission. We did not include data on emergency admissions or patient's patterns of inappropriate utilization that may begin to reveal behavioral patterns around healthcare. Other pertinent predictors of readmission seen in prior works including patient's social support are unfortunately unavailable in electronic medical records and therefore not included in these analyses. Further, only discharge data are used to develop the risk score for readmission. The statistical tests used in this study are association tests and they do not establish causality. Since the dataset comes from hospital in the upper Midwest, the patients are primarily Caucasian. No chart review information is used in the score. Future directions include refinement and validation of the current risk score. Specifically, placing the score into a live clinical environment to assess real-life effectiveness. Historically, impacts of regression to the mean (patient's eventual regress to the mean without intervention) have clouded the influence of predictive ability, especially in assessments applying pre- and post- studies. We intend on circumventing any confounding factors by implementing our validation through controlintervention environments. In collaboration with a large integrated healthcare delivery system we have the ability to evaluate the score with a relatively homogenous patient population where patients are separated by large distances. As such, any novel interventions such as score implementation in one setting and not another will not cross contaminate the control-intervention environments. Metrics indicating predictive success will be available through EMR data including decreased readmissions overtime.

It is evident that prescriptive algorithms are the future for analytics in healthcare. Past the promise of prediction, prescriptive approaches will fully engage providers in the use of big data and analytics. The current risk score with predictive and prescriptive capability provides a much-needed movement towards this work.

References

Andel, C., Davidow, S. L., Hollander, M., & Moreno, D. A. (2012). The economics of health care quality and medical errors. *Journal of Health Care Finance, 39(1)* 39.

Banoff, K. M., Milner, K., Rimar, J., Greer, A. E., & Canavan, M. (2016). Assessment of a novel tool for identifying hospitalized patients with heart failure at risk for 30-day readmission, high cost, and longer length of stay. *Nursing Economics, 34(4),* 172.

Bradley, E., Yakusheva, O., Horwitz, L. I., Sipsma, H., & Fletcher, J. (2013). Identifying patients at increased risk for unplanned readmission. *Medical Care, 51(9),* 761.

Centers for Medicare & Medicaid Services (CMS). (2012). *Readmissions reduction program*. Retrieved from http://go.cms.gov/1L93Lh4).

Cotter, P. E., Bhalla, V. K., Wallis, S. J., &Biram, R. W. (2012). Predicting readmissions: poor performance of the LACE index in an older UK population. *Age and Ageing, 41(6),* 784-789.

De Lissovoy, G. (2013). Big data meets the electronic medical record: a commentary on “identifying patients at increased risk for unplanned readmission”. *Medical Care, 51(9)*, 759-760.

Donzé, J., Aujesky, D., Williams, D., &Schnipper, J. L. (2013). Potentially avoidable 30-day hospital readmissions in medical patients: derivation and validation of a prediction model. *JAMA Internal Medicine, 173(8),* 632-638.

Falvey, J. R., Burke, R. E., Malone, D., Ridgeway, K. J., McManus, B. M., & Stevens-Lapsley, J. E. (2016). Role of physical therapists in reducing hospital readmissions: optimizing outcomes for older adults during care transitions from hospital to community. *Physical Therapy, 96(8)*, 1125-1134.

Gruneir, A., Dhalla, I. A., van Walraven, C., Fischer, H. D., Camacho, X., Rochon, P. A., & Anderson, G. M. (2011). Unplanned readmissions after hospital discharge among patients identified as being at high risk for readmission using a validated predictive algorithm. *Open Medicine, 5(2),* e104.

Judge, G. G., Griffiths, W. E., Carter Hill, R., &Lütkepohl, H. Tsoung-Chao Lee (1985). The Theory and Practice of Econometrics. New Nork, NY: *John Wiley & Sons*.

Kansagara, D., Englander, H., Salanitro, A., Kagen, D., Theobald, C., Freeman, M., &Kripalani, S. (2011). Risk prediction models for hospital readmission: A systematic review. *Journal of the American Medical Association, 306(15)*, 1688-2698

Kasotakis, G., Schmidt, U., Perry, D., Grosse-Sundrup, M., Benjamin, J., Ryan, C., & Bittner, E. A. (2012). The surgical intensive care unit optimal mobility score predicts mortality and length of stay. *Critical Care Medicine, 40(4),* 11221128.

Kripalani, S., Theobald, C. N., Anctil, B., &Vasilevskis, E. E. (2014). Reducing hospital readmission rates: current strategies and future directions. *Annual Review of Medicine, 65,* 471-485.

Ross, J. S., Mulvey, G. K., Stauffer, B., Patlolla, V., Bernheim, S. M., Keenan, P. S., &Krumholz, H. M. (2008). Statistical models and patient predictors of readmission for heart failure: a systematic review. *Archives of Internal Medicine, 168(13)*, 1371-1386.

Rothman, M.J., Rothman, S.I. &Beals, J. (2013). Development and validation of a continuous measure of patient condition using the electronic medical record. *Journal of Biomedical Informatics, 46(5),* 837-848.

Rothman, M.J., Solinger, A.B., Rothman, S.I., & Finlay, G.D. (2012). Clinical implications and validity of nursing assessments: A longitudinal measure of patient condition from analysis of the electronic medical record. BMJ Open, 2(e000849), 1-6.

Snowdon, M. (2014) White paper: The four steps of population health management. Optum

Tan, C., Loo, G, Pua, Y., Chong, H., Yeo, W., Ong, P., … Allison, G. (2014). Predicting discharge outcomes after total knee replacement using the Risk Assessment and Predictor Tool. *Physiotherapy, 100(2),* 176-181.

The Patient Factor (2016). World Health Organization's Ranking of the World's Health Systems/ August 16, 2016., Retrieved from http://thepatientfactor.com/canadian-health-care-information/world-health-organizationsranking-of-the-worlds-health-systems/

The World Bank Group. (2015). Health expenditure per capita (current US$). *The World Bank*. Retrieved August 26,

2015fromhttp://data.worldbank.org/indicator/SH.XPD.PCAP?order=wbapi_data_value_2013+wbapi_data_value+w bapi_data_value-last&sort=desc

Tsai, T. C., Joynt, K. E., Orav, E. J., Gawande, A. A., & Jha, A. K. (2013). Variation in surgical-readmission rates and quality of hospital care. *New England Journal of Medicine, 369(12),* 1134-1142.

Van Walraven, C., Dhalla, I., Bell, C., Etchells, E., Stiell, I., Zarnke, K., … Forster, A. (2010). Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. *Canadian Medical Association Journal, 182(6),* 551-557.

Wagener, G., Minhaz, M., Wang, S., Panzer, O., Wunsch, H., Playford, H., &Sladen, R. (2011). The Surgical Procedure Assessment (SPA) score predicts intensive care unit length of stay after cardiac surgery. *Journal of Thoracic Cardiovascular Surgery, 142(2)*, 443-450.

Wang, H., Robinson, R., Johnson, C., Zenarosa, N., Jayswal, R., Keithley, J., & Delaney, K. (2014). Using the LACE index to predict hospital readmissions in congestive heart failure patients. *BMC Cardiovascular Disorders, 14*, 1-8.

Appendix A:Risk Score for Cardiac Patient Readmissions

Cardiac Events (1 to 4 same; greater than 5 = 5)	5
Problem List (1 to 3 same; 4 to 5 = 4; greater than 5 = 5)	5
Length of Stay (1 to 3 same; 4 to 5 = 4; greater than 5 = 5)	5
No Shows (4 to 7 = 1; greater than 7 = 2)	2
HSD Rx (can take the values of 1, 2 or 3	3
Diastolic pressure (less than 68 = 1)	1

Male (yes = 1)	1
Tobacco user (yes =1)	1
Age > 69	1
Other Maximum score is 25.	1

Scale: 5 – 10 = Low Risk; 11 – 14 = Medium Risk; 15 & above = High Risk [This score is based on Univariate T-test and Logistic regression results]

Note: Patient data used in this study were de-identified and the project was approved by the Sanford Privacy Board and the USD Institutional Review Board.