1. **Introduction**

Hospital readmission is an adverse and frequent outcome of hospitalization. Impacting patients across all diseases and conditions, readmissions contribute to mortality and morbidity, and are costly. About 1 in every 7 patients discharged from the hospital will be readmitted within 30 days following discharge, with readmission rates for certain conditions like acute myocardial infraction (AMI) and heart failure (HF) exceeding 20%.1,2 To reduce readmissions, patients at an elevated risk are frequently assigned readmission-avoidance interventions to help manage post-acute care at home.3-7 A range of readmission-avoidance interventions are currently available for patients in need for enhanced discharge and post-discharge services. While standard discharge and post-discharge care protocols vary across hospitals, enhanced discharge services frequently include ordering medications and scheduling primary care visits prior to leaving the hospital and enhanced post-discharge services may include follow-up phone calls and in-person or virtual home visits. Patients may be assigned a combination of these services depending on their risk for a poor readmission outcome. Comprehensive multi-component interventions, especially ones that include in-person home health care (HHC) visits during the post-discharge period, have been shown to reduce 30-day readmission risk by as much as 20-30% while costing between $170 and $425 (in 2022 USD) per discharge.8,9

Accurately assessing a patient’s risk of readmission a*t the time of discharge* is important clinicians who determine what readmission-avoidance care plan is most appropriate for each patient’s successful transition from hospital to home-based care. Not every patient needs enhanced discharge and post-discharge care; a younger, healthier patient with strong family and social support may do well under standard discharge and post-discharge care. However, patients with certain risk factors (e.g., chronic conditions requiring multiple medications, needing assistance with activities of daily living, lacking family or social support) may require varying levels of enhanced discharge and post-discharge management tailored to each patient’s specific risk factors. Underestimating a patient’s risk for 30-day readmission at the time of hospital discharge can lead to a poor readmission outcome, while overestimating a patient’s risk may result misallocation of costly, labor-consuming care.

Readmission risk prediction modeling is one of the most widely published research areas—a systematic review10 reported over 7,500 academic peer-reviewed citations for readmission risk,11 and another specifically focusing on risk-prediction methods reporting 265 citations. Not including condition-specific models, 73 general risk prediction models have been proposed in the literature.11 Yet, few existing models are able to provide accurate information to enable risk-stratification of patients at the time of hospital discharge; the out-of-sample predictive performance (ability to correctly identify high-risk patients while also correctly exclude low-risk patients) of even the most sophisticated newest models leaves much to be desired (validation sample C-statistics range around 0.60 to around 0.75). A number of relatively recent studies applied machine learning and other big-data approaches to mine an abundance of patient data routinely collected and stored in hospitals’ electronic health records (EHR),12 yet their out-of-sample performance demonstrated a limited improvement over traditional regression models and rubrics (validation sample C-statitic range around 0.85).11 Notably, the few risk models most commonly used in clinical practice (e.g. LACE+,13 BOOST 8Ps,14 Charlson Comorbitity Index,15-17 Elixhauser Readmission Index15,18,19) rely on a small set of patient characteristics, with C-statistics hovering around 0.60, and lower in prospective validations. 10,20

In this study, we examine the possibility that, at the present time, the “gold standard” approach to readmission risk modeling suffers from an unrecognized treatment selection bias, leading to a systematic misclassification of some high-risk patients as low-risk (and vice versa). At the time of discharge, a patient’s true risk for becoming readmitted during the subsequent 30-day period is an unobserved latent variable. As we will discuss in detail later, treatment selection bias arises due to the use of the *observed 30-day readmission outcome* as a “realization” of the patient’s *latent true risk of becoming readmitted*—without accounting for non-random assignment of patients to readmission-avoidance interventions. All currently known readmission risk modeling methods, including machine learning and other data-mining methods, start with a data set of observed patient characteristics and 30-day readmission outcomes. Most then calibrate the risk model to put more weight on those variables that are strongly associated with 30-day readmission outcomes, and less weight, or eliminating, weak predictors of observed readmissions.10,20-22 However, patients who did not have a readmission within 30 days represent a mix of those who truly had a low readmission risk at the time of discharge and those who may have been at a high risk at the time of discharge but the risk was correctly identified by clinicians and appropriate discharge and post-discharge interventions were put in place to mitigate the risk. Being unable to differentiate between the two groups, a risk model calibrated to the observed readmission outcome would systematically underestimate the predictive power of patient characteristics that are easily observed by clinicians and correctly signal a patient’s elevated *risk* for readmission at the time of discharge. In fact, the better an observable patient characteristic at signaling a patient’s true readmission risk to clinicians, the more its weight in the prediction model could be underestimated (treatment selection bias). In a nutshell, the current readmission risk modeling approach may have a tendency of underestimating the importance of the most salient observable patient risk factors.

As an example, older age is perhaps one of the most easily observed patient characteristics and a risk factor for a number of adverse outcomes from post-operative complications, to medication errors and non-adherence, to falls. Most U.S. older adults have multiple risk factors (e.g., chronic conditions, limited ability to independently perform activities of daily living, and may live alone) further increasing their risk for unplanned readmissions and emergency room visits. In a systematic review of readmission risk models,8 older age was almost always considered as a patient risk variable in the beginning stages of model development; yet it was ultimately dropped from over a quarter of all final models due to its weak statistical association with the observed 30-day readmission outcome. Other patient characteristics that were frequently considered but ended up excluded from final risk models are patient race, marital status, and socioeconomic status.8 Is it possible that the predictive signal of these observable patient characteristics was muted by the formentioned treatment selection bias? After all, none of the models considered readmission-avoidance interventions.10,11 If this is the case, many readmission-prediction models currently used in clinical practice may be systematically understating the needs for readmission-avoidance services among some of the most vulnerable patient groups, thus potentially contributing to poor patient outcomes and systemic health inequities.

To our knowledge, the treatment selection bias in readmission risk modeling has received little to no attention in the literature, to date. Despite decades of research focusing on the right-hand side of the risk model (selection of predictor variables and/or the functional form of the model), none of the published studies to our knowledge have attempted to examine the left-hand side of the model—use of observed readmission outcomes as the dependent variable. This is an important and fundamental oversight that cannot be easily corrected by adjusting/controlling for readmission avoidance interventions on the right-hand-side of the model. Although doing so would result in more accurate predictions of 30-day readmissions (conditional on readmission-avoidance interventions that were implemented), it will do little to improve our ability to assess a patient’s true readmission risk at the time of discharge—a key to informing optimal discharge and post-discharge care for each patient.

Therefore, we carry out a theoretical and empirical investigation of the treatment-selection bias in readmission risk modeling. This bias is inherent to the prediction of hospital readmissions, as well as any adverse outcome whose risk of a future occurrence is both (at least) partly observable by clinicians and modifiable through intervention (mortality, hospital-aquired infections, falls). Because the root cause of the problem lies is the measurement of the right-hand-side (dependent) variable, it can confound even the most sophisticated current risk modeling approaches (e.g., logistic regression, box-cox proportional hazard model, various machine learning algorithms). Therefore, this study aims to analytically derive the treatment selection bias in readmission risk modeling, propose correction approaches, and provide a simple illustration of the correction method using electronic health records data from one acute health care facility.

1. **Theoretical derivation**

To demonstrate the treatment selection bias, we present a causal conceptual model (**Figure 1**) as the following four equations: 23

1. the patient’s true latent risk at the time of discharge is determined by a set of patient risk factors :

; [1]

1. based on patient’s observable characteristics, clinicians form their own assessment of the patient’s risk, , which is also latent and may or may not be the same as the patient’s true risk:

; [2]

1. clinicians’ assessment of patient risk influences their choice of the intensity of readmission-avoidance intervention assigned to the patient at discharge:

; [3]

1. and finally, the patient’s readmission outcome at 30-day post-discharge is determined by both the patient’s underlying risk of readmission at the time of dischage and the effectiveness of the patient’s assigned readmission-avoidance interventions :
2. . [4]

Assuming there are observations (i.e. patients) and x-variables (i.e., patient characteristics), we used the following notation:

* is a matrix of patient characteristics (n observations/rows on k variables/columns),
* is a vector of observations for a continuous [0, 1] latent variable for a patient’s readmission risk at the time of discharge,
* is a vector of observations for a continuous [0, 1] latent variable for clinicians’ assessment of the patient’s readmission risk at the time of discharge,
* is a vector of observations for a continuous variable measuring the intensity of readmission-avoidance intervention from none (0) to some feasible maximum assigned to the patent at the time of discharge,
* is a vector of observations for a 0/1 categorical variable with 1 indicating a patient had a readmission during the 30-day period after discharge, observed at 30-days post-discharge.
* is the logged likelihood operator (e.g. conditional probability, logged proportional hazard ratio); 24
* is a vector of regression parameters of the true readmission risk model; the coefficients represent the incremental increase in patient readmission risk at the time of discharge associated with patient characteristic .
* is a vector of regression parameters of clinicians’ risk assessment model; they may be the same or different from vector A depending on how well clinicians are able to access a patient’s risk from observable patient’s characteristics;
* is a scalar coefficient representing the incremental increase in the patient’s assigned treatment intensity per an increase in clinicians’ assessment of patient risk;
* is a scalar coefficient representing the incremental reduction in the likelihood of a patient’s 30-day readmission per an increase in treatment intensity ;
* are classical iid regression error terms with a zero (or constant) mean and constant variance.

The “true” readmission risk model is represented by Equation 1. Knowing the parameters of vector A would allow clinicians to assess a patient’s risk of readmission from observed patient characteristics. A direct estimation of the vector A would require observable data on for a representative sample of the patient population. However, the fundamental problem is that is an unobservable latent variable, rendering direct risk estimation impossible.25

The traditional approach to readmission risk modeling is to substitute unobserved risk with the likelihood of the observed 30-day readmission outcome, :

[5]

Note, that while coefficients of the estimated model are derived from the variance-covariance matrices of the patient characteristics variables with the observed outcome , the true coefficient estimates should, theoretically, be derived from the latent readmission risk variable: . By substituting the observed 30-day outcome for unobserved latent risk, the conventional modeling approach, therefore, implicitly assumes that the observed outcome is an unbiased realization of the latent risk, ). and, as such, is translated to clinical practice for readmission risk assessment and intervention assignment at discharge. 25

Yet, from the conceptual causal model (1) we know that the likelihood of the observed 30-day outcome is affected by endogenously assigned readmission-avoidance interventions . Therefore, the coefficients of the estimated model can be written as

Or:

[6]

From the above, it follows that, if interventions are at least somewhat effective at preventing 30-day readmissions ( and the treatment assignment is non-orthogonal to patient characteristics () , the traditional approach of using the observed 30-day outcome as the left-hand side variable in modeling produces biased (or “shrunk”) regression coefficient estimates.

[7]

Re-writing the above equation in an open matrix form, we can further examine how the magnitude of the treatment selection bias varies across the parameter estimates of individual patient characteristics.

[6’]

Here, it is clear that patient characteristics that more accurately signal a patient’s need for readmission-avoidance interventions (greater ) will have a larger “shrinking” bias in the estimate of the regression coefficient using the traditional approach.

It is important to note that a direct adjustment for the treatment-assignment variable as a control in the traditional risk prediction model with the observed outcome as the dependent variable (i.e., ) is not a solution to the treatment selection assignment bias. This is because any estimates of the coefficient vector obtained from a variance-covariance matrix of patient characteristics and observed outcomes will be similarly biased for estimating the true patient risk at discharge, for the reason discussed above. Although a direct covariate adjustment for may improve the performance of the model for predicting 30-day readmission outcomes, it will have little effect on the ability of the model to predict the risk of a 30 day readmission at the time of discharge.

In sum, because of non-random assignment of risk-avoidance interventions to high-risk patients at the time of discharge, the current approach to risk-prediction modelling suffers from treatment selection bias. The more effective are risk-mitigating treatments available to clinicians, the less valid the conventional approach will be for predicting pre-treatment risk; and the more a patient characteristic is viewed as a readmission risk factor by clinicians, the less predictive it will appear in a traditional readmission risk model. By the nature of the bias, any risk model calibrated to the observed 30-day outcome potentially suffers from the treatment selection bias, regardless of the selection of variables and controls, or the functional form (OLS, limited dependent variable, survival, etc.) of the right-hand-side of the equation, including machine learning approaches.

**III. Empirical illustration**

We provide an empirical illustration as a proof of concept that demonstrates treatment selection bias in predicting patients’ risk of 30-day readmission at the time of discharge. For this illustration, we use patient-level data on patient characteristics (X), whether or not a patient was discharged from the hospital with a referral to home health care (T), and the patient’s subsequent hospital utilization history during 30-days after discharge from our prior multi-site study of hospital readmissions.26,27

III.1 Data

We use EHR data from the Readiness Evaluation and Assessment Discharge Intervention (READI) study26 for adult (18+) hospital patients discharged to home-based care from medical, surgical, or medical-surgical units at 31 US hospitals between October 15, 2014 and March 31, 2016. The READI study was a cluster-randomized trial of a discharge teaching intervention; we included only patients discharged from the control units who received usual care (n=90,796). We further excluded patients with short (<24 hours) lengths of stay (n=8,708) and those missing data on key predictor variables (history of prior hospitalization, n=9,872, and severity or mortality scores n=25,581). The final sample for this illustrative analysis included n=44,948 patient discharges).

III.2 Variables and measures

For this illustrative analysis, we included a limited set of patient characteristics frequently used in readmission risk models, 10,21 one binary readmission-avoidance intervention variable, and 30-day same-hospital readmissions. The descriptive statistics of the sample are presented in **Table 1.**

* Patient characteristics: = [age, sex, marital status, race, ethnicity, history of prior hospitalizations within 30 days of admission, patient’s type of admission (medical/surgical), intensive care unit (ICU) stay during hospitalization, severity of illness score, mortality score, length of hospital stay, and patient’s major diagnostic category (MDC) based on the discharge diagnosis].
* Readmission avoidance intervention: assignment to home health care (HHC) services at the time of discharge based on hospital discharge disposition codes from EHRs. HHC is a common readmission reduction strategy in the US and internationally which was shown to prevent readmissions in two randomized-controlled trials. 8,9 The HHC intervention variable was defined as binary HHC assignment record, = [1 if patient discharged home with HHC; 0 otherwise].
* Observed 30-day readmission outcome was also a binary variable, = [1 if patient was readmitted to the same hospital within 30 days after discharge; 0 otherwise].

III.2 Approach

Because selective assignment of readmission-avoidance interventions to high risk patients is the root cause of the treatment selection bias, the aim of our illustrative analysis is to incorporate the patients’ HHC assignment status in the risk model.

We first estimate a logistic regression model with the observed 30-day readmission outcome as the left-hand-side variable and patient characteristics as right-hand side variables. We refer to this model as the traditional risk model, which we assert is biased due to non-random treatment selection. We also estimate an HHC-adjusted version of the traditional model to examine the impact of performing a direct covariate adjustment for readmission-avoidance intervention on the coefficients of the patient variables; as we note earlier, we expect that a direct covariate adjustment for HHC assignment will have very little effect on the parameters of the risk model.

Then, we illustrate two potential ways of addressing treatment selection bias in readmission risk modeling—an outcome-inflated model and a structural equations model.

*III.2.a Outcome-inflated model (OIM)*

The OIM approach attempts to reduce treatment selection bias by artificially negating the effect of HHC on 30-day readmissions (constraining in equation [7]). Based on the existing randomized-controlled trials, we assume that HHC reduces patients’ likelihood of readmission by 20-30% and simulate a hypothetical counter-factual (without HHC) 30-day readmission rate, in which the proportion of readmissions among patients with an HHC assignment is artificially inflated so that . Assuming the 20-30% effect of HHC assignment is accurate (**Appendix 1**), the inflated readmission variable is unbiased relative to true patient risk :

[8]

We simulated the model 1000 times each time replacing a new set of randomly selected non-occurrences among HHC patients with counterfactual occurrences. All replacements were at made at random, orthogonally to patient characteristics. In each iteration, we estimated the conventional readmission risk equation [4] and the OIM model using the bias-adjusted as the dependent variable:

[9]

Estimated parameters and standard errors were be derived as the means of the respective sampling distributions of iteration-specific estimated coefficients and their standard errors. We use a seemingly unrelated estimation28,29 to test the null hypothesis of equality of the estimated coefficients between the OIM and the traditional (baseline) model and reported the proportion of rejection samples as the p-value for the hypothesis of equality of the estimates.

The OIM approach requires that the true causal effect size of the intervention on the likelihood of 30-day readmissions is known. As sensitivity analyses, we used a range assumptions about HHC effectiveness (10%, 20%, 30%, 40%) (**Appendix 1**). In all analyses, we assumed there were no heterogeneous treatment effects and inflated observed readmissions orthogonally to patient characteristics .

*III.2.b Structural equation model (SEM)*

The idea of using an SEM is to exploit the observed HHC assignment as a signal of a high patient risk as assessed by clinicians. Doing so would allow to implicitly incorporate clinician knowledge and expertise thus augmenting information about patient risk contained in the observed 30-day outcome. We estimate the following structural equation model (SEM) using a multiple indicators multiple causes (MIMIC) structural model30 in MPlus:

SEM: [10]

The MIMIC model is a special case of SEM and consists of two parts: a predictive model and a measurement model (**Figure 2**); the predictive part models the structural relationship between latent risk and its predictors (patient characteristics ); the measurement part models and as observed realizations, or measurement indicators, of the unobserved risk. To the extent that assignment to HHC is correlated with clinicians’ assessment of patient risk (and clinician’s assessment correctly incorporates observable patient characteristics (, including intervention assignment as a measurement indicator can improve the estimate of latent risk .

The SEM approach requires at least partial independence among the measurement indicators. While otherwise structurally similar to the conceptual model, the SEM assumes structural parity between measurement indicators and , without explicitly modeling the causal effect of the intervention on the observed outcome. Therefore, it is assumed that the observed outcome is at least partly driven by risks not present or unknown to clinicians at the time of discharge.

III.3 Results

Descriptive statistics of the sample are presented in **Table 1**. The sample is 52% female, average age is 59.7 years old, 71% are medical patients, average length of stay is 4.05 days and 19% have an ICU admission during their hospital stay. Thirteen percent were discharged home with an HHC referral and 10% were readmitted within 30 days of discharge. Correlation between the HHC assignment and the observed outcome was 0.06 (P<0.001).

Regression results are presented in **Table 2** (patient demographic variables only) and **Appendix 2** (full set of results). Colums 1-2 in both tables show the estimation logistic regression results of the HHC assignment and 30-day readmissions models, respectively. Patient demographic variables (age, gender, marital status, race) were significant predictors of HHC assignment (Column 1), but contributed very little explanatory variance to the observed 30-day outcome model (Column 2). Clinical variables (prior hospitalization, severity of illness and mortality scores, diagnoses) were strong predictors in both HHC assignment and 30-day readmission models. Among the clinical variables, a few had oppositely-signed coefficients in the two models: surgical admission and an ICU stay during the encounter were positively associated with HHC assignment and negatively with readmissions, while a number of diagnostic groups were negatively associated with HHC assignment and positively with 30-day readmissions. These results are consistent with our fundamental premises that 1) HHC is selectively assigned to patients whom clinicians may perceive at a higher risk for 30-day readmissions (older, non-married, female, black patients) and 2) because these characteristics are not predictive of the 30-day observed outcome, the observed 30-day readmissions model may systematically underestimate the role of these observable patient characteristics in predicting readmission risk at the time of discharge.

Column 3 shows the results of direct covariate adjustment for HHC assignment in the 30-day readmission model. As expected, HHC assignment is positively associated with 30-day readmissions and slightly increases the predictive power of the model for the 30-day readmission outcome (AUC increases from 0.71 to 0.72 and R-squared increases from 0.09 to 0.10. Nevertheless, the coefficients of the patient variables remain qualitatively and quantitatively similar as in the traditional 30-day readmission model without adjusting for HHC assignment.

Columns 4-5 (4-7 in the Appendix 2) show the results of inflating the outcome variable in the alternative outcomes framework. Inflating the 30-day readmission rate in the HHC group increases the coefficients of patient demographic variables strongly associated with HHC assignment (Column 1) in a dose-response fashion. The more effective is HHC presumed to be at preventing 30-day readmissions (high δ), the higher and more significant are the coefficients of patient demographic variables. Additionally, the most salient demographic variables like age and race (high β’s) are impacted the most under all assumptions regarding the effectiveness of HHC assignment. Specifically, while the age categories do not predict the likelihood of observed outcome (Column 2), the coefficients increase in both magnitude and significance in the IOM models. The simulated results for the linear age variable (Figure 3) are also consistent with the premise that age is a significant predictor of the latent risk at discharge at HHC effectiveness levels of 0.5 and higher (the details of the analysis are provided in Appendix 3).

Column 6 (8-9 in Appendix 2) shows the SEM results. Using both the observed hospital utilization during 30 days post-discharge and HHC assignment at the time of discharge as the measurement indicators for the patient’s latent risk at the time of discharge, the predictive side of the model shows that patient demographic characteristics are strong predictors of latent risk. For comparison, and to examine the incremental contribution of HHC assignment to the measurement side of the model, we also estimated a SEM model without HHC assignment on the measurement side (Appendix 2 Column 8) Removing HHC assignment from the measurement side slightly worsens the fit of the model and renders the coefficients of most patient demographic variables non-significant.

Overall, these results illustrate the potential impact of the treatment selection bias on the estimated coefficients of patient characteristics in predicting patients’ latent risk at discharge. The more effective is the risk-avoiding intervention and the more salient are the patient risk factors, the less likely are the corresponding parameters to appear significant in models for the observed 30-day outcome.

**IV. Discussion**

The goal of readmission risk modeling methods is to estimate a patient’s risk of becoming readmitted at the time of discharge using the patients’ observable patient characteristics. In this paper, we point out a fundamental flaw of the traditional approach of using the observed readmission outcome to calibrate the model—readmission outcomes are observed with a 30-day lag, during which the patient’s risk of readmission may have been intervened with, and reduced, by clinicians. To the extent that clinicians’ assessments of a patient’s risk is at least partly correlated with the patient’s true readmission risk at the time of discharge, the patient’s observed 30-day readmission outcome is a biased realization of the patient’s true risk at the time of discharge. This can introduce a treatment selection bias in the estimates of the most salient predictors of patient risk at the time of discharge, resulting in poor performance of risk-assessment tools and models in clinical practice, misallocation of care, and poor patient outcomes.

To our knowledge, this study is the first to investigate treatment selection bias in readmission risk modeling mathematically, provide an empirical illustration of the problem, and preliminarily explore some potential methodological improvements. We demonstrated that the bias is most likely to occur when clinicians correctly identify observable patient risk factors, when risk-avoidance interventions are targeted more selectively to high-risk patients, and when more (or more effective) risk-avoidance interventions are available in clinicians’ disposal. When all three are true, most salient true predictors of readmission risk will be the most biased towards non-significance in a risk model calibrated to the observed 30-day outcomes. This bias can result in a systematic misclassifying high-risk patients as low-risk, and vice versa. In our preliminary illustrative results, this bias disproportionally affected older black female patients, which could have significant implications for the quality of care and outcomes in this vulnerable patient population.

We explored two potential approaches to addressing treatment selection bias in readmission risk modeling—outcome inflation and structural equations modeling. Both appear to have a potential for correcting the treatment selection bias and improving readmission risk modeling; yet both have limitations and strong, sometimes untestable, identifying assumptions. The outcome inflation approach offers computational simplicity and an intuitive interpretation of coefficient estimates (similar to a conventional logistic regression). Yet its primary weakness is the untestable assumption regarding the effect size of readmission-avoidance interventions on the likelihood of 30-day readmissions. This assumption becomes even more difficult to satisfy when multiple interventions are bundled as part of a patient’s readmission avoidance plan. The advantage of the structural equations modeling approach is its ability to account for multiple readmission avoidance interventions (e.g., assessment and mitigation of home for fall risk, use of occupational therapy, etc.). The reduction in treatment-selection bias is likely to be greater, and the precision of risk modeling higher, when a number of different risk-reduction interventions are available in the data, especially if their assignment is triggered by different observable patient characteristics. As we demonstrated, the model could also incorporate additional observed realizations of risk (e.g., occurrence and/or number of emergency department (ED) visits and hospital admissions for observation), potentially allowing for further improvements in pre-exposure risk prediction. One limitation of the structural equations modeling approach is its potential vulnerability to interdependence of measurement indicators—a testable but not always correctible problem. Another limitation is the potential of clinician bias or misinformation to skew observed intervention assignment in a direction opposite to the patient’s true risk. Finally, a limitation of both approaches is that their performance AUC, positive/negative predictive values difficult to validate (internally or externally) since true patient risk at the time of discharge is unobserved. Their comparative effectiveness against the traditional approach can only be inferred by improved clinical decision making and reduced readmissions when used in clinical practice.

All limitations notwithstanding, our study lays a conceptual foundation for a new dimension of scientific inquiry with a potential to, someday, significantly improve our ability to predict patient risk. Currently, the science of risk prediction focuses on incremental improvements to the right-hand-side of the model and revolves around either finding new patient variables or clusters of variables (limitations of daily living, expected support) or more sophisticated mathematical approaches to uncover more complex predictive patterns (machine learning methods). Yet, few studies to our knowledge tacked the left-hand-side of the model, and specifically questioned the assumption that the observed 30-day readmission outcome is an unbiased realization of patient risk at the time of discharge. Developing methods for testing and remedying the potential negative implications of violating the assumption will require a different suit of methods from ones currently being deployed in the field of readmission risk modeling. Notably, the conceptual framework explicated here is applicable to risk prediction for any adverse patient outcome amendable through appropriate clinical interventions (e.g., patient restrains for patient falls, prophylactic antibiotics for post-surgical infections, opioid administration for pain management). While this paper opens more questions than it answers, continued work in this area is necessary in order to assess and correct what is possibly a broad-scale detriment to our ability to predict patient risk.

**V. Conclusion**

An improved ability to predict the risk of readmission (or other significant outcomes such as mortality, HAIs, falls) is paramount to improving quality of patient-centered care. These outcomes are of interests and importance to patients and their caregivers because of the human and economic burden associated with them. Correcting the treatment selection bias in readmission risk modelling is a novel approach with a potential for improving the validity and trustworthiness of clinical decision-support systems to inform high-quality patient-centered care.

**References**

1. Hines AL, Barrett ML, Jiang HJ, Steiner CA. Conditions with the Largest Number of Adult Hospital Readmissions by Payer, 2011. Agency for Healthcare Research and Quality, Healthcare Cost and Utilization Project. Accessed 1/30, 2019. <https://www.hcup-us.ahrq.gov/reports/statbriefs/sb172-Conditions-Readmissions-Payer.pdf>

2. Barrett ML, Wier LM, H.J. J, Steiner CA. All-Cause Readmissions by Payers and Age, 2009-2013:. Agency for Healthcare Research and Quality, Healthcare Cost and Utilization Project. Accessed 1/30, 2019. <https://www.hcup-us.ahrq.gov/reports/statbriefs/sb199-Readmissions-Payer-Age.pdf>

3. Yakusheva O, Hoffman GJ. Does a Reduction in Readmissions Result in Net Savings for Most Hospitals? An Examination of Medicare’s Hospital Readmissions Reduction Program. *Medical Care Research and Review*. 2020;77(4):334-344. doi:10.1177/1077558718795745

4. Siclovan DM, Bang JT, Yakusheva O, et al. Effectiveness of home health care in reducing return to hospital: Evidence from a multi-hospital study in the US. *Int J Nurs Stud*. Jul 2021;119:103946. doi:10.1016/j.ijnurstu.2021.103946

5. Naylor MD, Brooten DA, Campbell RL, Maislin G, McCauley KM, Schwartz JS. Transitional care of older adults hospitalized with heart failure: a randomized, controlled trial. *Journal of the American Geriatrics Society*. May 2004;52(5):675-84. doi:10.1111/j.1532-5415.2004.52202.x

6. de Mestral C, Kayssi A, Al-Omran M, Salata K, Hussain MA, Roche-Nagle G. Home care nursing after elective vascular surgery: an opportunity to reduce emergency department visits and hospital readmission. *BMJ Quality &amp; Safety*. 2019;28(11):901-907. doi:10.1136/bmjqs-2018-009161

7. Xiao R, Miller JA, Zafirau WJ, Gorodeski EZ, Young JB. Impact of Home Health Care on Health Care Resource Utilization Following Hospital Discharge: A Cohort Study. *The American journal of medicine*. Apr 2018;131(4):395-407.e35. doi:10.1016/j.amjmed.2017.11.010

8. Naylor MD, Brooten DA, Campbell RL, Maislin G, McCauley KM, Schwartz JS. Transitional Care of Older Adults Hospitalized with Heart Failure: A Randomized, Controlled Trial. *Journal of the American Geriatrics Society*. 2004;52(5):675-684. doi:<https://doi.org/10.1111/j.1532-5415.2004.52202.x>

9. Advanced practice nurse directed transitional care reduced readmission or death in elderly patients admitted to hospital with heart failure. *Evidence Based Nursing*. 2004;7(4):116-116. doi:10.1136/ebn.7.4.116

10. Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: A systematic review. *JAMA*. 2011;306(15):1688-1698. doi:10.1001/jama.2011.1515

11. Zhou H, Della PR, Roberts P, Goh L, Dhaliwal SS. Utility of models to predict 28-day or 30-day unplanned hospital readmissions: an updated systematic review. *BMJ open*. Jun 27 2016;6(6):e011060. doi:10.1136/bmjopen-2016-011060

12. Yakusheva O, Bang JT, Hughes RG, Bobay KL, Costa L, Weiss ME. Nonlinear association of nurse staffing and readmissions uncovered in machine learning analysis. *Health services research*. Apr 2022;57(2):311-321. doi:10.1111/1475-6773.13695

13. van Walraven C, Dhalla IA, Bell C, et al. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*. Apr 6 2010;182(6):551-7. doi:10.1503/cmaj.091117

14. Hansen LO, Greenwald JL, Budnitz T, et al. Project BOOST: effectiveness of a multihospital effort to reduce rehospitalization. *Journal of hospital medicine*. Aug 2013;8(8):421-7. doi:10.1002/jhm.2054

15. Li B, Evans D, Faris P, Dean S, Quan H. Risk adjustment performance of Charlson and Elixhauser comorbidities in ICD-9 and ICD-10 administrative databases. *BMC health services research*. Jan 14 2008;8:12. doi:10.1186/1472-6963-8-12

16. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases*. 1987;40(5):373-83.

17. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *Journal of clinical epidemiology*. Jun 1992;45(6):613-9.

18. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Medical care*. Jan 1998;36(1):8-27.

19. Project HCCaU. Elixhauser Comorbidity Software, Version 3.7. <https://www.hcup-us.ahrq.gov/toolssoftware/comorbidity/comorbidity.jsp>

20. Smith LN, Makam AN, Darden D, et al. Acute Myocardial Infarction Readmission Risk Prediction Models: A Systematic Review of Model Performance. *Circulation Cardiovascular quality and outcomes*. Jan 2018;11(1):e003885. doi:10.1161/circoutcomes.117.003885

21. Mahajan SM, Heidenreich P, Abbott B, Newton A, Ward D. Predictive models for identifying risk of readmission after index hospitalization for heart failure: A systematic review. *European journal of cardiovascular nursing : journal of the Working Group on Cardiovascular Nursing of the European Society of Cardiology*. Dec 2018;17(8):675-689. doi:10.1177/1474515118799059

22. Ross JS, Mulvey GK, Stauffer B, et al. Statistical models and patient predictors of readmission for heart failure: a systematic review. *Archives of internal medicine*. Jul 14 2008;168(13):1371-86. doi:10.1001/archinte.168.13.1371

23. Hubert M. Blalock J. *CAUSAL INFERENCES IN NONEXPERIMENTAL RESEARCH*. New York : Norton, [1972, c1964]. 1964.

24. Long JS. *Regression Models for Categorical and Limited Dependent Variables.* . Sage Publicaitons; 1997.

25. Everitt B. *An Introduction to Latent Variables Models*. Chapman and Hall; 1984.

26. Weiss ME, Yakusheva O, Bobay KL, et al. Effect of Implementing Discharge Readiness Assessment in Adult Medical-Surgical Units on 30-Day Return to Hospital: The READI Randomized Clinical Trial. *JAMA network open*. Jan 4 2019;2(1):e187387. doi:10.1001/jamanetworkopen.2018.7387

27. Weiss MPI. READI (Readiness Evaluation And Discharge Interventions): Implementation as a Standard Nursing Practice for Hospital Discharge. Accessed 2/7, 2017. <http://www.marquette.edu/nursing/readi-index.shtml>

28. Clogg CC, Petkova E, Haritou A. Statistical methods for comparing regression coefficients between models. *American Journal of Sociology*. 1995;100:1261–1312.

29. Gallant AR. Seemingly unrelated nonlinear regressions. *Journal of Econometrics*. 1975;3:35-50.

30. Muthen BO. Latent Variable Modeling in Heterogeneous Populations. *Psychometrika*. 1989;54(4):557-585.

**Meta-Analyses And Individual Studies Of Demonstrating Effectiveness Of Treatments To Reduce Readmission Risk**

|  |  |  |  |
| --- | --- | --- | --- |
| Meta-analyses | Population | Treatment(s) | Effect size |
| Feltner, 20148 | adults hospitalized with heart failure | 53 studies of 47 RCTs of treatments for heart failure patients. Included studies were categorized as home-visiting programs (15 RCTs), structured telephone support (STS) (13 trials), telemonitoring (8 trials), outpatient clinic-based treatments (7 trials), and primarily educational treatments (4 trials). 30 days, 3–6 months | Effect sizes reported by treatment. Only home visiting and MD led heart failure clinics had evidence of effectiveness, For home visiting programs; High intensity (1 study):  0.34 (0.19 to 0.62)  Lower intensity (1 study): 0.89 (0.43 to 1.85) for 30 days. |
| Leppin et al., 20149 | adults hospitalized for medical-surgical cause for more than 24 hrs and discharged home | 42 RCTs – any treatment included. Treatments included 1-7 activities. Most common activities were case management, patient education, home visit, and self-management support. | RR=0.82 [95% CI 0.73 to 0.91]  Treatments that augmented patient capacity for self-care, RR-.68 (95%CI 0.53 to 0.86) |
| Braet, Weltens, & Sermeus, 201610 | adults (18 years or older) discharged from a medical or surgical ward.) | 51 RCTs that evaluated discharge treatments. The included treatments must have been performed – at least partly – by hospital professionals with the intention of easing the care transition out of the hospital to home, or to prevent or alleviate problems after hospital discharge. Disease specific approaches were not considered. | RR=0.77 [95% CI, 0.70-0.84]  Exploratory subgroup analysis found that treatments starting during hospital stay and continuing after discharge were more effective in reducing readmissions compared to treatments starting after discharge (between subgroup difference p=0.01). Multicomponent treatments were not more effective compared to single component treatments (between subgroup difference p=0.54). Treatments oriented towards patient empowerment were more effective compared to all other treatments (between subgroup difference p=0.02). |
| Individual RCTs | Patient population | Treatment classification per Braet, Weltens, & Sermeus, 20168 | Effect size |
| Hansen, 199511 | Discharged from geriatric ward | Transitional Care (home visits) | RR=0.30 (95% CI 0.16-0.57) |
| Naylor, 199012 | Elderly medical-surgical | Discharge planning/Transitional Care | RR=0.27 (95% CI 0.04-0.83) |
| Naylor, 199413 | Elderly cardiac surgical | Discharge planning/Transitional Care | RR=0.35 (95%CI 0.19-0.65) |
| Naylor, 200414 | Elderly medical-surgical | Discharge planning/Transitional Care | RR=0.34 (95% CI 0.19 to 0.62) |
| McDonald, 200215 | Heart failure | Multidisciplinary | RR=0.18 (95% CI 0.04 to 0.80) |
| Parry, 200916 | Elderly | Transitional Care | RR=0.30 (95% CI 0.11 to 0.85) |
| Huang, 200517 | Hip fracture | Discharge Planning | RR=0.31 (95% CI 0.11 to 0.89) |
| Lopez-Cabeza, 200618 | Heart failure | Education | RR=0.32 (95% CI 0.16 to 0.32) |
| Wong, 201119 | Admitted to medical unit | Transitional Care | RR=0.42 (95% CI 0.26 to 0.66) |
| Evans, 199320 | VA patients at risk | Discharge Planning | RR=0.70 (95% CI 0.57 to 087) |
| Rich, 199521 | Elderly heart failure | Multidisciplinary | RR=0.71 (95% CI 0.51 to 0.97) |
| Legrain, 201122 | Older adults | Discharge Planning | RR=0.71 (95% CI 0.54 to 0.93) |
| Melton, 201223 | Patients at risk | Tele | RR=0.79 (95% CI 0.64 to 0.96) |

**Appendix References:**

1. van Walraven C, Dhalla IA, Bell C, et al. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne. 2010;182(6):551-557.
2. Sieck, Carol K., "Correlation of the Boost Risk Stratification Tool as a Predictor of Unplanned 30-Day Readmission in Elderly Patients" (2017). Dissertations. 2855. [https://ecommons.luc.edu/luc\_diss/2855](https://urldefense.proofpoint.com/v2/url?u=https-3A__ecommons.luc.edu_luc-5Fdiss_2855&d=DwMFAg&c=S1d2Gs1Y1NQV8Lx35_Qi5FnTH2uYWyh_OhOS94IqYCo&r=PWuOTl1QBJXJsRn5i67lTXlO_ZV4JFMfiDy9peBL7D4&m=Ham68OiBFwa40etpFvEQQ8PW-AJveuXmF4xlX3kaLec&s=6epHSYxetIWnbG3Y6JhwV8UTIcElNJiG4eMNi1F3hSc&e=).
3. Stavem K, Hoel H, Skjaker SA, Haagensen R. Charlson comorbidity index derived from chart review or administrative data: agreement and prediction of mortality in intensive care patients. *Clin Epidemiol*. 2017;9:311-320. Published 2017 Jun 2. doi:10.2147/CLEP.S133624
4. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, Saunders LD, Beck CA, Feasby TE, Ghali WA. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care* 2005;43(11):1130-1139.
5. Moore BJ1, White S, Washington R, Coenen N, Elixhauser A.Identifying Increased Risk of Readmission and In-hospital Mortality Using Hospital Administrative Data: The AHRQ Elixhauser Comorbidity Index. Med Care. 2017 Jul;55(7):698-705. doi: 10.1097/MLR.0000000000000735.
6. Elixhauser Comorbidity Software, Version 3.7. Agency for HealthCare Research and Quality; 2018. [https://www.hcup-us.ahrq.gov/toolssoftware/comorbidity/comorbidity.jsp](https://urldefense.proofpoint.com/v2/url?u=https-3A__www.hcup-2Dus.ahrq.gov_toolssoftware_comorbidity_comorbidity.jsp&d=DwMFAg&c=S1d2Gs1Y1NQV8Lx35_Qi5FnTH2uYWyh_OhOS94IqYCo&r=PWuOTl1QBJXJsRn5i67lTXlO_ZV4JFMfiDy9peBL7D4&m=Ham68OiBFwa40etpFvEQQ8PW-AJveuXmF4xlX3kaLec&s=USJ7auN67SIrP0Hzt1m8gfQ6ceGgwfZb9rk0Y4OXypQ&e=).
7. Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: a 7 systematic review. JAMA. 2011;306(15):1688-1698.
8. Feltner C, Jones CD, Cene CW, et al. Transitional care treatments to prevent readmissions for persons with heart failure: a systematic review and meta-analysis. Annals of internal medicine. 2014;160(11):774-784.
9. Leppin AL, Gionfriddo MR, Kessler M, et al. Preventing 30-day hospital readmissions: a systematic review and meta-analysis of randomized trials. JAMA internal medicine. 2014;174(7):1095-1107.
10. Braet A, Weltens C, Sermeus W. Effectiveness of discharge treatments from hospital to home on hospital readmissions: a systematic review. JBI database of systematic reviews and implementation reports. 2016;14(2):106-173.
11. Hansen FR, Poulsen H, Sorensen KH. A model of regular geriatric follow-up by home visits to selected patients discharged from a geriatric ward: a randomized controlled trial. *Aging (Milan, Italy).* 1995;7(3):202-206.
12. Naylor MD. Comprehensive discharge planning for hospitalized elderly: a pilot study. Nurs Res. 1990; 39(3): 156-61. 67. Naylor M, Brooten D, Jones R, Lavizzo-Mourey R, Mezey M, Pauly M. Comprehensive discharge planning for the hospitalized elderly: a randomized clinical trial. Ann Intern Med. 1994; 120(12): 999-1006.
13. Naylor MD, Brooten DA, Campbell RL, Maislin G, McCauley KM, Schwartz JS. Transitional care of older adults hospitalized with heart failure: a randomized, controlled trial. Journal of the American Geriatrics Society. 2004;52(5):675-684.
14. Naylor MD, Brooten DA, Campbell RL, Maislin G, McCauley KM, Schwartz JS. Transitional care of older adults hospitalized with heart failure: a randomized, controlled trial. Journal of the American Geriatrics Society. 2004;52(5):675-684.
15. McDonald K, Ledwidge M, Cahill J, Quigley P, Maurer B, Travers B, et al. Heart failure management: Multidisciplinary care has intrinsic benefit above the optimization of medical care. J Card Fail. 2002; 8(3): 142-8.
16. Parry C, Min SJ, Chugh A, Chalmers S, Coleman EA. Further application of the care transitions treatment: results of a randomized controlled trial conducted in a fee-for-service setting. Home Health Care Serv Q. 2009; 28(2-3): 84-99.
17. Huang T, Liang S. A randomized clinical trial of the effectiveness of a discharge planning treatment in hospitalized elders with hip fracture due to falling. J Clin Nurs. 2005; 14(10): 1193-201.
18. Lopez Cabezas C, Falces Salvador C, Cubi Quadrada D, Arnau Bartes A, Ylla Bore M, Muro Perea N, et al. Randomized clinical trial of a postdischarge pharmaceutical care program vs. regular follow-up in patients with heart failure. Farm Hosp. 2006; 30(6): 328-42.
19. Wong FK, Ho MM, Yeung S, Tam SK, Chow SK. Effects of a health-social partnership transitional program on hospital readmission: a randomized controlled trial. Soc Sci Med. 2011; 73(7): 960-9.
20. Evans RL, Hendricks RD. Evaluating hospital discharge planning - a randomized clinical-trial. Med Care. 1993 Apr; 31(4): 358-70.
21. Rich MW, Beckham V, Wittenberg C, Leven CL, Freedland KE, Carney RM. A multidisciplinary treatment to prevent the readmission of elderlypati ents with congestive heart failure. *N Engl J Med*. 1995;333(18):1190-1195.
22. Legrain S, Tubach F, Bonnet-Zamponi D, Lemaire A, Aquino JP, Paillaud E, et al. A new multimodal geriatric discharge-planning treatment to prevent emergency visits and rehospitalizations of older adults: The optimization of medication in aged multicenter randomized controlled trial. J Am Geriatr Soc. 2011; 59(11): 2017-28.
23. Melton LD, Foreman C, Scott E,McGinnisM, Cousins M. Prioritized post-discharge telephonic outreach reduces hospital readmissions for select high-risk patients. *Am J Manag Care*. 2012;18(12): 838-844.