Prediction and Management of Readmission Risk for Congestive Heart Failure

Senjuti Basu Roy[1] Si-Chi Chin[1]
[1] *Center for Web and Data Science, Institute of Technology, The University of Washington - Tacoma, 1900 Commerce Street, Tacoma, WA 98402-3100, U.S.A.*
{*senjutib, scchin*}*@uw.edu*

Keywords: Hospital Readmission Risk Prediction, Readmission Risk Management, Predictive Modeling

Abstract: This position paper investigates the problem of 30-day readmission risk prediction and management for Congestive Heart Failure (CHF), which has been identified as one of the leading causes of hospitalization, especially for adults older than 65 years. The underlying solution is deeply related to using predictive analytics to compute the readmission risk score of a patient, and investigating respective risk management strategies for her by leveraging statistical analysis or sequence mining techniques. The outcome of this paper leads to developing a framework that suggests appropriate interventions to a patient during a hospital stay, at discharge, or post hospital-discharge period that potentially would reduce her readmission risk. The primary beneficiaries of this paper are the physicians and different entities involved in the pipeline of health care industry, and most importantly, the patients. This paper outlines the opportunities in applying data mining techniques in readmission risk prediction and management, and sheds deeper light on healthcare informatics.

1 INTRODUCTION

Early readmission is a profound indicator of the *quality of care* provided by the hospitals. The Centers for Medicare & Medicaid Services (CMS) recently began using readmission rate as a publicly reported quality metric. The estimated cost of unplanned readmission was 17.9 billion in 2004 (Jencks et al., 2009), and more than 27% of them were considered avoidable (Walraven et al., 2011). Readmission can result from a variety of reasons, including early discharge of patients, improper discharge planning, and poor care transitions. Studies have shown that appropriate interventions during the hospital stay, during or post-discharge plans like home based follow up, and patient education can improve the health outcome of the patients and reduce the readmission likelihood, especially in elderly patients, and decrease the overall medical costs (Naylor MD, 1999; Rich et al., 1995; Schneider et al., 1993; Phillips CO, 2004; Koelling et al., 2005).

To that end, this position paper focuses on *investigating analytical techniques to mitigate the readmission risk of the individuals and improve their overall health outcome*. While our vision is generic and applicable to any disease, for the purpose of illustration, our primary investigation hinges on Congestive Heart Failure (CHF), in particular. CHF is one of the leading causes of hospitalization, and studies show that many of these admissions are readmissions within a short window of time. Based on the 2005 data of Medicare beneficiaries, it has been estimated that 12.5% of Medicare admissions due to CHF were followed by readmission within 15 days, accounting for about $590 million in healthcare costs (Krumholz HM, 1997). More specifically, this paper emphasizes the 30-day readmission problem for CHF, as this time window is considered *clinically meaningful* by different healthcare services and standards. Identifying patients who have greater risk of readmission can guide implementation of appropriate interventions to prevent these readmission. The primary objective of this work is to investigate a *comprehensive framework to address the problem of readmission risk prediction and management* for CHF. We summarize the state of the art on the general problem of CHF risk prediction, identify their limitations, and outline future opportunities.

Readmission is common and costly. It can result from a variety of reasons, including early discharge of patients, improper discharge planning, and poor care transitions. Appropriate interventions or pre-discharge planning (Schneider et al., 1993) , and post discharge plans like home based follow up (Naylor MD, 1999) and patient education (Koelling et al., 2005) can reduce the readmission rates considerably and improve the health outcome of the patients. In particular, studies have shown that targeted interventions during the hospital stay, during or post-discharge, can reduce the readmission likelihood, especially in elderly patients, and decrease the overall medical costs (Naylor MD, 1999; Rich et al., 1995). However, to the best of our knowledge, there does not exist any effort that proposes an iterative analytical framework to predict 30-day readmission risk for CHF and to recommend appropriate personalized intervention strategies to the patients at different phases, which is the primary focus of our work.

This paper has two primary objectives: 1) first, summarize current research on predicting the 30-day readmission risk score (or percentage) of CHF patients; 2) second, outline the challenges and opportunities in designing personalized intervention strategies to assist decision making, such that the readmission risk of the patients gets reduced by a certain percentage. While the former task is *risk prediction*, the latter objective is referred to as *risk management*. Interestingly, suggesting appropriate interventions is tightly integrated with a patient's current phase – for example, the post-discharge interventions may only be limited to appropriate follow-ups or patient education, while physicians could suggest different procedures or surgery, if the intervention is being administered during her hospital stay. Existing research has studied different clinical risk prediction problems in silos. This paper is one of the first efforts to study the risk prediction and management problem in conjunction.

Our vision is to design the solution strategies using statistical and data mining techniques. For example, the risk prediction problem could be designed as a statistical classification or a regression (Han and Kamber, 2006) problem, with the objective to learn a mathematical function that correctly outputs the associated probability of an individual's 30-day risk of readmission, or correctly outputs the actual number of days until the next readmission will take place for the individual, using different factors that causes CHF readmission. On the other hand, the risk management problem could also be studied using statistical and data mining techniques. Given a set of possible interventions, if the risk prediction problem is designed as a regression problem, then to achieve a targeted (lower) risk score, the risk management problem could be solved by: 1) learning the reverse regression or *calibration* (Johnson and Wichern, 1988) of the intervention parameters that results in the intended targeted risk score, and 2) performing sequence mining (Han and Kamber, 2006) to suggest appropriate interventions.

2 Risk Predictions

In this section, we summarize current studies in predicting risk of hospital readmission and discuss the limitations of the field.

2.1 Applying Extensive Data Preprocessing to Improve Quality of Prediction

The quality of data determines the quality of predictions. This paper suggests to incorporate a wide variety of data preprocessing and predictive modeling techniques, i.e. missing value imputation, clustering, and classification (Han and Kamber, 2006), for improving the prediction of 30-day readmission risk for CHF patients. Real world clinical data are noisy and heterogeneous in nature, severely skewed, and contain hundreds of pertinent factors. They contain information on patients' socio-demographical characteristics, such as marital status and ethnicity; clinical data such as diagnosis, discharge information; and comorbidity factors[1]; other cost related factors pertaining to a particular hospital admission; lab results; procedures. The proposed solution relies on data mining and predictive analytics. Current research has investigated a wide range of techniques to that end – starting from simple Naive Bayes' Classifier, Support Vector Machine (Zolfaghar et al., 2013b; Zolfaghar et al., 2013a), Regression models (Kansagara D, 2011), to Ensemble of Multilayer classifier (Zolfaghar et al., 2013c).

A sophisticated and effective predictive model often requires a large set of attribute values that may not all be available (or known) at the time when a patient or a healthcare provider uses the risk assessment tool. To transform the limited inputs to the complete set of attribute values on which the predictive model is trained, the first task of the risk prediction model is to map the input values to a group (i.e., cluster) of patients who are most similar to the provided user profile. The model pre-computes the clusters based on different permutations of input attributes using the *k*-mode algorithm (Han and Kamber, 2006). To accommodate all possible scenarios, the model constructs $k * 2^n$ clusters, where n is the number of factors (i.e.,

[1]Comorbidities are specific patient conditions that are secondary to the patient's principal diagnosis and that require treatment during the stay.

attributes) used in the predictive model and k is the predetermined number of clusters. The model efficiently matches the given set of input attribute values to its closest cluster centroid using the index-based matching algorithm (Zolfaghar et al., 2013a). The pre-computed cluster centroids are used to *complete* the remaining missing attribute values.

Then the problem of computing risk score is studied as a classification problem, or as a regression problem. The primary objective is to design a mathematical function that learns the relationship between the attributes and the risk score. The relationship could be linear or non-linear. The risk could be some categories (High risk, Medium Risk, Low Risk), or it could be a positive real number. Given n factors, the task is to learn the function, which generates the appropriate risk score y.

$$y = f(x_1, x_2, \ldots, x_n)$$

2.2 Limitations of Current Risk Prediction Research

There are two primary limitations of current research on risk of readmission predictions. The first is the lack of intervention recommendations based on the results of risk predictions. The second limitation is the lack of customization of care management strategies for individual patients and for different healthcare systems. While most research has invested largely in predicting risk of readmission, these aspects are scarcely unexplored.

The value of risk prediction is massively limited if there is no guidance about selecting proper interventions to manage the risk. The knowledge of prediction results is only actionable if predictive modeling also enables the decision making about the interventions. Extensive number of studies investigate risk factors contributing to risk of readmission (Kansagara D, 2011). However, many primary factors, such as age, gender, and prior hospitalization history, can not be manipulated. The knowledge that age contributes to higher risk of readmission is not actionable since it is impossible to alter one's age to reduce the readmission risk. Therefore, predictive modeling should aim to enable *intervening* on different stages of care: upon admission, during the hospitalization, and at discharge.

Currently, most care managers in the hospitals adopt a holistic approach to intervene based on simple categorization of patients (e.g. high risk versus low risk patients). However, intervention strategies could be more effective if they were tailored to individual patients. For example, while patient have no family or other caregiver at home may benefit from home care and home visit, other patients may benefit more from medication review and prescriptions to lower their sodium level. Furthermore, intervention strategies should also be customized based on the needs of different hospitals and healthcare systems. For example, while some hospital may wish to over predict the number patients with risk of readmission in order to decrease the readmission rate, other hospitals may aim at precise predictions to efficiently allocate care resource and can afford to tolerate some instances of false negatives. Therefore, the problem of risk management intervention prediction and recommendation should be studied in conjunction with the cost of interventions and available resource in a hospital or a healthcare system.

The limitations provide the opportunities to expand the scope of the applying predictive modeling to the problem of readmission reduction. We outline some possible solutions to address the first limitation in Section 3.

3 Our Vision: From Risk Prediction to Risk Management

This section describes our vision of applying statistical and data mining techniques for readmission risk management. We first present a framework that incorporates predictive modeling to risk management feedback loop. We then suggest two methods – calibration and sequential mining – to design a sequence of interventions for risk management.

3.1 Incorporating Predictive Modeling to Risk Management Feedback loop

The objective of risk management is to develop appropriate medical interventions and care management strategies to reduce the risk score of an individual. Interventions could take place during hospitalization, at discharge time, or post-discharge. Patients should be treated and/or reached in a unique way in order to minimize the risk score. From the factors described above, our proposed solution identifies and selects a subset of intervention factors that are *actionable*, either during hospitalization, at time of discharge, or post discharge. Additionally, the model considers post-discharge interventions, including medication review and counseling by clinical pharmacist, dietary and social service consultation, coordination of home care and home visits, follow up with patients via telephone, use of tele-health in patient care, etc.

In Figure 1, steps A-J illustrate how analytics can

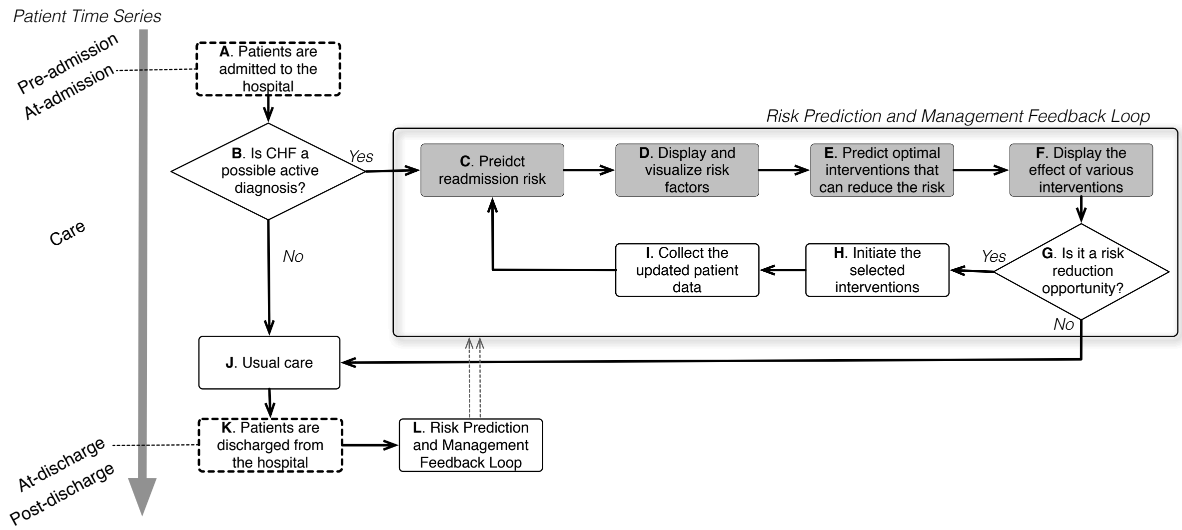

Figure 1: Feedback loop for predicting and managing risk of readmission. The shaded steps describe how the project fits in the process of care.

be integrated into the feedback loop in clinical practice to reduce risk of readmission. After identifying admitted patients with CHF (steps A,B), we first estimate the risk of readmission and analyze risk factors for patients of different characteristics (steps C, D). Then, our analysis also suggests interventions that can reduce the risk and their relative effect on risk reduction (steps E, F). Healthcare providers may then determine whether to initiate interventions and capture the outcome of the interventions (steps G-I). The outcome of the interventions produces a new risk estimate (step C). Given the updated risk assessment (steps C-F), healthcare provider could then re-evaluate whether the risk is minimized and whether further opportunity for risk reduction exists (step G). This way, the proposed framework contributes to the prediction and management of readmission risk for CHF patients.

3.2 Predictive Modeling Solutions to Risk Management

This section describes the adopted procedure to solve the risk prediction and the management problem, and the evaluation methods of the solutions. Imagine that the risk score of a patient is expressed as a function of different interventions. Table 1 describes three imaginary patient records with only five interventions. To manage patients' 30-day readmission risk, our overall process relies on understanding the respective risk score first, followed by learning the best interventions to reduce it.

As discussed in Section 2, the risk prediction problem could be solved using many different predictive analytics techniques. For the purpose of illustration, here, we consider a specific and effective model, namely multiple linear regression (Han and Kamber, 2006) and describe the risk management methodology using that.

A general additive multiple linear regression model relates the risk score of a patient as the dependent variable or outcome (y), to a set of k independent or predictor variables (i.e., interventions), $x_1, x_2, \ldots, x_k$. The model is expressed by the equation

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \quad (1)$$

From the patient data, using $x_1, x_2, \ldots\ldots, x_k$ and the observed risk score, multiple linear regression learns the coefficients $\alpha, \beta_1, \ldots, \beta_k$. That is, the objective is to learning the dependency between the independent variables and the outcome. Given a patient and a set of possible interventions, the task is, which subset of interventions to administer, such that her readmission risk score gets reduced (e.g., to 45%)? We propose two different techniques to that end.

3.2.1 Inverse Regression/Calibration

Given a regression equation such as Equation 1, the intended reduced risk score y' (e.g., to 45%), the coefficients $\alpha, \beta_1, \ldots, \beta_k$ and the function is known. The objective is to output the value of the independent variables $x_1, x_2, \ldots\ldots, x_k$ that results in y'. Note that this is inverse regression or the "multivariate calibration"(Johnson and Wichern, 1988; Hardin et al., 2003) problem in statistics. The proposed research

Ejection Fraction	Blood Pressure	Depression	Diabetes	Cardiac CT	Risk Score (y)
<40	160/100	Yes	Yes	No	75%
58	140/90	No	Yes	Yes	55%
51	170/110	Yes	Yes	No	87%

Table 1: A set of 3 simple patient data with 5 interventions

applies the Bayesian Approach for calibration. Equation 1 above for the n-sample calibrating data can be written as,

$$Y = B'X + E$$

where $Y(nx1)$ is the response matrix, $X(nx(k+1))$ is a fixed matrix of dependent factors, and B are the coefficients, and E is the error matrix.

A particular risk outcome follows the same assumption and could be expressed as,

$$y_0 = B'X^* + \alpha_0$$

Finding appropriate interventions here is analogous to finding the marginal posterior distribution of X^*. To do that, the process first calculates the joint prior, expressed as,

$$p(B, \Sigma, X^*) = p(B, \Sigma)p(X^*)$$

Furthermore, it is assumed that $p(X^*|Y) = p(X^*)$. Then the posterior distribution of X^* could be expressed by the joint density function. We omit further details for brevity, and refer to (Plessis and Merwe, 2004) for detailed discussion. The output of this method outputs the different predictors (i.e., interventions) which give rise to the desirable lower risk score.

3.2.2 Sequence Mining

An alternative investigation is to study the problem from the sequential pattern mining (Han and Kamber, 2006) perspective. The objective is to output the appropriate sequence of the interventions which leads to a desirable lower risk score. Many frequent itemset mining algorithms (such as Apriori or FP-Growth (Han and Kamber, 2006)) with appropriate adaptations could be applied to solve the problem. The basic idea is to treat patient history (such as provided in Table 1) and consider frequency of co-occurrence of the interventions (considering their ordering) and the outcomes (i.e., associated risks) to generate *association rules* between the likely outcome and the suggested interventions. The output would generate rules of the form "if a patient is treated and cured for *diabetes* followed by *depression*, and then her *Ejection fraction = 58*", she is likely to have risk score 45%.

Trade-offs exist between the two suggested methods described above. The former strictly relies on a specific function and probability distribution assumption to estimate the posterior. It introduces challenges as real world patient data exhibits severe randomness over the time. The latter is computationally expensive, especially when the attributes are continuous and the generated rules are required to have orders, such as ours. We explore both these paths and choose the winning approach for validations.

4 Conclusion

This position paper investigates the problem of 30-day readmission risk prediction and management for Congestive Heart Failure (CHF), which has been identified as one of the leading causes of hospitalization. Although current research has demonstrated the effort of predicting risk of readmission in silos, those solutions are mostly not applicable to design intervention strategies for personalized risk management. In this position paper, we envision that a horizon of opportunity could be unveiled, if these two problems are studied in conjunction and in an iterative manner. We believe that the solutions could be largely adapted from the computing domain, in particular by applying data mining and statistical analysis techniques. We note that novel solutions could be designed with such adaptations to support clinical and care management decision making processes to reduce the risk of readmission and improve the quality of care.

REFERENCES

Han, J. and Kamber, M. (2006). *Data mining: concepts and techniques*. Morgan Kaufmann.

Hardin, J. W., Schmeidiche, H., and Carroll, R. J. (2003). The regression-calibration method for fitting generalized linear models with additive measurement error. *Stata Journal*, 3(4):361–372.

Jencks, S. F., Williams, M. V., and Coleman, E. A. (2009). Rehospitalizations among patients in the medicare fee-for-service program. *New England Journal of Medicine*, 360(14):1418–1428.

Johnson, R. A. and Wichern, D. W., editors (1988). *Applied multivariate statistical analysis*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Kansagara D, E. H. (2011). Risk prediction models for hospital readmission: A systematic review. *JAMA*, 306(15):1688–1698.

Koelling, T. M., Johnson, M. L., Cody, R. J., and Aaronson, K. D. (2005). Discharge education improves clinical outcomes in patients with chronic heart failure. *Circulation*, 111(2):179–185.

Krumholz HM, P. E. (1997). Readmission after hospitalization for congestive heart failure among medicare beneficiaries. *Archives of Internal Medicine*, 157(1):99–04.

Naylor MD, B. D. (1999). Comprehensive discharge planning and home follow-up of hospitalized elders: A randomized clinical trial. *JAMA: The Journal of the American Medical Association*, 281(7):613–620.

Phillips CO, W. S. (2004). Comprehensive discharge planning with postdischarge support for older patients with congestive heart failure: A meta-analysis. *JAMA: The Journal of the American Medical Association*, 291(11):1358–1367.

Plessis, J. L. D. and Merwe, A. J. V. D. (2004). Inferences in multivariate bayesian calibration. *JSTOR*.

Rich, M. W., Beckham, V., Wittenberg, C., Leven, C. L., Freedland, K. E., and Carney, R. M. (1995). A multidisciplinary intervention to prevent the readmission of elderly patients with congestive heart failure. *New England Journal of Medicine*, 333(18):1190–1195.

Schneider, J. K., Hornberger, S., Booker, J., Davis, A., and Kralicek, R. (1993). A medication discharge planning program measuring the effect on readmissions. *Clinical Nursing Research*, 2(1):41–53.

Walraven, C. v., Bennett, C., Jennings, A., Austin, P. C., and Forster, A. J. (2011). Proportion of hospital readmissions deemed avoidable: a systematic review. *Canadian Medical Association Journal*, 183(7):E391–E402.

Zolfaghar, K., Agarwal, J., Sistla, D., Chin, S.-C., Roy, S. B., and Verbiest, N. (2013a). Risk-o-meter: an intelligent clinical risk calculator. In *KDD*, pages 1518–1521.

Zolfaghar, K., Meadem, N., Sistla, D., Chin, S.-C., Roy, S. B., Verbiest, N., and Teredesai, A. (2013b). Exploring preprocessing techniques for prediction of risk of readmission for congestive heart failure patients. In *Data Mining and Healthcare Workshop*.

Zolfaghar, K., Verbiest, N., Agarwal, J., Meadem, N., Chin, S.-C., Roy, S. B., Teredesai, A., Hazel, D., Amoroso, P., and Reed, L. (2013c). Predicting risk-of-readmission for congestive heart failure patients: A multi-layer approach. *CoRR*, abs/1306.2094.