Predicting unplanned hospital readmissions

Martin H RENNHACKKAMP [a] and Assoc Prof Graeme K HART [b]
[a] *Chief Data Scientist, PBT Australia*
[b] *Director Dept Intensive Care, Austin Health*

Abstract. The number of unplanned readmissions is an important quality improvement indicator for Australian hospitals, as it has a direct impact on costs and patient outcomes. It is a useful indicator for patient multi-morbidity to identify patients that should potentially have different treatment or discharge plans. The objective of the Health Insights Challenge project was to develop a predictive model to determine the likelihood of an unplanned hospital readmission.

The outcome was a validated analytical model that predicts the likelihood of an unplanned hospital readmission occurring within 30 days after a given in-patient episode. An 'all-cause' model does not pay attention to any specific disease or diagnosis groups. The predicted likelihood of an unplanned readmission occurring can assist care workers and clinicians in planning treatment and discharge processes.

Keywords. Patient readmission, patient admission, patient discharge, multimobidity, inpatient, episode of care, predictive analytics, hospital analytics.

Introduction

This paper describes the development, validation and outcomes of a predictive model that determines the likelihood of unplanned patient readmissions occurring within 30 days of discharge. It identities high risk inpatients for intervention during admission, hospitalisation or after discharge. It concludes with a discussion on the key learnings, alignment with the project's objectives, the impact of its potential implementation and recommendations for future work.

This work was performed as part of the Health Insights Challenge – a collaborative project between Prescient Business Technologies (PBT Australia), EntityThree, the Austin Health group of hospitals, the Department of Information Systems and Business Analytics at Deakin University and the Australian Centre for Health Innovation.

1. Background

Austin Health is a public provider of tertiary health services, education and research in Melbourne. It operates 980 beds across acute, sub-acute and mental health facilities [1]. During the 2014 financial year (1 July – 30 June), Austin Health had 95,142 inpatient admissions, 177,027 outpatient attendances and 75,366 emergency presentations respectively.

1.1. Problem definition

Austin Health's objectives for predicting the likelihood of 30-day unplanned readmissions were to:

- identify the most likely patient groups for 30-day unplanned readmissions;
- identify the drivers of 30-day unplanned hospital readmission;
- redesign the admission, treatment and discharge planning processes; and
- predict the impact of avoidable bed days on costs, throughput and waiting lists.

Austin Health uses the concept of a spell, which is a continuous series of inpatient episodes, where episodes may be separated by statistical separations to account for different cost centres.

Unplanned readmissions are measured in Victoria as a hospital outcome indicator by the Australian Commission on Safety and Quality in Health Care [3]. Austin Health defines an unplanned readmission based on the Dr Foster business rules [4]. It occurs when a patient with an emergency admission at a hospital on or before the 30th day after discharge is admitted to acute care where either the principal diagnosis for the readmission is related to the previous admission, or the previous admission has a surgical diagnosis code for any episode. Regular attenders (e.g. haemodialysis, chemotherapy, radiotherapy and mental health) are not considered unplanned readmissions.

1.2. Literature review

Preliminary research identified 294 relevant papers on 30-day all-cause readmissions. Five relevant papers, clearly identifying the variables, the type of model and the model validation approach, were used as reference case studies, as summarised in table 1 [5],[6],[7],[8],[9]. They used mostly administrative data, however, they were performed on different datasets of varying numbers of patients and episodes, and over different time periods.

Table 1. Literature survey summary

	Case Study 1	Case Study 2	Case Study 3	Case Study 4	Case Study 5
Authors	Shulan, M., Gao, K., and Moore, C.	Billings, J., Blunt, I., Steventon, A., Georghiou, T., Lewis, G., and Bardsley	Donzé, J., Aujesky, D., Williams, D., and Schnipper, J.L.	Choudhry, S.A., Li, J., Davi, D., Erdmann, C., Sikka, R., and Sutariya, B.	Robin, L.K., Harlen, D.H., Richard, W.M., Matthew, F.E., Douglas, S.W., and David, R.M.
Publication title	Predicting 30-day all-cause hospital readmission	Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30)	Potentially Avoidable 30-Day Hospital Readmissions in Medical Patients	A Public-Private Partnership Develops and Externally Validates a 30-Day Hospital Readmission Risk Predictive Model	Risk Factors for All-Cause Hospital Readmission Within 30 Days of Hospital Discharge
Significant Variables	87	88	7	49	5
Model(s) used	Logistic regression using stepwise refinement	Logistic regressions	Backward multivariable logistic regression	Logistic regressions	Logistic regression

The variables used in these models were compared to the data from Austin Health and a large correspondence gave confidence that Austin Health's dataset was suitable for this project.

2. Methodology

The project followed the SEMMA (Sample, Explore, Modify, Model and Assess) framework developed by SAS Institute [11].

Figure 1 illustrates SEMMA as embedded within the predictive modelling lifecycle used in this project. Relevant details from selected steps follow below.

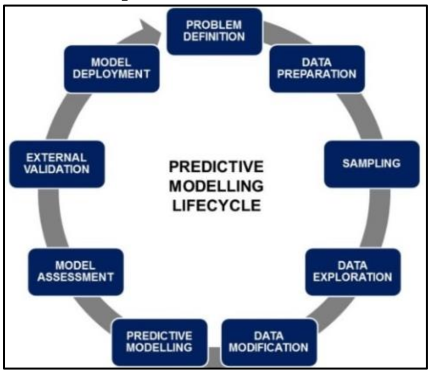

Figure 1: Predictive Modelling Lifecycle

2.1. Data preparation

Austin Health provided de-identified data for a 5-year period (2009 to 2014) depicted in Table
Table 2: Data Extracts (1 July 2009 to 31 October 2014):

Table 2: Data Extracts (1 July 2009 to 31 October 2014)

No.	Activity Data Tables	No. of Records
1	Emergency Department	382,936
2	Inpatient Spells	291,068
3	Inpatient Episodes	308,069
4	Inpatient Comorbidities	308,069

The following data preparation steps were performed:

- Episodes and spells were joined to allocate the spell attributes across the episodes.
- Episodes resulting from unplanned readmissions were linked to the indexed episodes, to enable the calculation of impact in terms of bed days and direct hospitalisation costs.
- Lookup tables were joined in to enable analysis and reporting on descriptive attributes.
- The number of Emergency Department presentations since the previous inpatient episode were determined and added as a new field on each inpatient episode record.
- A binary target variable for model training was added to indicate the episodes resulting in 30-day unplanned readmissions according to the Dr Foster's business rules [4].

This resulted in a dataset comprising 141 variables and 280,078 episodes, with the proportion of unplanned readmissions (4.27%) to non-readmitting episodes (95.73%) highly skewed.

2.2. Data exploration

An extensive data exploration exercise was performed, the results of which will be published elsewhere. The most important findings were:

- A majority of the patients having unplanned readmissions were between 50 – 89 years old, were diagnosed with chronic obstructive pulmonary disease (COPD), heart failure, ascites and other diagnoses affecting the heart, lungs and liver, and they had many and frequent inpatient episodes as well as many presentations at the emergency department.
- Many of the patients having unplanned readmissions had high Charlson scores, which is a measure of comorbidity [10].
- Many anecdotal indicators for unplanned readmissions were investigated, such as care at home, hospital in the home days, method of transport, interpreter required and so on, but the data exploration did not surface any positive evidence that these were relevant.

2.3. Predictive modelling

SAS Enterprise Miner was used to develop the models [11]. In order to provide opportunities for intervention, the models were trained and evaluated on the data available at admission. Stepwise, Forward and Main Effects regression, and one Decision Tree models were built. Using lift and the Receiver Operator Characteristic (ROC) index of the validation dataset as selection criteria, the Forward Regression model was the best performing model. It was named the Predictive model for All-Cause unplanned Readmissions (PACR).

Twenty variables were significant in the model, as illustrated in Table 3, calculated using a decision tree algorithm [12]. The top three predictors were counts of emergency presentations, in-hospital spells per lifetime and the accumulated length of stay over 5 years. Other important predictors include the admission ward, the Charlson Score and the admission unit.

Table 3: Predictor variables

Predictor Variables	Period
No of ED presentations	5 years
Total No of IP Spells	Lifetime
Total Length of Stay (TLOS) Days	5 years
Admission Ward	Episode
Charlson Score at Admission	Episode
Admission Unit	Episode
Length of Stay (LOS) Days	Episode
Length of Stay Type	Episode

Age	Episode
Renal Disease Flag at Admission	Episode
Mild Liver Disease Flag at Admission	Episode
Malignancy Disease Flag at Admission	Episode
COPD Disease Flag at Admission	Episode
Moderate – Severe Liver Disease Flag at Admission	Episode
No of Spells	5 years
Admission Source	Episode
Care Type	Episode
Cerebrovascular Disease Flag at Admission	Episode
Non Chronic Disease Flag at Admission	Episode
Dementia Disease Flag at Admission	Episode

2.4. Model assessment

The gains chart shows the percentiles of the patient episodes, ranked by their likelihood to have an unplanned readmission. As depicted in Figure 2, the Cumulative Captured Response Rate depicts the number of unplanned readmissions that were correctly identified at each percentile. At the 35th percentile, almost 75% of predicted readmissions were identified. Comparatively, a model with no predictive power (i.e. a random model) would be expected to predict approximately 5% of events in each percentile. The cumulative lift at the 5th percentile shows the model is four times more likely to correctly predict unplanned readmissions, as compared to random chance.

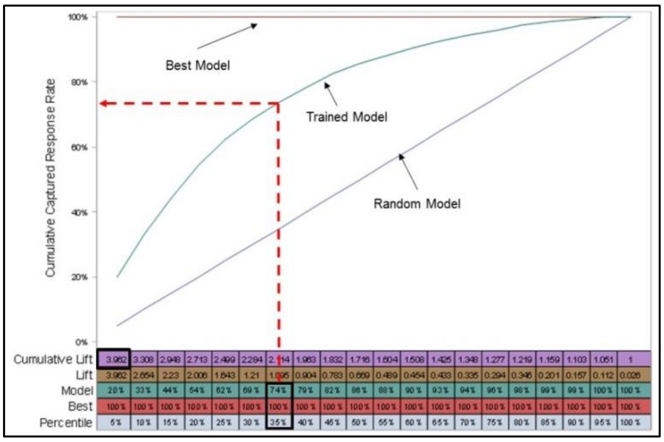

Figure 2: Model Gains Chart – Cumulative Captured Response Rate by Decile

For the models identified in the literature review studies, only the ROC statistic was published to give an indication of model performance [5],[6],[7],[8] and [9]. The PACR model came in second when benchmarked against this metric (Table 4). Note that this is only an indicative comparison; as the models were not evaluated against the same dataset, nor for the same period.

Table 4: Model Comparison

Model	ROC	Predictive ability
Predicting 30-day all-cause hospital readmission [5]	0.80	High
PACR model	0.77	Moderate
A Public-Private Partnership Develops and Externally Validates a 30-Day Hospital Readmission Risk Predictive Model [8]	0.76	Moderate
Potentially Avoidable 30-Day Hospital Readmissions in Medical Patients [7]	0.71	Moderate
Risk Factors for All-Cause Hospital Readmission Within 30 Days of Hospital Discharge [9]	0.70	Moderate
Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30) [6]	0.67	Low

2.5. External validation

To avoid bias and to ascertain the model's discriminative ability in an actual operational setting, PACR was tested on an out-of-sample dataset, which consisted of data for the first eight months of the 2015 financial year. It consisted of 41,666 inpatient episode records, with their corresponding emergency department presentations. As depicted in Table 5, the model had a marginally higher specificity (74%), as compared to its sensitivity (70%) for this dataset.

Table 5: Key Model Assessment Metrics

Key Metrics	%	Description
Sensitivity	70%	Probability of identifying patients at risk of unplanned readmissions within 30 days of discharge (event)
Specificity	74%	Probability of identifying patients not at risk of unplanned readmissions within 30 days of discharge (non-event)
Accuracy	74%	Accurate classification of ‘events’ and ‘non-events’ as a % of total episodes
Overall Error	26%	Misclassification rate

2.6. Model deployment

Patients predicted to have unplanned readmissions were categorised using a risk stratification indicator. The top 1/6 were scored as highly likely, the next 1/3 as medium likely and the remaining 1/2 with a low likelihood, based on their predicted likelihood, as illustrated in Figure 3.

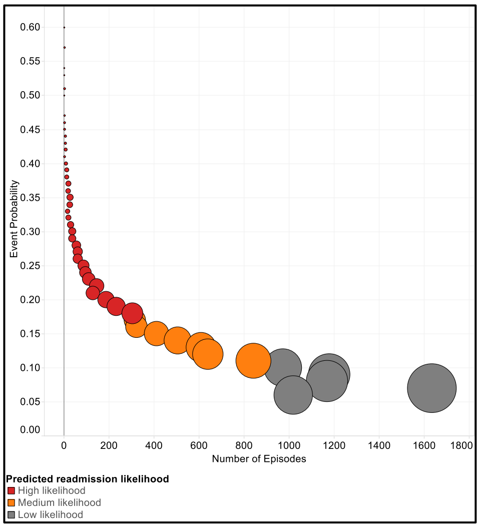

Figure 3: Readmission likelihood stratification

Using this stratification, clinicians can prioritise their attention on the predicted episodes. **Error! Reference source not found.** illustrates the impact of paying attention to each episode predicted to be succeeded by an unplanned readmission in the 8-month validation dataset, using the likelihood stratification introduced in Figure 3. Between 955 patients highly likely to have unplanned readmissions, they had a total of 1747 episodes over the 8 month period. By focusing on an average of 7 cases per day, the highly likely to readmit patients can all be considered for specialised treatment or planning. Should these interventions be successful, 5,509 bed-days of resulting readmissions could potentially be saved (shown as the “readmission LOS days” in figure 4). This represents a theoretical potential saving as not all such interventions would necessarily be successful.

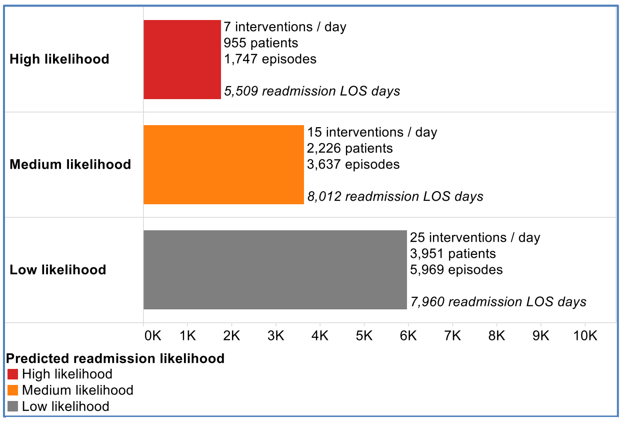

Figure 4: Potential impact

3. Discussion

A sensitivity score of 70% indicates that the model has a relatively good discriminative ability for predicting unplanned readmissions – even when only using administrative data.

The inclusion of Charlson comorbidity score at admission is an interesting point. As an indicator of disease complexity, it significantly contributes to the model.

3.1. Alignment with objectives

The most likely patients for 30-day unplanned readmissions were identified through data exploration.

The most likely to readmit episodes were identified through the significant variables of the model. The model outputs indicate which values or value bands are most significant [12].

Discussions are underway with Austin Health regarding field trials to utilise the model outputs to redesign admission, treatment and discharge planning processes.

The number of potentially avoidable bed days and their direct hospital costs were shown. These form the basis of the business case to justify the redesign of treatment and discharge processes. A quick simulation study also showed the impact on throughput and waiting lists.

3.2. Model's potential impact

Unplanned readmission is a key quality improvement indicator. Managing highly likely unplanned readmissions can reduce their related costs; increase the capacity for new patients needing care; streamline processes to improve operations; provide higher and more targeted quality of care for patients requiring care; and indirectly improve the hospital's reputation in the industry.

3.3. Recommended future work

Validating the model on an out-of-sample dataset has given good insight on the model's predictive ability and its potential benefits in an operational setting. The next step would be to validate the effectiveness of the model through an on-site field trial at the Austin Health hospitals.

Prediction models that use the Charlson comorbidity index need to be compared to models that do not, as applied to the same datasets. Many Australian hospitals do not use Charlson scores.

Although 30 days is an accepted metric in Victoria, some hospitals and health insurers elsewhere use 28 days. The impact of the length of the interval before an unplanned readmission on the model's outcomes should also be investigated. Readmissions that occurred just outside the 30 day cut-off are also "medically and financially speaking" just as worthy of intervention.

A thorough simulation study will better show the impact that reducing unplanned readmissions have on patient throughput and waiting times.

Additional improvements could further boost the model's discriminative ability, for example, by incorporating additional clinical and pharmacy variables.

4. Acknowledgements

Thank you to executive sponsors Mark Petty, Fiona Webster, and analysts Ray Robbins and Ronald Ma of Austin Health. Also thanking PBT Australia, Deakin University (Ali Tammadoni), the Australian Centre for Health innovation and EntityThree for collaborating on this project.

Special mention is due to the review board: Dr. Chris Bain, currently Chief Health Information Services Manager at Mercy Public Hospitals; Shmuley Goldberg, former Director of Corporate Development for Victorian Comprehensive Cancer Centre (VCCC), Robert Heyes, Workforce Analyst at the Victorian Department of Health and Professor Damminda Alahakoon, currently Professor in Business Analytics at La Trobe University.

References

[1] Austin Health Victoria, *About Us,* [Online] Available at: http://www.austin.org.au/about-us/, 2015.

[2] Health Information Management Association Australia, *FAQ - What is ICD-10-AM, ACHI and ACS?,* [Online] Available at: www.himaa2.org.au/education/?q=node/81, 2015.

[3] State Government of Victoria, *Victorian Health Service Performance Framework 2014-15*, [Online] Available at: http://docs.health.vic.gov.au/docs/doc/184F7ACEBB1167FECA257D3800093A93/$FILE/1408007_VPMF%20web2.pdf., 2015.

[4] Dr Foster Intelligence, *Quality Investigator Victoria Business Rules*, March, 2014.

[5] M. Shulan, K. Gao and C. Moore, Predicting 30-day all-cause hospital readmissions, *Health Care Management Science*, **16**,2 (2013), 167-175, Springer Science + Business Media, New York, 2013.

[6] J. Billings, I. Blunt, A. Steventon, T. Georghiou, G. Lewis and M. Bardsley, Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30), *BMJ Open*, **2**,4 (2012).
[7] J. Donzé, D. Aujesky, D, Williams and J.L. Schnipper, Potentially avoidable 30-day hospital readmissions in medical patients: derivation and validation of a prediction model, *JAMA Internal Medicine*, **173**,8 (2013) 632-638.
[8] S.A. Choudhry, J. Li, D. Davi, C. Erdmann, R. Sikka and B. Sutariya, A public-private partnership develops and externally validates a 30-day hospital readmission risk prediction model, *Online Journal Of Public Health Informatics*, **5**,2 (2013) 219.
[9] L.K. Robin, D.H. Harlen, W.M. Richard, F.E. Matthew, S.W. Douglas and R.M. David, Risk Factors for All-Cause Hospital Readmission Within 30 Days of Hospital Discharge, *Journal of Clinical Outcomes Management*, **20**,5 (2013) 203-214.
[10] M.E. Charlson, P. Pompei, K.L. Ales and C.R. MacKenzie, A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation, *Journal of Chronic Diseases*, **40**,5 (1987) 373–383.
[11] SAS Institute Inc., *Data Mining Using SAS® Enterprise Miner^{TM}: A Case Study Approach*, Third Edition, SAS Institute Inc. Cary NC, 2013.
[12] M. Maldonado, *Tip: How to Interpret SAS Rapid Predictive Modeler Results*, [Online] Available at: https://communities.sas.com/docs/DOC-6813, 2014.