Digital Health Innovation for Consumers, Clinicians, Connectivity and Community
A. Georgiou et al. (Eds.)

doi:10.3233/978-1-61499-666-8-67

Predicting Unpanned Return to Hospital for Chronic Disease Patients

Sankalp KHANNA[1], Norm GOOD and Justin BOYLE
CSIRO Australian e-Health Research Centre, Brisbane, Australia

Abstract. Preventing unplanned returns, including readmissions and re-presentations to the emergency department is increasingly becoming a performance target for hospitals across the globe. Significant successes have been reported from interventions put in to place by hospitals to reduce their incidence. However, despite several risk stratification algorithms being proposed in recent years, there is limited use of these algorithms in hospital services to identify patients for enrolment into these intervention programs. This study identifies constraints limiting the practical use of such algorithms. We also develop and validate models that focus on clinically relevant patient cohorts and are thus better suited to practical deployment in hospitals, while still offering good predictive ability.

Keywords. Bed occupancy, hospital bed capacity, crowding

Introduction

As hospitals around the world struggle to meet increasing demand despite constrained budgets, reducing potentially preventable chronic disease readmissions, often credited with being responsible for a significant portion of health spending [1],[2], is increasingly climbing the priority list of hospital administrators. Recent efforts to measure readmission rates and penalise hospitals based on these [3],[4] have further driven efforts to identify high risk patients and driving targeted interventions to reduce the incidence of them returning to hospital as unplanned readmissions or emergency department (ED) presentations.

While it has been shown that no single intervention implemented alone significantly reduces 30-day readmission rates [5], several intervention programs have reported success in reducing the incidence of unplanned revisits to hospital among enrolled patients [5],[6]. In Australia, the HARP program, introduced and run by the Victorian Department of Health, has been the most successful to date, claiming 35% fewer ED presentations and 52% fewer emergency admissions among recruited patients [7].

In contrast, while several models have been proposed over recent years for predicting the risk of hospital readmission [8]–[12], the performance of most has been found to be sub-optimal [10] and very few are actively used in hospital settings for risk stratification. Development of the popular PARR tool developed in the UK [8] was discontinued within a few years of its launch. The Australian HARP program itself

[1] Corresponding Author: Sankalp Khanna, CSIRO Australian e-Health Research Centre, Brisbane, Australia; E-mail: Sankalp.Khanna@csiro.au

utilised a combination of techniques ranging from nurse instinct to paper based scoring tools to nominate patients for enrolment in the program [7].

We recently developed several models that utilised cohort population and clinical data and demonstrated that they were capable of precisely identifying chronic disease patients with a high risk of rehospitalisation within 30 days [13]. In discussing the practical application of these to drive interventions, we identified that several patient cohorts included in our analysis, and in other models proposed around the world, were of low relevance to clinicians seeking to identify patients for these interventions. Of these, Dialysis patients were the only cohort that we had specifically removed in our previous modelling. Also, while our algorithms focussed on readmission within 30 days, clinicians were also keen on identifying patients that might re-present to ED within 30 days following discharge from hospital.

This study was undertaken with the objective of employing administrative and clinical information for patients belonging to a lower socio-economic region of Queensland, Australia, to develop and validate prediction models for identifying chronic disease patients with a high risk of unplanned non-routine readmission, re-presentation, or return (readmission or re-presentation) within 30 days of discharge from hospital. It is hoped these changes will better inform and support discharge planning and community interventions aimed at reducing potentially preventable readmissions of chronic disease patients.

1. Methods

The study re-employed data reported in [13], comprising administrative inpatient and emergency department data from 2005-2010 for patients residing in a lower socio-economic area of Queensland, Australia. The study focused on the patient cohort that had at least one chronic disease admission (identified by ICD-10 codes, see Table I) during the analysis period. Ethics approval for this research was obtained from the Queensland Health Metro South Health Services District.

Table 1. List of ICD-10 diagnosis codes used to identify chronic disease patients for the study period

Diagnosis Code Block	Description
E11*	Type 2 Diabetes Mellitus
I25*	Chronic Ischaemic Heart Disease
I50*	Heart Failure
I60*	Subarachnoid Haemorhage
I61*	Intracerebral Haemorrhage
I62*	Other Nontraumatic Intracranial Haemorrhae
I63*	Cereral Infarction
I64*	Stroke, Not Specified as Haemorrhage or Infarction
J44*	Other Chronic Obstructive Pulmonary Disease
J45*	Asthma
J46*	Status Asthmaticus
N18*	Chronic Kidney Failure
Z49*	Care Involving Dialysis

Following data linkage, the data was cleaned by removal of incomplete/inconsistent records. All episodes identified as routine, i.e. Chemotherapy (DRG=”R63Z”), Routine Rehabilitation same day (DRG=”Z60C”) and Renal Dialysis (DRG=”L61Z”) were excluded from the modelling, though counts of such episodes were included as variables for patients attending for other illnesses. Obstetric admissions were removed as they were heavily weighted to return in less than 30 days. Episodes representing index (first) admissions and where patients died during the episode were also removed. The resulting dataset, comprised 39797 episodes, which were divided into test and training sets at each iteration for modelling and independent validation. Microsoft Excel 2007, R 3.1.1 and Matlab 7.13.0 were employed for data manipulation and statistical analysis.

Potential predictor variables included patient demographic data, diagnosis codes (ICD and DRG) and their interaction with the inpatient and emergency departments, outpatient clinics, and some limited hospital affiliated community health services. Four response variables were selected for the modelling – Return as an admitted patient, Return as an admitted patient through emergency, Re-presentation at the emergency department, and Return to hospital as either an admission or emergency presentation. 30 days was chosen as the benchmark return time as it is well regarded as the optimal choice for the purpose [14]–[16], and is also employed by the metrics used for evaluating and comparing hospital performance [3], [4]. The four resulting response variables were RA30 - Return as an admitted patient within 30 days of discharge, RA30E - Return as an admitted patient through emergency within 30 days of discharge, RP30 – Re-presentation at ED within 30 days of discharge, and RU30 - Return to hospital as either an admission or emergency presentation within 30 days of discharge.

Three algorithms were employed for model development. - generalised estimating equations (GEE), artificial neural networks (ANN), and random forests (RF). The GEE approach [17] was chosen because of its established efficacy in using weighted combinations of observations to extract the appropriate amount of information from correlated data. This was deemed to be a good fit given repeat encounters by patients within the dataset. Initial variable selection was done using random forests [18]. The top 15 variables selected here were fed to a stepwise binomial regression model to select a significant variable set. This served as input to a GEE model that employed Patient ID as the grouping variable. ANN [19] and RF [18] were employed given their established superiority in pattern recognition from large complex data. Various combinations of variables, including some informed by the variable selection employed by the GEE modelling, were explored for developing ANN and RF models.

Sensitivity and specificity were calculated for each encounter and the Receiver Operator Characteristic (ROC) graph was generated for each evaluation. The c-statistic, representing the area under the ROC curve, was then calculated for each model as a measure of discrimination and used to compare the performance of the models.

2. Results

Variable selection revealed overlap across the various response variables evaluated, with 6 variables being returned as important across all response variables. Table 2 lists the significant variables selected for each response variable in the GEE modelling.

The performance of the various models for each response variable is presented in Figure 1. The performance of GEE models was average, achieving between 62% and

64% area under the curve in the ROC analysis. ANN models reported higher levels of performance, with the area under the curve in the ROC analysis ranging between 72% and 78%. The best performance was reported by the RF models, achieving between 74% and 82% area under the curve in the ROC analysis. Figure 2 presents the ROC curve for the RF model employing RP30 as the response variable. Generally, the performance was highest for these models predicting re-presentation to ED within 30 days of discharge from hospital.

Table 2. Significant variables used in GEE models

PREDICTOR VARIABLE	RESPONSE VARIABLE			
	RA30	RA30E	RP30	RU30
Days since last discharge (ReturnedIn)	✓	✓	✓	✓
DRG family of current admission (DRGcat)	✓	✓	✓	✓
Admission Index of current admission in system history (VisitID)	✓	✓	✓	✓
DRG family of previous admission (PREV.DRGcat)	✓			
Standardised Admission Unit of current admission (cadm_stnd_unit)	✓	✓	✓	✓
Primary Diagnosis ICD-10 chapter of current admission (PD1cat)	✓	✓		
Standardised Admission Unit of previous admission (PREV.cadm_stnd_unit)	✓	✓	✓	✓
Length of Stay of current admission (LOS)	✓	✓	✓	✓
Number of Outpatient Department visits in the past 180 days (OPD180)	✓			✓
Number of ICD-10 Diagnosis codes attached to this admission (NUMICD)	✓	✓		✓
Age at time of current admission (Age)		✓	✓	
Number of Outpatient Department visits in the past 365 days (OPD365)			✓	
Number of Emergency Department presentations in the past 365 days (EDIS365)		✓	✓	✓
Primary Diagnosis ICD-10 chapter of previous admission (PREV.PD1cat)		✓	✓	✓
Number of ICD-10 Diagnosis codes attached to previous admission PREV.NUMICD		✓		
Number of Emergency Department presentations in the past 180 days (EDIS180)			✓	
Number of Emergency Department presentations in the past 90 days (EDIS90)			✓	✓
Length of Stay of previous admission (PREV.LOS)			✓	✓

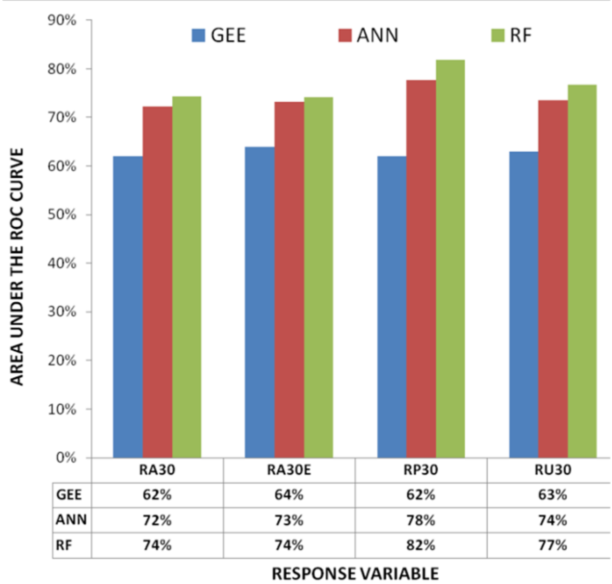

Figure 1. Comparing Model Performance

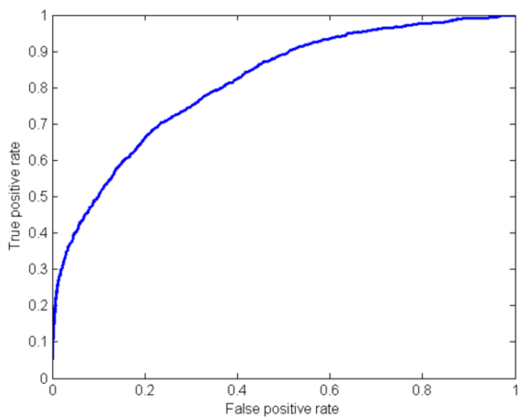

Figure 2. ROC Curve (Random Forest model for RP30)

3. Discussion

The key contribution of this study is the development and validation of risk stratification algorithms that focus on clinically relevant cohorts of patients. By excluding admissions that are operationally less relevant, it is hoped that the algorithms will be more meaningful for clinicians and possibly lead to these being incorporated in discharge planning. The predictors employed for the analysis represent a good mix of administrative and clinical parameters that can be easily accessed from existing information systems without requiring additional input from clinical staff.

The ROC analysis demonstrates that while removing routine admissions, which are likely easier to predict given their nature, reduces the efficacy of the models when compared to previous models [13], reasonably good models can still be developed for predicting clinically relevant return to hospital. The performance of ANN and RF models was found to be significantly superior to GEE models, and RF models outperformed both other approaches for all response variables. A better understanding of the difference in performance and discriminative ability between the algorithms could however be gained from additional analysis of correctly and incorrectly predicted encounters, and from further testing on datasets from other hospital services. Predicting the return of a patient to ED offered the highest performance across the response variables with both ANN and RF models, which was somewhat expected given that the patient cohort comprised chronic disease patients with complex care needs, a group more prone to emergency presentations and admissions via the ED.

The developed models exhibit good discrimination ability while serving the task of addressing only clinically relevant cohorts of patients. We are currently discussing a trial of these models at a local health service and incorporating information from pharmacy and general practice to further improve predictive power.

4. Limitations

The study focused on patients primarily belonging to a specified lower socio-economic region in the state. Additional analysis needs to be undertaken before the models developed as part of this study can be applied outside this region. The study was unable to account for patients that may have visited hospitals interstate or overseas, or not returned because they died. Because of limited collection of outpatient and community health data, the efficacy of these indicators could not be fully exploited and more work needs to be done to improve models with information from these data sources. The model does not employ clinical and non-clinical indicators during the current visit. The process would benefit from research into how these indicators could be integrated.

References

[1] Kimberly Swartz, "Projected Costs of Chronic Diseases." Health Care Cost Monitor, 13-Apr-2011.

[2] D. E. Bloom, E. Cafiero, E. Jané-Llopis, S. Abrahams-Gessel, L. R. Bloom, S. Fathima, A. B. Feigl, T. Gaziano, A. Hamandi, M. Mowafi, D. O'Farrell, E. Ozaltin, A. Pandya, K. Prettner, L. Rosenberg, B. Seligman, A. Z. Stein, C. Weinstein, and J. Weiss, "The Global Economic Burden of Noncommunicable Diseases," Program on the Global Demography of Aging, PGDA Working Paper 8712, 2012.

[3] R. Bhalla and G. Kalkut, "Could Medicare readmission policy exacerbate health care system inequity?" *Ann. Intern. Med.*, vol. 152, no. 2, pp. 114–117, Jan. 2010.

[4] S. Kangovi and D. Grande, "Hospital readmissions-not just a measure of quality," *JAMA*, vol. 306, no. 16, pp. 1796–1797, Oct. 2011.

[5] L. O. Hansen, R. S. Young, K. Hinami, A. Leung, and M. V. Williams, "Interventions to reduce 30-day rehospitalization: a systematic review," *Ann. Intern. Med.*, vol. 155, no. 8, pp. 520–528, Oct. 2011.

[6] S. Kripalani, C. N. Theobald, B. Anctil, and E. E. Vasilevskis, "Reducing hospital readmission rates: current strategies and future directions," *Annu. Rev. Med.*, vol. 65, pp. 471–485, 2014.

[7] Victorian Government Department of Human Services, "Improving care. Hospital Admission Risk Program. Public report," 2006. [Online]. Available: http://www.health.vic.gov.au/harp/downloads/improvingcare.pdf. [Accessed: 7-Mar-2016].

[8] J. Billings, J. Dixon, T. Mijanovich, and D. Wennberg, "Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients," *BMJ*, vol. 333, no. 7563, p. 327, Aug. 2006.
[9] C. van Walraven, I. A. Dhalla, C. Bell, E. Etchells, I. G. Stiell, K. Zarnke, P. C. Austin, and A. J. Forster, "Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community," *CMAJ*, vol. 182, no. 6, pp. 551–557, Apr. 2010.
[10] D. Kansagara, H. Englander, A. Salanitro, D. Kagen, C. Theobald, M. Freeman, and S. Kripalani, "Risk prediction models for hospital readmission: a systematic review," *JAMA*, vol. 306, no. 15, pp. 1688–1698, Oct. 2011.
[11] Canadian Institute for Health Information, "Early Identification of People at-risk of Hospitalization: Hospital Admission Risk Prediction (HARP)," 2013.
[12] S. Hao, Y. Wang, B. Jin, A. Y. Shin, C. Zhu, M. Huang, L. Zheng, J. Luo, Z. Hu, C. Fu, D. Dai, Y. Wang, D. S. Culver, S. T. Alfreds, T. Rogow, F. Stearns, K. G. Sylvester, E. Widen, and X. B. Ling, "Development, Validation and Deployment of a Real Time 30 Day Hospital Readmission Risk Assessment Tool in the Maine Healthcare Information Exchange," *PLoS ONE*, vol. 10, no. 10, p. e0140271, 2015.
[13] S. Khanna, J. Boyle, and N. Good, "Precise prediction for managing chronic disease readmissions," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2014, pp. 2734–2737, Aug. 2014.
[14] P. Halfon, Y. Eggli, G. van Melle, J. Chevalier, J. B. Wasserfallen, and B. Burnand, "Measuring potentially avoidable hospital readmissions," *J Clin Epidemiol*, vol. 55, no. 6, pp. 573–587, Jun. 2002.
[15] T. Heggestad and S. E. Lilleeng, "Measuring readmissions: focus on the time factor," *Int J Qual Health Care*, vol. 15, no. 2, pp. 147–154, Apr. 2003.
[16] E. L. Y. Wong, A. W. L. Cheung, M. C. M. Leung, C. H. K. Yam, F. W. K. Chan, F. Y. Y. Wong, and E.-K. Yeoh, "Unplanned readmission rates, length of hospital stay, mortality, and medical costs of ten common medical conditions: a retrospective analysis of Hong Kong hospital data," *BMC Health Serv Res*, vol. 11, p. 149, 2011.
[17] J. Hardin, J. Hilbe, "Generalized Estimating Equations, Second Edition," *CRC Press*. 2013
[18] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
[19] [19] G. Zhang, B. Eddy Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: The state of the art," *International Journal of Forecasting*, vol. 14, no. 1, pp. 35–62, Mar. 1998.