Improving decision making in the management of hospital readmissions using modern survival analysis techniques

Todd, James ; Gepp, Adrian; Stern, Steven; Vanstone, Bruce

Decision Support Systems

DOI:
https://doi.org/10.1016/j.dss.2022.113747

Published: 01/05/2022

Peer reviewed version

Improving Decision Making in the Management of Hospital Readmissions using Modern Survival Analysis Techniques

Abstract

Hospital readmissions lead to unnecessary demand for healthcare resources, greater financial costs, and poorer patient outcomes. These consequences have led hospitals to attempt to identify high-risk patients with predictive models, but research has rarely focused on survival analysis techniques, model applications, and performance measures. This study establishes the uses of survival models to support managerial decision-making for readmissions. First, machine learning and statistical survival techniques are applied, ten of which have not been used in previous readmission research. Secondly, applications of survival models in a decision support capacity are proposed, relating to intervention targeting, follow-up care customisation, and demand forecasting. Thirdly, performance measures for the proposed applications are determined and used for empirical model assessment. These performance measures have not been applied in previous readmission research. The empirical assessment is based on adult admissions to the Emergency Department of Gold Coast University Hospital (n = 46,659) and Robina Hospital (n = 23,976) in Queensland, Australia. The relevant aspects of performance were determined to be discrimination and calibration, as measured by time-dependent concordance and D-Calibration respectively. A range of discrimination and calibration combinations can be achieved by different models, with the Recursively Imputed Survival Tree, Cox regression, and hybrid Cox-ANN techniques being most promising. Survival approaches linking techniques, proposed applications, and performance measurement should be given greater consideration in future healthcare research and in institutions aiming to manage readmissions.

Keywords: Predictive analytics; Hospital readmissions; Survival analysis; Machine learning; Performance measurement

1 Introduction

Unplanned and early readmissions put patients at greater risk of adverse outcomes, burdens limited hospital resources, and imposes costs on the healthcare system. Readmissions may also indicate underlying issues in the quality of care being provided to patients before and after their discharge [19].

The US Hospital Readmissions Reductions Program (HRRP) introduced in 2012 is the most prominent example of healthcare policy targeting readmissions, under which hospital risk-adjusted readmission rates for certain conditions are linked to funding [10]. Healthcare policies targeting readmissions have similarly been implemented in Germany, Denmark, and England [42]. Most recently, Australia's Independent Hospital Pricing Authority [32] has developed a pricing model adjusting funding for admission episodes based on readmission outcome, condition and complexity. Such policy aims to incentivise hospitals to improve quality of care, communication and management of high-risk patients to reduce readmissions.

While the usage of financial penalties have been critiqued in some cases [19, 37, 42, 70], there is agreement that many readmissions are avoidable [61, 72], through better clinical management or discharge planning [7]. Research has found robust interventions to be effective, though resource requirements make it important to identify high-risk patients for intervention targeting [41]. Accordingly, much research has focused on the development of predictive models relating patient-specific factors to readmission risk. These models are intended to serve as decision support systems for hospitals. Unlike risk adjustment models in healthcare policy, they are restricted to data available at the time when decisions are made, commonly discharge time.

Many predictive models have been proposed to quantify the risk of readmission given a patient's available information, though these have often been characterised by unimpressive performance. Most such predictive models have taken a classification approach in which readmission status is considered at a single time point, generally 30 days. This approach allows for straightforward application of well-established techniques, easily interpretable predictions, and facilitates standardised performance comparisons across hospitals. Reflecting the focus on supporting administrative rather than clinical decisions, less interpretable machine learning classification techniques have also been applied. Survival approaches have primarily been used in inferential readmission research aiming to identify risk factors, with predictive research employing survival techniques being much rarer. Fewer still have considered practical applications specific to survival approaches and associated performance measures. This work (i) identifies a range of applicable survival techniques, (ii) proposes applications of survival models to support managerial decision-making, and (iii) determines performance measures suitable for assessing the survival models for these applications. The primary contribution of this work is in linking these three elements of developing decision support tools for hospital administrators, with this link lacking in prior research. To operationalise this, the following research question is considered: "How well can various survival modelling techniques capture aspects of hospital readmission risk over time relevant to managerial decision-making?"

In addressing this question, four practical applications of survival models to support managerial decision-making for readmissions are proposed:

- **Dynamic Risk Ranking (DRR)**: To facilitate allocation of limited resources for interventions and patient management, patients can be stratified by risk of readmission with a survival model. This stratification is dynamic in that it can be updated for patients who have already been discharged, rather than being limited to the time of discharge as in classification models.
- **Elevated Risk Period (ERP) and Elevated Risk Period Probability (ERPP)**: The ERP application of a survival model assesses the length of time before a patient's risk of readmission reaches some acceptable level and thus how long they are of interest for post-discharge management decisions. The ERPP application assesses the

probability of readmission within the ERP. These applications allow for differences in risk profiles between patients (rather than considering a single time point for all patients) and context-specific customisation of how acceptable risk levels are defined.

- **Expected Readmissions**: Given survival curves from a well-fitted model, it is straightforward to calculate the expected number of readmissions in a period conditional on patients being readmission-free up to the start of the period. Forecasting of aggregate readmissions supports planning and resource allocation decisions.

Additionally, ten machine learning survival techniques which have not been investigated in prior readmission research are identified and empirically evaluated. The empirical assessment is based on adult admissions to the Emergency Department of Gold Coast University Hospital (n = 46,659) and Robina Hospital (n = 23,976) in Queensland, Australia. on two emergency department populations. Unlike many prior studies using survival models, this evaluation is based on measures of discrimination and calibration that directly relate to the desirable features of models in the proposed applications. This empirical assessment of machine learning techniques and comparison with more interpretable statistical techniques demonstrates the range of alternatives available for readmission modelling. It also allows for consideration of whether there is a loss of predictive power from more interpretable techniques and, if so, how much.

The remainder of this paper is set out as follows. Section 2 summarises key prior research in the field of readmission prediction. Section 3 describes the data used in this work and its processing. Section 4 details the modelling techniques considered, describes the model selection process, and discusses appropriate performance measures. Section 5 presents and discusses the performance of the final models with respect to the research question, which are discussed further in Section 6. Finally, Section 7 highlights the key contributions made, suggests directions for future research, and discusses relevant limitations.

2 Summary of Key Related Research

Motivated by the various costs associated with readmissions, many studies have aimed to develop and validate predictive models to support decisions regarding interventions and clinical management. Most have adopted a classification approach in which readmission is a binary outcome determined by patient status at a fixed time point. The most common time point has been 30 days, which matches the definition of readmissions used in the US HRRP [10]. Beyond matching policy definitions, considering outcomes as binary has allowed for application of well-established techniques, most often logistic regression [6]. Less commonly, studies have used survival models to predict readmissions, with such models more frequently seen in studies investigating risk factors. Survival models do not require that readmission status be considered only at a fixed time point and instead aim to model risk across time. The most common survival technique for readmission prediction has been the Cox regression model and related variations [25, 46], as noted in a recent review [6].

Motivated by a desire to improve on the performance of existing statistical readmission models [38], machine learning techniques have increasingly been considered [6]. This has been further motivated by their lack of distributional assumptions and their greater ability to capture highly non-linear and complex relationships compared to traditional techniques. Prominent machine learning techniques have included artificial neural networks (ANNs) [2, 24, 36, 65, 68], support vector machines (SVMs) [8, 56, 69], random forests [15, 20, 27], and decision trees [48, 63]. In general, more complex techniques have been found to improve on logistic regression, though generalisation of results is made difficult by differences in datasets, patient groups, and conditions across studies [6]. Under survival approaches, the only machine learning techniques applied have been random survival forests (RSFs) [28, 46].

Classification models are often intended to assist in stratifying patients by risk of readmission. Accordingly, model performance is typically assessed by the area under the receiver operating characteristic curve (AUC), which measures the ability of a model to discriminate between positive and negative observations. Survival models are not restricted to a fixed time and risk predictions can be calculated conditional on the patient being readmission free for some period. This allows for alternative model applications and thus requires alternative appropriate performance measures. One such measure is Harrell's concordance index [29], which has been used to assess the discrimination of survival models applied for risk stratification [25, 46]. It was developed in the context of the Cox regression model, however, and relies on the assumption of time-invariant risk rankings, which machine learning models may not provide.

As stated in the introduction, this work applies a wide range of previously unconsidered machine learning survival techniques, suggests survival-specific model applications, and employs appropriate performance measures. This is motivated by two characteristics of prior research.

The first characteristic is the increased interest in machine learning techniques for readmission prediction. This has almost exclusively been seen for classification approaches, with encouraging results, despite the motivation for such techniques being equally applicable to survival approaches. This work explores the potential value of a wide range of previously unconsidered machine learning survival techniques in addition to RSFs and Cox regression.

The second characteristic of prior research relates to the absence of studies combining survival models, survival-specific applications, and appropriate performance measurement. Where survival models were used, evaluation of predictive performance was often cursory [13], or absent [34, 55]. Where predictive performance was assessed, this was often based on prediction at discrete points [1] or how classification models would be applied [43, 64, 67], despite some studies mentioning applications specific to survival models such as dynamic risk ranking [28]. Other studies used survival models but did not identify potential survival-specific applications [3, 46, 47]. A final study directly applied regression techniques by only considering readmission times for 30-day readmissions and used regression performance measures [21].

This work aims to address this lack through the proposal of several survival-specific model applications and identification of model performance measures appropriate to these applications. This reflects the view that survival approaches should be seen as complementary rather than competitive with classification approaches. Of the four applications proposed, prior research has only considered DRR [28] and performance measures did not appropriately consider predictions of risk over time or the possibility of time-varying risk rankings. The authors are not aware of any readmission research which has considered the remaining three applications of ERP, ERPP, and Expected Readmissions.

3 Data

The data used in this work consists of costing data for hospital discharges of adult patients admitted to the Emergency Department (ED) of Gold Coast University Hospital (GCUH) (n = 46,659) and Robina Hospital (RH) (n = 23,976), both of which service the Gold Coast region of Australia. These relate to adults discharged in the period ranging from April 30th, 2016 to April 30th, 2018. The hospitals are treated as separate datasets given the goal of developing institution-specific decision support tools. Additionally, the two hospitals service different patient populations and treating them separately allows for results to be compared. For both hospitals, data was longitudinally split into training and test sets containing 70% and 30% of the data respectively, as shown in Table 1.

Table 1. Train and Test Data - Size and Dates

Hospital	Split	Quantity	Start Date	End Date
GCUH	Train	32,661	2016-04-30	2017-09-30
	Test	13,998	2017-09-30	2018-04-30
RH	Train	16,783	2016-04-30	2017-09-29
	Test	7,193	2017-09-29	2018-04-30

Patient discharges were also excluded if discharge was to another hospital, as details of patient care and effective discharge date are unknown, or if discharge was against medical advice, as consistent with existing literature and measurement under relevant healthcare policy [10, 32].

Table 2. Features Used in Modelling

Feature	Feature Description
AdmitWardCode1 (*Derived Feature*)	An aggregated version of the AdmitWardCode field. This derived field is described in Appendix A. AdmitWardCode: WardCode patient is admitted to.
Age	Age of a patient calculated at the time of discharge.
ED_NumPresPrevYear	Number of ED presentations that occurred during the year prior to the current admission.
ED_NumPresSincePrevAdm	Number of ED presentations that occurred since the patient's previous inpatient admission via ED.
ED_NumPresSincePrevAdmALL	Number of ED presentations that occurred since the patient's previous inpatient admission (via Outpatients, ED etc.).
GenderCode	Gender of a patient (M or F).
iGC (*Derived Feature*)	A grouped version of the Postcode field specifying the region of the Gold Coast the patient's home address is in. This derived field is described in Appendix A.
Inpat_NumAdmPrevYearALL	The number of all inpatient admissions (via Outpatients, ED etc.) that occurred during the year prior to the current admission.
Inpat_PrevAdmLOSPrevYear	Length of stay of previous inpatient admission via ED in days.
Inpat_PrevAdmLOSPrevYearALL	Length of stay of previous inpatient admission (via Outpatients, ED etc.) in days.
Inpat_TimeSincePrevAdmALL	Days since the previous inpatient admission (via Outpatients, ED etc.) that occurred during the year prior to the current row's admission date.
Inpat_TotalAdmInICU	Number of Inpatient Admissions that the patient had in the ICU within the previous year.
Inpat_TotalAdmInICUALL	Number of Inpatient Admissions (via Outpatients, ED etc.) when the patient was in ICU within the previous year from the current row's admission date.
Inpat_TotalTimeAdmPrevYear	Cumulative length of stay in days as an inpatient admission via ED during the year prior to the current row's admission date.
Inpat_TotalTimeAdmPrevYearALL	Cumulative length of stay in days as an inpatient admission (via Outpatients, ED etc.) within hospital during the year prior to the current row's admission date.
LOSCalc (*Derived Feature*)	Difference in days between the time of inpatient admission and time of inpatient discharge.
Outp_NumApptPrevYear	Number of outpatient appointments that occurred during the year prior to the current row's admission date.
Outp_NumApptSincePrevAdm	Number of outpatient appointments that occurred since the patient's previous inpatient admission via ED.
Outp_NumApptSincePrevAdmALL	Number of outpatient appointments that occurred since the patient's previous inpatient admission (via Outpatients, ED etc.)

Table 3. Descriptive Statistics (Full Data)

	GCUH	RH
Data		
Total admissions	46,659	23,976
Readmissions in 30 days	14.41%	15.65%
Censored Observations	61.62%	58.02%
Selected Features used in Modelling		
Age: Mean (SD)	59.16 (20.50)	66.48 (19.94)
Female (%)	48.13%	52.25%
Region		
Inner Gold Coast	62.94%	74.51%
Outer Gold Coast	24.14%	17.86%
Other	12.92%	7.63%
Length of Stay: Mean	4.53	3.98
Inpatient Admissions in Previous Year: Mean (Median)	1.30 (0)	1.41 (0)
Outpatient Appointments in Previous Year: Mean (Median)	5.42 (1)	4.41 (0)
ED Presentations in Previous Year: Mean (Median)	1.94 (1)	2.16 (1)

To avoid consideration of planned and routine admissions, discharges were considered to have resulted in an unplanned readmission if readmission type was coded as Acute and readmission status was coded as Emergency. Considered data features related to prior use of health services, sociodemographic factors, and length of stay for the initial admission. All features and descriptions are shown in Table 2.

Descriptive statistics for the dataset are shown in Table 3, which further supports the decision to consider the two hospitals separately. RH is characterised by patients who are older, are admitted for shorter times, have less frequent inpatient and outpatient admissions, and are more often from the inner Gold Coast region.

Additional problem-specific processing of the data was carried out before the application of all techniques outlined in Section 4.1. This is detailed in Appendix A.

4 Methods

4.1 Techniques Considered

In determining the techniques considered in this work, the focus was on exploring the performance of a wider range of techniques than considered in prior research. Accordingly, both Cox regression and RSFs were included. Other techniques were selected based on a review of major machine learning categorisations, which included decision trees, ensembles, SVMs, and ANNs. Within these categories, techniques adapted for survival data were identified. Techniques were not included if they did not provide predictions of risk over time, or if they were improved upon in a later variation. Fully parametric techniques were not included as they entail statistical constraints beyond that of Cox regression and the consideration of machine learning techniques here and in prior work has been motivated in part by their non-parametric nature.

From decision trees, survival trees under a log-rank splitting rule and under a one-step likelihood approach [44] are considered. Direct extensions of decision trees to survival data have been achieved via modification of splitting rules, and these two variations have been among the most common employed. Doubly robust Censoring Unbiased Regression Trees (CURTs) [52] are also considered. This extension of trees to censored data is based on data transformations rather than modified splitting rules, with the doubly robust transformation being more robust than the alternative inverse probability of censoring weighting (IPCW) transformation [52]. From ensembles, RSFs [33], doubly robust Censoring Unbiased Regression Ensembles (CURE) [53], Recursively Imputed Survival Trees (RIST) [71], and Bayesian Additive Regression Trees (BART) [51] are considered. The doubly robust transformation is used for the CURE technique rather than the IPCW transformation for the same reason as already stated. Excluding those using IPCW transformations [30] or only bootstrapped aggregation of survival trees [31], no other ensembles of trees were identified in the literature. For ANNs, three extensions to survival data were identified and considered. These were a time-coded ANN, multiple time point ANN, and hybrid Cox-ANN. The considered time-coded ANN is based on the principles set out by

Biganzoli, Boracchi and Marubini [9]. A recent implementation of a multiple time point ANN is used, termed Nnet-survival [22], as well as a recent implementation of a hybrid Cox-ANN, termed Cox-nnet [11, 62]. A fourth, single time point extension of ANNs to survival data was also identified [14, 35], but not included as this extension predicted risk at a single time point. Lastly, while several extensions of SVMs to survival data were identified [16, 17, 23, 39, 40, 49, 50, 57-60], none produced risk over time predictions by default or with straightforward modifications. Of these machine learning survival techniques, only RSFs are known to have been applied to readmission prediction [28, 46].

It should be noted that the CURT and CURE techniques do not offer probabilistic outputs as part of their original algorithms, but they are included because this can be simply remedied. This is achieved by summarising terminal nodes with Kaplan-Meier functions.

4.2 Performance Measures

Considering the four applications of survival models outlined in the introduction, the two relevant aspects of performance are discrimination and calibration. Model discrimination is of primary importance for DRR, while ERP, ERPP, and Expected Readmissions also require a discriminative model to ensure these applications account for the differences in patient characteristics. Model calibration is of primary importance for ERP, ERPP, and Expected Readmissions to ensure the underlying risk predictions are reliable. While the relative value of discrimination and calibration will depend in practice upon the specific application and context, these aspects of model performance are most relevant for the proposed applications supporting managerial decision-making.

The most common measure of discrimination for survival models is Harrell's concordance index, also known as the c-index [29]. This measure considers the temporal aspect of survival data by comparing model predictions only on observation pairs where one observation is known to have experienced the event before the other, known as comparable pairs. It does not, however, allow for the assigned risk ranking of observations to vary over time. While this is suitable for proportional models such as Cox regression, it is inappropriate for machine learning models which may produce time-varying risk rankings. A more appropriate, time-dependent concordant index was proposed by Antolini, Boracchi and Biganzoli [5]. In this time-dependent concordance index, for a comparable pair of observations, the model's predictions are concordant if the observation experiencing the event was assigned a high probability of event occurrence at the time of the event. While this measure has been applied in other areas of health analytics [18, 45, 66], the authors are not aware of its usage within readmission research.

While calibration measures such as the Hosmer-Lemeshow test are well established and appropriate for predicting n-day readmissions (where n is constant), they lack a direct extension to risk over time predictions. Motivated by the prognostic value of individualised survival curves and the need for tests of their calibration, a measure termed "D-Calibration" has been proposed [4, 26]. The core idea of the D-calibration measure is that the model producing individual survival functions acts as a mapping of observed event times on the interval $[0, \infty)$ to survival probabilities on the interval $[0,1]$. It is then expected that the proportion of probabilities in a subset $[a, b]$ of the interval $[0,1]$ will be equal to the width of the interval for a well-calibrated model. This idea leads to a straightforward application of the χ^2 test to assess the null hypothesis that the model is D-Calibrated.

These identified measures of discrimination and calibration appropriately capture the desired characteristics of survival models in the proposed applications. A third measure, however, is introduced to supplement time-dependent concordance and D-Calibration, as neither are appropriate for determining the final hyperparameter settings for the machine learning techniques considered in this work. Using either measure in isolation would result in a final model that did not reflect the need for both aspects of model performance. Accordingly, the Integrated Brier Score (IBS) is used for model selection. IBS is commonly used in survival modelling contexts and has the attractive feature of considering both discrimination and calibration, albeit in a distinct and fixed manner, and can be expressed as a sum of these two components [54]. The more specialised time-dependent concordance index and D-Calibration measures are then used in conjunction with IBS for richer evaluation of the final models associated with each technique.

4.3 Model Tuning and Selection

To facilitate comparability and reproducibility, five-fold cross-validation and minimum IBS is used to determine the final hyperparameter settings for each machine learning model. Standard grid search approach is employed, with the hyperparameter values considered and used available in Appendix B~~Appendix A~~. The only exception to this was the BART technique, for which no hyperparameters were varied. This was driven by previous findings that excellent performance is achieved by the default hyperparameter settings [12, 51] and by BART being extremely computationally intensive with respect to runtimes and memory requirements. When using the R programming language, generating predictions for the training data of RH took 6.89 hours and the prediction object was 138.8Gb. Additional implementation details for the ANN techniques are provided in Appendix C~~Appendix B~~. Where model predictions are only available at discrete time points, such as for decision trees and some ANN techniques, survival curves were linearly interpolated.

4.3.1 Cox Regression

Given that statistical models make assumptions about the nature of the underlying data, adjustments to the training data are an inherent part of a sophisticated model implementation. This is relevant as many of the features exhibit high positive skew with large outliers. The presence of extreme outliers or skewness leading to sparse regions in the predictors is problematic because of the large effect on coefficient estimates. As would be the case in practice, the data are adjusted prior to model fitting. This was only done for Cox regression, as machine learning models are purported to be more flexible and better able to handle such data characteristics without requiring extensive pre-processing. To ensure reproducibility, adjustments for numeric data were made according to two rules:

Rule 1: Let $o_{1,j}, o_{2,j}, \dots, o_{U,j}$ be the U unique and ordered values of the j-th covariate. If the relative frequency of $o_{1,j}$ is greater than 85%, the variable is transformed with the equation $x_{i,j}^* = \mathbf{1}(x_{i,j} = o_{1,j}) \times o_{1,j} + \mathbf{1}(x_{i,j} > o_{1,j}) \times o_{2,j}$ where $\mathbf{1}$ is the indicator function taking a value of 1 if the condition is satisfied and 0 otherwise.

Rule 2: If the combined relative frequency of $o_{u,j}, \dots, o_{u+4,j}$ is less than $1/U$ and u is the minimum value for which this condition is true, the variable is transformed with the equation $x_{i,j}^* = \mathbf{1}(x_{i,j} \leq o_{u,j}) \times x_{i,j} + \mathbf{1}(x_{i,j} > o_{u,j}) \times o_{u,j}$.

The effects of the modifications are shown in Table 4.

Table 4. Statistical Model Data Transformations

Feature	**Upper Bounds - GCUH**	**Upper Bounds - RH**
Age	105 → 95	107 → 97
ED_NumPresPrevYear	74 → 11	76 → 11
ED_NumPresSincePrevAdm	38 → 1	23 → 1
ED_NumPresSincePrevAdmALL	38 → 1	23 → 1
Inpat_NumAdmPrevYearALL	34 → 8	34 → 7
Inpat_PrevAdmLOSPrevYear	195 → 14	150 → 12
Inpat_PrevAdmLOSPrevYearALL	195 → 16	154 → 14
Inpat_TimeSincePrevAdmALL	365 → 162	365 → 203
Inpat_TotalAdmInICU	6 → 1	7 → 1
Inpat_TotalAdmInICUALL	6 → 1	7 → 1
Inpat_TotalTimeAdmPrevYear	273 → 31	152 → 30
Inpat_TotalTimeAdmPrevYearALL	297 → 44	270 → 37
LOSCalc	303 → 22	489 → 20
Outp_NumApptPrevYear	140 → 30	185 → 27
Outp_NumApptSincePrevAdm	105 → 12	103 → 8
Outp_NumApptSincePrevAdmALL	114 → 12	86 → 9

Term selection was performed systematically by considering main effects, interactions, and polynomial terms. As it is unrealistic to define a candidate variable set considering all possible effects of each type, a greedy-style approach to determining the terms to include in a final model was used. This involves the application of stepwise procedures to the training data in three stages.

1. All covariates are considered as main effects. A hybrid forward and backward stepwise procedure using the Akaike Information Criterion (AIC) beginning from a full model is applied to identify a reduced set of main effects.

2. Main effects that remain after Stage 1 are considered in addition to all their possible pairwise interactions. A similar stepwise procedure is then applied using the Bayesian Information Criterion (BIC).

3. For each numeric feature with more than ten unique values retained after Stage 2, squared and cubic terms are considered using another BIC-based stepwise procedure to identify the final Cox regression model.

As stepwise procedures aim to maximise an information criterion intended to proxy for out-of-sample performance, using cross-validation procedures as well is unnecessary. This makes the model development procedure distinct from that used for machine learning techniques but reflects the lack of hyperparameters relevant to Cox regression beyond the information criterion being maximised in the stepwise procedures.

4.3.2 Discretisation of Time

For the time-coded and multiple time point ANNs, risk is predicted for discrete time intervals rather than as a truly continuous variable. These techniques necessitate the definition of intervals. While there is little guidance in the literature on how these intervals should be defined for time-coded models, there is some evidence that the multiple time point ANN is insensitive to how intervals are defined [22]. Time intervals are defined in this work to reflect the problem-specific emphasis on the time soon after discharge where most readmissions occur. Each interval has an approximately equal number of observed events for a pre-specified number of intervals. The number of intervals is set to 40 for the time coded ANN and treated as a hyperparameter for Nnet-survival.

4.3.3 Modifications to CURT and CURE Algorithms

To implement the CURT and CURE algorithms, the code used was provided by the primary author of the publications proposing them [52, 53]. Modifications made to the algorithms are briefly described for completeness. Firstly, the original algorithms of CURT and CURE did not provide survival predictions by default. This research modified the code to compute Kaplan-Meier functions to summarise terminal nodes, which were also averaged between trees in the case of CURE. Second, the CURT code automatically selected tree depth using a simulation approach with a quadratic loss function. A model using this method was included in the results (CURT V1), as well as a second version (CURT V2) in which depth was treated as a hyperparameter in the five-fold cross-validation process.

5 Results

The performance of the final models for GCUH and RH are presented in Table 5 and Table 6. The results are presented ordered by time-dependent concordance, analogous to the emphasis on AUC in readmission literature, and again ordered by IBS, which was used for model selection. This facilitates comparisons between the four scenarios, corresponding to the two bases for ranking and the two hospitals. The p-value results from the test of D-Calibration are considered only in terms of whether a model is calibrated, as the omnibus nature of the underlying χ^2 test makes ranking these values inappropriate.

These results are briefly described in terms of each of the three measures individually. A summary of the measure results is then provided before the discussion.

Concordance performance varied within a tight band for each hospital. Excluding the worst four models for GCUH and RH, the range of concordance values were 1.186% and 0.699% respectively. Much lower performance was seen for the survival tree and CURT models, with concordance values at least 2.227% and 2.480% lower than all other models for GCUH and RH, respectively. RH appears to be a more complex problem characterised by lower performance in general with respect to concordance (and IBS as will be mentioned below). Notably, machine learning models demonstrated slightly improved relative performance on the more complex problem.

All models were found to be D-calibrated at the 5% level of significance apart from the Nnet-survival model on GCUH. This is encouraging as it indicates the identified techniques can produce suitably calibrated models.

When considering the results with respect to IBS, the worst four models are less distinct and no longer entirely made up of the individual tree models. In particular, the survival tree using a one-step likelihood splitting function is ranked sixth for both hospitals and the modified CURT is eighth on RH. As when considering concordance, there is some between-hospital consistency for the top performers with RIST and RSF being common to both. Also consistent with consideration of concordance-ranked results, the Cox regression model exhibited lower relative performance on the more complex problem.

To summarise, models were found to be D-calibrated on both hospitals with only one exception. This poorly calibrated model for GCUH, Nnet-survival, also demonstrated the greatest discrimination on this hospital, highlighting the expected trade-off between discrimination and calibration and need to measure both aspects. When comparing hospitals, it was noted that RH is more complex than GCUH and is also characterised by less competitive performance of the statistical survival model (Cox regression). In terms of the measure used for model ranking, some variation was observed in both the best and worst models when using concordance versus IBS. Most notable is the variation in the worst performing models, where concordance ranking found individual tree models to be substantially worse than all others on both hospitals but IBS ranking of these four models was less severe. Finally, while differences related to hospital and measures considered manifested in notable differences in relative model performance, some models demonstrated strong performance in all instances, most notably the RIST model.

Table 5. Final Model Performance for GCUH

	Ordered by Concordance				Ordered by IBS			
Rank	**Method**	**Time-Dependent Concordance**	**D-Calibration *p*-value (k=10)**	**IBS**	**Method**	**Time-Dependent Concordance**	**D-Calibration *p*-value (k=10)**	**IBS**
1	Nnet-survival	72.1235%	3.7668%	0.1243	RSF	70.9374%	73.2305%	0.12050
2	Cox Regression	72.0204%	93.8873%	0.1218	RIST	71.3790%	99.6067%	0.12096
3	Cox-nnet	71.6616%	5.3971%	0.1224	Cox Regression	72.0204%	93.8873%	0.12177
4	Time-Coded ANN	71.5640%	47.5766%	0.1233	CURE	71.2976%	66.4773%	0.12191
5	RIST	71.3790%	99.6067%	0.1210	Cox-nnet	71.6616%	5.3971%	0.12242
6	CURE	71.2976%	66.4773%	0.1219	Survival Tree (Likelihood)	68.7104%	99.2736%	0.12326
7	BART	71.2200%	72.2484%	0.1239	Time-Coded ANN	71.5640%	47.5766%	0.12331
8	RSF	70.9374%	73.2305%	0.1205	BART	71.2200%	72.2484%	0.12389
9	Survival Tree (Likelihood)	68.7104%	99.2736%	0.1233	CURT V2	68.3680%	97.6678%	0.12392
10	Survival Tree (Log Rank)	68.4682%	77.6016%	0.1239	Survival Tree (Log Rank)	68.4682%	77.6016%	0.12394
11	CURT V2	68.3680%	97.6678%	0.1239	Nnet-survival	72.1235%	3.7668%	0.12426
12	CURT V1	0.0000%	99.9787%	0.1491	CURT V1	0.0000%	99.9787%	0.14915

Table 6. Final Model Performance for RH

Rank	Ordered by Concordance				Ordered by IBS			
	Method	Time-Dependent Concordance	D-Calibration *p*-value (k=10)	IBS	Method	Time-Dependent Concordance	D-Calibration *p*-value (k=10)	IBS
1	CURE	70.0901%	51.4138%	0.1326	Nnet-survival	69.3910%	26.9462%	0.1300
2	RIST	70.0790%	94.5840%	0.1311	RIST	70.0790%	94.5840%	0.1311
3	Cox-nnet	69.9858%	12.3572%	0.1322	RSF	69.5290%	99.4790%	0.1312
4	Cox Regression	69.9082%	97.5226%	0.1328	Cox-nnet	69.9858%	12.3572%	0.1322
5	Time-Coded ANN	69.8737%	83.3974%	0.1342	CURE	70.0901%	51.4138%	0.1326
6	BART	69.6933%	79.0386%	0.1337	Survival Tree (Likelihood)	65.9155%	87.1188%	0.1328
7	RSF	69.5290%	99.4790%	0.1312	Cox Regression	69.9082%	97.5226%	0.1328
8	Nnet-survival	69.3910%	26.9462%	0.1300	CURT V2	66.9108%	99.4329%	0.1330
9	CURT V2	66.9108%	99.4329%	0.1330	BART	69.6933%	79.0386%	0.1337
10	Survival Tree (Likelihood)	65.9155%	87.1188%	0.1328	Time-Coded ANN	69.8737%	83.3974%	0.1342
11	Survival Tree (Log Rank)	65.0441%	71.1968%	0.1345	Survival Tree (Log Rank)	65.0441%	71.1968%	0.1345
12	CURT V1	55.8811%	79.9270%	0.1367	CURT V1	55.8811%	79.9270%	0.1367

6 Discussion

The above results provide an empirical demonstration of the ability of various survival modelling techniques to capture the aspects of model performance relevant for managerial decision-making. This section considers the results with respect to machine learning and statistical techniques, development of models for the proposed applications, and influence of performance measures.

The variability in model rankings as a function of both hospitals and basis for ranking have several implications. Focusing first on comparisons between the two hospitals, the Cox regression model had slightly worse relative performance on the more complex problem represented by RH for both ranking metrics. This is consistent with more general expectations regarding machine learning techniques being most promising for more complex problems. The relative ranking of the Cox-nnet and Cox regression models is consistent with the expectation of better machine learning performance on more complex problems. Cox-nnet represents a machine learning (ANN) extension of the statistical Cox model. This machine learning extension ranked below the statistical model on the less complex problem (GCUH) for both ranking measures, but this was reversed for the more complex problem (RH).

This variability is also relevant for the applications being proposed, particularly as the aspects of model performance being measured were motivated by these applications. Focusing on ERP, ERPP, and Expected Readmissions, these three applications consider the actual probabilities produced by the underlying model to identify acceptable levels of risk, probabilities of readmission, and expected readmissions respectively. If the underlying model is not well-calibrated, it cannot be reliably used for these applications. Further, to effectively improve and support administrative decision-making, the underlying models must also account for patient-specific characteristics in produced survival functions. The results demonstrate that a range of models may be suitable for these applications, being both well-calibrated and with relatively high discrimination. Across both bases for ranking and hospitals, RIST was most consistently high-performing, with Cox regression and Cox-nnet also of note. The best models in each scenario, however, were not consistent. As it appears unlikely that any single technique will be optimal across applications and settings, an institution aiming to apply a survival model in one of these applications should consider a breadth of models, with this work's results providing an informed starting point. A similar conclusion is relevant for the DRR application. In its simplest form, only model discrimination is important for DRR. This changes little, with model ranking between hospitals when only considering discrimination also being variable. It should also be noted that some element of calibration is likely to be desirable in a model applied for DRR, as this would support cost-benefit analyses for prospective interventions and improvement measurement for prior interventions. Again, driven by the variability in discrimination ranking and by the likely requirement for some level of calibration, institutions should consider a range of potential models to assess the range of discrimination-calibration combinations available for this application. This is particularly pertinent in the healthcare setting, where the magnitude of financial and patient welfare costs makes marginal improvement important.

Linked to the need for a context- and application-specific balance of both discrimination and calibration, the use of IBS for model selection and evaluation bears discussion. It has previously been noted that the IBS equation can be formulated as a sum of a calibration and discrimination component [54], making it a useful measure given these are the aspects of model performance determined to be relevant for managerial decision-making. It does not, however, explicitly report the contribution of these components. When considering concordance, the two survival tree models and the two CURT models performed notably worse than all other models, but this was less pronounced when ranked based on IBS. In particular, the survival tree using a splitting function based on a one-step likelihood was ranked sixth on both GCUH and RH. This is relevant and surprising because while almost all models exhibited acceptable D-calibration this model was characterised by notably lower discrimination. This may indicate that future research should consider modification of the IBS measure to adjust the relative balance between calibration and discriminations components, depending on the model applications considered. For example, the RH survival tree is ranked sixth in terms of IBS but only generates 20 unique survival curves which may be insufficient for certain applications. The emphasis placed on calibration by the unadjusted IBS measure and its use in model selection may also have been a contributor to almost all models being D-calibrated.

7 Conclusion

The major contribution of this work has been to identify relevant survival techniques for a range of practical applications supporting managerial decision-making for readmissions, as well as determining appropriate performance measures linked to these applications. This involved the proposal of four applications of survival models to support decision-making, three of which have not been suggested in prior research. Facilitating this, ten previously unconsidered machine learning techniques were identified and empirically assessed in terms of performance measures determined to be appropriate for these applications. Key conclusions of this work are:

- The relevant aspects of survival model performance for practical applications supporting managerial decision-making are the discrimination and calibration of risk over time predictions. Appropriate measures capturing these aspects are time-dependent concordance and D-calibration, neither of which have been used in prior readmission research.
- Many machine learning survival techniques are applicable for readmission modelling but have not been considered in previous readmission research.
- Machine learning survival techniques can improve on the most common statistical survival technique, particularly on more complex readmission problems, but no single technique is expected to consistently offer the best performance across applications and contexts.

- Survival techniques, both machine learning and statistical, can capture relevant aspects of readmission risk for a variety of applications supporting managerial decision-making.

It is expected that the suggested applications, which complement current classification model applications, should motivate greater consideration of survival techniques in future readmission research. In particular, the RIST, Cox-nnet, and Cox regression techniques should be prominent in future research, though considering a wide range of techniques is important to achieve the best combinations of discrimination and calibration in different settings. Secondary contributions of this work are in the provision of empirical findings comparing various machine learning survival techniques and in adding to the readmission research specific to Australia.

Future research should expand on this work in several areas. First, future research should establish the generalisability of the presented findings in terms of region, data sources, and cohort definitions. Secondly, the use of IBS for model selection implicitly assigns a relative weighting to calibration and discrimination, which could be modified to account for context-specific needs. Thirdly, as an initial proposal, the DRR and ERP model applications were considered in the general sense, and so detailed recommendations were not made as to how they should be implemented. Future research should aim to establish guidelines for the practical implementation of the proposed applications and assess the value derived from them.

Acknowledgements: The authors thank Healthcare Logic Pty Ltd, who provided the data and supported the project. We would also like to thank Jon Arni Steingrimsson, who shared code which was used in the training of Censoring Unbiased Regression Tree and Censoring Unbiased Regression Ensemble models.

Funding: This work was supported by an Australian Government Research Training Program Scholarship (James Todd) and Healthcare Logic Pty Ltd via Australian Innovation Connections Grants (ICG0043 and ICG000945).

8 Appendices

Appendix A – Feature Recoding

Prior to model construction, the Postcode field and AdmitWardCode fields were modified to reduce their dimensionality. The Postcode field was transformed to represent whether the patient's home address was from the inner city, outer city, or other.**Error! Reference source not found.**

Table A.1 iGC Feature Definition

iGC Field Values	Corresponding Postcodes Values
InnerGC	4214-4220, 4226-4230
OuterGC	4208-4210, 4212, 4221, 4223-4225, 4270-4272, 4275
Other	All others

Similarly, the AdmitWardCode field detailed the ward code the patient was admitted to. This field contained 70 unique values across both hospitals, with the seven most frequent values making up almost 90% (88.28%) of all observations. The possible values differ between the two hospitals and thus the recoding for this field was done for each hospital separately. For each hospital, the relative frequency of values was generated using the training data. All codes with a relative frequency below 5% were collected in an "Other" category.

Appendix B – Search Grids

In this appendix, the search grid of hyperparameters considered for the various machine learning techniques in this work are shown. The final hyperparameter values for the final models are bolded.

Table B.1 Search Grid Hyperparameters (Survival Trees)

Model Type	Parameters Varied	Parameter Values Considered	GCUH	RH
Survival Tree – One Step Likelihood	Cost-complexity parameter	0.00010, 0.00015, 0.00020, …, 0.0090	0.0004	0.001
		0.00100, 0.00200, 0.00300, …, 0.01000		
Survival Tree – Log Rank Statistic	Node depth	2, 3, 4, …, 20	7	6

Table B.2 Search Grid Hyperparameters (CURT V1)

Model Type	Parameters Varied	Parameter Values Considered	GCUH	RH
CURT	Model for conditional survival function	Survival Tree – Log Rank Statistic	Survival Tree – Log Rank Statistic	Survival Tree – Log Rank Statistic
		Random Survival Forest		
		Log-normal AFT		
		Log-logistic AFT		

Table B.3 Search Grid Hyperparameters (CURT V2)

Model Type	Parameters Varied	Parameter Values Considered	GCUH	RH
CURT	Model for conditional survival function	Survival Tree – Log Rank Statistic	Survival Tree – Log Rank Statistic	Survival Tree – Log Rank Statistic
		Random Survival Forest		
		Log-logistic AFT		
		Log-normal AFT		
	Cost-complexity parameter	0.000010, 0.000015, 0.000020, …, 0.000095	0.000035	0.00045
		0.000100, 0.000150, 0.000200, …, 0.000950		
		0.00100, 0.00200, 0.00300, …, 0.01000		

Table B.4 Search Grid Hyperparameters (Random Survival Forest)

Model Type	Parameters Varied	Parameter Values Considered	GCUH	RH
Random Survival Forest	Number of trees	500, 750, 1000	1000	750
	Covariates considered at each split	1, 2, 3, …, 8	3	3
	Terminal node size	3, 15	15	15

Table B.5 Search Grid Hyperparameters (CURE)

Model Type	Parameters Varied	Parameter Values Considered	GCUH	RH
CURE	Model for conditional survival function	Survival Tree – Log Rank Statistic Random Survival Forest	Random Survival Forest	Survival Tree – Log Rank Statistic
	Number of trees	100, 250, 500, 750, 1000	750	500
	Covariates considered at each split	1, 2, 3, …, 8	5	6
	Terminal node size	3, 10, 20	20	20

Table B.6 Search Grid Hyperparameters (RIST)

Model Type	Parameters Varied	Parameter Values Considered	GCUH	RH
RIST	Number of trees	30, 40, 50, 60, 70	60	60
	Covariates considered at each split	3, 5, 7	7	5
	Terminal node size	10, 20, 30, 40, 50, 100	20	20
	Imputation cycles	1, 2, 3	2	1

Table B.7 Parameters used in the BART Model (GCUH)

Model Type	Parameter	Parameter Values Considered
BART (GCUH)	Number of trees	50
	Draws from the posterior	200
	Burn-in sample	250
	Thinning	10

Table B.8 Parameters used in the BART Model (RH)

Model Type	Parameter	Parameter Values Considered
BART (RH)	Number of trees	50
	Draws from the posterior	500
	Burn-in sample	250
	Thinning	10

Table B.9 Search Grid Hyperparameters (Nnet-survival)

Model Type	Parameters Varied	Values Considered	GCUH	RH
Nnet-survival	Hidden layers and nodes	1 layer, 5 nodes	2 layers, 15 and 10 nodes	2 layers, 15 and 10 nodes
		1 layer, 10 nodes		
		1 layer, 15 nodes		
		2 layers, 10 and 10 nodes		
		2 layers, 15 and 10 nodes		
	Epochs	100, 200, 300, …, 1500	600	1100
	Mini-batch size	128, 256, 512	256	128
	Regularisation penalty (L2)	$\exp(-4)$, $\exp(-5)$, $\exp(-6)$	$\exp(-5)$	$\exp(-5)$
	Intervals	20, 30, 40	20	40

Table B.10 Search Grid Hyperparameters (Time-Coded ANN)

Model Type	Parameters Varied	Values Considered	GCUH	RH
Time-Coded ANN	Hidden layers and nodes	1 layer, 5 nodes 1 layer, 10 nodes 1 layer, 15 nodes 1 layer, 20 nodes 2 layers, 10 and 10 nodes 2 layers, 15 and 10 nodes	1 layer, 10 nodes	1 layer, 20 nodes
	Epochs	100, 200, 300, …, 1500	700	1300
	Mini-batch size	128, 256, 512, 1024, 2048, 4096, 8192	2048	2048
	Regularisation penalty (L2)	$\exp(-4)$, $\exp(-5)$, $\exp(-6)$	$\exp(-6)$	$\exp(-6)$

Table B.11 Search Grid Hyperparameters (Cox-nnet)

Model Type	Parameters Varied	Values Considered	GCUH	RH
Cox-nnet	Hidden layers and nodes	1 layer, 8 nodes 1 layer, 14 nodes 1 layer, 21 nodes 2 layers, 5 and 5 nodes 2 layers, 7 and 4 nodes	1 layer, 14 nodes	2 layers, 7 and 4 nodes
	Epochs	50, 100, 200, 500, 600, 1000	1000	1000
	Regularisation penalty (L2)	$\exp(-5)$, $\exp(-6)$	$\exp(-6)$	$\exp(-6)$
	Batch normalisation	Yes, No	No	No

Appendix C – ANN Implementation Details

Several other details of model implementation were not varied but should be specified:

- Activation function – The ReLU activation function was used in the hidden layers of all candidate Nnet-survival models and time-coded ANN models, and the sigmoid activation function was used in the output layer to ensure predictions were in the range [0,1]. To remain consistent with the original implementation [11], the hyperbolic tangent activation function was used for the Cox-nnet model rather than ReLU.
- Data Processing – All data were converted to a numeric format for model training and all covariates were standardised based on the entire training data for each hospital through the usual approach for networks: $x_{i,j}^{*} = \left(x_{i,j} - \mu_j\right)/\sigma_j$. The same standardisation parameters were used for training and test data.
- Variables – For each hospital, variables with low predictive value were not included. The variables used in ANN construction were those which were included in any form in the final Cox regression model or would have been included in a logistic regression model using the three-stage process described in Section 4.3.1~~4.4.1~~ under AIC or BIC. This included 14 out of 19 available variables.

9 References

[1] A. Alaeddini, J.E. Helm, P. Shi, S.H.A. Faruqui, An integrated framework for reducing hospital readmissions using risk trajectories characterization and discharge timing optimization, IISE Transactions on Healthcare Systems Engineering, 9(2) (2019) 172-185.
[2] S. Alajmani, H. Elazhary, Hospital readmission prediction using machine learning techniques: A comparative study, International Journal of Advanced Computer Science and Applications, 10(4) (2019) 212-220.
[3] A. Alassaad, H. Melhus, M. Hammarlund-Udenaes, M. Bertilsson, U. Gillespie, J. Sundström, A tool for prediction of risk of rehospitalisation and mortality in the hospitalised elderly: secondary analysis of clinical trial data, BMJ Open, 5(2) (2015) e007259.
[4] A. Andres, A. Montano-Loza, R. Greiner, M. Uhlich, P. Jin, B. Hoehn, D. Bigam, J.A.M. Shapiro, N.M. Kneteman, A novel learning algorithm to predict individual survival after liver transplantation for primary sclerosing cholangitis, PLoS One, 13(3) (2018) e0193523.
[5] L. Antolini, P. Boracchi, E. Biganzoli, A time-dependent discrimination index for survival data, Stat Med, 24(24) (2005) 3927-3944.
[6] A. Artetxe, A. Beristain, M. Graña, Predictive models for hospital readmission risk: A systematic review of methods, Computer Methods and Programs in Biomedicine, 164(2018) 49-64.
[7] Australian Commission on Safety and Quality in Healthcare, Avoidable Hospital Readmissions, in, (2019).
[8] C. Baechle, A. Agarwal, R. Behara, X.Q. Zhu, Ieee, Latent Topic Ensemble Learning for Hospital Readmission Cost Reduction, in: 2017 International Joint Conference on Neural Networks, (IEEE, New York, 2017), pp. 4594-4601.
[9] E. Biganzoli, P. Boracchi, E. Marubini, A general framework for neural network models on censored survival data, Neural Networks, 15(2) (2002) 209-218.
[10] Centers for Medicare and Medicaid Services, Hospital Readmissions Reduction Program (HRRP), in, (2020).
[11] T. Ching, X. Zhu, L.X. Garmire, Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data, PLoS Computational Biology, 14(4) (2018) e1006076.
[12] H.A. Chipman, E.I. George, R.E. McCulloch, BART: BAYESIAN ADDITIVE REGRESSION TREES, The Annals of Applied Statistics, 4(1) (2010) 266-298.
[13] J.M. Corrigan, J.B. Martin, Identification of factors associated with hospital readmission and development of a predictive model, Health Serv Res, 27(1) (1992) 81-101.
[14] M. De Laurentiis, S. De Placido, A.R. Bianco, G.M. Clark, P.M. Ravdin, A Prognostic Model That Makes Quantitative Estimates of Probability of Relapse for Breast Cancer Patients, Clinical Cancer Research, 5(12) (1999) 4133.
[15] M. Deschepper, K. Eeckloo, D. Vogelaers, W. Waegeman, A hospital wide predictive model for unplanned readmission using hierarchical ICD data, Computer Methods and Programs in Biomedicine, 173(2019) 177-183.
[16] A. Eleuteri, A.F.G. Taktak, Support vector machines for survival regression, in: 8th International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics, (2012), pp. 176-189.
[17] L. Evers, C.-M. Messow, Sparse kernel methods for high-dimensional survival data, Bioinformatics, 24(2008) 1632-1638.
[18] S. Farrell, A. Mitnitski, K. Rockwood, A. Rutenberg, Generating synthetic aging trajectories with a weighted network model using cross-sectional data, Scientific Reports, 10(1) (2020) 19833.
[19] C. Fischer, H.F. Lingsma, P.J. Marang-van de Mheen, D.S. Kringos, N.S. Klazinga, E.W. Steyerberg, Is the readmission rate a valid quality indicator? A review of the evidence, PLoS One, 9(11) (2014) e112282.
[20] J. Futoma, J. Morris, J. Lucas, A comparison of models for predicting early hospital readmissions, Journal of Biomedical Informatics, 56(2015) 229-238.
[21] A. Garmendia, M. Graña, J.M. Lopez-Guede, S. Rios, Neural and statistical predictors for time to readmission in emergency departments: A case study, Neurocomputing, 354(2019) 3-9.
[22] M.F. Gensheimer, B. Narasimhan, A scalable discrete-time survival model for neural networks, PeerJ, 7(2019) e6257-e6257.

[23] Y. Goldberg, M. Kosorok, Support vector regression for right censored data, Electron. J. Stat., 11(1) (2017) 532-569.
[24] M. Graña, J.M. Lopez-Guede, J. Irazusta, I. Labayen, A. Besga, Modelling hospital readmissions under frailty conditions for healthy aging, Expert Systems, (2019).
[25] M. Grzyb, A. Zhang, C. Good, K. Khalil, B. Guo, L. Tian, J. Valdez, Q. Gu, Multi-task cox proportional hazard model for predicting risk of unplanned hospital readmission, in: 2017 Systems and Information Engineering Design Symposium, SIEDS 2017, (2017), pp. 265-270.
[26] H. Haider, B. Hoehn, S. Davis, R. Greiner, Effective Ways to Build and Evaluate Individual Survival Distributions, Journal of Machine Learning Research, 21(2020) 1-63.
[27] A. Hammoudeh, G. Al-Naymat, I. Ghannam, N. Obied, Predicting hospital readmission among diabetics using deep learning, in: Procedia Computer Science, (2018), pp. 484-489.
[28] 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, 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, 10(10) (2015) e0140271-e0140271.
[29] F.E. Harrell, Jr., K.L. Lee, D.B. Mark, Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors, Stat Med, 15(4) (1996) 361-387.
[30] T. Hothorn, P. Bühlmann, S. Dudoit, A. Molinaro, M.J. Van Der Laan, Survival ensembles, Biostatistics, 7(3) (2006) 355-373.
[31] T. Hothorn, B. Lausen, A. Benner, M. Radespiel-Tröger, Bagging survival trees, Statistics in Medicine, 23(1) (2004) 77-91.
[32] Independent Hospital Pricing Authority, National Efficient Price Determination 2021-22, in, (2021).
[33] H. Ishwaran, U.B. Kogalur, E.H. Blackstone, M.S. Lauer, Random Survival Forests, The Annals of Applied Statistics, 2(3) (2008) 841-860.
[34] S.F. Jencks, M.V. Williams, E.A. Coleman, Rehospitalizations among patients in the Medicare fee-for-service program, N Engl J Med, 360(14) (2009) 1418-1428.
[35] J.M. Jerez-Aragonés, J.A. Gómez-Ruiz, G. Ramos-Jiménez, J. Muñoz-Pérez, E. Alba-Conejo, A combined neural network and decision trees model for prognosis of breast cancer relapse, Artificial Intelligence in Medicine, 27(1) (2003) 45-63.
[36] S. Jiang, K.S. Chin, G. Qu, K.L. Tsui, An integrated machine learning framework for hospital readmission prediction, Knowledge-Based Systems, 146(2018) 73-90.
[37] R.P. Jindal, D.K. Gauri, G. Singh, S. Nicholson, Factors influencing hospital readmission penalties: Are they really under hospitals' control?, Decision Support Systems, 110(2018) 58-70.
[38] D. Kansagara, H. Englander, A. Salanitro, D. Kagen, C. Theobald, M. Freeman, S. Kripalani, Risk prediction models for hospital readmission: A systematic review, JAMA - Journal of the American Medical Association, 306(15) (2011) 1688-1698.
[39] F.M. Khan, V. Bayer-Zubek, Support vector regression for censored data (SVRc): A novel tool for survival analysis, in: IEEE International Conference on Data Mining, (2008), pp. 863-868.
[40] D.-H. Kim, H.-C. Jeong, Weighted LS-SVM Regression for Right Censored Data, Communications for Statistical Applications and Methods, 13(3) (2006) 765-776.
[41] S. Kripalani, C.N. Theobald, B. Anctil, E.E. Vasilevskis, Reducing hospital readmission rates: current strategies and future directions, Annu Rev Med, 65(2014) 471-485.
[42] S.R. Kristensen, M. Bech, W. Quentin, A roadmap for comparing readmission policies with application to Denmark, England, Germany and the United States, Health Policy, 119(3) (2015) 264-273.
[43] H.M. Krumholz, S.I. Chaudhry, J.A. Spertus, J.A. Mattera, B. Hodshon, J. Herrin, Do Non-Clinical Factors Improve Prediction of Readmission Risk?: Results From the Tele-HF Study, JACC: Heart Failure, 4(1) (2016) 12-20.
[44] M. Leblanc, J. Crowley, Relative risk trees for censored survival data, Biometrics, 48(2) (1992) 411.

[45] J.D. Long, J.A. Mills, Joint modeling of multivariate longitudinal data and survival data in several observational studies of Huntington's disease, BMC Medical Research Methodology, 18(1) (2018) 138.
[46] B. Padhukasahasram, C.K. Reddy, Y. Li, D.E. Lanfear, Joint impact of clinical and behavioral variables on the risk of unplanned readmission and death after a heart failure hospitalization, PloS one, 10(6) (2015) e0129553-e0129553.
[47] L. Pereira, C. Choquet, A. Perozziello, M. Wargon, G. Juillien, L. Colosi, R. Hellmann, M. Ranaivoson, E. Casalino, Unscheduled-return-visits after an emergency department (ED) attendance and clinical link between both visits in patients aged 75 years and over: a prospective observational study, in: PloS one, (2015), pp. e0123803.
[48] H.N. Pham, A. Chatterjee, B. Narasimhan, C.W. Lee, D.K. Jha, E.Y. Fai Wong, S. Ellyanti, Q.H. Nguyen, B.P. Nguyen, M.C.H. Chua, Predicting hospital readmission patterns of diabetic patients using ensemble model and cluster analysis, in: Proceedings of 2019 International Conference on System Science and Engineering, ICSSE 2019, (2019), pp. 273-278.
[49] J. Shim, C. Hwang, Support vector censored quantile regression under random censoring, (2009).
[50] P.K. Shivaswamy, W. Chu, M. Jansche, A Support Vector Approach to Censored Targets, in: Seventh IEEE International Conference on Data Mining (ICDM 2007), (2007), pp. 655-660.
[51] R.A. Sparapani, B.R. Logan, R.E. McCulloch, P.W. Laud, Nonparametric survival analysis using Bayesian Additive Regression Trees (BART), Statistics in Medicine, 35(16) (2016) 2741-2753.
[52] J.A. Steingrimsson, L. Diao, A. Molinaro, R.L. Strawderman, Doubly robust survival trees, Statistics in Medicine, 35(20) (2016) 3595-3612.
[53] J.A. Steingrimsson, L. Diao, R.L. Strawderman, Censoring Unbiased Regression Trees and Ensembles, Journal of the American Statistical Association, 114(525) (2019) 370-383.
[54] E.W. Steyerberg, A.J. Vickers, N.R. Cook, T. Gerds, M. Gonen, N. Obuchowski, M.J. Pencina, M.W. Kattan, Assessing the performance of prediction models: a framework for traditional and novel measures, Epidemiology, 21(1) (2010) 128-138.
[55] A.D. Tulloch, A.S. David, G. Thornicroft, Exploring the predictors of early readmission to psychiatric hospital, Epidemiology and Psychiatric Sciences, 25(2) (2016) 181-193.
[56] L. Turgeman, J.H. May, A mixed-ensemble model for hospital readmission, Artificial Intelligence in Medicine, 72(2016) 72-82.
[57] V. Van Belle, K. Pelckmans, J.A.K. Suykens, S. Van Huffel, Support Vector Machines for Survival Analysis, in: Proceedings of the Third International Conference on Computational Intelligence in Medicine and Healthcare, (2007).
[58] V. Van Belle, K. Pelckmans, J.A.K. Suykens, S. Van Huffel, Survival SVM: A practical scalable algorithm, in: European Symposium on Artificial Neural Networks, (2008), pp. 89-94.
[59] V. Van Belle, K. Pelckmans, J.A.K. Suykens, S. Van Huffel, Additive survival least-squares support vector machines, Statistics in Medicine, 29(2) (2010) 296-308.
[60] V. Van Belle, K. Pelckmans, S. Van Huffel, J.A.K. Suykens, Support vector methods for survival analysis: a comparison between ranking and regression approaches, Artificial Intelligence in Medicine, 53(2) (2011) 107-118.
[61] C. van Walraven, C. Bennett, A. Jennings, P.C. Austin, A.J. Forster, Proportion of hospital readmissions deemed avoidable: a systematic review, Cmaj, 183(7) (2011) E391-402.
[62] D. Wang, Z. Jing, K. He, L.X. Garmire, Cox-nnet v2.0: improved neural-network based survival prediction extended to large-scale EMR data, Bioinformatics, (2021).
[63] H. Wang, Z. Cui, Y. Chen, M. Avidan, A.B. Abdallah, A. Kronzer, Predicting Hospital Readmission via Cost-Sensitive Deep Learning, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 15(6) (2018) 1968-1978.
[64] L. Wang, B. Porter, C. Maynard, C. Bryson, H. Sun, E. Lowy, M. McDonell, K. Frisbee, C. Nielson, S.D. Fihn, Predicting Risk of Hospitalization or Death Among Patients With Heart Failure in the Veterans Health Administration, The American Journal of Cardiology, 110(9) (2012) 1342-1349.
[65] P. Wolff, M. Grana, S.A. Ríos, M.B. Yarza, Machine Learning Readmission Risk Modeling: A Pediatric Case Study, BioMed Research International, 2019(2019).
[66] F. Yan, X. Lin, R. Li, X. Huang, Functional principal components analysis on moving time windows of longitudinal data: dynamic prediction of times to event, Journal of the Royal Statistical Society: Series C (Applied Statistics), 67(4) (2018) 961-978.

[67] S. Yu, F. Farooq, A. van Esbroeck, G. Fung, V. Anand, B. Krishnapuram, Predicting readmission risk with institution-specific prediction models, Artificial Intelligence in Medicine, 65(2) (2015) 89-96.
[68] J. Zhang, S.S. Lam, S. Poranki, A classification model for hospital readmission using combined neural networks, in, (Institute of Industrial Engineers, 2013), pp. 1088-1097.
[69] J. Zhang, S.W. Yoon, M.T. Khasawneh, K. Srihari, S. Poranki, Hospital readmission prediction using swarm intelligence-based support vector machines, in: IIE Annual Conference and Expo 2013, (2013), pp. 1522-1531.
[70] S. Zheng, A. Hanchate, M. Shwartz, One-year costs of medical admissions with and without a 30-day readmission and enhanced risk adjustment, BMC Health Services Research, 19(1) (2019) 155.
[71] R. Zhu, M.R. Kosorok, Recursively Imputed Survival Trees, Journal of the American Statistical Association, 107(497) (2012) 331-340.
[72] H.M. Zolbanin, D. Delen, Processing electronic medical records to improve predictive analytics outcomes for hospital readmissions, Decision Support Systems, 112(2018) 98-110.