

# BMJ Open Development of a risk predictive scoring system to identify patients at risk of representation to emergency department: a retrospective population-based analysis in Australia

Euijoon Ahn,<sup>1</sup> Jinman Kim,<sup>1,2</sup> Khairunnessa Rahman,<sup>3</sup> Tanya Baldacchino,<sup>2</sup> Christine Baird<sup>3</sup>

**To cite:** Ahn E, Kim J, Rahman K, *et al.* Development of a risk predictive scoring system to identify patients at risk of representation to emergency department: a retrospective population-based analysis in Australia. *BMJ Open* 2018;**8**:e021323. doi:10.1136/bmjopen-2017-021323

► Prepublication history and additional material for this paper are available online. To view, please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2017-021323>).

Received 22 December 2017  
Revised 22 June 2018  
Accepted 22 August 2018



© Author(s) (or their employer(s)) 2018. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>School of Information Technologies, University of Sydney, Sydney, New South Wales, Australia

<sup>2</sup>Nepean Telehealth Technology Centre, Nepean Hospital, Penrith, New South Wales, Australia

<sup>3</sup>Integrated Care Initiative, Nepean Hospital, Penrith, New South Wales, Australia

## Correspondence to

Dr Jinman Kim;  
[jinman.kim@sydney.edu.au](mailto:jinman.kim@sydney.edu.au)

## ABSTRACT

**Objective** To examine the characteristics of frequent visitors (FVs) to emergency departments (EDs) and develop a predictive model to identify those with high risk of a future representations to ED among younger and general population (aged ≤70 years).

**Design and setting** A retrospective analysis of ED data targeting younger and general patients (aged ≤70 years) were collected between 1 January 2009 and 30 June 2016 from a public hospital in Australia.

**Participants** A total of 343 014 ED presentations were identified from 170 134 individual patients.

**Main outcome measures** Proportion of FVs (those attending four or more times annually), demographic characteristics (age, sex, indigenous and marital status), mode of separation (eg, admitted to ward), triage categories, time of arrival to ED, referral on departure and clinical conditions. Statistical estimates using a mixed-effects model to develop a risk predictive scoring system.

**Results** The FVs were characterised by young adulthood (32.53%) to late-middle (26.07%) aged patients with a higher proportion of indigenous (5.7%) and mental health-related presentations (10.92%). They were also more likely to arrive by ambulance (36.95%) and leave at own risk without completing their treatments (9.8%). They were also highly associated with socially disadvantage groups such as people who have been divorced, widowed or separated (12.81%). These findings were then used for the development of a predictive model to identify potential FVs. The performance of our derived risk predictive model was favourable with an area under the receiver operating characteristic (ie, C-statistic) of 65.7%.

**Conclusion** The development of a demographic and clinical profile of FVs coupled with the use of predictive model can highlight the gaps in interventions and identify new opportunities for better health outcome and planning.

## INTRODUCTION

Emergency departments (EDs) are designed to manage acute episodic medical diseases or injury. However, an important proportion of patients return to EDs frequently and

## Strengths and limitations of this study

- Limited researches have been carried out with a focus on understanding frequent visitors (FVs) to emergency departments (EDs) or the prediction of those with high risk of a future representation to EDs among younger and general population.
- This study examined the demographic patterns and clinical conditions of FVs to EDs and derived a risk predictive scoring system to target younger and general patients (aged ≤70 years), that is not restricted by certain chronic diseases or older age groups.
- This study was strengthened by using all available data collected during ED presentations.
- This is a retrospective population-based analysis and the risk predictive scoring system was derived using data from a single hospital.

unexpectedly. Representations to EDs have been directly associated with increased utilisation of inpatient care, multiple admissions and intensive care.<sup>1</sup> Improving the management of these high cost patients is therefore important for better health outcome and healthcare planning. A core strategy to reduce potentially preventable representations is the development of a predictive model to identify those at risk of health deterioration and hospitalisation accurately, and therefore tailored integrated intervention can be provided before substantial avoidable representations have been incurred.<sup>2</sup>

Many predictive models have been developed to target and calibrate resources for interventions for at-risk patients, thereby reducing the overall cost of the interventions.<sup>2</sup> These models used routinely collected data to understand the characteristics of those patients, and then identify people at high risk

of a future admission or readmission.<sup>3</sup> The development of such predictive models has been of great interest in recent years. However, these models were only specifically designed to predict patients with high risk of a future readmission rather than a future representation to EDs, often targeting only older patients with certain chronic diseases.<sup>4–8</sup> Limited researches have been carried out with a focus on understanding of frequent visitors (FVs) to EDs or the prediction of a future representation to EDs. While both prediction of readmissions and representations can benefit patients and hospital together, specific characteristics that need to be considered in design of predictive modelling are greatly different from each other.

To understand the distinct characteristics of FVs to EDs, we first examined the demographic patterns and clinical conditions of FVs using a large scale (7.5 years) of hospital ED data. Using such large scale of ED data, we sought to understand and describe distinct patterns of the FVs. The characteristics we learnt were then used for the development of a predictive model, which we named as a Risk Predictive Scoring System for ED (RPSS-ED). The RPSS-ED is designed to target younger and general patients, who are not restricted by certain chronic diseases or older age groups. It is easy to use and relies only on a small number of variables that are easily collected from the electronic medical record system. We suggest that the development of a demographic and clinical profile of FVs coupled with the use of predictive model will highlight the gaps in interventions and identify new opportunities for better health outcome and planning.

## METHODS

### Patient and public involvement

Patients and public were not involved in the development of the research questions or in the design of the study. Dissemination of the general results (no personal data) would be made on demand.

### Study setting and population

This was a retrospective analysis of an ED data collected from the Nepean hospital in New South Wales (NSW), Australia. Nepean Blue Mountains Local Health District (NBMLHD) covers both urban and semi-rural areas, covering approximately 9179 km. The estimated resident population of NBMLHD in 2011 is 345 564, which includes an indigenous community (2.6%).<sup>9</sup> The number of younger aged profiles and indigenous people have been steadily increasing in recent years. The number of ED presentations is projected to increase by 33% in 2022 along with increases in mental healthcare, rehabilitation and recovery, cancer care and renal dialysis.<sup>9</sup> The increasing populations of both younger and elderly people introduce new and unique challenges in healthcare demands, planning and service delivery.

Data were extracted from Health Information Exchange system, which included all available ED records from the hospital from the period 1 January 2009 to 30 June 2016.

Recent studies suggest that the FVs to ED are highly associated with both elderly age groups and with younger age groups<sup>10 11</sup> and discuss the importance of understanding the younger age groups.<sup>12</sup> In addition, identification of FVs from older age group (aged >70 years) was relatively trivial as many of them were already FVs of the hospital, often suffering from certain chronic diseases. Such patients also tend to have distinct characteristics with a skewed distribution, introducing confounding (eg, survival) bias in the analysis. Identifying FVs from younger and general population (aged ≤70 years), however, is more challenging as their characteristics are more complex and heterogeneous. We therefore targeted younger and general patients aged ≤70 years in our analysis. We included all patients' information (including any chronic conditions) if they visited the ED during this period. A total of 343 014 ED presentations were identified from 170 134 individual patients.

Our data contain demographic information including age, sex, marital status, indigenous status, patient post-code and county of birth. Other clinical variables including referral source (eg, self-referred, general practice and specialist), mode of arrival (eg, ambulance or others), presenting problem and mode of separation (eg, admitted to ward, not a critical care ward) were collected. Triage categories with scale of 1–5 (as defined by Australasian Triage Scale) were used. EDs diagnoses were categorised into subgroups based on headers of International Statistical Classification of Diseases and Related Health Problems, 10th revision of Australian modification.

We adopted the definition of FVs to be patients attending to EDs more than four times annually.<sup>13 14</sup>

We used SAS V.9.4 and Matlab 2017a for data analysis/manipulation and model development, respectively.

### Population analysis of FVs to ED

The outcomes of FVs with different number of attendances to ED were compared with those of all visitors and non-FVs to identify the distinct patterns of FVs, consistent with other studies.<sup>10 11</sup>

### Development of a risk predictive scoring system for ED

Descriptive statistics were used to identify a subset of candidate predictors for multivariable statistical analysis. We excluded candidate predictors with <10 expected events to avoid model instability; variables with variance inflation factors for multicollinearity exceeding a threshold of 10; those with unstructured keywords (eg, presenting problem). The continuous predictors are initially calibrated to find appropriate subgroups (eg, age bins) that maximise the differences or similarities of characteristics using maximum likelihood monotone coarse classifier algorithm.<sup>15</sup> We used a mixed-effects model to understand the complexities (ie, correlations) within individual patient. We selected candidate predictors for multivariable logistic regression by testing the bivariable association of each fixed-effects predictor with outcomes at the 5% significance level. The patient-specific variations

**Table 1** A mixed-effects model used to derive the risk score

Variable	Coefficients	SE	T-Stat	P values
Age	0.630	0.079	7.959	<0.001
Sex	0.362	0.426	0.850	0.394
Marital status	0.504	0.066	7.572	<0.001
Indigenous status	1.006	0.076	13.169	<0.001
Mode of arrival	0.077	0.040	1.906	<0.05
Triage category	-0.012	0.062	-0.205	0.837
Mode of separation	0.088	0.054	1.631	0.102
Referred to on departure	0.176	0.058	2.994	<0.001
ICD-10	0.298	0.025	11.594	<0.001

ICD-10, International Statistical Classification of Diseases and Related Health Problems, 10th revision.

were measured using a random-effects intercept. We then applied the coefficients to the fixed-effects variables and translated into a point-scoring system. The points are scaled for simpler interpretation by setting a target level of 10 points with a target odds level of 2 and points-to-double-odds of 1. We divided the data into 10 groups from the training data, fitting the model on 90% (ie, derivation group) and using the model to predict remaining 10% of the data (ie, internal validation group). This process was repeated 10 times using each set (ie, 10-fold cross-validation). Variables included in our RPSS-ED are shown in [table 1](#).

We also evaluated our model using a separate external validation dataset (ie, the dataset was not included during the model derivation stage). They were patients of the Nepean hospital who were receiving tailored integrated intervention. The clinicians manually selected patients 1) with multiple health and social care needs or 2) who have presented to ED 10 times in the past year. There were total 77 enrolled patients with a total of 3142 presentations since 2009–2016, with average of 40 presentations per patient. We computed the risk scores associated with each patient using the derived model.

## RESULTS

### Characteristics of FVs compared with visitors to EDs

The characteristics of FVs compared against those of all visitors to EDs from January 2009 to June 2016 are shown in [table 2](#). Given the number of repeat representations in the database, the unit of analysis for clinical variables was EDs presentation rather than individual cases. Demographic information, such as age, sex and marital status, were analysed at individual level. The frequent representations (4+ per annum) were characterised by young adulthood (20–39 years) to late-middle (40–59 years) aged patients, accounting for 32.53% and 26.07%, respectively. [Figure 1](#) shows the age-specific analysis based on annual frequency

of attendance to EDs. The age group of 20–39 years was the highest with FVs who attended 9+ times per annum. As expected, the younger aged groups of 0–19 years had least contribution to FVs, having consistently smallest proportion compared with other age groups. Marital status has shown to be another important characteristic of frequent representations; they were highly related to people who have been widowed (2.16%), divorced (6.52%) or separated (4.13%). Frequent representations were also shown to be strongly associated with acuity presentations (52.34% in triage category 1, 2 or 3) and higher proportion of mental and behaviour disorders (10.92%) or endocrine and metabolic diseases (1.81%). Use of ambulance or police was also highly linked with frequent representations, accounting for 36.95%. Many FVs either did not wait to complete their treatment or leave the EDs at their own risk. FVs are highly associated with indigenous people with proportion of 5.7% compared with non-FVs of 3.08%. Similarly, consistent demographic patterns and clinical conditions were reported from the very frequent FVs (10+ per annum).

[Table 3](#) shows the patterns of FVs based on the time of arrival to ED. The data were further split into six hourly group of the day, day of the week and seasonal groups. Compared with the characteristics of non-FVs, the very FVs are more likely to come to ED on Wednesday (14.24%), Thursday (14.4%) or Friday (14.28%) of the week or between 18:00 and 12:00 hours (36.47%). We also noted that the autumn was the busiest season with proportion of over 27% across all four different cohorts.

### Risk predictive scoring system for EDs

[Table 1](#) summarises the coefficients learnt for our selected nine variables using the mixed-effects model. The full list of variables with corresponding points are available on the online supplementary appendix. Total score associated with individual ED presentation can be calculated by summing all corresponding points of the variables. The lower total score indicates higher risk of representation to EDs. The two most important metrics in evaluating the performance of a predictive model are the accuracy of detecting FVs (true positive rate or sensitivity) and the accuracy of detecting non-FVs (false positive rate). The performance of the model from internal validation is therefore measured using receiver operating characteristics (ROC) curve and the Kolmogorov-Smirnov (K-S) plot ([figure 2](#)). Using 10-fold cross-validation, our derived model achieved area under the ROC curve (AUROC) of 65.7%. A bigger the size of AUROC curve indicates higher overall accuracy of the predictive model (an AUROC of 0.5 indicates no discrimination and so the higher the curve above the diagonal the better the predictive accuracy). Similarly, the K-S plot is a common statistic used to measure the predictive power of the scoring system. It shows the distribution of FVs and the distribution of non-FVs on the same plot. The key statistic of interest is to identify the maximum difference between these two distributions (sensitivity minus false positive). The score

**Table 2** Clinical and demographic characteristics of FVs to ED compared with those of non-FVs and whole visitors

A total of 343 014 ED presentations from 170 134 individual patients	Whole population	FVs (39979 presentations) (4+ per annum)	Very FVs (7570 presentations) (10+ per annum)	Non-FVs (303035 presentations) (<4 per annum)
Total presentations since 2009–2016	Percentage (frequency)			
<b>Demographic</b>				
0–9	18.02 (30 651)	11.93 (641)	2.22 (7)	18.05 (30 520)
10–19	14.93 (25 406)	12.47 (670)	10.48 (33)	14.93 (25 258)
20–39	31.80 (54 102)	32.53 (1747)	40 (126)	31.82 (53 811)
40–59	24.24 (41 244)	26.07 (1400)	34.92 (110)	24.24 (40 988)
60–70	11.01 (18 729)	17 (913)	12.38 (39)	10.97 (18 548)
Male	52.64 (89 582)	49.1 (2637)	46.98 (148)	52.65 (89 041)
Female	47.34 (80 548)	50.9 (2734)	53.02 (167)	47.35 (80 082)
Indigenous	2.97 (5056)	5.7 (306)	9.54 (30)	3.08 (5214)
<b>Arrival mode</b>				
Ambulance (1)	24.22 (83 086)	36.95 (14 771)	55.73 (4218)	22.54 (68 315)
Private car (3)	73.28 (251 365)	59.45 (23 766)	39.77 (3010)	75.11 (227 599)
Police (7)	1.32 (4524)	1.95 (780)	2.66 (201)	1.24 (3744)
<b>Triage category</b>				
1	0.58 (1996)	0.74 (296)	0.96 (73)	0.56 (1700)
2	14.76 (50 642)	17.48 (6987)	18.08 (1369)	14.41 (43 655)
3	29.28 (100 433)	34.12 (13 642)	34.45 (2608)	28.64 (86 791)
4	43.61 (149 591)	37.11 (14 836)	34.8 (2634)	44.47 (134 755)
5	11.76 (40 338)	10.55 (4217)	11.7 (886)	11.92 (36 121)
<b>Marital status</b>				
1 (never married)	53.39 (90 492)	50.05 (2687)	49.52 (156)	53.33 (89 873)
2 (widowed)	1.20 (2027)	2.16 (116)	2.86 (9)	1.21 (2031)
3 (divorced)	3.56 (6040)	6.52 (350)	10.79 (34)	3.61 (6076)
4 (separated)	2.19 (3709)	4.13 (222)	6.98 (22)	2.20 (3707)
5 (married)	37.86 (64 163)	36.58 (1964)	29.52 (93)	37.98 (63 995)
<b>Mode of separation</b>				
Admitted and discharged as inpatient within ED	7.28 (24 972)	9.39 (3753)	11.6 (878)	7.00 (21 219)
Admitted to ward, not a critical care ward	31.05 (106 519)	35.73 (14 284)	26.99 (2043)	30.44 (92 235)
Admitted: left at own risk	0.35 (1208)	0.86 (343)	1.41 (107)	0.29 (865)
Admitted: to critical care ward	0.92 (3141)	1.26 (503)	1.1 (83)	0.87 (2638)
Departed: did not wait	1.76 (6031)	2.4 (958)	3.41 (1)	1.67 (5073)
Departed: left at own risk	5.01 (17 183)	6.54 (2613)	8.82 (258)	4.81 (14 570)
Departed: treatment completed	52.38 (179 657)	42.63 (17 042)	45.44 (668)	53.66 (162 615)
<b>Diagnostic category (ICD-10 v6)</b>				
Certain infectious and parasitic diseases	4.27 (12 967)	3.01 (1079)	1.41 (96)	4.44 (1188)
Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	0.4 (1204)	0.74 (267)	0.25 (17)	0.35 (937)
Diseases of the circulatory system	3.44 (10 439)	3.93 (1410)	2.28 (155)	3.38 (9029)
Diseases of the digestive system	5.49 (16 656)	5.46 (1957)	4.03 (274)	5.49 (14 699)
Diseases of the ear and mastoid process	0.70 (2130)	0.42 (150)	0.24 (16)	0.74 (1980)

Continued

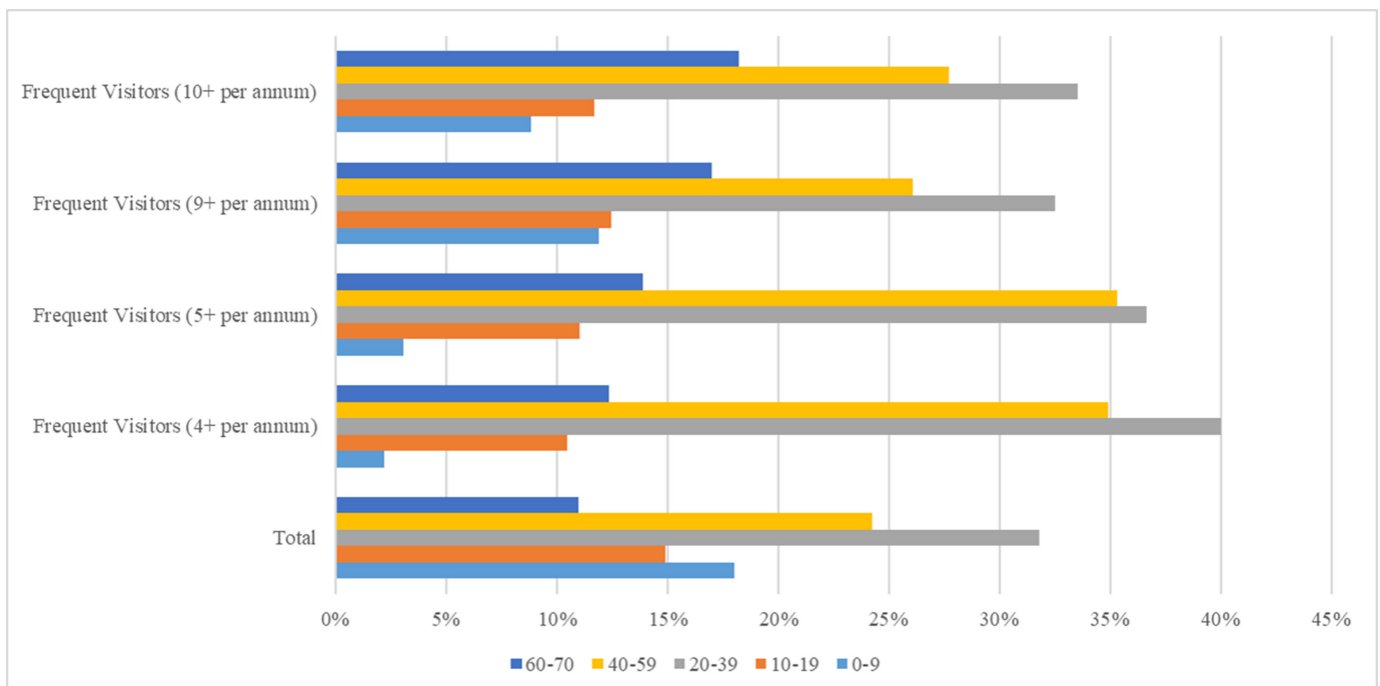
**Table 2** Continued

A total of 343 014 ED presentations from 170 134 individual patients	Whole population	FVs (39979 presentations) (4+ per annum)	Very FVs (7570 presentations) (10+ per annum)	Non-FVs (303035 presentations) (<4 per annum)
Total presentations since 2009–2016	Percentage (frequency)			
Diseases of the eye and adnexa	0.81 (2452)	0.35 (126)	0.15 (10)	0.87 (2326)
Diseases of the genitourinary system	4.45 (13 486)	4.32 (1548)	2.77 (188)	4.46 (11 938)
Diseases of the musculoskeletal system and connective tissue	6.10 (18 498)	5.72 (2051)	6.73 (457)	6.15 (16 447)
Diseases of the nervous system	1.09 (3298)	1.15 (414)	1.35 (92)	1.08 (2884)
Diseases of the respiratory system	7.54 (22 861)	7.88 (2825)	5.43 (369)	7.49 (20 036)
Diseases of the skin and subcutaneous tissue	3.06 (9275)	3.02 (1081)	1.97 (134)	3.06 (8194)
Endocrine, nutritional and metabolic diseases	0.90 (2735)	1.81 (650)	2.55 (173)	0.78 (2085)
External causes of morbidity and mortality	0.25% (748)	0.5 (179)	0.85 (58)	0.21 (569)
Factors influencing health status and contact with health services	2.19 (6629)	2.44 (873)	3.2 (217)	2.15 (5756)
Injury, poisoning and certain other consequences of external causes	26.14 (79 288)	12.94 (4639)	10.01 (680)	27.91 (74 649)
Mental and behavioural disorders	5.00 (15 176)	10.92 (3915)	16.11 (1094)	4.21 (11 261)
Neoplasms	0.45 (1365)	0.81 (290)	0.32 (22)	0.40 (1075)
Pregnancy, childbirth and the puerperium	1.86 (5637)	2.04 (731)	0.87 (59)	1.83 (4906)

ED, emergency department; FV, frequent visitor; ICD-10, International Statistical Classification of Diseases and Related Health Problems, 10th revision.

at which the maximum is achieved is used for our cut-off threshold score in our model. The maximum difference achieved in our experiment was 22.60% with cut-off total

score of 16.22 (figure 2). At a risk score threshold of 16.22, our RPSS-ED had a sensitivity of 65.94% of correctly identifying FVs, and 43.35% were incorrectly classified as FVs



**Figure 1** Percentage of age groups by annual frequency of attendance to emergency departments.

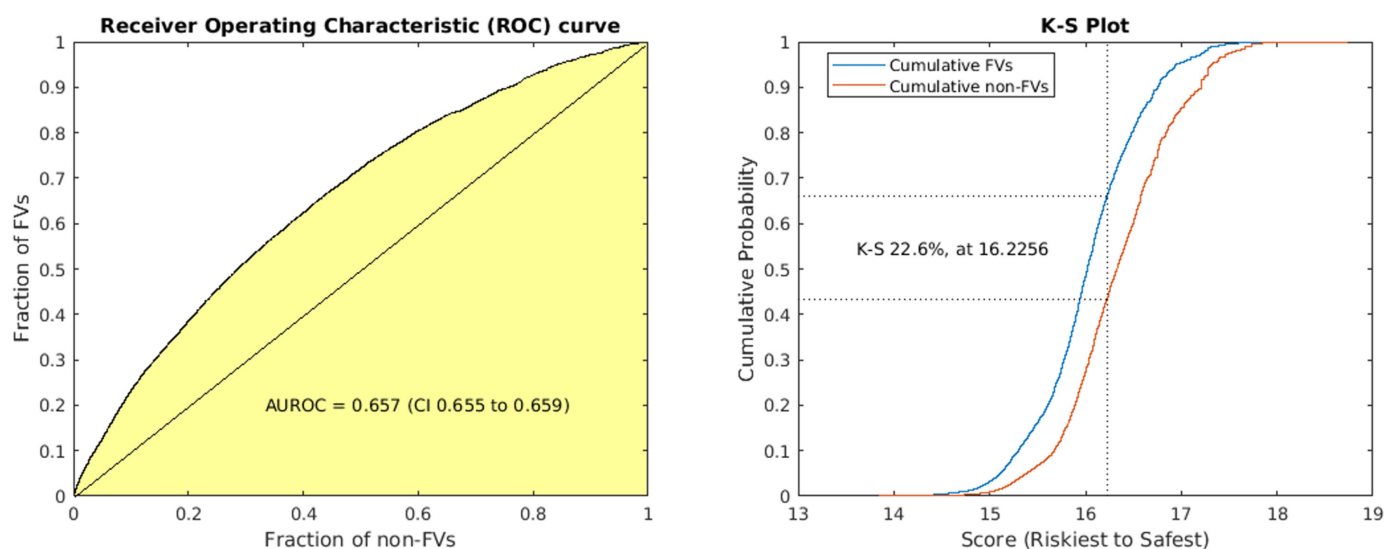
**Table 3** The patterns of FVs based on the time of arrival to ED

	Whole population	FVs (total 39979 presentations, 4+ per annum)	Very FVs (total 7570 presentations, 10+ per annum)	Non-FVs (303035 presentations, <4 per annum)
<b>Total presentations since 2009–2016</b>	<b>Percentage (frequency)</b>			
<b>Time of arrival analysis</b>				
00:00 to 06:00	10.58 (36 274)	11.46 (4583)	10.7 (810)	10.46 (31 691)
06:00 to 12:00	35.77 (122 698)	34.45 (13 773)	17.82 (1349)	24.15 (73 188)
12:00 to 18:00	23.95 (82 162)	22.45 (8974)	35.01 (2650)	35.94 (140 616)
18:00 to 24:00	29.7 (101 880)	31.64 (12 649)	36.47 (2761)	29.45 (89 231)
<b>Day of arrival analysis</b>				
Sunday	15.68 (53 776)	14.18 (5669)	14.36 (1087)	15.88 (48 107)
Monday	14.90 (51 114)	15.45 (6176)	14.48 (1096)	14.83 (44 938)
Tuesday	14.14 (48 505)	14.60 (5838)	14.64 (1108)	14.08 (42 667)
Wednesday	13.68 (46 936)	14.02 (5604)	14.24 (1078)	13.64 (41 332)
Thursday	13.87 (47 581)	14.17 (5667)	14.4 (1090)	13.83 (41 914)
Friday	13.63 (46 746)	14.37 (5746)	14.28 (1081)	13.53 (41 000)
Saturday	14.10 (48 356)	13.20 (5279)	13.61 (1030)	14.22 (43 007)
<b>Seasonal variation analysis</b>				
Summer (December to February)	24.43 (83 789)	24.11 (9637)	25.23 (1910)	24.47 (74 152)
Autumn (March to May)	27.39 (93 935)	27.45 (10 973)	27.66 (2094)	27.38 (82 962)
Winter (June to August)	24.82 (85 125)	24.62 (9343)	24.33 (1842)	24.84 (75 282)
Spring (September to November)	23.37 (80 165)	23.83 (9526)	22.77 (1724)	23.31 (70 639)

The data were further split into six hourly group of the day, day of the week and seasonal variations  
ED, emergency department; FV, frequent visitors.

(table 4). With a lower risk score threshold, for instance, 15.38, the rate of incorrectly identified FVs dropped to 5%, but the model had a lower accuracy of identifying FVs and non-FVs (sensitivity).

For the separate external validation, we computed the total risk scores based on current total enrolled patients' corresponding points of variables. Using the cut-off threshold derived from the model (16.22),



**Figure 2** The performance of Risk Predictive Scoring System for ED (RPSS-ED) using receiver operating characteristics (ROC) curve and Kolmogorov-Smirnov (K-S) plot. The blue and red line indicate the cumulative distribution of frequent visitors (FVs) and non-FVs, respectively. AUROC, area under the ROC curve.

**Table 4** Performance of the predictive model at different risk score thresholds

Evaluation metrics	Risk thresholds					
	15.3840	95% CI	16.2256	95% CI	16.6111	95% CI
Sensitivity	0.1254	0.1220 to 0.1288	0.6594	0.6546 to 0.6643	0.8530	0.8494 to 0.8566
Specificity	0.9494	0.9492 to 0.9508	0.5665	0.5648 to 0.5683	0.3188	0.3172 to 0.3205
False positive rate	0.05	0.0498 to 0.0502	0.4335	0.4318 to 0.4352	0.6811	0.6795 to 0.6827
Positive predictive value	0.2353	0.2297 to 0.2409	0.1572	0.1562 to 0.1584	0.1331	0.1326 to 0.1337
Negative predictive value	0.8985	0.8982 to 0.8989	0.9313	0.9304 to 0.9323	0.9464	0.9452 to 0.9477
Accuracy	0.8599	0.8587 to 0.8611	0.5766	0.5750 to 0.5783	0.3772	0.3756 to 0.3789

it achieved average positive sensitivity of 86.40% in detecting FVs to EDs. Since all enrolled patients were already FVs to the hospital, false positive rate was not computed.

## DISCUSSION

This study has identified the demographic patterns and clinical conditions of FVs to EDs and developed a predictive model that is specifically designed to identify patients with high risk of a future representation to ED. To the best of our knowledge, this is the first work on focusing on the FVs to EDs, which is different from well-established existing works<sup>2 16 17</sup> that mainly address prediction of patients with high risk of a future readmission.

Our findings indicate that FVs were characterised by young adulthood to late middle-aged patients with a higher proportion of indigenous population. Unlike FVs in older age group (60–70 years) who often have chronic illnesses such as diabetes and heart diseases, FVs in young adulthood (20–39 years) to late middle-aged (40–59 years) patients are highly associated with mental health-related diseases along with alcohol and drug-related diagnosis. Our findings are consistent with previous outcomes.<sup>10 18</sup> Additionally, we also found that the FVs are more likely to arrive by ambulance and leave at own risk without completing their treatments. We also observed that FVs are highly associated with socially disadvantaged groups such as people who have been divorced, widowed or separated. This suggests that these groups may be the focus of certain interventions to reduce preventable representations. The identified patterns of FVs based on the time of arrival to ED also provide important implications in relation to ED management and strategic planning (eg, staff allocation and prediction of the number of beds required) to improve overall health outcome.<sup>19</sup>

We have developed a risk predictive scoring system only using a limited set of variables that were easily obtained from electronic medical record system, which allows integration to current medical systems. The performance of our predictive model was favourable with an AUROC of 65.7% (95% CI 0.655 to 0.659) and with a sensitivity of 65.94% for a risk score threshold of 16.22. More reliable results of 86.40% in detecting FVs to EDs were achieved from the separate external validation group (ie, current

total enrolled patients). We attribute this to the size of external validation group which is much smaller than the internal validation group. For example, a systematic review of predictive risk models for readmission shows that AUROC ('C-statistic') are ranging from 50% to 72%.<sup>20</sup> The use of RPSS-ED can potentially remove manual at-risk patient searches and therefore allows clinicians to focus more on patient care and service delivery.

The model has a few identified limitations. The performance of the model could have been improved by including more variables including social factors (eg, unemployment and previous trauma); however, these variables were not available. The model was developed based on a single hospital data; expanding the data across multiple hospitals will also help better understand diversified patterns of FVs and increase the performance of the model. Although our database is comprehensive and complete, some missing data or inconsistent coding at data entry can be problematic in model learning as it could lead to underprediction or overprediction. More consistent and accurate data are expected to improve the predictive modelling. The model cannot be applied to predict FVs in older age group (aged >70 years), since it was designed to predict FVs among younger and general population (aged ≤70 years). Our model is designed to accommodate additional data (ie, other characteristics of FVs), which can be used to identify more emerging risks of representation. There are a lot of opportunities in improving the model by linking ED data with general practice data, inpatient data and outpatient data.

**Contributors** EA, JK, TB and CB were responsible for the conceptualisation of this project. KR, EA and CB were responsible for creation of the datasets in this project. EA and JK were responsible for statistical analysis and development of the predictive modelling. All authors contributed to the preparation of the manuscript.

**Funding** This research was supported by New South Wales (NSW) Health, Australia as part of the Integrated and Intensive Care Management Across Sector Collaboration (IICMASC) project. The resulting analyses and models are being used by the Nepean Blue Mountains Local Health District (NBMLHD).

**Competing interests** None declared.

**Patient consent** Not required.

**Ethics approval** Human Research Ethics Committees (HRECs) at the Nepean Blue Mountains Local Health District (NBMLHD), NSW, Australia.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** No additional data are available.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## REFERENCES

1. Enard KR, Ganelin DM. Reducing preventable emergency department utilization and costs by using community health workers as patient navigators. *J Healthc Manag* 2013;58:412.
2. Billings J, Dixon J, Mijanovich T, et al. Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients. *BMJ* 2006;333:327.
3. Billings J, Blunt I, Steventon A, et al. Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30). *BMJ Open* 2012;2:e001667.
4. Hamner JB, Ellison KJ. Predictors of hospital readmission after discharge in patients with congestive heart failure. *Heart Lung* 2005;34:231–9.
5. Baylor University Medical Center. *Risk factors for 30-day hospital readmission in patients? 65 years of age. Baylor university medical center proceedings*, 2008.
6. Gupta S, Tran T, Luo W, et al. Machine-learning prediction of cancer survival: a retrospective study using electronic administrative records and a cancer registry. *BMJ Open* 2014;4:e004007.
7. Louis DZ, Robeson M, McAna J, et al. Predicting risk of hospitalisation or death: a retrospective population-based analysis. *BMJ Open* 2014;4:e005223.
8. Shen Y, Tay YC, Teo EWK, Ewk T, et al. Association between the elderly frequent attender to the emergency department and 30-day mortality: a retrospective study over 10 years. *World J Emerg Med* 2018;9:20–5.
9. Nepean Blue Mountains Local Health District. *Nepean blue mountains local health district healthcare services plan 2012 to 2022*. Nepean Blue Mountains Local Health District: Kingswood, 2013.
10. Dinh MM, Berendsen Russell S, Bein KJ, et al. Trends and characteristics of short-term and frequent representations to emergency departments: A population-based study from New South Wales, Australia. *Emerg Med Australas* 2016;28:307–12.
11. Jelinek GA, Jiwa M, Gibson NP, et al. Frequent attenders at emergency departments: a linked-data population study of adult patients. *Med J Aust* 2008;189:552.
12. Berry JG, Gay JC, Joynt Maddox K, et al. Age trends in 30 day hospital readmissions: US national retrospective analysis. *BMJ* 2018;360:k497.
13. Locker TE, Baston S, Mason SM, et al. Defining frequent use of an urban emergency department. *Emerg Med J* 2007;24:398–401.
14. Markham D, Graudins A. Characteristics of frequent emergency department presenters to an Australian emergency medicine network. *BMC Emerg Med* 2011;11:1.
15. Thomas LC, Edelman DB, Crook JN. *Credit scoring and its applications: SIAM*, 2002.
16. van Walraven C, Dhalla IA, Bell C, et al. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. *CMAJ* 2010;182:551–7.
17. Ontario HQ. *Early identification of people at risk of hospitalization: Hospital Admission Risk Prediction (HARP) - a new tool for supporting providers and patients*. Toronto: Health Quality Ontario, 2013.
18. LaCalle EJ, Rabin EJ, Genes NG. High-frequency users of emergency department care. *J Emerg Med* 2013;44:1167–73.
19. Trzeciak S, Rivers EP. Emergency department overcrowding in the United States: an emerging threat to patient safety and public health. *Emerg Med J* 2003;20:402–5.
20. Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: a systematic review. *JAMA* 2011;306:1688–98.