BMJ Open

Characterising and predicting persistent high-cost utilisers in healthcare: a retrospective cohort study in Singapore

Sheryl Hui Xian Ng [1,2] Nabilah Rahman [1,2] Ian Yi Han Ang [1,2]
Srinath Sridharan [1] Sravan Ramachandran,[1] Debby Dan Wang [1]
Astrid Khoo,[2] Chuen Seng Tan [1] Mengling Feng,[1]
Sue-Anne Ee Shiow Toh [2,3,4] Xin Quan Tan[1,2]

¹Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore
²Regional Health System Office, National University Health System, Singapore, Singapore
³Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
⁴Singapore Population Health Improvement Centre (SPHERiC), National University Health System, Singapore, Singapore

Correspondence to
Dr Xin Quan Tan;
kyletanxq@gmail.com

ABSTRACT

Objective We aim to characterise persistent high utilisers (PHUs) of healthcare services, and correspondingly, transient high utilisers (THUs) and non-high utilisers (non-HUs) for comparison, to facilitate stratifying HUs for targeted intervention. Subsequently we apply machine learning algorithms to predict which HUs will persist as PHUs, to inform future trials testing the effectiveness of interventions in reducing healthcare utilisation in PHUs.

Design and setting This is a retrospective cohort study using administrative data from an Academic Medical Centre (AMC) in Singapore.

Participants Patients who had at least one inpatient admission to the AMC between 2005 and 2013 were included in this study. HUs incurred Singapore Dollar 8150 or more within a year. PHUs were defined as HUs for three consecutive years, while THUs were HUs for 1 or 2 years. Non-HUs did not incur high healthcare costs at any point during the study period.

Outcome measures PHU status at the end of the third year was the outcome of interest. Socio-demographic profiles, clinical complexity and utilisation metrics of each group were reported. Area under curve (AUC) was used to identify the best model to predict persistence.

Results PHUs were older and had higher comorbidity and mortality. Over the three observed years, PHUs' expenditure generally increased, while THUs and non-HUs' spending and inpatient utilisation decreased. The predictive model exhibited good performance during both internal (AUC: 83.2%, 95% CI: 82.2% to 84.2%) and external validation (AUC: 79.8%, 95% CI: 78.8% to 80.8%).

Conclusions The HU population could be stratified into PHUs and THUs, with distinctly different utilisation trajectories. We developed a model that could predict at the end of 1 year, whether a patient in our population will continue to be a HU in the next 2 years. This knowledge would allow healthcare providers to target PHUs in our health system with interventions in a cost-effective manner.

BACKGROUND

The disproportionate concentration of the healthcare utilisation and expenditure in a small group of users is a common phenomenon observed worldwide.[1–3] In the USA, the

Strengths and limitations of this study

► The study utilised a large longitudinal database of 9 years, allowing the development of a stable definition of PHUs.
► The prediction models developed contained features commonly available in most administrative databases, increasing applicability in other health systems.
► Internal validation using a wide range of hyperparameters, as well as an additional round of external validation, ensured the robustness of the models selected.
► The PHU definition did not include patients with only high outpatient utilisation and no inpatient utilisation, limiting generalisability of our findings to HUs with at least one inpatient admission.
► External validation of the final prediction model was carried out on an external data set which did not exclude patients who died and did not have polypharmacy score as a feature, resulting in findings which must be interpreted with caution.

Agency for Healthcare Research and Quality reported that 1% of users in 2008 accounted for nearly 20% of all healthcare expenditure, and that 5% of users accounted for nearly 50% of expenditure.[4] In Ontario, Canada, patients in the top 5% of expenditure accounted for 61% of combined hospital and nursing home expenditure and almost all of mental health and rehabilitation costs.[5 6] The definition of high utilisers (HU) ranges from the top 1% to 20% in healthcare expenditure of each cohort. These HUs have been characterised as a population with high levels of chronic disease and prevalence of mental illness,[1] and have been the traditional target for interventions designed to reduce healthcare utilisation.[7] Unfortunately, heterogeneity in the definition of HUs and intervention design, as well as a lack of high-quality evidence in scientific literature, renders evaluation of effectiveness of these interventions difficult.[7–10]

Available robust evidence on these interventions is mixed, with some studies suggesting the need for identifying and targeting of specific patient groups for an intervention to be effective.[7 9 11] High utilisation in the bulk of HUs tends to taper off over time and naturally regress to lower utilisation, and could also potentially explain why interventions have not been successful in reducing cost or utilisation when compared with controls.[12 13] A subgroup of patients, who do not exhibit such a regression-to-the-mean effect and experience prolonged high utilisation, are termed persistent HUs (PHUs), and have been defined as individuals who achieve HU status for more than two consecutive years.[2 14] Various studies suggest only 21%–40% of HUs are PHUs who remain in the top 5%–10% of expenditure 2 years later.[2 15–19] Identifying this subgroup of patients with persistent high cost for active targeted interventions could achieve better results in long-term reduction of healthcare utilisation and cost to the healthcare system.

The public healthcare system in Singapore serves around 80% of acute inpatient needs, and public hospitals receive government subvention to provide subsidised care to the resident population.[20] Patients may receive government subsidies, insurance coverage or funding from safety net programme before they pay out-of-pocket for acute care.[20 21] PHUs would hence impose a resource burden for both the government and the health system. Moreover, due to constraints on capacity from an ageing population, the government has put in place measures to prepare for the population's long-term care needs and move care to the community setting.[22] Hence, developing interventions for PHUs would reduce avoidable admissions and shift suitable patients to be cared for in the community setting, optimising the use of finite resources and improving quality of care for patients.

Given the prolonged high utilisation by PHUs, there is opportunity for early identification of these patients for intervention to shorten the period of high utilisation. However, it is only meaningful to identify these patients early on before they have truly reached PHU status, and so predictive models would best serve this purpose. To our knowledge, only two studies have developed models for prediction of PHU status within a HU population which incorporate an assessment of model performance.[19 23] Models from both studies achieved an area under curve (AUC) score exceeding 0.90, denoting outstanding predictive abilities.[24] With the ubiquitous adoption of electronic health records and data sets of increasing size and complexity in recent years,[25–27] machine learning algorithms have been utilised widely due to their ability to learn and identify patterns in the data in absence of parametric assumptions.[28] While some studies have attempted to predict HUs using machine learning,[29–32] none of these have been designed to predict persistence. Building a prediction model for PHUs using machine learning methods would not only aid in the prospective identification of PHUs but also contribute to the limited literature on predictive modelling for persistence. Furthermore, most studies on HUs and persistence have been based in the USA[1]; our study originating from a developed multi-ethnic Asian country will add to the literature in this field.

With the objectives of stratifying HUs for targeted interventions and contributing to the gap in literature on prediction of PHUs, our study aims to characterise PHUs and subsequently predict PHU status within a HU population using machine learning algorithms. We will divide the study population into three groups: (i) PHUs, (ii) transient high utilisers (THUs), (iii) non-HUs, describe these groups and compare their characteristics to better understand the profiles associated with persistence. Subsequently, we will use machine learning techniques to develop models that distinguish PHUs from a pool of HUs using data from the first year observed. We will then perform both internal and external validation of the model to evaluate the performance of the models and demonstrate the feasibility of predicting PHUs for potential targeted interventions. The findings from this study will add invaluable insight into the study of persistence in high healthcare utilisation, and inform future trials to test effectiveness of interventions in reducing healthcare utilisation in a persistently high utilisation group.

METHODS
Description of data source
We examined data from 2005 to 2013 comprising demographics, utilisation, billing, medication and diagnosis data for both inpatient and outpatient care in National University Hospital, a 1000-bed Academic Medical Centre (AMC) in Singapore for this retrospective study. Details on the data set and its preparation have been reported separately.[33] The study cohort consisted of 549 109 patients aged 21 years and above who had at least one record of either an inpatient admission, specialist consultation, therapy visit or emergency attendance between 1 January 2005 and 31 December 2013.

Definitions of PHUs and THUs
Due to lack of access to standardised cost data, we used expenditures as an estimate, a common approach adopted in HU studies.[1] In our study, we defined expenditure as the total price of a service charged to a patient before any government subsidies or third-party payments (eg, payments by insurers or employers), adjusted to 2015 prices. While these billed charges may not be reflective of the true cost of service delivery, in absence of actual cost data, these charges serve as a good proxy for cost as they are often correlated with the actual cost, and systematically higher by a specified cost-to-charge ratio.[34 35] We computed the expenditure for each patient over all inpatient and outpatient settings in the hospital for each observation year and classified patients as HUs if they incurred expenditure in the top 10%, or at least SG$8150 within the year. Details on the derivation of the HU cut-off are provided in online supplementary file 1. With our key objective of identifying a group of patients
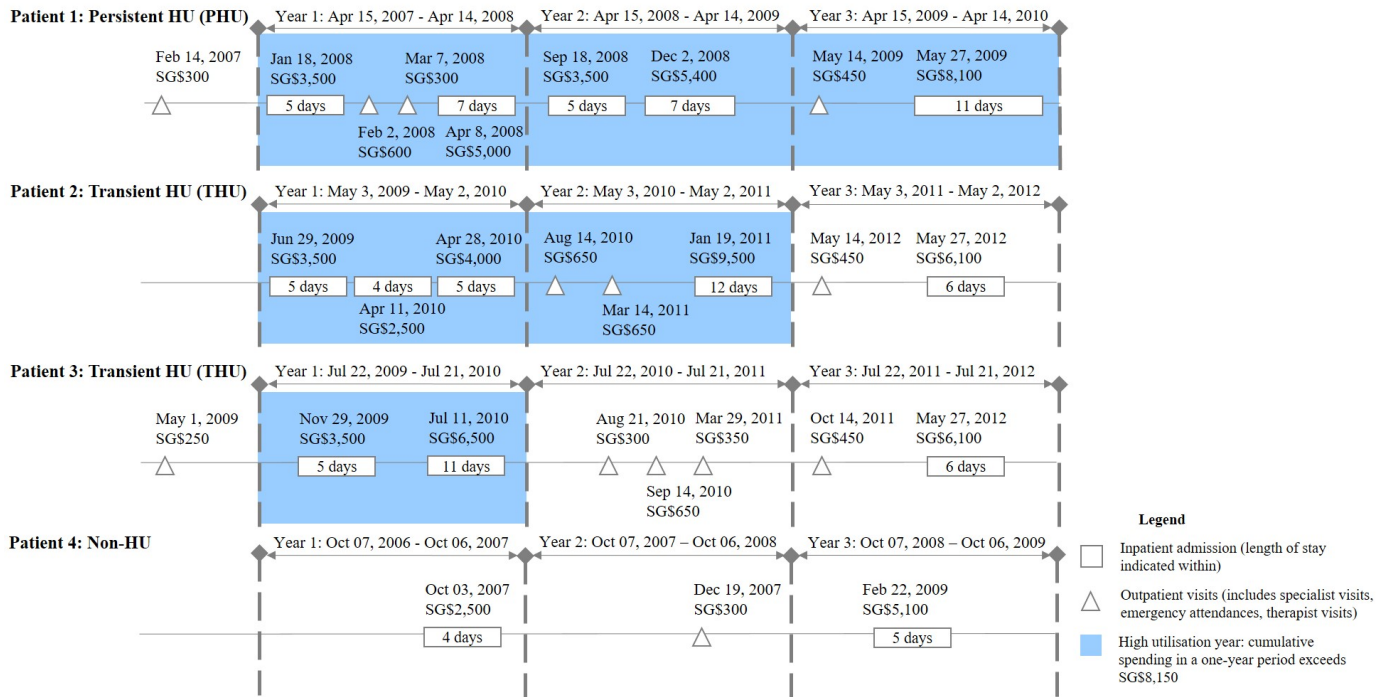
Figure 1 Illustration of persistent high utiliser (PHU)/transient high utiliser (THU) cohort generation using three hypothetical patients.

with extended high utilisation that could potentially be targeted for intervention, we defined PHUs in our study as those who met the HU criteria for three consecutive years, and THUs as patients who met the HU criteria for at least 1 year, but less than three consecutive years. Those who did not meet the criterion during the study period were classified as non-HUs.

To divide the population into groups, we first determined if a patient met the HU criteria at any point in time, by examining each inpatient episode of the patient. If the patient had accumulated healthcare expenditure of SG$8150 in the 1 year prior to and inclusive of the discharge date, the patient would be classified as a HU (figure 1). Only the first discharge date that fulfilled the criterion was considered. That discharge date would then be fixed as the end of Year 1 (Y1) for that patient; Year 2 (Y2) would be the next year from that date, and Year 3 (Y3) the following year. HUs who had expenditures of at least SG$8150 in Y2 and Y3 were classified as PHUs, while the remaining HUs were classified as THUs. For non-HUs, the discharge date of the index admission was defined as the end of Y1, and Y2 and Y3 were correspondingly constructed. In identifying HUs, Y1 was defined as the first 365-day period fulfilling the HU criterion as opposed to looking at expenditure in a calendar year. A calendar year was not used since patients who made contact with the hospital in the later part of the calendar year would be less likely to fulfil the HU criteria as they would have had less time to accumulate expenditure, thus introducing bias.

In defining PHUs and THUs, we excluded patients who had no inpatient admissions throughout the study period, as well as patients who recorded their first contact in 2012 and 2013 due to non-availability of their respective Y2 and/or Y3 data. We also excluded patients who died in Y1, as these patients would not have been a candidate at the end of Y1 for prediction of Y2 and Y3 HU status. However, we included patients that died in Y2 and Y3, given that we are trying to predict HU status using only Y1 data, and excluding those that died in Y2 and Y3 would introduce bias by leaving only healthier individuals in the pool. We included non-HUs as a separate cohort for descriptive comparison as well.

Comparison of PHU, THU and non-HU groups

We first described the profiles of the three groups defined in this study. Patient demographics at first visit as well as clinical complexity during Y1 were reported. Clinical complexity was measured by the Charlson Comorbidity Index (CCMI) and polypharmacy score (PPS). PPS measured the maximum number of unique dispensed medications a patient was prescribed in a visit, and demonstrated convergent and criterion validity when compared with CCMI and measures of healthcare utilisation.[33] In addition, we reported their expenditure, utilisation (inpatient, length of stay (LOS) and outpatient visits), as well as common diagnoses and subspecialties visited in Y1. To compare between PHU and THUs, as well as between PHU and non-HUs, we used χ^2 and two-sided Wilcoxon rank-sum tests, respectively, to compare categorical and continuous characteristics.[36] Within each set of comparisons with a control group, we adjusted the p values using Bonferroni's correction,[37] and assessed statistical significance using a threshold of 0.05. Thereafter, we

examined the utilisation patterns of the PHU, THU and non-HU groups over a 3-year period.

Feature extraction for prediction

To construct our predictive models to identify PHUs from a HU population, we first considered the same patient-level information from the descriptive analyses as features, or predictors, in the models, with the exception of mortality. In addition, we included bill components in the model. For each visit billed to the patient, we identified the different bill components for that particular visit. The bill components included expenditure on prescription drugs, laboratory and radiology services, surgical procedures, specialised investigations (non-routine diagnostics), doctor consultations, therapy (physiotherapy, occupational therapy or speech therapy), miscellaneous procedures as well as consumables, implants and appliances. We subsequently aggregated the visit-level expenditure information into a total amount spent per patient for Y1, and calculated the proportion of the Y1 bill that each component accounted for. We reviewed all features and subsequently dropped features which had missing categorical values (eg, housing type), exhibited collinearity or had zero variance. A description of features and hyperparameters used in our model is tabulated in online supplementary file 1.

Predictive model building and validation

Given the size and complexity of our data set, we opted to use machine learning algorithms for the prediction of PHUs as they have been shown to be adept at handling large volumes of high-dimensional data in healthcare settings to enable high-quality predictions.[38–40] We tested three algorithms: penalised regression, support vector machine and extreme gradient boosting (XGBoost), and evaluated their performance using standard metrics of sensitivity and specificity, as well as AUC with a two-sided DeLong's test.[41 42] For internal validation of our model, we performed repeated cross-validation with model construction on 70% of data and evaluation of performance on the remaining 30%. As the data set was imbalanced with a small number of PHUs relative to the THUs, the THUs were under-sampled within each training iteration of cross-validation.[43 44]

In addition, to increase applicability of the model as a predictive tool to be used as point of discharge, we identified a final model that would achieve similar predictive performance using only a subset of features. We tested sub-models containing different number of features with the highest variable importance against the full model using a two-sided DeLong's test. This was to identify a final sub-model with the lowest number of features that had an AUC not significantly different from that of the full model.[42] We reported the top features ranked by relative importance, or the improvement in accuracy the feature brought to its branch relative to the highest ranked, which would range from 0 to 100.

We then performed external validation for the final model. An independent data set of adult patients from the same administrative database for 2013 to 2016 was extracted. The HU classification process detailed in figure 1 was applied to this validation set and the final model was fitted to this data set to assess the performance of the algorithm when applied to new data. Due to lack of information on the exact medications patients were billed for in the external validation set, we performed both internal and external validation of the final model without PPS as a feature. The model without PPS as a feature was derived using the same procedure as the original final model. The same model performance metrics were reported. We used RStudio and the package *caret* for the computation described earlier.[45 46]

Patient and public involvement

The study was designed and conducted without patient or public involvement.

RESULTS

Comparison of characteristics of PHUs, THUs and non-HUs

A total of 152 497 patients were included in the study (figure 2). There were 5094 PHUs, 62 159 THUs and 85 244 non-HUs identified (table 1). PHUs were statistically different from both THUs and non-HUs in all demographics, disease complexity and utilisation compared. Overall median expenditure across all settings was highest in the PHUs, followed by the THUs and non-HUs (PHUs: $15,015; THUs: $13,727; non-HUs: $2,942). However, median inpatient expenditure was highest in the THUs (PHU: $10,743; THU: $12,226; non-HU: $2,395). Notably, PHUs and THUs had median ages 19 and 15 years greater than the non-HUs, respectively, at the point of first contact during the study period. Median values of CCMI were highest for the PHUs (PHUs: 2; THUs: 1; non-HUs: 0), while the median PPS values were similar for PHUs and THUs. The percentage of patients that died during the study period was also highest in the PHU group (43.5%) compared with the THU group (19.7%) and the non-HU group (5.8%).

As seen in online supplementary file 2, diabetes, hypertension and lipid disorders were found to be prevalent conditions across all three groups. However, the highest prevalence rates were consistently seen in the PHUs and lowest in the non-HUs. In addition, visits to the cardiac subspecialty were also common to the groups. We also observed a segment of patients unique to the PHUs who were likely to be undergoing treatment for kidney diseases, from the high prevalence of chronic kidney disease (25.4%) and visits to the nephrology subspecialty (24.2%).

Examining trends in total expenditure over the first 3 years, the PHUs' expenditure generally increased, whereas the THUs and non-HUs' spending decreased drastically from Y2 onwards (figure 3). Inpatient expenditure and visits also followed such a similar pattern for
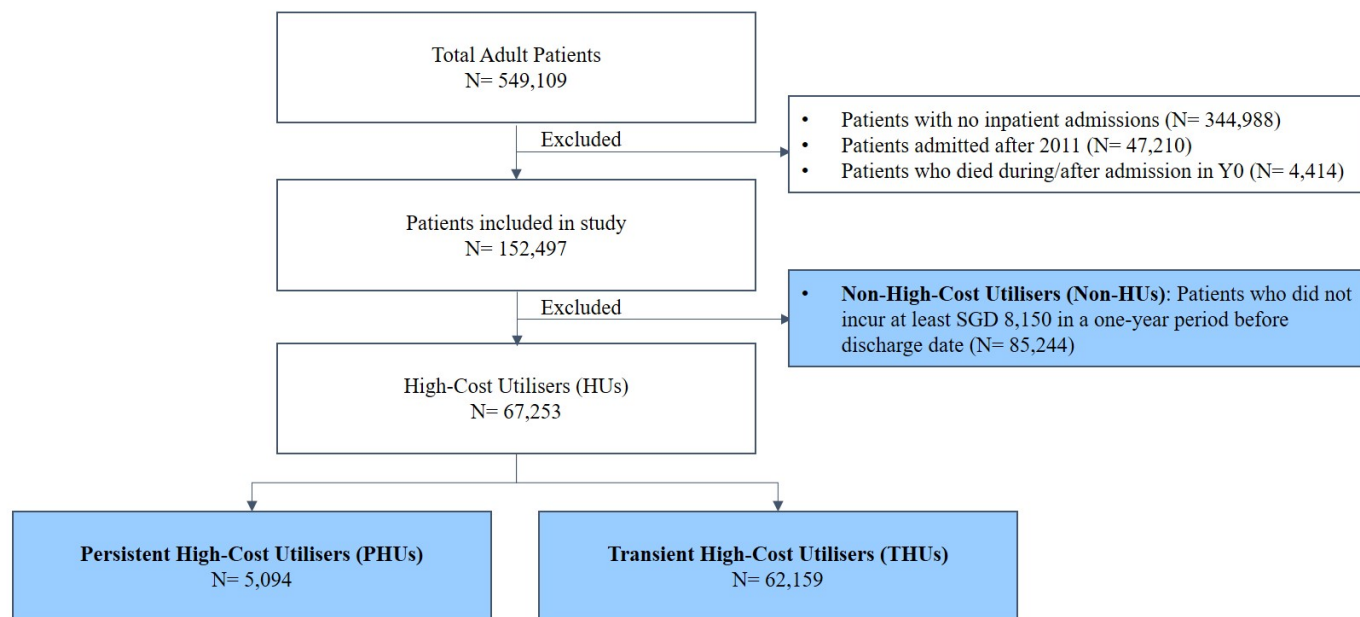
Figure 2 Inclusion and exclusion criteria for persistent high utiliser (PHU)/transient high utiliser (THU) cohort.

all groups. Importantly, we observed that the drop off in inpatient utilisation for THUs in Y2 and Y3 was very significant; the median inpatient utilisation across bill, LOS and visits was all zero in Y2 and Y3 (online supplementary file 3). In contrast, outpatient utilisation peaked in Y2 across all groups, with PHUs on average incurring above 20 visits in Y2 and Y3 while THUs and non-HUs had few or zero visits for both years.

Comparison of performance of candidate predictive models

Table 2 summarises the performance of algorithms used in this study for the prediction of PHUs from a pool of 67 253 patients at the end of Y1. Of the three algorithms, XGBoost performed the best in terms of AUC (83.4%, 95% CI: 82.4% to 84.4%) and sensitivity (78.7%, 95% CI: 76.5% to 80.7%). The AUC for the XGBoost model was also statistically different from the other two models compared (p<0.001). This full model was based on 514 features (online supplementary file 4), and we identified a final model with reduced number of features that exhibited similar model performance (p=0.28), which used only 51 features with the highest variable importance from the full model (figure 4). The top features for the final model were related to comorbidity, outpatient expenditure and prescription expenditure, followed by age, laboratory and radiology services, specialised investigations, therapy, doctor consultations, surgical procedures as well as consumables, appliances and implants. A description of the HUs in the study population and the external validation cohort are provided in online supplementary file 5. The AUC for the validated final model without PPS during internal validation was 83.1% (95% CI: 82.1% to 84.1%), while the AUC during external validation was slightly lower at 79.8% (95% CI: 78.8% to 80.8%) (online supplementary file 6).

DISCUSSION
Principal findings

In our study, we defined three groups of patients: PHUs who were HU for 3 years consecutively, THUs who were only HU for 1 or 2 years and non-HUs who did not qualify as an HU at any time. We described the demographics, utilisation and disease profiles of the groups and found that PHUs were distinctly different from THUs and non-HUs. Subsequently, we sought to predict which HUs would go on to be PHUs at the end of 1 year. We applied different machine learning algorithms and developed a model of excellent predictive performance. We then validated a final model with fewer features but of comparable performance for potential implementation as a tool to be used at point of discharge. The most important features in the final model were related to comorbidity, outpatient expenditure and prescription expenditure.

We identified 7% of our HU population to be PHUs, which is markedly lower than findings from other published studies on 3-year persistence.[2 19 23] This could be explained by the fact that we did not capture expenditure outside the hospital setting unlike these other studies. Hence, our PHU definition may not be directly comparable to studies that incorporate home care and nursing care costs in their computation of healthcare expenditure to determine persistence.[6 17] Comparing the disease profiles of the three groups of patients, the highest prevalence of lipid disorders, diabetes, hypertension and cardiac conditions was observed in the PHUs. Coupled with the older age and higher CCMI, PHUs exhibited a more medically complex profile compared with THUs and non-HUs. Of note, PHUs have been found to be older than THUs in some studies[19 23 47] and younger in others,[16 17 47 48] and this lack of uniform association of age with persistence may be attributable to

Table 1 Demographics, disease complexity and utilisation of PHUs, THUs and non-HUs

	PHU (n=5094)	THU (n=62 159)	Non-HU (n=85 244)	PHU vs THU*	PHU vs non-HU*
Patients (% of total population)	3.3%	40.8%	55.9%		
Number of years of utilisation, median (IQR)	4 (3–6)	3 (1–5)	2 (2–4)	***	***
Year 1 expenditure and utilisation					
Total expenditure, median (IQR)	$15 015 ($10 762–$23 778)	$13 727 ($10 125–$20 207)	$2942 ($1742–$4811)	***	***
Inpatient expenditure, median (IQR)	$10 743 ($7009–$18 293)	$12 226 ($8600–$18 577)	$2395 ($1311–$4082)	***	***
Outpatient expenditure, median (IQR)	$2256 ($550–$5878)	$672 ($306–$2015)	$333 ($220–$595)	***	***
Inpatient visits, median (IQR)	1 (1–2)	1 (1–1)	1 (1–1)	***	***
Inpatient length of stay, median (IQR)	9 (5–15)	8 (4–13)	2 (2–4)	***	***
Outpatient visits, median (IQR)	8 (2–17)	3 (1–9)	1 (1–3)	***	***
Year 1 demographics and clinical complexity					
Age, median (IQR)	60 (50–71)	56 (42–69)	41 (30–57)	***	***
Charlson Comorbidity Index, median (IQR)	2 (2–4)	1 (0–2)	0 (0–1)	***	***
Polypharmacy score, median (IQR)	16 (11–23)	16 (11–22)	7 (4–11)	***	***
Sex, N (%)				**	***
Female	2480 (48.7%)	27 841 (44.8%)	43 836 (51.4%)		
Male	2614 (51.3%)	34 318 (55.2%)	41 408 (48.6%)		
Ethnicity, N (%)				***	***
Chinese	3414 (67.0%)	39 113 (62.9%)	48 450 (56.8%)		
Indian	470 (9.2%)	6941 (11.2%)	11 646 (13.7%)		
Malay	727 (14.3%)	8001 (12.9%)	13 158 (15.4%)		
Others	483 (9.5%)	8104 (13.0%)	11 990 (14.1%)		
Housing type, N (%)	n=4636	n=52 161	n=71 839	***	***
Studio/1–2 room	263 (5.7%)	2330 (4.5%)	3148 (4.4%)		
3–5 room and larger	3772 (81.4%)	42 878 (82.2%)	59 770 (83.2%)		
Private	601 (13.0%)	6953 (13.3%)	8921 (12.4%)		
Mortality, N (%)					
During study period	2217 (43.5%)	12 244 (19.7%)	4966 (5.8%)	***	***
During Year 2	—	6128 (9.9%)	1959 (2.3%)	***	***
During Year 3	932 (18.3%)	1640 (2.6%)	794 (0.9%)	***	***

*χ^2 and Wilcoxon rank-sum tests were used to perform comparisons for categorical and continuous variables, respectively.
**p<0.01; ***p<0.001.
Non-HU, non-high-cost utiliser; PHU, persistent high-cost utiliser; THU, transient high-cost utiliser.

the heterogeneity in populations studied. In contrast, the higher comorbidity burden in PHUs is consistent with most studies on persistence.[14 17 19 23] We also found that the percentage of PHUs that died in our cohort was high (43.5%), especially when compared with the THU group (19.7%) and the non-HU group (5.8%). This is not unexpected given the PHUs had higher comorbidity.
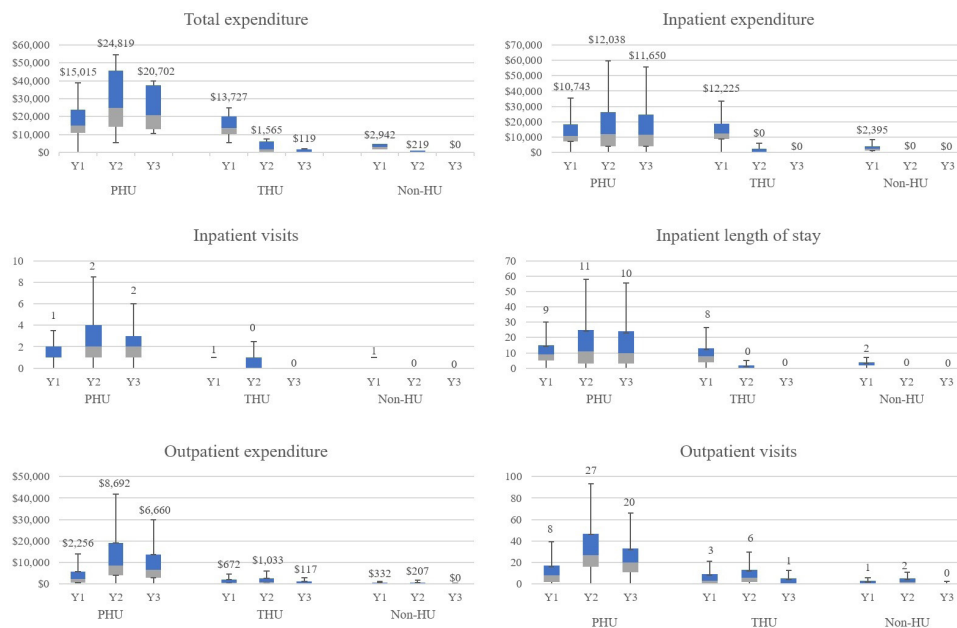
Figure 3 Yearly utilisation for three observation years, with yearly medians labelled. Non-HU, non-high utiliser; PHU, persistent high utiliser; THU, transient high utiliser.

In addition, the PHUs had greatly different expenditure and utilisation trajectories from the THUs. We observed that THUs' overall utilisation and expenditure fell significantly in Y2 and Y3 from Y1, illustrating the phenomenon of regression to the mean. This trend could have been due to an acute medical need in the THUs which tapered off after the initial period of high utilisation, and has been noted in other HU populations.[13 16 19 49 50] In contrast, PHUs' utilisation and expenditure continued to increase, motivating the need to target PHUs specifically for intervention instead of HUs in general.

Strengths

Our study utilised a large longitudinal database over a 9-year period, which allowed us to study the persistence of high utilisation in a large population over a sufficiently long period. We were thus able to develop a stable definition for PHUs and a predictive model that performed well in identification of PHUs. Furthermore, the approach we took in preparing and selecting features and developing the predictive model allows for easy deployment

in a real-world setting. The model can be applied at the point of discharge by drawing on data retrospectively, and contains features already commonly available in most administrative databases. In addition, we performed internal validation using repeated cross-validation and tested three models with a wide range of values for each parameter to ensure the robustness of the model selected. We also validated our final model on an external data set with model performance similar to results from internal validation, illustrating the generalisability and applicability of our predictive tool in other patient cohorts.

Limitations

There are some limitations arising from our methodology. We defined PHUs using the index admission as a frame of reference for accumulation of resource use, and so this definition cannot be applied to patients who incurred sizeable outpatient utilisation without any inpatient admissions. Our examination of persistent high users would hence only be generalisable to patients with conditions and care needs relevant to the inpatient

Table 2 Summary of the performance metrics of predictions made at the end of Year 1

Model	Number of features	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)
Penalised regression	102	72.7% (70.3% to 74.9%)	76.0% (75.4% to 76.7%)	81.7% (80.6% to 82.7%)
Support vector machine	514	71.2% (68.9% to 73.5%)	75.4% (74.8% to 76.0%)	80.4% (79.3% to 81.5%)
XGBoost	514	78.8% (76.7% to 80.9%)	71.9% (71.2% to 72.5%)	83.4% (82.4% to 84.4%)
XGBoost	51	78.7% (76.5% to 80.7%)	72.3% (71.6% to 72.9%)	83.2% (82.2% to 84.2%)

AUC, area under curve.

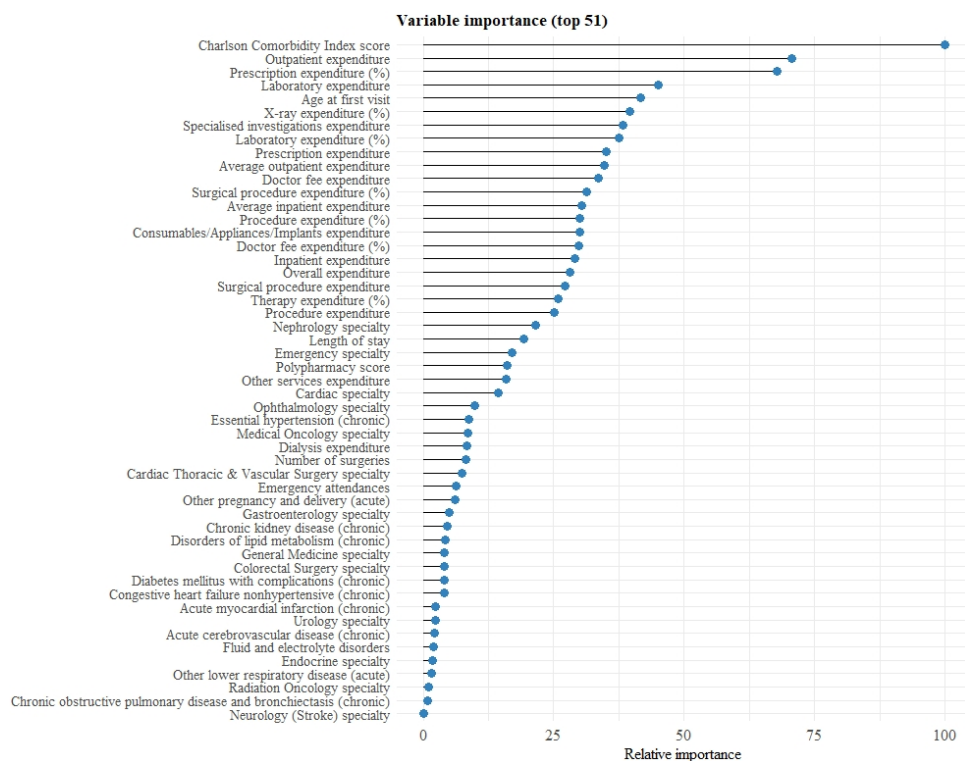**Variable importance (top 51)**

Figure 4 Variable importance for 51 features in final model.

setting. As the observation year was anchored at the first inpatient admission which met the HU criteria, it also did not include instances where patients met the PHU criteria more than once during the study period. However, in the real-world setting, this is less of an issue as patients who have met the criteria once would already be identified as a high-risk patient by the healthcare system that they seek care from. Furthermore, as the 12% of patients who died in Y2 were included in the analysis, potential bias may arise. While these patients may have exhibited Y1 characteristics sufficient to be predicted as PHUs, they did not survive long enough to fulfil the PHU definition. As a result, the predictive tool we have developed would have a higher false positive rate, and the model AUC reported would be a conservative estimate, which would be caveats to keep in mind in application of the tool.

In addition, the model was designed and validated on a single AMC. Previous research in Singapore revealed low cross-utilisation across multiple hospitals by a general patient population,[51] suggesting low selection and misclassification bias in our findings due to the use of data from only one hospital. Furthermore, prevalence rates of PHUs in studies outside Singapore incorporating data from multiple hospitals are similar to our findings at 1%–4%.[15–17 19] This would suggest that cross-utilisation rates of THUs and non-HUs across more than one hospital would likely be low, and correspondingly bias arising from this limitation would be modest. Generalisation of the model to other healthcare systems would be dependent on data and feature availability in those settings. However, the most important features in our models are common billing components and clinical

specialties routinely recorded within administrative databases, or easily generated from International Classification of Diseases diagnosis codes, increasing the ease of applying our model in other health systems.

Also, we were unable to examine in detail the differences in socioeconomic characteristics between the groups as only housing type was available as a proxy. As persistent high utilisation has been noted to be concentrated in those of lower socioeconomic status or minority races,[4 16 17 47] information such as household income, employment status or receipt of financial support would greatly increase the robustness of our comparisons.

Patients who died in Y1 in the external validation set were not removed due to lack of integration of death data, and hence the performance of the external validation must be interpreted with caution. As detailed medication was also not integrated into the external validation data set, we were unable to validate our final model with PPS as a feature. However, since the validation of the modified final model without PPS was good with an AUC of 80%, it appeared that the models developed perform reasonably well as long as most important features were present.

Implications for clinicians and policymakers

Our findings suggest that our PHU population is medically complex and tend to have recurrent high utilisation across multiple settings, and interventions should be tailored to this profile. Integrated care approaches involving case management, care coordination and disease management could potentially reduce healthcare use by this group of high risk patients.[7] In addition, given the high mortality rate, integrating palliative care into

interventions targeted at PHUs could also be a suitable approach to manage cost and improve quality of life in this group.[52] The distinctly different utilisation trajectories between PHUs and THUs reaffirm the necessity of stratifying the HU population to identify the correct subgroup of patients, to effectively deliver interventions for these high need patients.[53]

Our findings also suggest that there are ways to optimise the way we select patients for intervention. At present, most intervention programmes that target high utilisation patients enrol patients based on a one-off assessment of whether they incurred utilisation above a specified threshold. Such enrolment criteria would not be able to distinguish PHUs and THUs. An evaluation of a post-discharge intervention programme for HUs where enrolment was based on utilisation in Y1, and intervention was delivered in Y2, did not show an improvement of Y2 healthcare utilisation over matched controls.[54] Both the HUs and control patients in this study showed similar decrease in their high utilisation over time, illustrating the phenomenon of regression to the mean. This finding supports our observation on the need for refinement of enrolment criteria to better identify and target subgroups with increased risk of prolonged utilisation (ie, PHUs). The predictive tool we have developed in this paper will facilitate the targeting of PHUs from a pool of HUs that health systems already routinely identify and intervene on. A more targeted approach would allow healthcare systems to focus such resource intensive interventions on the group that is most likely to require and benefit from these interventions.

Unanswered questions and future research

Prediction of PHUs to identify and target them early for intervention is thus crucial. In our study, we were able to show that our models can predict prospectively which patients in our health system are likely to be PHUs with protracted high levels of healthcare utilisation. While we hypothesise that these PHUs would be better targets for healthcare utilisation reducing interventions as compared with HUs as a larger group, follow-up studies would need to be conducted to validate this. Given the scope of the study, we have not tested if the PHUs respond differently to various intervention programmes aimed at reducing utilisation. Future work will involve collaborations with groups beyond the AMC in this study that are responsible for the delivery of interventions to high utilisation groups. These collaborations would test the utility of our approach in identifying PHUs and providing these patients with different targeted interventions. Separately, testing the predictive tool in different populations and with varying HU thresholds would validate our methodology in different contexts. These further studies would test both the ability of the model to identify suitable targets in a real-world setting, and the impact of interventions targeting this group of PHUs as opposed to THUs and controls.

CONCLUSION

Persistent high-cost utilisers are a high-risk subgroup of patients with complex clinical needs and high expenditure and utilisation. While interventions could potentially accrue the largest benefits if this subgroup is targeted effectively, few to date have focused on identifying and intervening on this group. In our study, we constructed a definition for the identification of PHUs from a HU population, profiled the PHUs and compared them with the THUs and non-HUs. We demonstrated that the HU population could be stratified to reveal a subgroup of patients with recurrent high utilisation over time, which was distinct from the majority of patients who had transient high utilisation. Subsequently, we developed a machine learning model with the ability to predict at the end of 1 year, whether a patient in our population will continue to be a HU in the next 2 years, which is a novel approach to the prediction of persistence in healthcare utilisation. This knowledge would allow healthcare providers to more effectively target patients who have persistent utilisation, with existing or new interventions to make a greater impact in a cost-effective manner.

ORCID iDs
Sheryl Hui Xian Ng http://orcid.org/0000-0001-5046-2666
Nabilah Rahman http://orcid.org/0000-0002-0279-3169
Ian Yi Han Ang http://orcid.org/0000-0003-1124-3764

Srinath Sridharan https://orcid.org/0000-0001-9754-4925
Debby Dan Wang https://orcid.org/0000-0002-3755-8943
Chuen Seng Tan http://orcid.org/0000-0002-6513-2309
Sue-Anne Ee Shiow Toh http://orcid.org/0000-0003-1570-4417

REFERENCES

1 Wammes JJG, van der Wees PJ, Tanke MAC, *et al*. Systematic review of high-cost patients' characteristics and healthcare utilisation. *BMJ Open* 2018;8:e023113.
2 Tamang S, Milstein A, Sørensen HT, *et al*. Predicting patient 'cost blooms' in Denmark: a longitudinal population-based study. *BMJ Open* 2017;7:e011580.
3 Hartmann J, Jacobs S, Eberhard S, *et al*. Analysing predictors for future high-cost patients using German SHI data to identify starting points for prevention. *Eur J Public Health* 2016;26:549–55.
4 Cohen S, Yu W. The concentration and persistence in the level of health expenditures over time: estimates for the U.S. population,2008-2009 2012.
5 Ontario Association of Community Care Access Centres. *Ideas and opportunities for bending the health care cost curve - Advice for the Government of Ontario*, 2010.
6 Rais S, Nazerian A, Ardal S, *et al*. High-cost users of Ontario's healthcare services. *Healthc Policy* 2013;9:44–51.
7 Lee JY, Muratov S, Tarride J-E, *et al*. Managing high-cost healthcare users: the International search for effective evidence-supported strategies. *J Am Geriatr Soc* 2018;66:1002–8.
8 Iovan S, Lantz PM, Allan K, *et al*. Interventions to decrease use in prehospital and emergency care settings among super-utilizers in the United States: a systematic review. *Med Care Res Rev* 2019:1077558719845722.
9 Bleich SN, Sherrod C, Chiang A, *et al*. Systematic review of programs treating High-Need and high-cost people with multiple chronic diseases or disabilities in the United States, 2008-2014. *Prev Chronic Dis* 2015;12:E197.
10 Baker JM, Grant RW, Gopalan A. A systematic review of care management interventions targeting multimorbidity and high care utilization. *BMC Health Serv Res* 2018;18:1–9.
11 Peikes D, Chen A, Schore J, *et al*. Effects of care coordination on hospitalization, quality of care, and health care expenditures among Medicare beneficiaries. *JAMA* 2009;301:603–18.
12 Chakravarty S, Cantor JC. Informing the design and evaluation of Superuser care management initiatives. *Med Care* 2016;54:860–7.
13 Johnson TL, Rinehart DJ, Durfee J, *et al*. For many patients who use large amounts of health care services, the need is intense yet temporary. *Health Aff* 2015;34:1312–9.
14 Longden T, Wong CY, Haywood P, *et al*. The prevalence of persistence and related health status: an analysis of persistently high healthcare costs in the short term and medium term. *Soc Sci Med* 2018;211:147–56.
15 Wodchis WP, Austin PC, Henry DA. A 3-year study of high-cost users of health care. *Can Med Assoc J* 2016;188:182–8.
16 Yoon J, Chee CP, Su P, *et al*. Persistence of high health care costs among Va patients. *Health Serv Res* 2018;53:3898–916.
17 Figueroa JF, Zhou X, Jha AK. Characteristics and spending patterns of persistently high-cost Medicare patients. *Health Aff* 2019;38:107–14.
18 Hirth RA, Gibson TB, Levy HG, *et al*. New evidence on the persistence of health spending. *Med Care Res Rev* 2015;72:277–97.
19 Hwang W, LaClair M, Camacho F, *et al*. Persistent high utilization in a privately insured population. *Am J Manag Care* 2015;21:309–16.
20 Ministry of Health Singapore. Singapore's healthcare system, 2018. Available: https://www.moh.gov.sg/our-healthcare-system/healthcare-services-and-facilities/hospital-services [Accessed 21 Jun 2019].
21 Ministry of Health Singapore. Healthcare schemes and subsidies, 2019. Available: https://www.moh.gov.sg/cost-financing/healthcare-schemes-subsidies [Accessed 1 Jul 2019].
22 Lim W, Wong S, Leong I, *et al*. Forging a frailty-ready healthcare system to meet population ageing. *Int J Environ Res Public Health* 2017;14:1448.
23 Chang H-Y, Boyd CM, Leff B, *et al*. Identifying consistent high-cost users in a health plan: comparison of alternative prediction models. *Med Care* 2016;54:852–9.
24 Hosmer DW, Lemeshow S. Assessing the fit of the model. In: *Applied logistic regression*. New York: John Wiley and Sons, 2000: 160–4.
25 Santosh A. Shinde M, P. Raja Rajeswari D. Intelligent health risk prediction systems using machine learning: a review. *IJET* 2018;7:1019–23.
26 Ahmad P, Qamar S, Qasim Afser Rizvi S. Techniques of data mining in healthcare: a review. *Int J Comput Appl* 2015;120:38–50.
27 Jothi N, Rashid Nur'Aini Abdul, Husain W. Data mining in healthcare – a review. *Procedia Comput Sci* 2015;72:306–13.
28 Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods* 2018;15:233–4.
29 Moturu ST, Johnson WG, Liu H. Predictive risk modelling for forecasting high-cost patients: a real-world application using Medicaid data. *Int J Biomed Eng Technol* 2010;3:114.
30 Yang C, Delcher C, Shenkman E, *et al*. Machine learning approaches for predicting high cost high need patient expenditures in health care. *Biomed Eng Online* 2018;17:1–20.
31 Frost DW, Vembu S, Wang J, *et al*. Using the electronic medical record to identify patients at high risk for frequent emergency department visits and high system costs. *Am J Med* 2017;130:601.e17–601.e22.
32 Izad Shenas SA, Raahemi B, Hossein Tekieh M, *et al*. Identifying high-cost patients using data mining techniques and a small set of non-trivial attributes. *Comput Biol Med* 2014;53:9–18.
33 Rahman N, Wang DD, Ng SH-X, *et al*. Processing of electronic medical records for health services research in an academic medical center: methods and validation. *JMIR Med Inform* 2018;6:e10933.
34 Riley GF. Long-Term trends in the concentration of Medicare spending. *Health Aff* 2007;26:808–16.
35 Schousboe JT, Paudel ML, Taylor BC, *et al*. Estimating true resource costs of outpatient care for Medicare beneficiaries: standardized costs versus Medicare payments and charges. *Health Serv Res* 2016;51:205–19.
36 du Prel J-B, Röhrig B, Hommel G, *et al*. Choosing statistical tests: part 12 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* 2010;107:343–8.
37 Rice WR. Analyzing tables of statistical tests. *Evolution* 1989;43:223–5.
38 Bhardwaj R. A study of machine learning in healthcare. In: *2017 IEEE 41st annual computer software and applications conference*, 2017: 236–41.
39 Callahan A, Shah NH. *Machine learning in healthcare*. Elsevier Inc, 2017.
40 Kennedy EH, Wiitala WL, Hayward RA, *et al*. Improved cardiovascular risk prediction using nonparametric regression and electronic health record data. *Med Care* 2013;51:251–8.
41 Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit* 1997;30:1145–59.
42 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–45.
43 Sun Y, Wong AKC, Kamel MS. Classification of imbalanced data: a review. *Intern J Pattern Recognit Artif Intell* 2009;23:687–719.
44 Haixiang G, Yijing L, Shang J, *et al*. Learning from class-imbalanced data: review of methods and applications. *Expert Syst Appl* 2017;73:220–39.
45 RStudio Team. RStudio: integrated development for R, 2015. Available: http://www.rstudio.com
46 Kuhn M. Caret: classification and regression training, 2018. Available: https://cran.r-project.org/web/packages/caret/index.html
47 Robst J. Developing models to predict persistent high-cost cases in Florida Medicaid. *Popul Health Manag* 2015;18:467–76.
48 Ronksley PE, McKay JA, Kobewka DM, *et al*. Patterns of health care use in a high-cost inpatient population in Ottawa, Ontario: a retrospective observational study. *CMAJ Open* 2015;3:E111–8.
49 Coughlin TA, Long SK. Health care spending and service use among high-cost Medicaid beneficiaries, 2002–2004. *Inquiry* 2009;46:405–17.
50 Davies A, Ariti C, Georghiou T, *et al*. *Evaluation of complex health and care interventions using retrospective matched control methods*. London, 2015.
51 Saxena N, You AX, Zhu Z, *et al*. Singapore's regional health systems-a data-driven perspective on frequent admitters and cross utilization of healthcare services in three systems. *Int J Health Plann Manage* 2017;32:36–49.
52 Singer AE, Goebel JR, Kim YS, *et al*. Populations and interventions for palliative and end-of-life care: a systematic review. *J Palliat Med* 2016;19:995–1008.
53 Vuik SI, Mayer EK, Darzi A. Patient segmentation analysis offers significant benefits for integrated care and support. *Health Aff* 2016;35:769–75.
54 Ang IYH, Tan CS, Nurjono M, *et al*. Retrospective evaluation of healthcare utilisation and mortality of two post-discharge care programmes in Singapore. *BMJ Open* 2019;9:e027220.