

BMJ Open A retrospective cohort study of high-impact users among patients with cerebrovascular conditions

Ahsan Rao,¹ Alice Jones,¹ Alex Bottle,¹ Ara Darzi,² Paul Aylin¹

To cite: Rao A, Jones A, Bottle A, *et al.* A retrospective cohort study of high-impact users among patients with cerebrovascular conditions. *BMJ Open* 2017;7:e014618. doi:10.1136/bmjopen-2016-014618

► Prepublication history for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2016-014618>).

Received 12 October 2016
Revised 30 March 2017
Accepted 28 April 2017



CrossMark

¹Faculty of Medicine, Dr Foster Unit, Imperial College London, Dorset Rise, UK

²Faculty of Medicine, Global Health, Imperial College London, London, UK

Correspondence to

Ahsan Rao; a.rao@imperial.ac.uk

ABSTRACT

Objective To apply group-based trajectory modelling (GBTM) to the hospital administrative data to evaluate, model and visualise trends and changes in the frequency of long-term hospital care use of the subgroups of patients with cerebrovascular conditions.

Design A retrospective cohort study of patients with cerebrovascular conditions.

Settings Secondary care of all patients with cerebrovascular conditions admitted to English National Hospital Service hospitals.

Participants All patients with cerebrovascular conditions identified through national administrative data (Hospital Episode Statistics) and subsequent emergency hospital admissions followed up for 4 years.

Main outcome measure Annual number of emergency hospital readmissions.

Results GBTM model classified patients with intracranial haemorrhage (n=2605) into five subgroups, whereas ischaemic stroke (n=34 208) and transient ischaemic attack (TIA) (n=20 549) patients were shown to have two conventional groups, low and high impact. The covariates with significant association with high-impact users (17.1%) among ischaemic stroke were epilepsy (OR 2.29), previous stroke (OR 2.18), anxiety/depression (OR 1.63), procedural complication (OR 1.43), admission to intensive therapy unit (ITU) or high dependency unit (HDU) (OR 1.42), comorbidity score (OR 1.36), urinary tract infections (OR 1.32), vision loss (OR 1.32), chest infections (OR 1.25), living alone (OR 1.25), diabetes (OR 1.23), socioeconomic index (OR 1.20), older age (OR 1.03) and prolonged length of stay (OR 1.00). The covariates associated with high-impact users among TIA (20.0%) were thromboembolic event (OR 3.67), previous stroke (OR 2.51), epilepsy (OR 2.25), hypotension (OR 1.86), anxiety/depression (OR 1.63), amnesia (OR 1.62), diabetes (OR 1.58), anaemia (OR 1.55), comorbidity score (OR 1.39), atrial fibrillation (OR 1.27), living alone (OR 1.25), socioeconomic index (OR 1.13), older age (OR 1.04) and prolonged length of stay (OR 1.02). The high-impact users (0.5%) among intracranial haemorrhage were strongly associated with thromboembolic event (OR 20.3) and inversely related to older age (OR 0.58).

Conclusion GBTM effectively assessed trends in the use of hospital care by the subgroups of patients with cerebrovascular conditions. High-impact users persistently had higher annual readmission during the follow-up period.

Strengths and limitations of this study

- Patients with cerebrovascular conditions are known to have high unplanned readmission rate as evaluated using national administrative data, but limited evidence exists on the predictors and long-term hospital care use among their high-impact users.
- A novel application of group-based trajectory modelling to national administrative data was used to model and visualise trends in the long-term hospital care use of subgroups in the patient population, especially high-impact users.
- The model identified additional groups among patients with non-traumatic intracranial haemorrhage, which were not categorised by earlier studies.
- The selection of cases in this retrospective study may lead to a degree of selection bias, and the data did not include information on the use of accident and emergency care.

INTRODUCTION

Various efforts have been made to reduce readmission rate.¹ The programmes to decrease readmissions in cerebrovascular conditions have focused on discharge planning and community support of discharged patients.^{2,3} They have been resource intensive because they require integration of services from various health professionals, training of staff and large investment for administrative structure build-up.⁴ It has been recommended that these interventions should be targeted to the subgroups of patients, especially those who use most of the health resources and are at a higher risk of readmission.⁵

The classification of the patient population into different subgroups based on the hospital care use had been arbitrary and variable with no standard methodology.^{6,7} In previous studies, patients have been broadly categorised into two groups: high-impact and low-impact users.⁸ High-impact users are a small proportion of the total patient population but resource consumption is significantly

higher.⁸ The main focus of previous studies was to identify predictive factors associated with the high-impact user by conducting logistic regression model analysis.⁷⁻⁹ The statistical model only measured change over two time points and had limited ability to assess dynamic developmental changes over time.¹⁰ Similarly, in the other statistical tests, such as t-test, ANOVA and multiple regressions, the data are pooled from all the individuals in the study population and any change in the dependent variable is studied over two time points.¹⁰ However, methods such as cluster analysis categorise population into subgroups based on similar properties but do not focus on the identifying risk factors associated with the patient subgroups.^{11 12}

A robust methodology is required to model and visualise changes in the frequency of healthcare use in different subgroups in the patient population so that interventions developed to reduce readmission rate are cost-effective, parsimonious and its effects are long lasting.^{4 13} Some evidence suggest that more than two traditional groups exist, each with unique characteristics, if a population is observed for a longer period of time.^{14 15} In a recent study, five subgroups of patients were identified based on the pattern of recovery following stroke, each with a distinct prognosis and risk of mortality.¹⁶ Similar observations are present in patients with other conditions such as surgery and heart failure.^{17 18} It has been suggested that more than half of the high-risk users decreased their use of the healthcare visits after 1 year and a smaller proportion of the high-risk patients persistently and increasingly used healthcare services throughout the 4-year follow-up.¹⁹

Group-based trajectory modelling (GBTM) has been used in social and psychological sciences to understand changes in behaviour in the population and categorise pupils based on common developmental pathways.^{20 21} This growth model is novel in medical research but can be used to study variation in the long-term progression of the disease and its impact on the use of healthcare resources.¹³ The model has the ability to identify the subgroups in the population with similar progression of the outcomes and recognise the covariates associated with each group.²¹ Unlike the other models, it evaluates development of each group over multiple time points by analysing repeated measurement of the same outcome.¹³ The aim of the study is to evaluate the long-term progression of patients with cerebrovascular conditions based on their healthcare resource use by conducting GBTM of the longitudinal data. We hypothesise that GBTM can be used to categorise subgroups in a patient population and assess the trends in frequency of the hospital care use.

METHODS

Database

Data from Hospital Episode Statistics (HES) were used for this retrospective cohort study. It is the collection of patient administrative data managed by the Department of Health, Government of England. It covers information on inpatient hospital stays in public National Hospital

Service (NHS) hospitals as well as information on private patients treated in these hospitals.²² All emergency cases are admitted and initially treated in these hospitals.²² Each hospital admission is recorded as a 'spell' consisting of a number of 'consultant episodes', which denotes period of care under different consultants during their hospital admission.²³ If the patient admission includes transfer to another hospital before one is discharged, the whole period of care is recorded under 'superspell'. For each patient, information from their superspell was obtained, such as primary diagnosis, number of secondary diagnoses, primary operation, admission date, discharge date, length of stay (LOS), discharge destination, admission source etc. The primary diagnosis and the list of the secondary diagnosis are recorded using ICD-10 classification, whereas OPCS (Office of Population Censuses and Surveys) 4.7 coding is used for primary and list of secondary procedures. Each patient had a HES unique identifier that was used to recognise further hospital episodes. Ethical approval was obtained through the Health and Social Care Information Centre while obtaining the access to the pseudonymised patient administrative data.

Study population

All adult patients over the age of 18 years who had cerebrovascular conditions in the year 2010 were included in the study. The patient cohort was identified using specific ICD-10 codes (International Classification of Diseases) for all the index admissions in the year 2010. Similar codes were used in earlier studies: ischaemic stroke (I63x), transient ischaemic attack (TIA) (G45x, H34x) and non-traumatic intracranial haemorrhage (subarachnoid haemorrhage (I60x), intracerebral haemorrhage (I61x), other non-traumatic intracranial haemorrhage (I62x)).²⁴⁻²⁷ Once the patients were identified, the previous 10 years of HES data were examined to identify any history of previous stroke event admitted to an English NHS hospital. Patients with a history of stroke were retained in the data analysis and previous stroke was used as a covariate to assess its association with the subgroups. The patients were followed up for at least 4 years. The data were retrieved for each patient every time they were admitted to an English NHS hospital and therefore recorded in HES. ICD-10 and OPCS 4.7 codes were used to identify the main diagnosis and procedure associated with each hospital admission. The primary outcome was the number of emergency hospital readmissions every year.

Covariates

Various covariates have been analysed for their association with increased readmission rate. The list of these covariates was retrieved from previous clinical and population-based studies on the outcomes of the patients with cerebrovascular conditions.²⁴⁻²⁹ They can be broadly divided into patient characteristics, disease management, disease-associated adverse conditions

and hospital-acquired conditions (HACs). Age, gender, socioeconomic deprivation, comorbidity score, history of stroke and living alone were evaluated as patient characteristics. The patient characteristics include patient demographics and medical history. The conditions included in the medical history were adapted from previous clinical studies, which have shown these conditions having an impact on the readmission rate.³⁰ Disease management factors included discharge to nursing home, thrombolysis, disease-related procedures (carotid endarterectomy, craniectomy, carotid stenting, aneurysm repair, and craniotomy), the use of other procedures (tracheostomy, intubation, urinary catheterisation, invasive or non-invasive ventilation, percutaneous endoscopic gastrostomy or nasogastric tube insertion, and renal replacement therapy) and LOS. The management factors used in this study had been previously assessed using administrative data.³¹ Some of them were shown to have an impact on the short-term readmission rate, but their association with high-impact users was not assessed. The effect of the use of procedures was also assessed because they may lead to short-term and long-term complications, which may impact readmission rate.³² Each individual procedure was low in number; hence, they were grouped as disease-related and other procedures for valid comparison analysis. The disease-associated adverse events comprised hearing loss, vision loss, paralysis, cranial nerve palsy, speech and swallowing disorders, amnesia or coma, hydrocephalus, hypotension, hypertension, atrial fibrillation, renal failure, depression and anxiety, dementia, epilepsy or seizure and thromboembolic event (recurrent stroke, myocardial infarction, acute coronary syndrome, acute limb ischaemia and pulmonary embolism). These disease-related events showed the impact of a particular type of stroke-related disability on the readmission rate.²⁸ Other medical conditions included in the category had been shown by clinical studies to have high prevalence in patients with stroke.³³ Thromboembolic event consisted of a group of conditions of similar pathological process, which we expected to alter the readmission rate in the same manner. As assessed by previous studies, HACs consisted of procedural complications (bleeding, skin infection, foreign body complications), drug errors and side effects, trauma, falls and fractures, pulmonary embolism, deep venous thrombosis, pressure ulcers, pneumothorax, metabolic disorders, infections (urinary and chest infection, gastroenteritis and cellulitis) and blood transfusion reactions.^{24 27} ICD-10 codes were used to identify cases who had medical complications, whereas OPCS 4.7 codes were used to identify specific procedures carried out on patients with cerebrovascular conditions. The population-weighted quintiles of Carstairs deprivation score were used to classify patients according to their deprivation levels.³⁴ The quintile ranges from 1 to 6, where a score of 5 is defined as the most deprived residences and 6 means not known (missing postcode). Charlson score was used to calculate the comorbidity burden associated with each patient.³⁵ Higher score was

associated with the severity of comorbidity. Charlson score for the admission of cerebrovascular condition was obtained from the sum of score of past medical problems as listed in previous studies.³⁵ In addition, association of medical history of vascular conditions with the outcome was also assessed, such as history of stroke, ischaemic heart disease, peripheral vascular disease and diabetes. In the model, the impact of the covariate is measured as the change in the probability of the membership to the subgroups. The covariate with the positive impact on the subgroup will increase the number of patients with the presence of a covariate in that subgroup. Certain covariates may be inter-related, but each of them was shown to have an impact on the readmission rate independent of the other factors.³⁰ Hence, they were separately included in the analysis. Adjusted ORs were calculated by the use of multinomial logistic regression where the low-impact group was used for comparison.

Statistical analysis

A Statistical Analysis Software macro, 'proc traj', was used for GBTM to assess and predict systematic changes in the outcome for each individual in the study population.²¹ It is a semi-parametric model that relies on repeated measures of the outcome over time. It relaxes the assumption of one trajectory for one population and allows each subgroup to follow its own trajectory. The model was amended for it to be used for the administrative data. The format of administrative data was changed from compilation of spells for hospital admissions annually to longitudinal distribution for each patient. The number of annual readmissions was calculated for each patient. The information on each covariate was extracted from the data. The use of zero-inflated Poisson analysis with the continuous outcome data based on the total number of annual readmissions for each patient formulated a good fit model with trajectory of the hospital care use based on mean readmission rate.

GBTM with zero-inflated Poisson analysis was used to categorise individuals into different subgroups based on the continuous outcome data.²⁰ The outcome was the annual number of emergency readmissions for each patient. In order to determine the optimum number of subgroups within a population, the model was chosen based on the following criteria: smallest value of Bayesian Information Criteria (BIC), smallest value of Akaike Information Criteria (AIC), each trajectory with significant parameter estimates ($p < 0.05$), largest value for average posterior probability for each group and minimum of 5% of patient population in a subgroup trajectory. For each group, the mean of the posterior probability of the individuals of more than 0.7 was used to indicate adequate internal reliability. Models with trajectory groups ranging from 2 to 6 were tested. The trajectory shape for each group was assessed with different types of parameters (polynomial order), with the order of complexity increasing from intercept and linear to quadratic and cubic. The model with the highest number of groups, yet

parsimonious, was selected. The analysis was conducted on three patient populations: ischaemic stroke, TIA and non-traumatic intracranial haemorrhage.

The likelihood of an individual belonging to a group is calculated by posterior probability.¹³ The average posterior probability of the group was the mean of the posterior probability assigned to the group members. The membership of an individual to a trajectory is based on the probability but not certainty; hence, the conventional cross-group comparisons to assess correlation between the covariates and the groups cannot be used. Instead, the multinomial logistic regression model was used to assess the impact of covariates on the probability of group membership while controlling for other confounding factors.²¹ The group with persistently lowest use of the hospital care use was labelled as 'low-impact users' and used as a reference group. The association of each covariate was measured as the OR, with 95% CI, of the impact of that covariate on the probability of membership in the specified group relative to the stable low-impact group.

RESULTS

Ischaemic stroke

The patient population (n=34208) consisted of 51% men and 49% women, with a mean age of 72.17 (SD 13.37). The mean LOS was 15.35 (SD 22.47), 14.18% of the patients lived alone and 5.38% of them were discharged to a nursing home. The mortality rate was 5.8% (n=4853) by the end of the follow-up period. The trajectory modelling based on the hospital care use identified two subgroups of the patient population (BIC=-127 547, AIC=-127 509): group 1 (low-impact users, n=28 358 (82.9%)) and group 2 (high-impact users, n=5849 (17.1%)) (figure 1). The covariates associated with the high-impact users when compared with the low-impact users are listed in table 1. The overall mortality was significantly lower in high-impact users (n=195 (3.2%) vs 1798 (6.4%), p<0.001) when compared with the low-impact group.

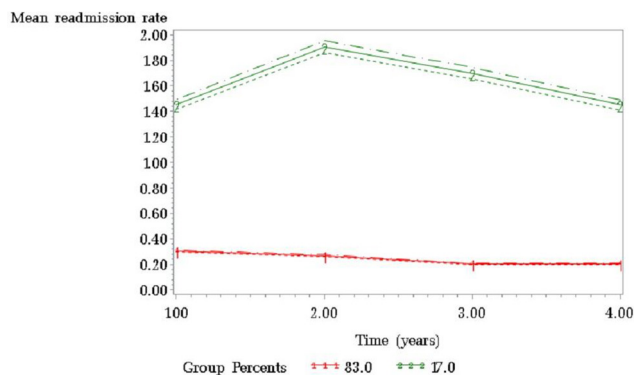


Figure 1 Trajectory pathways of subgroups of patients with ischaemic stroke. The horizontal axis starts with annual readmission rate at year 1 and the dotted lines represent 95% CIs for each subgroup.

Transient ischaemic attack

The patient population (n=20 549) consisted of 49% men and 51% women, with a mean age of 72.25 (SD 13.63). The mean LOS was 2.96 (SD 6.12) and 10.9% of the patients lived alone. On admission, 23.2% of the patients suffered hospital-acquired complications and 1.00% of them were discharged to a nursing home. The mortality rate was 4.60% (n=945) by the end of the follow-up period. The trajectory modelling based on the hospital care use identified two subgroups of the patient population (n=20 549, BIC=-76 497, AIC=-76 462): group 1, low-impact users (n=16 439, 80.0%); and group 2, high-impact users (n=4110, 20.0%) (figure 2). The covariates associated with the high-impact users when compared with the low-impact users are listed in table 2. High-impact users had significantly higher mortality (n=328 (5.4%) vs 617 (4.3%), p<0.001) when compared with the low-impact group.

Intracranial haemorrhage

The patient population (n=2605) consisted of 63% men and 37% women, with a mean age of 72.25 (SD 13.63). The mean LOS was 10.82 (SD 17.62) and 10.9% of the patients lived alone. On admission, 23.2% of the patients suffered hospital-acquired complications and 3.34% of them were discharged to a nursing home. The mortality rate was 5.57% (n=145) by the end of the follow-up period. The trajectory modelling based on the hospital care use identified five subgroups of the patient population (n=2605, BIC=-2704.19, AIC=-2683.7): group 1 (n=1391, 53.4%), group 2 (n=745, 28.6%), group 3 (4.9%), group 4 (n=328, 12.6%) and group 5 (high-impact group, n=13, 0.5%) (figure 3). Most of the patients were members of group 1 with the least readmission rate in the follow-up. They were considered as low impact and other groups were compared with it to assess the association of covariates. Group 2 was significantly associated with non-Caucasian ethnicity (n=731 (92.9%) vs 180 (12.6%), OR 39.2 (29.4 to 52.5), p<0.001), stroke LOS (mean 14.8 (SD 22.6) days vs 23.5 (SD 40.2), OR 0.99 (0.99 to 0.99), p=0.003), socioeconomic index (mean 2.9 (SD 1.5) vs 3.0 (SD 1.4), OR 0.78 (0.73 to 0.83), p<0.001), epilepsy (n=38 (4.8%) vs 94 (6.6%), OR 0.45 (0.31 to 0.66), p=0.03) and hypertension (n=285 (36.2%) vs 653 (45.6%), OR 0.66 (0.55 to 0.80), p=0.03). They had a slightly higher readmission rate than low-impact users. Group 3 was significantly associated with history of stroke (n=35 (20.6%) vs 165 (11.5%), OR 1.82 (1.40 to 2.36), p=0.02) and epilepsy (n=21 (12.3%) vs 94 (6.6%), OR 1.80 (1.32 to 2.46), p=0.05). They had a progressive rise in readmission rate with time. Group 4 was significantly associated with anxiety and depression (n=44 (22.2%) vs 152 (10.6%), OR 1.92 (1.43 to 2.56), p=0.02), number of hospital-acquired complications (mean=0.9 (SD 0.9) vs 0.4 (SD 0.7), OR 1.67 (1.40 to 1.97), p=0.002) and socioeconomic index (mean 3.4 (SD 1.3) vs 3.0 (SD 1.4), OR 1.20 (1.11 to 1.30), p=0.02). They had a high readmission

Table 1 Association of various covariates with the high-impact users when compared with the low-impact users in patients with ischaemic stroke

Covariates	Low-impact (n (%) or mean (SD))	High-impact (n (%) or mean (SD))	OR (95% CI)	p Value
Patient characteristics				
History of stroke	2615 (9.3)	1339 (21.6)	2.18 (2.10 to 2.27)	<0.001
Ischaemic heart disease	1532 (5.5)	659 (10.6)	1.84 (0.90 to 3.78)	0.39
Anxiety/depression	1114 (3.9)	373 (6.0)	1.63 (1.52 to 1.75)	<0.001
Charlson comorbidity score	1.5 (0.8)	1.9 (1.1)	1.36 (1.34 to 1.39)	<0.001
Living alone	3719 (13.3)	1134 (18.3)	1.25 (1.20 to 1.30)	<0.001
Diabetes	4679 (16.7)	1694 (27.3)	1.23 (1.16 to 1.31)	0.001
Socioeconomic index	3.0 (1.4)	3.3 (1.4)	1.20 (1.19 to 1.21)	<0.001
Anaemia	576 (2.1)	183 (2.9)	1.12 (1.00 to 1.25)	0.29
Hypertension	15 770 (56.30)	3837 (61.9)	1.06 (1.03 to 1.09)	0.09
Atrial fibrillation	6267 (22.4)	1681 (27.1)	1.05 (1.01 to 1.09)	0.18
Age	71.4 (13.6)	75.4 (11.6)	1.03 (1.03 to 1.03)	<0.001
Female sex	13 287 (47.4)	3078 (49.7)	0.95 (0.92 to 0.98)	0.12
Non-Caucasian ethnicity	3897 (13.9)	743 (11.9)	0.94 (0.93 to 0.95)	<0.001
Renal failure	252 (0.9)	93 (1.5)	0.90 (0.78 to 1.05)	0.49
Dementia	1052 (3.8)	210 (3.4)	0.59 (0.54 to 0.65)	<0.001
Stroke management factors				
Procedural complication	440 (1.6)	168 (2.7)	1.43 (1.27 to 1.62)	0.002
Admission to ITU/HDU	460 (1.6)	248 (4.0)	1.42 (1.27 to 1.58)	0.001
Length of stay	20.8 (30.7)	26.7 (35.1)	1.00 (1.00 to 1.01)	<0.001
Discharge to nursing home	1522 (5.4)	319 (5.2)	0.61 (0.56 to 0.66)	<0.001
Stroke-associated adverse conditions				
Epilepsy	620 (2.2)	314 (5.1)	2.29 (2.10 to 2.51)	<0.001
Vision loss	361 (1.3)	149 (2.4)	1.32 (1.17 to 1.49)	0.02
Amnesia	297 (1.1)	104 (1.7)	1.17 (1.02 to 1.35)	0.26
Speech and swallowing disorders	660 (2.4)	186 (3.0)	1.08 (0.97 to 1.21)	0.45
Cranial nerve palsy	304 (1.1)	84 (1.4)	1.08 (0.92 to 1.27)	0.61
Hearing loss	475 (1.7)	145 (2.3)	1.06 (0.94 to 1.20)	0.58
Incontinence	857 (3.1)	264 (4.3)	1.05 (0.96 to 1.15)	0.58
Thrombolysis	1779 (6.3)	338 (5.5)	1.01 (0.94 to 1.08)	0.92
Thromboembolic event	1758 (6.3)	684 (11.0)	0.85 (0.41 to 1.75)	0.82
Paralysis	1025 (3.7)	342 (5.5)	0.68 (0.62 to 0.74)	<0.001
Hospital-acquired conditions				
Urinary tract infection	1283 (4.6)	463 (7.5)	1.32 (1.22 to 1.43)	0.009
Chest infection	1184 (4.2)	451 (7.3)	1.25 (1.14 to 1.36)	0.01
Total no of hospital-acquired complications	0.34 (0.62)	0.51 (0.71)	0.99 (0.94 to 1.04)	0.78
Pulmonary embolism/deep vein thrombosis	236 (0.8)	27 (0.4)	0.58 (0.29 to 1.16)	0.43

rate in the beginning but rapidly declined during the follow-up period. Group 5 was significantly associated with thromboembolic event (n=6 (35.3%) vs 87 (6.1%), OR 20.3 (9.6 to 42.9), p<0.001) and age (mean=60.0 (SD 18.6) vs 68.6 (SD 17.7), OR 0.58 (0.46 to 0.73), p=0.01).

Group 5 was labelled as high-impact users because they had a persistently high readmission rate throughout the follow-up period. The mortality rate during the follow-up period was significantly high in group 4 (n=37 (18.7%)) when compared with group 1 (n=81 (5.6%)),

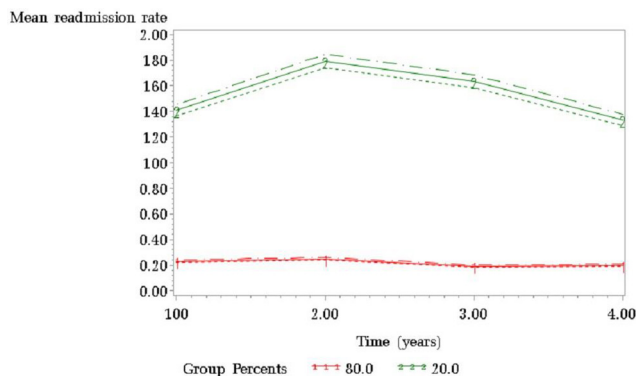


Figure 2 Trajectory pathways of subgroups of patients with TIA. The horizontal axis starts with annual readmission rate at year 1 and the dotted lines represent 95% CIs for each subgroup.

group 2 (n=27 (3.4%)), group 3 (n=0) and group 5 (n=0) ($p<0.001$).

DISCUSSION

On the basis of the long-term hospital care use, the ischaemic stroke and TIA populations consisted of two subgroups, while five subgroups were identified in the intracranial haemorrhage patient population. The majority of the patients were low-impact users and the use of hospital care was minimal and stable when compared with the other subgroups. The high-impact users had a persistently high readmission rate throughout the follow-up period. Among stroke and TIA patients, a significant proportion of patients were high-impact users. Older age, cardiovascular conditions, increased comorbidity burden, poor socioeconomic status, prolonged LOS, mental health conditions, epilepsy and living alone were common risk factors associated with high-impact users of stroke and TIA patients. Persistent high-impact users among intracranial haemorrhage were young patients who had a thromboembolic event. Among intracranial haemorrhage, a subgroup (group 3) had a significant rise in readmission rate. They were associated with a history of stroke and epilepsy. Group 4 among intracranial haemorrhage had a rapid decline in readmission rate due to high mortality rate. They had an increased number of hospital-acquired complications and were associated with poor socioeconomic index.

GBTM has the advantage over other models used to study longitudinal data.²¹ The expected trajectory of each subgroup is based on repeated observations over time. It assumes that the subgroups are part of the same population but each follows different developmental pathways. It does not pre-empt the number of groups but uses the statistical device for approximating the unknown distribution of trajectories across the population.²¹ The best fit model for ischaemic stroke and TIA had conventional two subgroups: low-impact and high-impact groups. However, the intracranial haemorrhage patient population were shown to three further subgroups, other than two

conventional groups. One of the subgroups, group 4, had a significant number of hospital-acquired complications that led to high mortality, which caused a rapid decline in the readmission rate. Medical history of stroke and epilepsy among patients with intracranial haemorrhage puts them at risk of progressive rise in readmission rate as shown by subgroup 3. Group 2 had a slightly increased readmission rate than the low-impact group. These patients had low socioeconomic status, had non-Caucasian ethnicity and had cardiovascular risk factors. Further research should focus on these subgroups with a significant number of patients to explore causes of mortality and readmissions. This may help to assess if these causes can be prevented to improve survival and reduce readmission rate in these patients.

The characteristics of other established models used to study the trajectory of subgroups follow assumptions that may not make it as competent a model as GBTM to study the long-term clinical outcomes.¹³ Growth curve models consist of hierarchical modelling and latent curve growth analysis.¹³ Yet different but they share common properties.¹³ They assume normal and continuous distribution of the patient population. They estimate the average trend in the development for the whole population. All the individuals follow a similar trend, and the subgroups are formed based on variation from the average trend. However, growth mixture modelling assumes that the population consists of two or more distinct groups.¹³ Outcome of each group is calculated as a separate component of the model. Each group has its own mean and variation. In all these models, the number of subgroups is finite and based on the pre-determined hypothesis of the study.¹³ In contrast, GBTM assumes that different trajectories in a population exist due to inter-individual variation within the same population.²¹ The number of groups is not finite and their trajectories are based on actual observation of outcome.³⁶ It is particularly beneficial for trajectories of unknown shapes and where there is a likelihood of unpredicted observations. The individuals within the group are more homogeneous in characteristics, and the variability is assessed by comparison of the groups rather than the individuals. It provides the advantage of quantifying probability of an individual to be associated with the group by measuring group membership probability score.³⁶

The covariates with significant association with the high-impact user had been shown in previous work to increase short-term readmission rate.³¹ Previous stroke and history of epilepsy were strong risk factors for being a high-impact user among ischaemic stroke and TIA patients. It can be inferred from the data that elderly patients with vision loss, previous stroke, diabetes, mental health disorders and increased comorbidity, who had prolonged LOS for stroke, admission to HDU/ITU, procedural complications as well as infections, had a higher chance of becoming long-term high-impact users among patients with stroke. Cardiovascular conditions and mental health disorders had been shown to influence readmission

Table 2 Association of various covariates with the high-impact users when compared with the low-impact users in patients with TIA

Covariates	Low-impact (n (%) or mean (SD))	High-impact (n (%) or mean (SD))	OR (95% CI)	p Value
Patient characteristics				
History of stroke	1477 (10.2)	1695 (27.9)	2.51 (2.36 to 2.66)	<0.001
Hypotension	85 (0.6)	98 (1.6)	1.86 (1.49 to 2.32)	0.004
Anxiety/depression	503 (3.5)	296 (4.9)	1.63 (1.46 to 1.82)	<0.001
Diabetes	2146 (14.8)	1596 (26.3)	1.58 (1.42 to 1.77)	<0.001
Anaemia	231 (1.6)	213 (3.5)	1.55 (1.35 to 1.79)	0.002
Charlson comorbidity score	1.4 (0.7)	1.8 (1.0)	1.39 (1.35 to 1.43)	<0.001
Atrial fibrillation	1969 (13.6)	1425 (23.5)	1.27 (1.20 to 1.35)	<0.001
Living alone	1297 (8.9)	955 (15.7)	1.25 (1.16 to 1.34)	0.001
Socioeconomic index	2.9 (1.4)	3.2 (1.3)	1.13 (1.12 to 1.14)	<0.001
Female sex	7120 (49.2)	3353 (55.2)	1.05 (1.01 to 1.09)	0.23
Age	70.1 (13.8)	77.4 (11.6)	1.04 (1.04 to 1.04)	<0.001
Hypertension	7491 (51.7)	3497 (57.6)	0.94 (0.90 to 0.97)	0.13
Dementia	627 (4.3)	427 (7.0)	0.92 (0.84 to 1.02)	0.39
Ischaemic heart disease	736 (5.1)	709 (11.7)	0.46 (0.25 to 0.84)	0.19
Non-Caucasian ethnicity	5753 (39.7)	391 (6.1)	0.11 (0.10 to 0.12)	<0.001
Stroke management factors				
Admission to ITU/HDU	109 (0.7)	149 (2.4)	1.43 (1.16 to 1.77)	0.09
Procedural complications	124 (0.9)	80 (1.3)	1.40 (1.13 to 1.75)	0.11
Length of stay	2.6 (6.1)	4.8 (10.3)	1.02 (1.01 to 1.02)	<0.001
Stroke-associated adverse conditions				
Thromboembolic event	747 (5.2)	723 (11.9)	3.67 (2.01 to 6.69)	0.03
Epilepsy	235 (1.6)	229 (3.8)	2.25 (1.95 to 2.59)	<0.001
Amnesia	95 (0.7)	97 (1.6)	1.62 (1.31 to 1.99)	0.02
Speech and swallowing disorders	182 (1.3)	129 (2.1)	1.36 (1.15 to 1.62)	0.06
Hearing loss	164 (1.1)	146 (2.4)	1.25 (1.05 to 1.48)	0.18
Cranial nerve palsy	72 (0.5)	41 (0.70)	1.23 (0.92 to 1.65)	0.47
Vision loss	154 (1.1)	115 (1.9)	1.13 (0.94 to 1.35)	0.49
Paralysis	267 (1.8)	241 (3.9)	1.02 (0.89 to 1.17)	0.87
Incontinence	111 (0.8)	75 (1.2)	0.92 (0.75 to 1.14)	0.69
Hospital-acquired conditions				
Urinary tract infection	257 (1.8)	274 (4.5)	1.95 (1.67 to 2.29)	<0.001
Chest infection	156 (1.1)		1.63 (1.32 to 2.01)	0.02
Total no of hospital-acquired complications	0.21 (0.41)	0.36 (0.57)	0.84 (0.76 to 0.91)	0.06

rate.³¹ If the patients with stroke have multiple complications during their stay in hospital, it leads to further insult resulting in long-term readmissions.²⁷ Although similar risk factors were found to have a significant association with high-impact users among TIA, having a thromboembolic event puts them into a higher risk of readmission. This finding correlates to the previous clinical work as patients with TIA are at higher risk of thromboembolic event, and efforts are made to prevent it.³⁷ If they suffer

from this event, then their morbidity rises.³⁷ In this study, various subgroups with distinct morbidity and mortality were identified among patients with intracranial haemorrhage. Those with low socioeconomic status and multiple hospital-acquired complications had the worse mortality rate. However, young patients who had a thromboembolic event were likely to survive but had compromised quality of life due to multiple long-term readmissions. Hence, the study identified potentially preventable

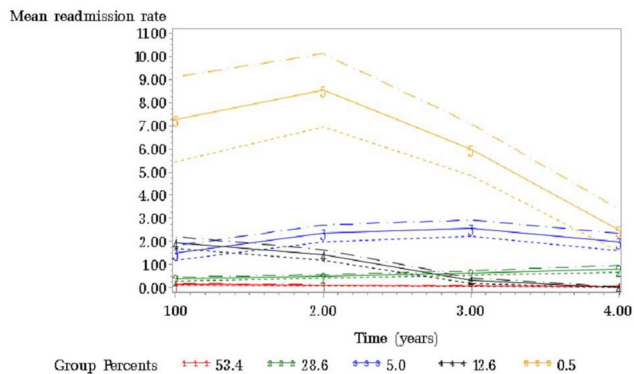


Figure 3 Trajectory pathways of subgroups of patients with intracranial haemorrhage. The horizontal axis starts with annual readmission rate at year 1 and the dotted lines represent 95% CIs for each subgroup.

factors to restrain a patient becoming a high-impact user by focusing on patient management factors, reduction in hospital-acquired conditions and infections, and prevention of thromboembolic events.³³

All the risk factors associated with high-impact users seem to interact with each other and form a vicious cycle around the patient. The risk of further cerebrovascular events increases after stroke, and recurrent stroke is one of the most common causes of readmission in the patient population.^{28 37} The patients with recurrent stroke obtain lesions at multiple sites leading to progressive cognitive decline and poor functional health.³⁸ Having a seizure or an epilepsy after stroke may indicate the severity of stroke.³⁹ It has been related to the involvement of multiple sites, larger lesions, hippocampus region and cortical damage. The patients with cerebrovascular conditions are at a high risk of suffering from urinary and respiratory infections.^{28 40} Urinary infections are also related to the use of urinary catheter and incontinence in the patients with stroke.⁴¹ Respiratory infections due to aspirations are commonly associated with patients with cerebrovascular conditions suffering from immobility and dysphagia causing aspiration pneumonia.^{40 42} Depression has been linked with severe disability after stroke, damage to the neural circuit of mood regulation, anatomical location of stroke and cognitive impairment.⁴³ The prolonged LOS allows multiple adverse factors to interact and cause functional health decline in a patient.⁴⁴

The data analysis had certain limitations. Identification of the patient cohort and the covariates was based on ICD and OPCS coding, which are prone to coding errors in the administrative data collection.⁴⁵ We have tried to use all possible codes that define the condition to include most cases accurately. It is important to consider the issues of big data quality and structuring when interpreting the results. The primary diagnosis associated with each hospital episode is defined by the condition that have incurred the most cost in patient care, which may be different from the condition for which the patient was initially admitted to the hospital. The information on the covariates is obtained from the secondary diagnosis

listed in each hospital spell. It does not distinguish the condition that occurred during the hospital stay or in the past. The data are collected by administrative staff who lack clinical knowledge and are not involved in patient management. Each data may not collect the same information in a similar way compared with other datasets, which makes comparison between different datasets difficult. The model uses repeated observation of a single outcome measure, which may not provide a complete picture of hospital care use. A separate modelling will be required to classify patients based on other outcomes of hospital use, such as cumulative annual LOS, outpatient visits etc. Selection of the cases in a retrospective cohort study may lead to a degree of selection bias. The assessment of healthcare use was based on inpatient hospital stay. It did not include the use of outpatient and accident and emergency services by the subgroups because the national HES data on the other hospital services are limited. Moreover, the focus of the study was to assess trends in inpatient hospital care use. The model produces trajectories of groups but does not have the same ability to predict the trajectory of an individual in the study sample. A minimum of three repeated measures is required for the model to predict a simple trajectory of a subgroup, which can be difficult. The classification of the subgroups is based on one type of outcome. Since very few studies have been conducted on long-term hospital care use with GBTM analysis, comparison of the study results with previous studies was limited.²⁸

With the help of the novel application of the statistical model to hospital administrative data, the study has attempted to categorise the patient population and observe frequency of the hospital care use in the subgroups. It has been shown that the hospital administrative data can be transformed to longitudinal data and repeated measurement of readmissions can be statistically modelled to form groups based on long-term hospital care use. Each individual in the population is provided with the probability score for its membership to the group. Additional analysis can be performed to assess other outcomes in the groups, such as mortality rate, the use of outpatient services, and accident and emergency care use. Further research is required to check its applicability in other administrative data sets, medical conditions and different sample sizes. It will be interesting to find out whether common causes of readmissions in the high-impact users differ from the other groups and if there is a common sequence of causes of readmission in the subgroup. The study suggests that the focus of the health policy makers should be to predict and reduce the use of healthcare resources in the high-impact users. Equally, the recognition of other subgroups and their associated risk factors is vital in cutting cost because they follow a different trajectory than high-impact users, have high mortality and rise in readmission rate.

Correction notice This paper has been amended since it was published Online First. Owing to a scripting error, some of the publisher names in the references were replaced with 'BMJ Publishing Group'. This only affected the full text version,

not the PDF. We have since corrected these errors and the correct publishers have been inserted into the references.

Contributors AR acquired, analysed and interpreted the data and drafted and revised the manuscript. AB conceived and designed the study, and revised the manuscript. AK acquired the data and drafted and revised the manuscript. AD revised the manuscript and supervised the study. PA interpreted the data, revised the manuscript and supervised the study. AR is the guarantor and had full access to all of the data (including statistical reports and tables) in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Competing interests All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare no support from any organisation for the submitted work and no other relationships or activities that could appear to have influenced the submitted work. PA is the co-director of Dr Foster Unit at Imperial College London, which is principally funded via a research grant by Dr Foster Intelligence, an independent healthcare information company and joint venture with the Department of Health.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data 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 and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- Mistiaen P, Francke AL, Poot E. Interventions aimed at reducing problems in adult patients discharged from hospital to home: a systematic meta-review. *BMC Health Serv Res* 2007;7:47.
- Teasell RW, Foley NC, Bhogal SK, et al. Early supported discharge in stroke rehabilitation. *Top Stroke Rehabil* 2003;10:19–33.
- Kwan J, Sandercock P. In-hospital care pathways for stroke. *Cochrane Database Syst Rev* 2003;34:587–8.
- Burke RE, Guo R, Prochazka AV, et al. Identifying keys to success in reducing readmissions using the ideal transitions in care framework. *BMC Health Serv Res* 2014;14:1.
- Burke RE, Coleman EA. Interventions to decrease hospital readmissions: keys for cost-effectiveness. *JAMA Intern Med* 2013;173:695–8.
- Moe J, Kirkland S, Ospina MB, et al. Mortality, admission rates and outpatient use among frequent users of emergency departments: a systematic review. *Emerg Med J* 2016;33:230–6.
- Howell S, Coory M, Martin J, et al. Using routine inpatient data to identify patients at risk of hospital readmission. *BMC Health Serv Res* 2009;9:1.
- Bottle A, Aylin P, Majeed A. Identifying patients at high risk of emergency hospital admissions: a logistic regression analysis. *J R Soc Med* 2006;99:406–14.
- Billings J, Blunt I, Stevenon 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.
- Curran PJ, Hussong AM. The use of latent trajectory models in psychopathology research. *J Abnorm Psychol* 2003;112:526–44.
- Everitt BS, Landau S, Leese M, et al. *Cluster analysis (Wiley Series in Probability and Statistics)*: Wiley, 2011.
- Bland JM, Altman DG. Survival probabilities (the Kaplan-Meier method). *BMJ* 1998;317:1572.
- Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. *Annu Rev Clin Psychol* 2010;6:109–38.
- Reinders I, Murphy RA, Martin KR, et al. Body mass index trajectories in relation to change in lean mass and physical function: The Health, Aging and Body Composition Study. *J Am Geriatr Soc* 2015;63:1615–21.
- Gunn J, Elliott P, Densley K, et al. A trajectory-based approach to understand the factors associated with persistent depressive symptoms in primary care. *J Affect Disord* 2013;148:338–46.
- Arling G, Ofner S, Reeves MJ, et al. Care trajectories of veterans in the 12 months after hospitalization for acute ischemic stroke. *Circ Cardiovasc Qual Outcomes* 2015;8:S131–S140.
- Daoust R, Emond M, Bergeron E, et al. Risk factors of significant pain syndrome 90 days after minor thoracic injury: trajectory analysis. *Acad Emerg Med* 2013;20:1139–45.
- Kheirbek RE, Alemi F, Citron BA, et al. Trajectory of illness for patients with congestive heart failure. *J Palliat Med* 2013;16:478–84.
- Dent A, Hunter G, Webster AP. The impact of frequent attenders on a UK emergency department. *Eur J Emerg Med* 2010;17:332–6.
- Jones BL, Nagin DS. Advances in group-based trajectory modeling and an SAS procedure for estimating them. *Social Methods Res* 2007;35:542–71.
- Nagin D. Group-based modeling of development. *Harvard University Press* 2005.
- Arora S, Charlesworth A, Kelly E, et al. *Public payment and private provision*. 2013.
- Aylin P, Bottle A, Jen MH, et al. HSMR mortality indicators. *Imperial College Technical Document* 2010.
- Rahman M, Neal D, Fargen KM, et al. Establishing standard performance measures for adult stroke patients: a nationwide inpatient sample database study. *World Neurosurg* 2013;80:699–U299.
- Brown DL, Morgenstern LB, Majersik JJ, et al. Risk of fractures after stroke. *Cerebrovasc Dis* 2008;25:95–9.
- Lichtman JH, Jones SB, Leifheit-Limson EC, et al. 30-day mortality and readmission after hemorrhagic stroke among medicare beneficiaries in joint commission primary stroke center-certified and noncertified hospitals. *Stroke* 2011;42:3387–91.
- Lichtman JH, Leifheit-Limson EC, Jones SB, et al. Preventable readmissions within 30 days of ischemic stroke among Medicare beneficiaries. *Stroke* 2013;44:3429–35.
- Bravata DM, Ho SY, Meehan TP, et al. Readmission and death after hospitalization for acute ischemic stroke: 5-year follow-up in the medicare population. *Stroke* 2007;38:1899–904.
- Bail K, Berry H, Grealish L, et al. Potentially preventable complications of urinary tract infections, pressure areas, pneumonia, and delirium in hospitalised dementia patients: retrospective cohort study. *BMJ Open* 2013;3:e002770.
- Rao A, Barrow E, Vuik S, et al. Systematic review of hospital readmissions in stroke patients. *Stroke Res Treat* 2016;2016:1–11.
- Lichtman JH, Leifheit-Limson EC, Jones SB, et al. Predictors of hospital readmission after stroke: a systematic review. *Stroke* 2010;41:2525–33.
- Lichtman JH, Leifheit-Limson EC, Jones SB, et al. Preventable readmissions within 30 days of ischemic stroke among Medicare beneficiaries. *Stroke* 2013;44:3429–35.
- Keyhani S, Myers LJ, Cheng E, et al. Effect of clinical and social risk factors on hospital profiling for stroke readmission: a cohort study. *Ann Intern Med* 2014;161:775–51.
- Carstairs V. Deprivation indices: their interpretation and use in relation to health. *J Epidemiol Community Health* 1995;49(Suppl 2):S3–S8.
- Bottle A, Aylin P. Comorbidity scores for administrative data benefited from adaptation to local coding and diagnostic practices. *J Clin Epidemiol* 2011;64:1426–33.
- Nagin DS. Analyzing developmental trajectories: a semiparametric, group-based approach. *Psychol Methods* 1999;4:139–57.
- Giles MF, Rothwell PM. Risk of stroke early after transient ischaemic attack: a systematic review and meta-analysis. *Lancet Neurol* 2007;6:1063–72.
- Pendlebury ST, Rothwell PM. Prevalence, incidence, and factors associated with pre-stroke and post-stroke dementia: a systematic review and meta-analysis. *Lancet Neurol* 2009;8:1006–18.
- Myint PK, Staufenberg EF, Sabanathan K. Post-stroke seizure and post-stroke epilepsy. *Postgrad Med J* 2006;82:568–72.
- Martino R, Foley N, Bhogal S, et al. Dysphagia after stroke: incidence, diagnosis, and pulmonary complications. *Stroke* 2005;36:2756–63.
- Poisson SN, Johnston SC, Josephson SA. Urinary tract infections complicating stroke: mechanisms, consequences, and possible solutions. *Stroke* 2010;41:e180e184.
- Lewsey J, Jhund PS, Gillies M, et al. Temporal trends in hospitalisation for stroke recurrence following incident hospitalisation for stroke in Scotland. *BMC Med* 2010;8:1.
- Whyte EM, Mulsant BH. Post stroke depression: epidemiology, pathophysiology, and biological treatment. *Biol Psychiatry* 2002;52:253–64.
- Creditor MC. Hazards of hospitalization of the elderly. *Ann Intern Med* 1993;118:219–23.
- O'Malley KJ, Cook KF, Price MD, et al. Measuring diagnoses: ICD code accuracy. *Health Serv Res* 2005;40:1620–39.