A Clustering-Based Patient Grouper for Burn Care

DOI:
10.1007/978-3-030-33617-2_14

A Clustering-Based Patient Grouper for Burn Care

Chimdimma Noelyn Onah[1], Richard Allmendinger[1], Julia Handl[1], Paraskevas Yiapanis[2], Ken W. Dunn[3]

[1] University of Manchester
[2] Medical Data Solutions and Services
[3] University Hospital South Manchester

Abstract. Patient casemix is a system of defining groups of patients. For reimbursement purposes, these groups should be clinically meaningful and share similar resource usage during their hospital stay. In the UK National Health Service (NHS) these groups are known as health resource groups (HRGs), and are predominantly derived based on expert advice and checked for homogeneity afterwards, typically using length of stay (LOS) to assess similarity in resource consumption. LOS does not fully capture the actual resource usage of patients, and assurances on the accuracy of HRG as a basis of payment rate derivation are therefore difficult to give. Also, with complex patient groups such as those encountered in burn care, expert advice will often reflect average patients only, therefore not capturing the complexity and severity of many patients' injury profile. The data-driven development of a grouper may support the identification of features and segments that more accurately account for patient complexity and resource use. In this paper, we describe the development of such a grouper using established techniques for dimensionality reduction and cluster analysis. We argue that a data-driven approach minimises bias in feature selection. Using a registry of patients from 23 burn services in England and Wales, we demonstrate a reduction of within cluster cost-variation in the identified groups, when compared to the original casemix.

Keywords: Patient Casemix, Clustering, Data Driven.

1 Motivation

The NHS serves a wide population with varied demographic and medical histories, with the aim of providing health interventions to the population who need them. The provision and maintenance of these interventions is constrained by scarce resources and cost containment [1]. The pressure from binding budget constraints, and thus the need to control costs, has induced a shift in favor of prospective payments over retrospective payment systems.

The use of patient-level payment system transfers all cost burden to the payer, since the reimbursement is based on the real costs. In the context of such a system, even profit maximizing providers may be insufficiently motivated to decrease costs. In contrast, prospective payment systems (PPSs) determine the provider's payment rates ex ante without any link to the real costs of the individual provider [2]. This payment system is increasingly being adopted over retrospective systems, as it encourages cost containment and a shared burden with the providers. There is wide adoption of PPS globally, with approximately 70% of all OECD countries and more than 25 low-and middle-income countries having adopted some sort of casemix system for reimbursement purposes [3, 4].

Here, a casemix is a system of defining cohorts of related patients, which comprise cases that are homogenous by resource consumption pattern and at the same time, clinically similar. In the NHS, the National Casemix Office (NCO) is commissioned to develop and maintain a set of casemix groupings, called HRG (health resource group). This is a type of PPS where payment rate is determined as the average patient cost in each HRG. HRGs are generated using nationally mandated patient-level data, which primarily includes age, complications and comorbidities, diagnosis and procedures. Adopted in acute care, the groups are generated by transcribing expert advice into if-else rules, with the aim of capturing differing patient severity and length of stay (LOS).

Any reimbursement methodology based on generalizations across patient groups (i.e. determining payment rate as an average of cost in each HRG) will have weaknesses regarding its ability to fairly work across a variety of settings and HRGs are no exception to this. The use of LOS as an (imperfect) indicator of resource use contributes further to this weakness – it is known to be unreliable particularly for the case of surgical patients [5]. Finally, the identification of relevant factors based on expert advice alone carries the risk of ignoring other unknown (or less well established) factors that may account for the case complexity of certain patient sub-groups.

Our core hypothesis is that in-depth analysis of the available data should be used in conjunction with expert input to develop an evidence-based model that comprehensively captures the complexity of care provided by such services, and accurately classifies patients into homogeneous groups with respect to costs and patient characteristics. This dual approach was previously not possible due to a lack of availability of extensive patient-level cost data, and the resulting primary dependence on expert advice.

Our research aims to provide evidence for this hypothesis. First, we explore the accuracy of current HRGs in terms of actual resource usage. Second, we describe an analytical approach to the development of an alternative, data-driven grouper. Throughout our analysis, we use burn care as a base case. Burn services are selected as an example of a specialized service, which deals with rare and complex conditions and by necessity operates at high expenditure. Burn services are to be open regardless of the number of patients admitted, with a minimum number of staff, and they rely on the use of highly specialist equipment and interventions. We expect that the complex characteristics of this setting make them particularly sensitive to the impact of weaknesses in the current HRG classification.

The remainder of this paper is structured as follows. The next section introduces the data sets used to explore HRGs and generate the data-driven groups. We then introduce the analysis pipeline adopted, which includes data pre-processing, dimensionality reduction and the deployment of clustering approaches in two separate steps. In

Section 3, we discuss the results, using visualizations and within cluster variation of costs to identify improvements. The final section includes a conclusion and discussion of future work.

2 Methodology

2.1 Data

This study uses comprehensive anonymized patient-level data that is nationally mandated for all burn units in England and Wales. The data covers a time period from 2003 to 2019 and captures 164 features for just over 100,000 patients. This includes features such as demographic characteristics (age, gender), burn characteristics (depth, total burn surface area, burn site, locality, type, source, category and injury group), pre-existing conditions (self-harm, alcohol usage, asthma, clotting disorder etc.), time from injury to admission, patient-level cost, LOS and index of multiple deprivation (IMD).

To highlight current variation in HRGs and as a benchmark for model performance, we use the 2017/18 average patient-level cost by HRG open data released by NHS Improvement. This is limited to one year as PLICS adoption was introduced just in 2017/18 data collection cycle. This data is at the burn service level and so represents average patient level cost in each service.

2.2 Analysis Pipeline

Step 1: Selecting relevant features and cases. To ensure the use of quality features that reflect the clinical and cost differences of patients, the features selected for clustering were those identified as statistically significant in predicting patient-level cost and patient outcome. Cost prediction accuracy was improved with the removal of non-survivals, which LOS and cost less compared to survivals with similar burn characteristics. Thus, is in line with the current grouper, the following analysis focuses on survival cases only. All cases with missing data were deleted, leaving just over 80,000 cases and 24 features after feature selection. Table 1 lists these features.

Table 1. Selected Features

Feature type (count)	Feature
Demographic (3)	Gender, Age, Index of Multiple Deprivation (IMD)
Burn characteristics (17)	Total burn surface area (TBSA); Presence of inhalation; Site of burn (leg; upper limb (UL); torso and thorax; face, hands, feet and perineum (FHPP); head and hand (HH); face, hands and feet (FHF)); Type of injury (contact, cold, flame, electrical, scald, chemical, friction, flash, radiation)
Comorbidity (2)	Number existing disorders, significance of existing disorder
Cost Features (2)	Adjusted LOS, Patient-level cost

We implement further dimensionality reduction to minimise noise, data complexity and reduce redundancy. Dimensionality reduction also helps reduce processing time and mitigates against the curse of dimensionality [6]. Linear discriminant analysis (LDA), a supervised approach to dimensionality reduction, is adopted. Here, this method is preferred over unsupervised dimension reduction models such as principal component analysis (PCA), as we wish to identify components that maximise cost separation rather than percent of variance alone.

Step 2: Deriving target feature for LDA. We derive a set of target classes for the LDA using a cluster analysis on multiple cost features, to reduce sensitivity to a single cost measure. This is achieved by using cost features: patient level cost and adjusted LOS as the target space. The target feature is then generated using k-means clustering algorithm (k = 38, same as number of HRGs) to partition the two-dimensional target space defined by adjusted LOS and patient-level cost.

Step 3: Segmentation by age. The current grouper splits the data into young patients (<16 years old) and older patients (>=16 years old). This reflects the burn care pathway, designed to treat pediatrics separately from adults as young age is identified as a significant complicator. The 2001 National Burn Care Review Report [8] highlights the unpredictable complication of seemingly simple burn injuries especially for pediatric patients. It argues and mandates the need for separate burn units for children and adults, due to the peculiar needs of children such as play specialist, teachers, family counselors and intensive psychosocial support. In line with the current grouper, we therefore further split the data by age group.

Step 4: Dimensionality reduction using LDA. The comorbidity details, demographic and burn characteristics listed in Table 1 are used as the input features for the LDA. We retain the first two LDA components. Therefore, the output of this analysis is a projection from the original feature space into a two-dimensional manifold spanned by orthogonal components that maximise separation by the target feature constructed in Step 2. This is done on each segment derived in Step 3.

Step 5: Segmentation into homogeneous patient groups. With these pre-processing and dimensionality reduction steps completed, an unsupervised clustering method is deployed to derive homogenous patient groups. This paper uses an unsupervised clustering method, as we assume that the true class of patients are unknown. The use of a supervised method, for example, using cost labels may create groups that are homogenous in terms of cost only. This therefore does not meet the clinical relevance criteria.

In particular, we deploy an agglomerative hierarchical clustering (HAC) algorithm using the LDA components generated on each age segments (<16 years old and >=16 years old) as input data to generate 13 and 25 patient groups respectively. The group numbers reflect the number of segments generated by the current grouper, to facilitate comparison.

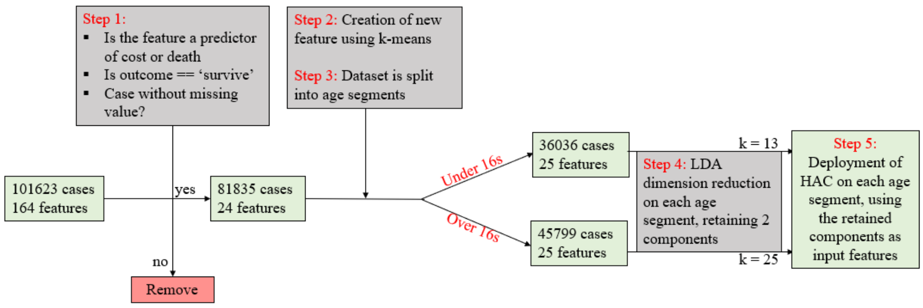

Fig. 1. Summary of analysis pipeline

We use the Euclidean distance as the metric used to compute the distance between data points. Average link is used as the linkage criterion, i.e. the distance between clusters is calculated as the average of the pairwise distances between all observations in the two clusters. At each step in the hierarchy, the algorithm merges the pair of clusters that minimises this criterion. Average link HAC clustering and all preprocessing steps were implemented using the relevant scikit-learn module on python, with other parameters left as default [9]. Figure 1 depicts a summary of the analytical pipeline.

3 Results and Analysis

We explore the patient-level cost by HRG, as generated by the National Casemix office. In Figure 2, we visualize this data using boxplots that show the distribution of patient-level cost broken up by HRG. The wider the boxplot, the more variable are the costs within that group. There are 25 adult HRGs, with JB40A being the most complex and JB71A the least complex in terms of injury severity. Whereas, the child segment includes 13 HRGs, with JB50A as the most complex and JB71B the least complex in terms of injury severity. These HRGs shown in Figure 2 are split into the two age segments and ordered in decreasing order of injury complexity.

Overall, we would expect the boxplots to slowly go down in terms of patient-level cost (for adults and children) as the injury complexity reduces. While this trend is generally apparent, there is scope to achieve a better match between injury complexity and costs for both adult and children. The results show low within-group homogeneity and we also observe a low degree of heterogeneity between groups. For a model that accurately reflects resource usage and complexity, we would expect these two aspects to be more pronounced. For example, JB40B, a more complex group than JB41A should have the highest average cost amongst the two, but we observe that this is not the case.

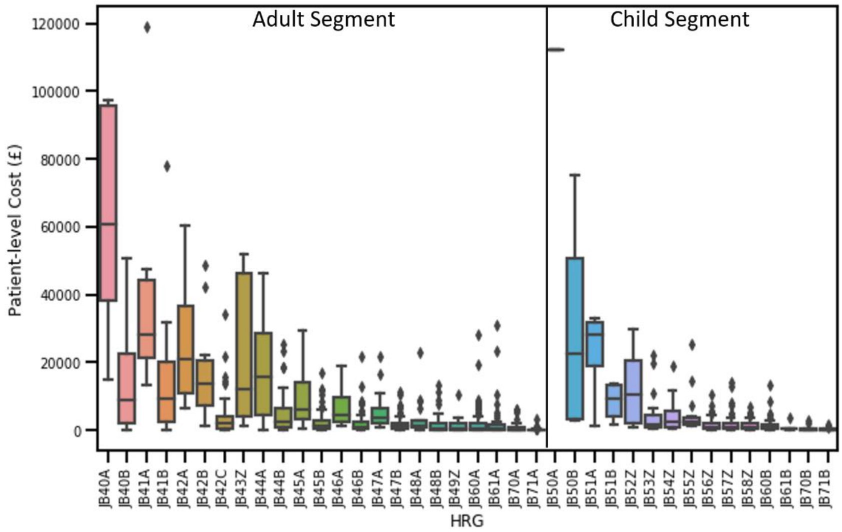

Fig. 2. 2017/18 HRG by patient-level cost. Ordered in decreasing order of injury complexity

Thus, this exploratory analysis appears to support the need for a grouper with a better performance at minimizing within-group cost variance and maximising between-group separation. Due to limitations of the dataset used, we do not have access to HRG grouping at a patient level and are therefore not able to link back costs to individual patient characteristics.

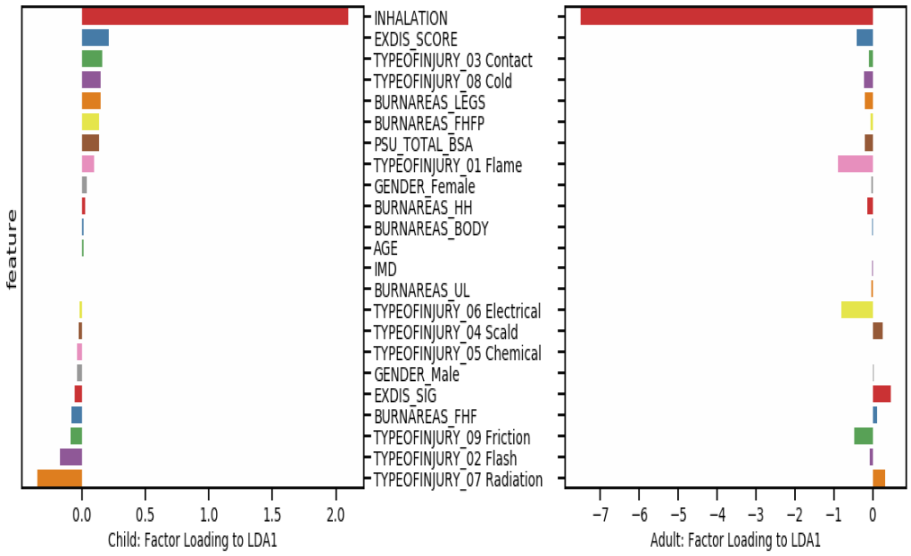

Fig. 3. Factor loadings on 1st linear discriminant: Child vs Adult Segment

One of the key steps in the analysis pipeline was the split of data on age group reflecting the burn care pathway. The importance of this approach is highlighted in Table 2 and Figure 3. Table 2 shows that child groups, although with lower adjusted LOS, have higher average costs relative to comparable adult groups. In Figure 3, feature contribution to separation in data captured by LDA across both age groups are different. Where presence of inhalation was the most important in both groups, burns caused by radiation have the second highest factor loading in the child segment whereas injury caused by flame is second highest factor in the adult segment. It is clear that the feature contributions in the two groups differ and this supports the need to split data by age group.

The suitability of LDA in class separation was evaluated by deploying a train test model, to enable the calculation of an accuracy score. For the adult and child segments, we get a score of 73% and 72% respectively. This highlights the ability of non-cost features in capturing resource usage thus the suitability of the LDA components generated.

The resultant groups derived using our data-driven pipeline, are shown in Figure 4. In particular, the model identifies well-defined high cost clusters in both adult (Adult0, Adult8, Adult13 and Adult17) and child (Child8, Child3) groups. Investigating further, we compare the within cluster variance (calculated using the coefficient of variation of patient-level cost in each group) between HAC generated groups and HRGs. Figure 5 highlights the improvement obtained from the data-driven model: as evident from the histogram, the HAC grouper tends to produce groups with a higher cost similarity. Specifically, the HAC groups (green) on average have a lower within cluster variation (0.89) compared to that of the HRGs at 1.19. The overall results therefore suggest an improved differentiation between groups.

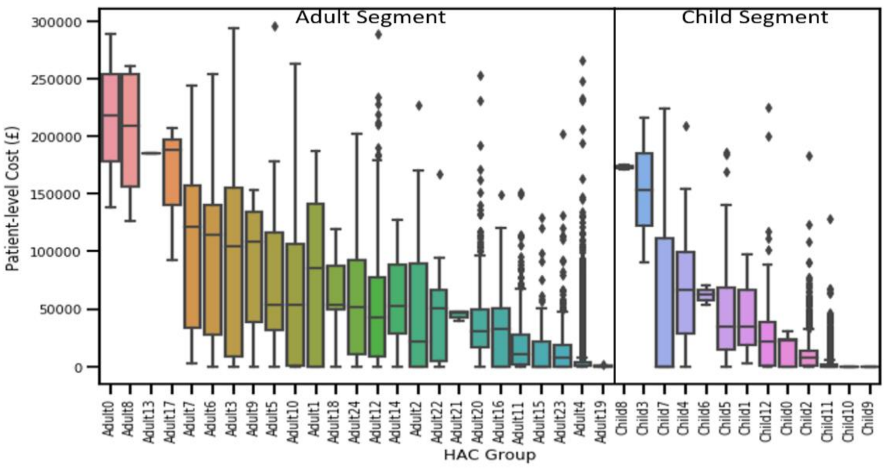

Fig. 4. Identified groups by patient-level cost ordered by decreasing average cost

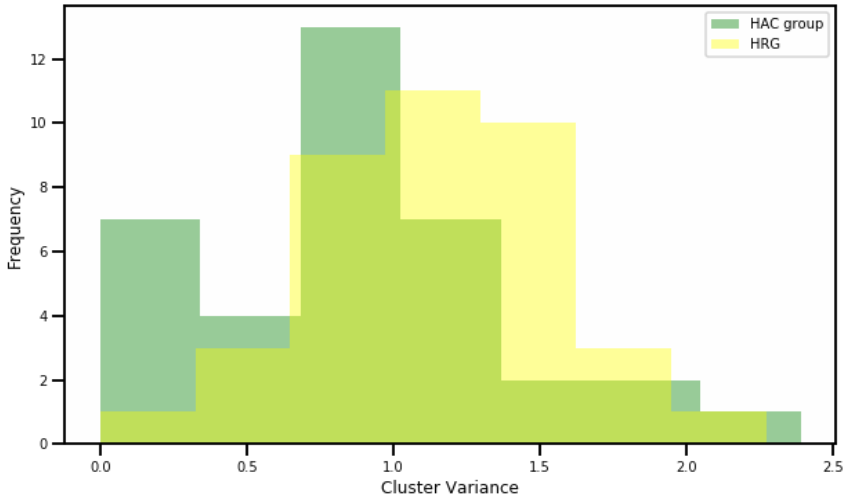

Fig. 5. Within cluster variation of patient-level cost in HRG vs HAC

We further explore the patient characteristics of the HAC groups. Table 2 provides some illustrative examples that highlight the interpretability of some of the resulting clusters, based on patient characteristics. For example, when comparing the clusters Adult3 and Adult12, these have very similar average age, but Adult3 has the more severe burns (TBSA), higher LOS and cost, and so the necessity to have separate groups. In the child group, comparing clusters Child2 and Child11, we see a similar pattern, i.e. similarity with respect to average age but differentiation by severity of burn, LOS and cost.

Table 2. A sample* of HAC groups by patient characteristics (average).

Cluster Label	Age	TBSA	Adjusted LOS	Existing Disorder (Significance)	Patient-level cost (£)
Adult3	41.02	50.59	13.71	0.27	99184.74
Adult12	42.99	31.30	9.08	0.49	56979.04
Adult20	46.61	19.49	9.28	0.79	39006.15
Adult11	67.64	5.28	7.36	2.68	19037.41
Child5	5.13	34.86	5.77	0.10	45585.72
Child12	4.85	25.10	5.08	0.07	32656.89
Child2	3.87	10.56	2.63	0.08	10595.17
Child11	3.87	2.56	0.88	0.03	1921.42

*Purposive sampling

Meanwhile comparing Adult20 and Adult11, there have different patient characteristics and thus different average patient-level cost. Child5 and Child10 though with similar adjusted LOS, Child5 has a higher TBSA, higher score with respect to the severity of existing disorders and thus a higher average patient-level cost. The observed similarity in adjusted LOS but higher difference in patient-level cost in this child5 and Child10 also supports the need for using more than one feature space in deriving the LDA target feature. These results highlight the effectiveness of the data-driven HAC grouper in generating groups with homogenous patient characteristics.

4 Conclusion and Future Work

HRGs are used as the basis for cost reimbursement in health systems. When designing HRGs, the accurate reflection of resource usage is imperative in ensuring different settings get fair reimbursements. The collection of patient-level cost, at a national scale, has created the possibility of generating improved data-driven groups. We have been able to highlight that improvements can be made in identifying patient case mix suitable for payment rate derivation. This was done with the implementation of a data driven process on burn care data to identify significant features and the subsequent use of a clustering algorithm. With the adopted analysis pipeline, we showed a visible difference in the distribution of patient-level cost variance when HAC groups are compared to HRGs.

There could be further reduction in within cluster variance with the use of state-of-the-art clustering algorithms that simultaneously consider Step 2, 3 and 4 of our analysis. A further limitation of this paper is the lack of certain features in the patient-level data such as current patient level HRG labels, in-hospital complications and interventions. The unavailability of patient-level data labelled with current HRGs, limited our ability to evaluate performance of current HRGs in terms of patient characteristics. Finally, the lack of complications and interventions data could limit the homogeneity by patient characteristics in the HAC identified groups.

Future work will be aimed at exploring changes to our analytical model, including the consideration of different approaches to dimensionality reduction and cluster analysis, as well as the inclusion of expert opinion in feature selection and group validation. This will be done using comprehensive anonymized patient-level data set. An improved analytical pipeline developed and validated using burn care data can be applied to other injury type. The adoption in other injury type would include the use of expert and data-driven feature selection phase, but with the aim of a standardised pipeline, the same dimensionality reduction and cluster analysis technique would be implemented.

References

1. M. Hopfe, G. Stucki, R. Marshall, C. D. Twomey, T. B. Üstün, and B. Prodinger, “Capturing patients’ needs in casemix: a systematic literature review on the value of adding functioning information in reimbursement systems,” BMC Health Serv. Res., vol. 16, no. 1, p. 40, Dec. 2015.

2. M. Jegers, K. Kesteloot, D. De Graeve, and W. Gilles, “A typology for provider payment systems in health care,” Health Policy (New. York)., vol. 60, no. 3, pp. 255–273, Jun. 2002.
3. I. Mathauer and F. Wittenbecher, “Hospital payment systems based on diagnosis-related groups: experiences in low- and middle-income countries,” Bull. World Health Organ., vol. 91, no. 10, pp. 746-756A, Oct. 2013.
4. W. Fischer, “The DRG Family: State of affairs,” 2007.
5. A. Street, C. Kobel, T. Renaud, and J. Thuilliez, “How Well Do Diagnosis- Related Groups Explain Variations in Costs or Length Of Stay Among Patients And Across Hospitals? Methods for Analysing Routine Patient Data,” Health Econ., vol. 21, pp. 6–18, Aug. 2012.
6. R. W. Sembiring, J. M. Zain, and A. Embong, “Dimension Reduction of Health Data Clustering,” Int. J. New Comput. Archit. their Appl., vol. 1, no. 3, pp. 1041–1050.
7. A. Chris, Python Machine learning Cookbook: Practical Solutions from preprocessing to deep learning. 2018.
8. National Burn Care Review Committee, “National Burn Care Review Committee Report Standards and Strategy for Burn Care A Review of Burn Care in The British Isles Contents,” 2001.
9. F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” J. Mach. Learn. Res., vol. 12, no. Oct, pp. 2825–2830, 2011.