# Patient segmentation: Identifying clinically relevant groups of similar cost

Chimdimma Noelyn Onah[1]

Supervisors: Richard Allmendinger[2], Julia Handl[2], Paraskevas Yiapanis[3], Ken Dunn[4]

[1]School of Social Sciences, University of Manchester, [2]Alliance Manchester Business School, University of Manchester, [3]School of Computing, Maths & Digital Tech Manchester Metropolitan University, [4]Manchester Burn Service, National Health Service

## 1. Introduction

The National Health Service (NHS) serves a wide population with varied demography and medical history. The accelerating medical advances have led to greater scope to treat chronic illness. Consequently, the hopes and expectations of patients have also risen, which, in turn, can burden the NHS with rising costs. Rising costs coupled with funding constraints have fuelled the popularity and adoption of payment reforms that promote the "payment by results" (PbR) pricing system.

The PbR system pays providers a standard national tariff for each patient seen or treated. This tariff is calculated for each clinically meaningful group, i.e. groups of patients that are expected to consume similar level of resources based on shared characteristics. These patient groups are also known as Health Resource Groups (HRGs). Expert interviews have shown that current HRGs do not fully meet the requirement of similar costs. This is particularly problematic for specialist services which are complex, have low patient volume and operate at high expenditure levels. This research project uses the example of burn services to investigate this further.

## 2. Motivation

The profile of burn patients varies in complexity, this leads to substantial variation in the cost of treating individual cases.
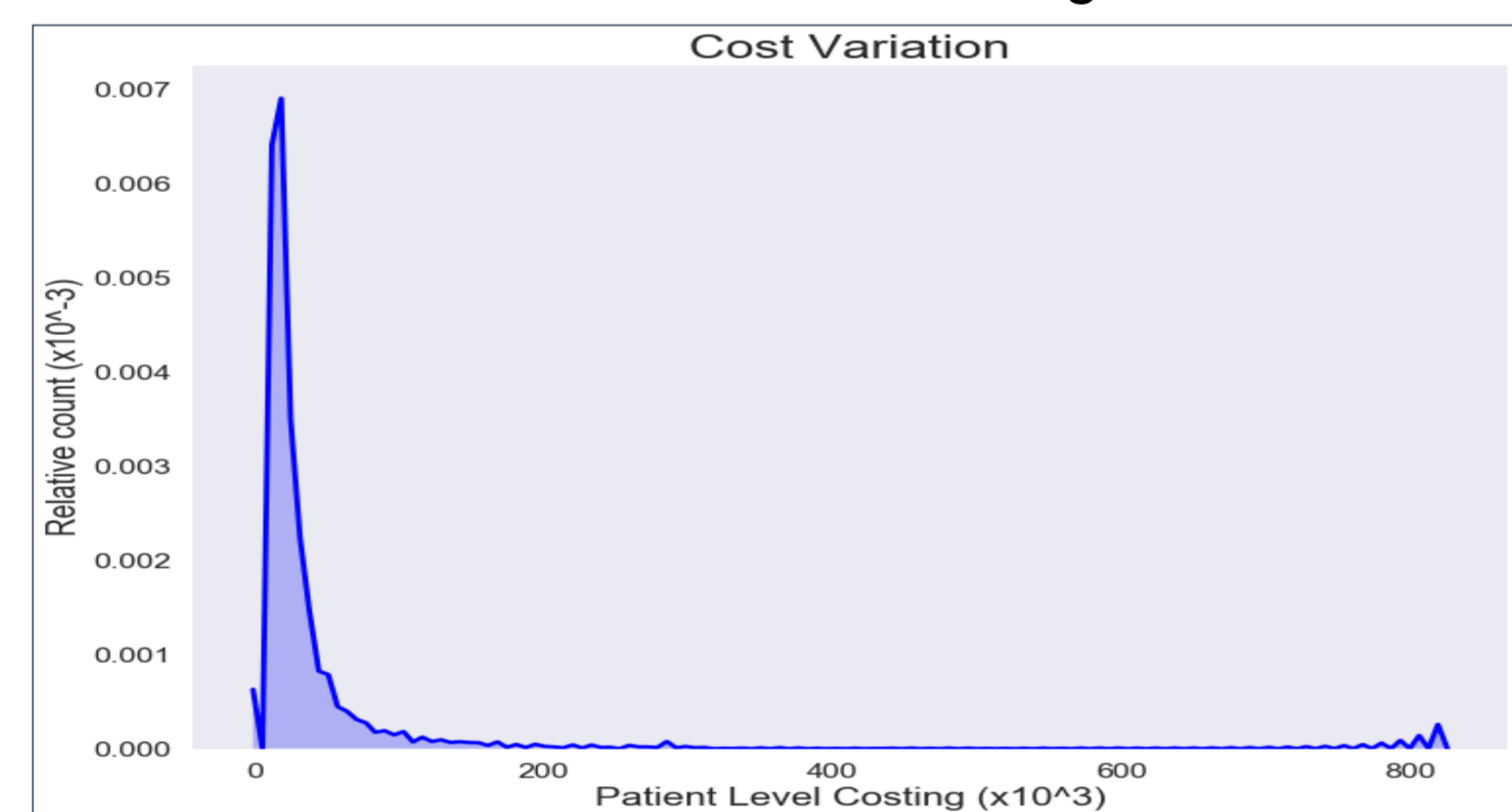


This variation in cost and complexity makes developing HRGs that are simple enough to use and also accurately reflect the costs of care for each group a tough order.
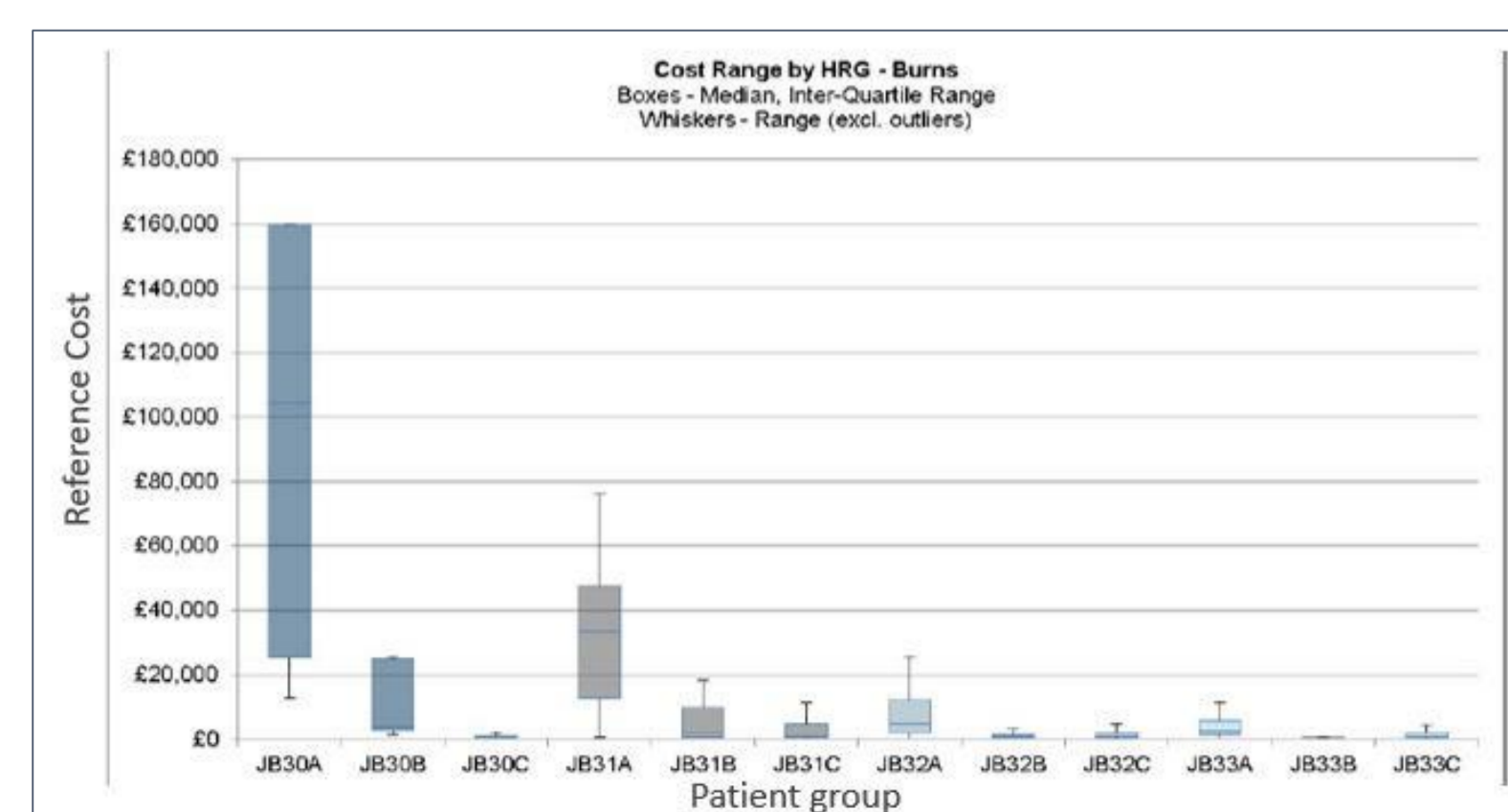


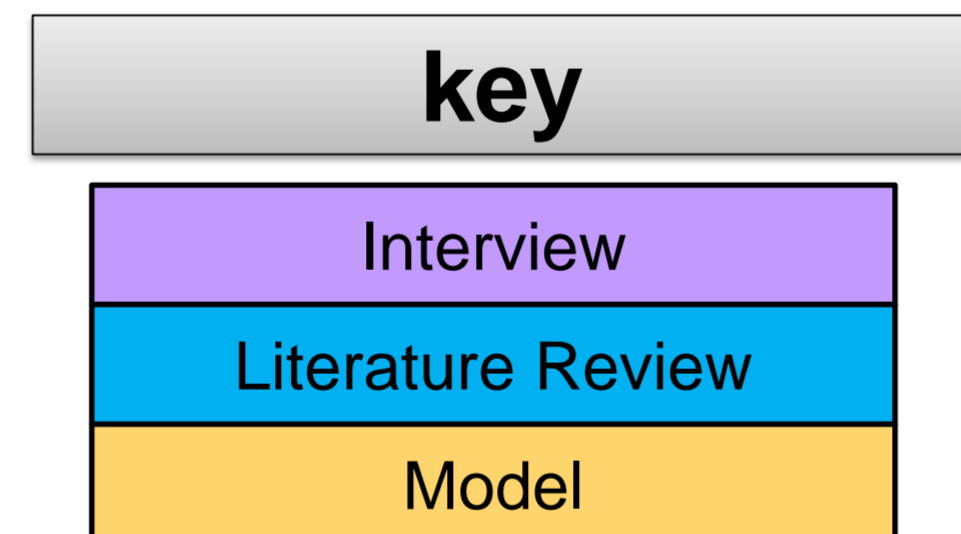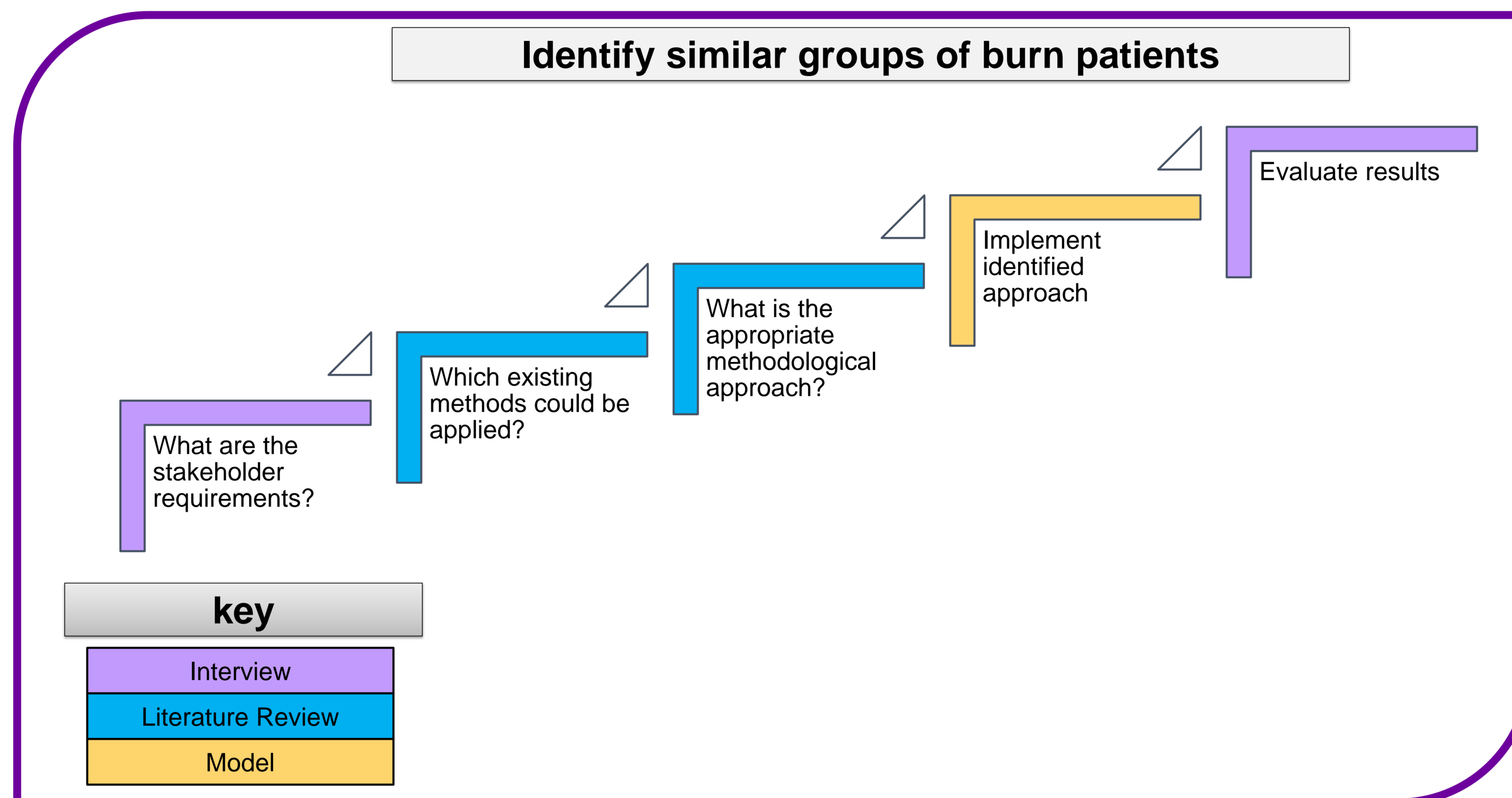Source: HRG4+ 2014/15 Reference Costs Appraisal of Subchapter JB Burns Procedures and Disorders
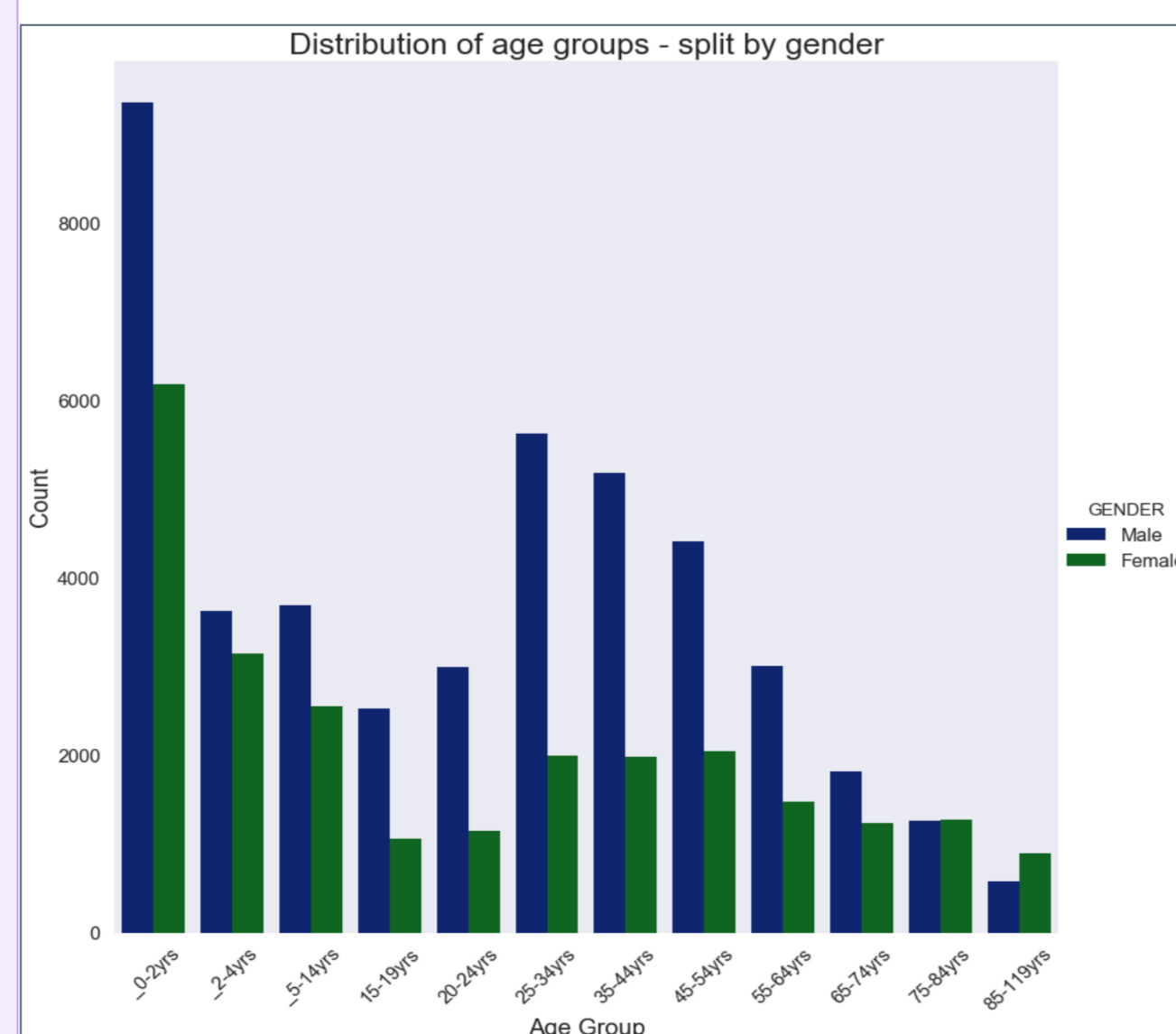
## 3. Research Question

This research aims to
❑ Identify homogenous groups of burn patients that are clinically relevant and are similar in cost.

A clinically relevant group of patients share the same patient characteristics (i.e. similar burn severity and demography).

To achieve this, the workflow depicted in the Figure below would be adopted. It would entail understanding the requirements, criteria of an appropriate group, identification of appropriate methods. The use of qualitative and quantitative data for modelling and evaluation of the results.



**Identify similar groups of burn patients**

- Evaluate results
- Implement identified approach
- What is the appropriate methodological approach?
- Which existing methods could be applied?
- What are the stakeholder requirements?

**key**
- Interview
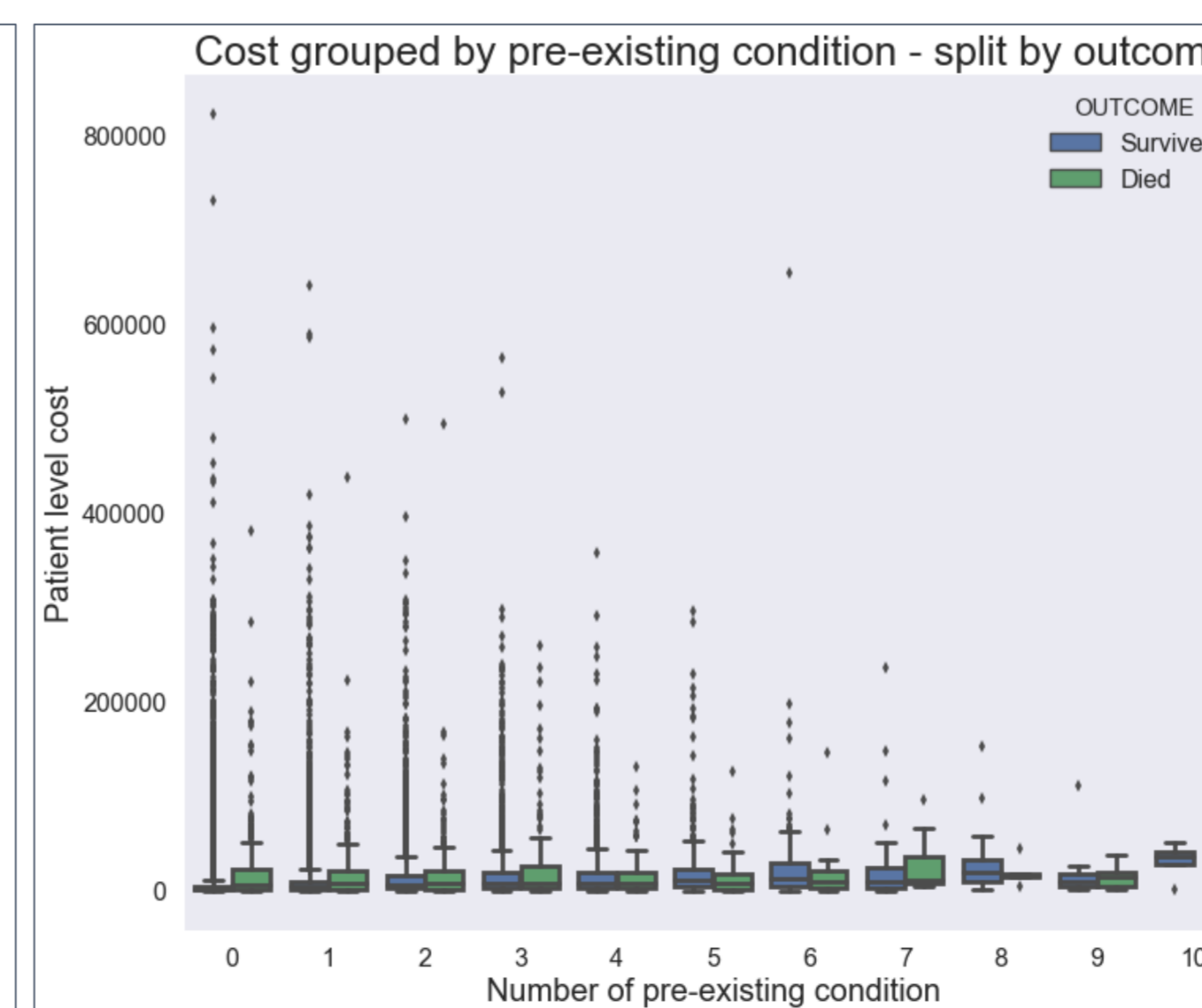- Literature Review
- Model

## 4. Patient Profile



a. The most likely patient is under 2 and male. Overall, there are more males than females except for over 75 year olds.
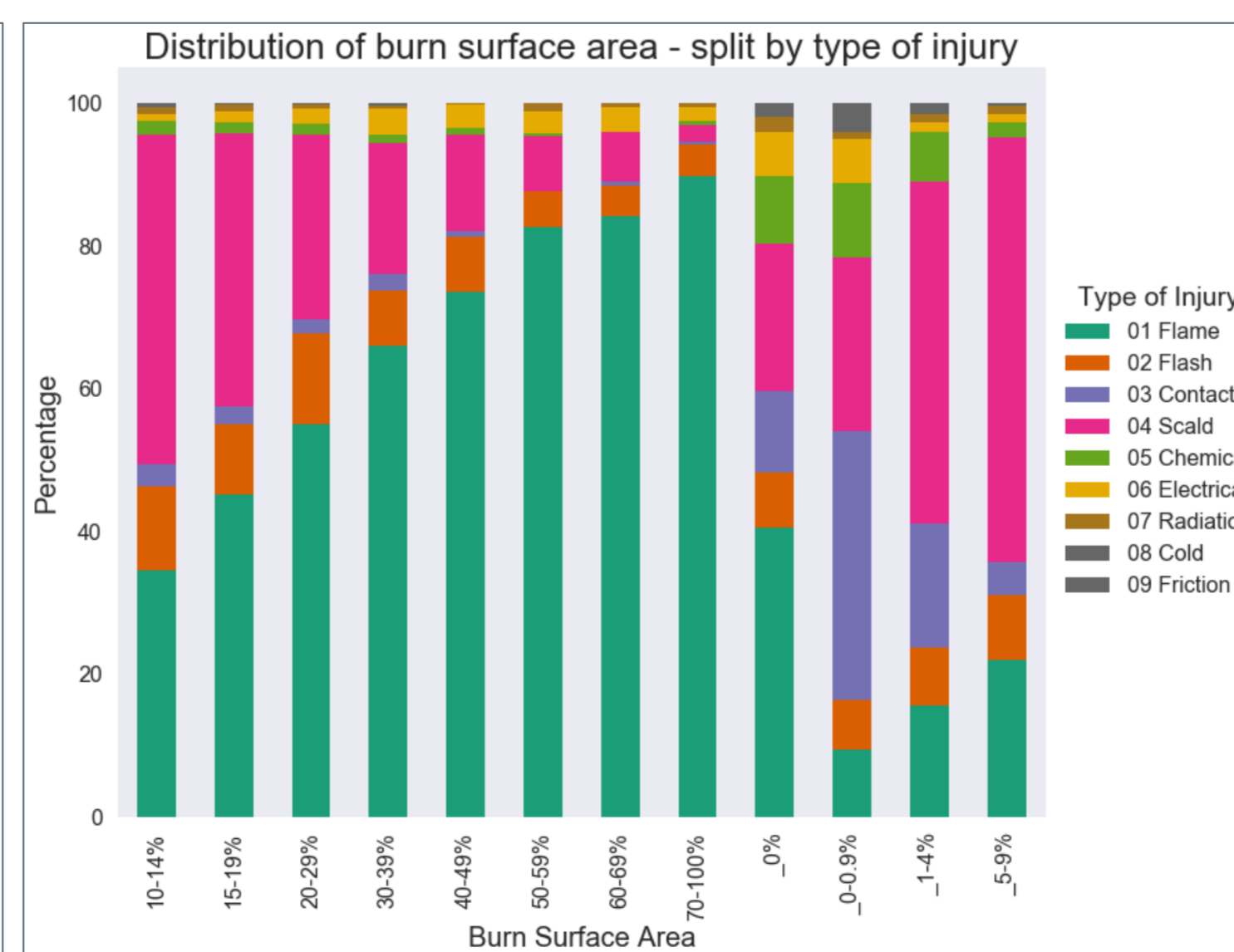
b. Most cases have a burn surface area between 10% and 14% and caused by a scald. The most severe cases are as a result of flame.

c. Cost do not increase with increase in number of pre-existing conditions. It also does not have a strong relationship with outcome.

## 5. Cluster Analysis

An attempt to group the burn patients using k-means was made. Data pre-processing was carried out to remove sparse variables, outlier cases and standardise the dataset.

Principal component analysis (PCA) and T-Distributed Neighbouring Entities (t-SNE) were applied to reduce the dimension of the standardised dataset before further analysis were performed. PCA performed better at differentiating the cases.
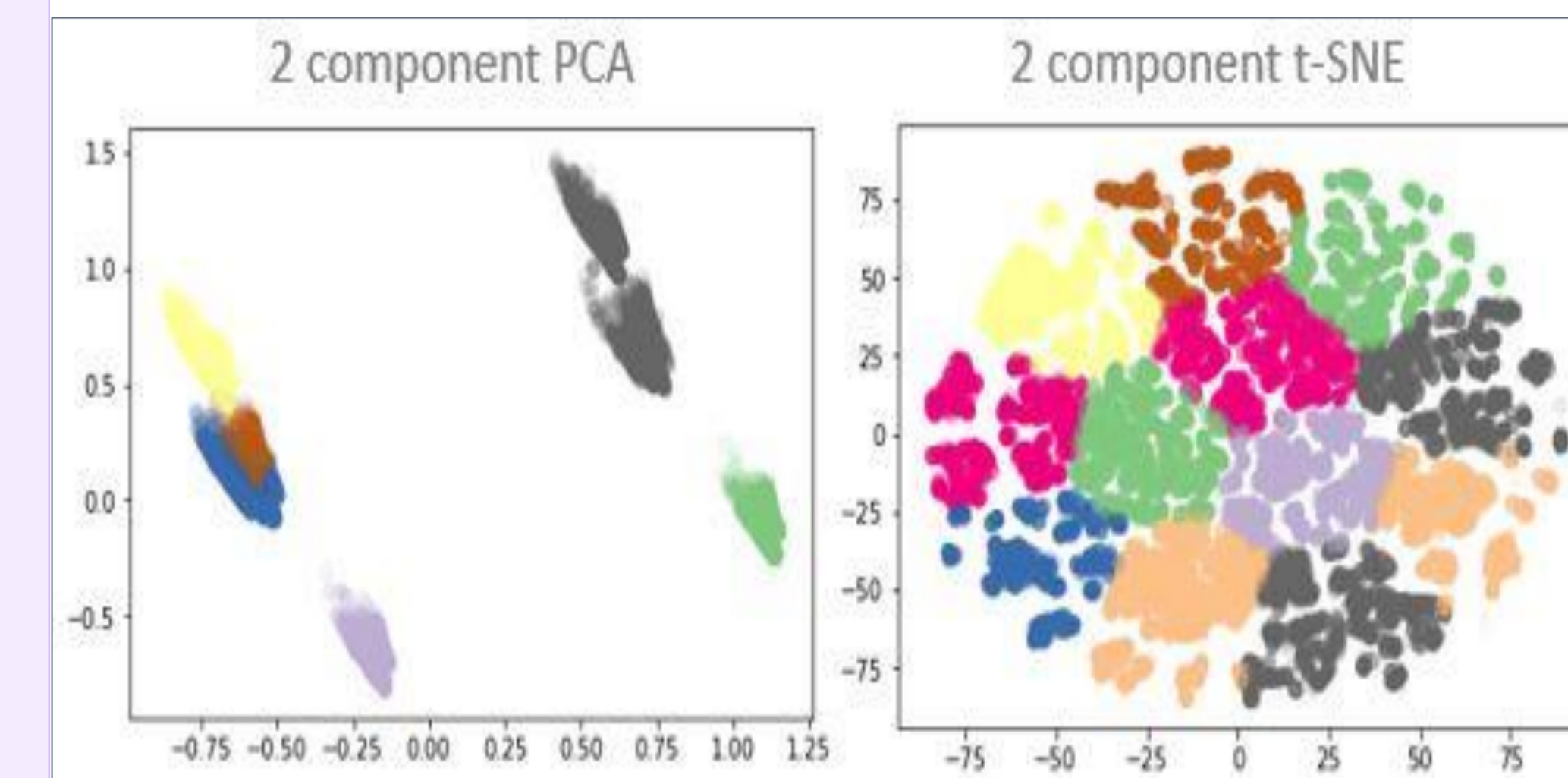


To improve results, clustering experiments were performed by using different subsets of the data. These models was built on the two principal components.
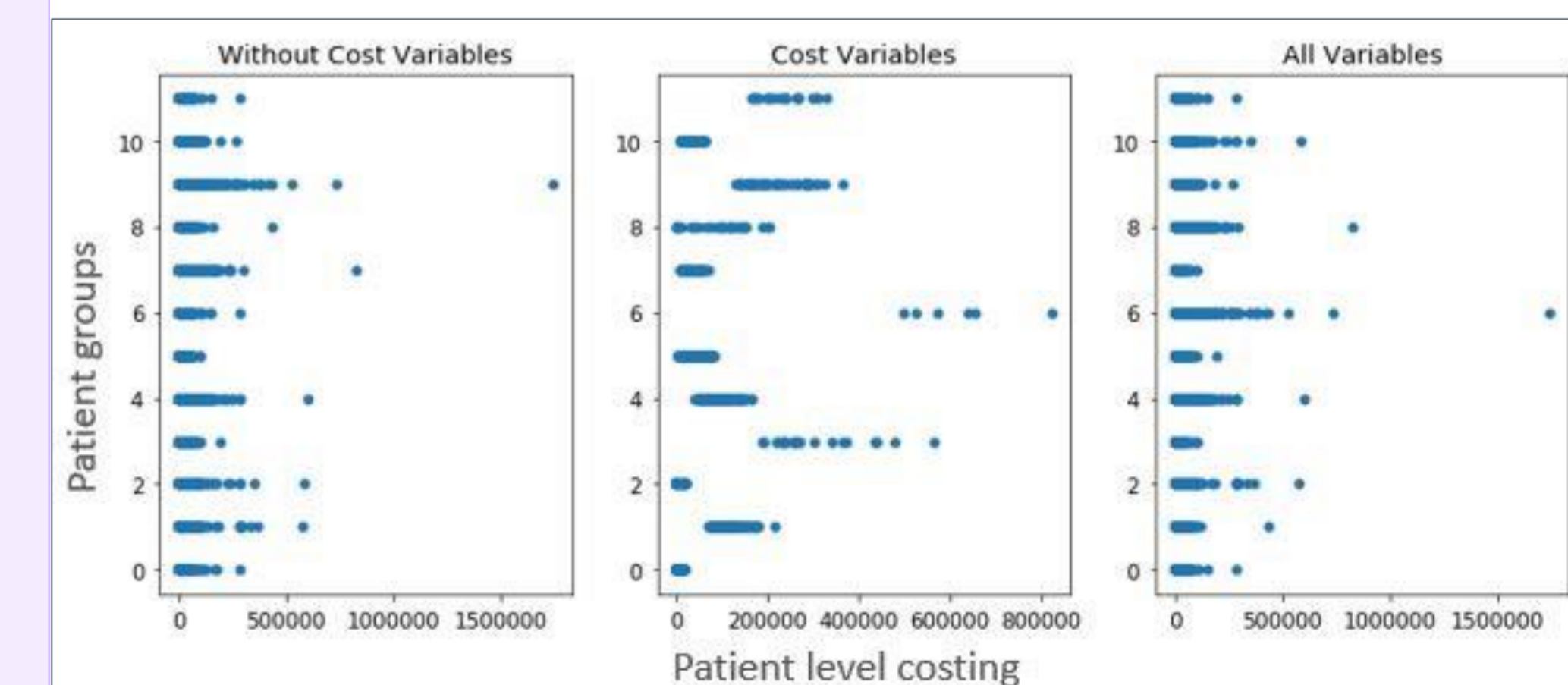


## 6. Future Work

The above figures illustrate that clustering by cost variables is able to achieve well-defined groups that exhibit clear homogeneity with respect to total cost. This suggests that better weighting schemes are needed to obtain clusters that are well defined with respect to both cost factors and additional patient-level characteristics.

Future work would be aimed at
1. Consulting with stakeholders on inclusion of further variables.
2. Using multi-objective clustering techniques to produce better results.

Email: chimdimma.onah@manchester.ac.uk
Twitter: @OnahNoelyn

MANCHESTER 1824
The University of Manchester

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Data Analytics & Society