Segmentation of Clinical Commissioning Groups in England based on the prescribing patterns of their constituent General Practices

Ransome Mpini[1], Philip Worral[1] Helen Dickens[2] & Mark Bass [2]

[1]University of Westminster, [2]GlaxoSmithKline plc

Project Background

This project aims to enable GlaxoSmithKline to develop an understanding of how drugs are prescribed within England's Clinical Commissioning Groups (CCGs). The project groups CCGs based on the prescribing patterns they exhibit. The first part of the project aimed to cluster GP practices based on the drugs they prescribed (18 clusters in total). For each CCG, the cluster each of its practices fell into was identified and could be calculated as a relative proportion. Following this, CCGs were then clustered.

Data and Methods

GP practices are put into cluster groups based on their prescribing patterns for respiratory ailments. The resulting practice clusters are then analysed at CCG level to better understand the distribution of practice clusters within the CCGs. This gave an indication of the cohesiveness of a CCG where a cohesive CCG would have evidence of fewer practice clusters and CCGs with a wider variety of practice clusters is evidence of a less cohesive CCG.

Openly available datasets were used and these included the GP Prescribing dataset and the Quality and Outcomes Framework (QOF) dataset. Variables of interest in the GP prescribing dataset included the British National Formulary, which uniquely identifies each prescribed drug and it's Net Ingredient Cost of each prescribed drug at a GP practice. Another variable of interest was the Practice List Size, which was obtained from the QOF dataset. The practice list size values were used to normalize the Net Ingredient Cost figures so as to make them comparable between practices of different sizes.

The drugs prescribed at each GP practice were treated as the variables to describe the particular practice. The original prescribing dataset contains just over 10300 unique drugs across the GP's, although a reduced dataset of 551 drugs (focusing only on the respiratory drugs) was used for final analysis

Key Findings

A cluster analysis using the 'Ward' clustering technique revealed that CCGs fall into 3 main clusters. The first is made up of smaller CCGs. The most common practice cluster in this group constitutes around 37% or below of the total practices in the CCG. The second cluster identified is made up of smaller CCGs but these tend to have higher ratios for the most common practice cluster. The third cluster identified is composed of larger CCGs where the most prevalent practice cluster groups have very small ratios and these tend to have the highest ratios for less prevalent practice cluster groups.

Complications were faced during the analytic phases of this project mainly caused by the sheer size of the GP Prescribing dataset that constituted just over 9383 GP practices (observations) described by just over 10,300 unique drugs prescribed at each GP practice which were treated as the variables describing the GP practices. A decision was made to focus on the respiratory chapter (BNF Chapter 3) reducing the total number of drugs (variables) to 551.

Value of the Research

This kind of work has not to the best of my knowledge been attempted and the result of this project will be a big contribution to GSK analytic process. It has the potential to enable them to better understand CCG cluster groups and also enable them to use this result as input to their core analytics process. Insight gained from this project will enable GSK to increase their competitive advantage through a better understanding of CCGs. GSK will be able to use evidence based, tailored messages when engaging CCGs thus maximizing returns on investment.
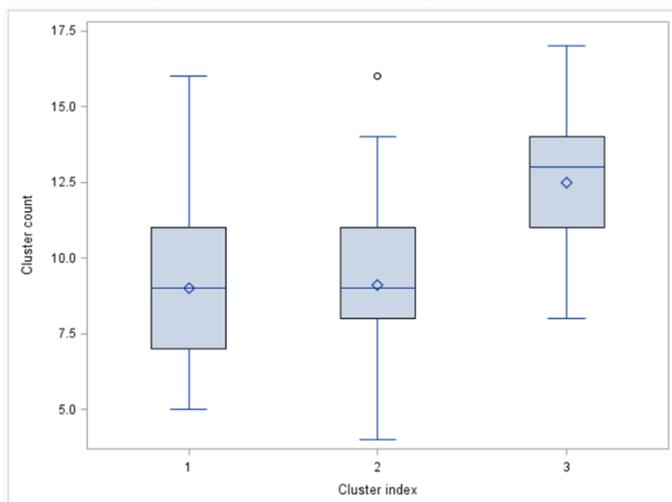
Practice cluster count distribution in CCG clusters
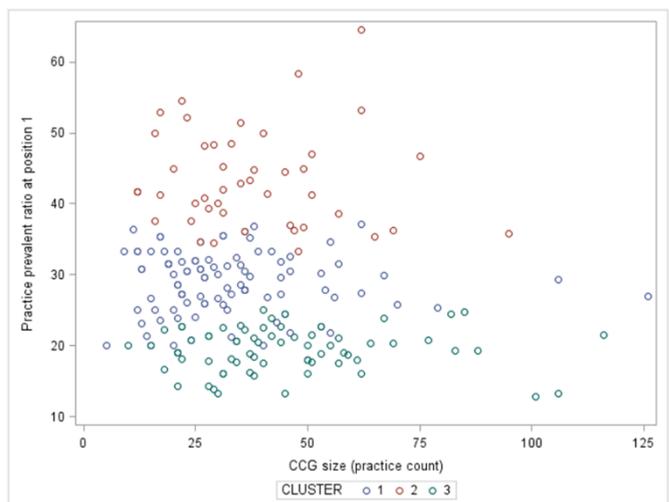
Plot of 1st positioned practice cluster ratio in CCG by CCG size