

Q-Step Internship: Royal College Data for General Medical Council

Ryan Khurana, BA Philosophy, Politics, and Economics



The General Medical Council Education and Standards team is responsible for ensuring that UK Medical Education produces the top quality

General Practitioners and Specialists to provide care for UK residents. As an intern, I was tasked with data analysis in compiling a database of academic progression for all doctors and specialists in the country.

Objectives

The aim of the project was to compile a comprehensive database of all exam data from the Royal Colleges throughout the country such that it could be accessed by third parties in a clear to use way.

Various exams were used in assessing which specialisms doctors were entering most. The scores and frequency of various specialist exams as well as exam data from medical schools across the country were all to be put into the UKMED database, a project to present a non-identifiable sampling of all those in the medical profession with data regarding race, gender, age, background, and test scores.

The project was commissioned by the GMC in 2010 and the process is ongoing as the legal sensitivity of the data required approval from the Royal Colleges and more data is being collected. 2017 is the expected publication date.

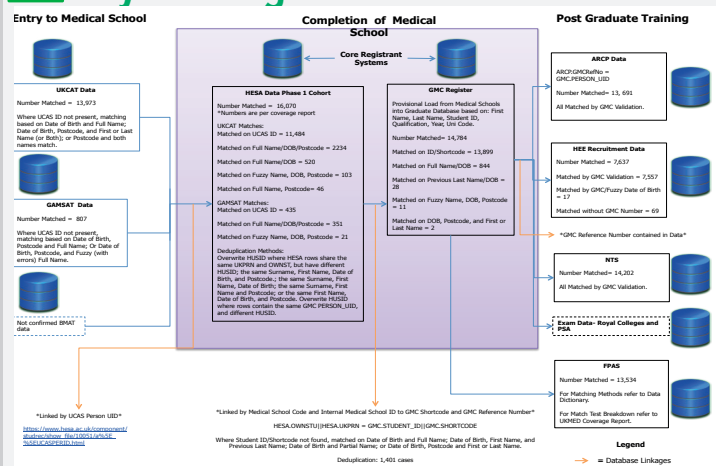
Method

I was involved in the modelling of the project which took place in several stages. The first involved cross referencing the data sharing agreements provided by the various Royal Colleges with the exam data we required, and with the data received. This was mapped in Excel such that revised data sharing agreements could be compiled if too much data was present, or requests could be sent out for the missing data if it was not enough to conduct analysis. Following this data mining was required in SQL in order to match the individuals from the data provided to Siebel, the GMC's internal database. This was to ensure consistency in the data received, to weed out duplicates, and to ensure the collection would involve the relevant timeframe.

SPSS multiple regression analysis was then conducted to model the exam data on the various demographic data we received.

Finally, a Tableau report was compiled which presented histograms of frequencies and test scores, as well as the regression outputs from SPSS. The entire process was model built, such that data adjustments would automatically update the entire product chain.

Key Findings



The project was one which involved building the models and algorithms that would allow analysis and output to be conducted in the future. As the database is not ready nor published, there are no conclusions currently to be made given the sample we received. Testing was done to ensure that given data sets previously analysed, such as the reliability of aptitude tests, that we received the same results. As this was the case we could have a reasonable degree of confidence that our models were an accurate representation of the data received, and that when the finalised collection is prepared for publication, the results that third parties received would be from a trustworthy source.

Conclusion

- The modelling development within SQL and SPSS allowed the Tableau report to live update to all new data inputs. When tested using the UK-CAT 12 Study to replicate their results, we received an identical output. This results in our modelling being robust and adaptable.

- It is clear that when the final database is compiled, that third party users will be able to have accurate reflections of trends within medical exam data in the UK, which can inform hospitals and other services on which variables to aid in more apt selection. This data will also provide the government with a method of improving recruitment to more struggling specialisms.