

# **ANALYSIS OF NHS DENTAL STATISTICS AND THE EFFECT OF DEPRIVATION, 2018 - 2019**

## **INTRODUCTION**

Over the twelve-month period, ending on 30 June 2019, NHS dentists received around 22 million adults and 7 million children (NHS Digital 2019). This report will analyse NHS dental data between 31 March 2018 and 31 March 2019, before exploring the relationships with the 2019 English Indices of Multiple Deprivation. The data used for this report was acquired from two sources, namely the NHS Digital Dental Statistics for the 2018/19 period (<https://digital.nhs.uk/data-and-information/publications/statistical/nhs-dental-statistics/2018-19-annual-report-pas>), and the United Kingdom (UK) Government's English Indices of Deprivation for 2019 (<https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019>).

This report aims to analyse the single data table formed from the collation of these sources to answer the three research questions below.

Collating data from these sources into a single data table will allow further analysis, which can be used to answer the three research questions which form the basis of this report.

1. Is there a relationship between the age and number of patients seen by a dentist over this period?
2. What is the effect of deprivation on this relationship?
3. Using 2019 data, which type of Clinical Commissioning Group (CCG) will face the greatest shortage of dentists in 2031?

## **DATA CHARACTERISATION**

This section will give a detailed description of the data to be analysed in this report. The data set has a size of about 1.1MB; it contains 14,516 records and ten variables. **Table 1** shows a sample of the data table.

*Table 1: A Record from the created Data Table*

| CCG Code | CCG ONS Code | CCG Name       | Patient Type | Age Band | Patients Seen | Population | Age Group | Dentist Count | IMD average score |
|----------|--------------|----------------|--------------|----------|---------------|------------|-----------|---------------|-------------------|
| 07R      | E38000027    | NHS Camden CCG | Child        | 5        | 1299          | 3178       | 35-44     | 48            | 20.131            |

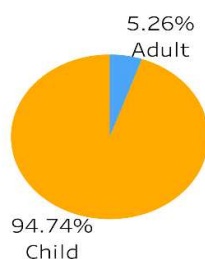
More information about each variable will be provided below.

**CCG NAME:** This is the name given to a Clinical Commissioning Group (CCG) in England. From NHS Digital's data dictionary, a CCG is a clinically led group that encompasses all the General Practice (GP) groups in their geographical area. They work with patients and healthcare professionals and in partnership with local communities and local authorities. In the data set, there are 191 distinct and uniformly distributed CCG names, which are all associated with variables, "CCG CODE" and "CCG ONS CODE" respectively. The high number of distinct values in the variable leads to it having an exceptionally high cardinality.

**CCG CODE:** The CCG code is a unique 3-character code that begins with two digits and ends with a letter. As with the "CCG Name" variable, the code also has 191 distinct values in our dataset. It is a unique code attached to each CCG name. It is used to reference the location.

**CCG ONS CODE:** The CCG Office of National Statistics (ONS) Codes are 9-character codes that are used to represent the geographical location of a clinical commissioning group. The ONS codes in the data set begin with "E38" and end with a 6-digit number. From the ONS website, the codes represent a sub-region in England. Along with the CCG name and code, the ONS code has 191 distinct values in the data table.

**PATIENT TYPE:** The values of the patient type variable are categorical data. There are two categories in this variable, and they are “Adult” and “Child”. **Figure 1** shows the distribution of both categories in the data set. Patient types are related to the “CCG NAME”, “CCG CODE” and, “CCG ONS CODE” variables.



*Figure 1: Distribution of Patient Type Categories*

**AGE BAND:** This is another form of categorical data in our data set; it divides the number of patients seen over 12 months into 19 distinct age groups, ranging from age 0 to age 18+. Each age band category in our data set appears 764 times. It should be emphasized that the ‘18+’ Age band and the ‘Adult’ patient types are the same value and are perfectly correlated to one another.

**PATIENTS SEEN:** This variable, of the numeric data type, provides information on the number of patients seen over a certain period by a CCG over 12 months (1 April 2018 – 31 March 2019) for children and over 24 months (1 April 2017 – 31 March 2019) for adults. It is linked heavily with the “CCG CODE”, “CCG NAME”, “AGE BAND”, and “PATIENT TYPE” variables respectively. Every adult patient seen recorded received dental care for 24 months, with each child receiving dental care for 12 months.

**POPULATION:** This numerical variable contains data on the estimated population of distinct age bands in each CCG over 12 months. The variable comprises values for the population of age bands under the control of CCGs. Each population estimate is acquired from the ONS.

**AGE GROUP:** A variable in the data set that contains four individual categories that contain information about the ages of dentists that work at each CCG. The categories range from “Under 35” to “55+”. Each category can be found 3,629 times in the data set. This variable is also highly correlated to the “DENTIST COUNT”, which will be explored next.

**DENTIST COUNT:** This variable relates with the “AGE GROUP” and tells us the number of dentists in each age group category in the 2018-2019 period.

**IMD AVERAGE SCORE:** The average Index for Multiple Deprivation (IMD) score is a measure of the degree of multiple deprivation within a geographical boundary, in this case, a CCG. The IMD is calculated using a weighted combination of seven domains: income, employment, education, health, crime, barriers to housing and services, and living environment. The IMD is then constructed using 39 different indicators, producing a generalised measure of deprivation. The larger the average score, the more deprived the CCG location is. Each value in this variable is numeric and connected to the variable “CCG ONS CODE”.

A sample of the dataset is shown in Table 2.

*Table 2: Summary statistics of numerical variables in the data set, rounded to 2 d.p.*

| VARIABLES         | SUMMARY       |               |          |         |                    |
|-------------------|---------------|---------------|----------|---------|--------------------|
|                   | Maximum value | Minimum value | Mean     | Median  | Standard deviation |
| PATIENTS SEEN     | 497950.00     | 5.00          | 7969.00  | 1768.00 | 30262.00           |
| POPULATION        | 962934.00     | 747.00        | 15326.00 | 3001.00 | 60114.00           |
| DENTIST COUNT     | 258.00        | 4.00          | 44.00    | 34.00   | 34.00              |
| IMD AVERAGE SCORE | 52.14         | 7.18          | 21.91    | 20.83   | 7.98               |

## DATA QUALITY

The quality of the data set is ensured through several data processing methods. The IMD uses a weighted combination of seven distinct areas. This allows for the wider effects of deprivation to be captured within society, and as such a more correct representation of an area's population. The reliability and robustness are ensured through the consistent and uniform application of a methodology, aided with 'shrink estimation'. Shrink estimation overcomes issues associated with small area data where small sample sizes result in unreliable estimates with extreme standard errors. This uncertainty is reduced by estimating scores for smaller areas using the mean score of a larger area alongside the score for the smaller area. The result is an improvement in the reliability of this data, and therefore, an improvement in the data quality. The remaining nine variables are all collected by the NHS Dental Services. Unlike the IMD, these nine variables are not calculated indexes, resulting in different data quality methods to ensure accuracy and reliability. The Patients Seen data set is published by the NHS as final data within both the quarterly and annual reports. Data is collected from the FP17 form and is dated at the point of process. Duplication of records is reduced by omitting patients with the same surname, initials, sex, and date of birth. Whilst individuals can still be included with a change of name or practice, this issue is only likely to impact the overall count data by 1-2%.

The process of data quality can be split into two areas: completeness and correctness. Within each, there are subsections ensuring a rigorous analysis of data quality. Overall, this data set appears to be largely complete with a high degree of correctness, however, there are a few issues. The dataset is composed of 191 unique NHS CCGs, each with 76 associated rows, with a single group of unallocated data with 19 rows. Whilst this suggests complete spatial coverage as this contains the same number of CCGs as the total over the period 2018 – 2019, it also suggests inconsistency. This is also highlighted in the presence of missing values. "POPULATION", "DENTIST COUNT", and "IMD AVERAGE SCORE" variables are all missing 19 values, all located within the unallocated grouping. This is expected, as each of these three variables directly depends on an individual CCG, which is not provided within the unallocated grouping. Additionally, the difference between the records associated with a single CCG and the unallocated grouping suggests the presence of duplicity. Further analysis shows the data is repeated four times for both the age group and dentist count variables, and 19 times for the "IMD AVERAGE SCORE" variable. When accounting for this, only 25% of the data set is distinct.

When taking the correctness issues into account, the dataset appears to be largely consistent, as displayed in **Figure 2**. Using the range values for each of the numerical variables suggests there are no special or default values within the data. This trend is repeated for the "DENTIST COUNT" variable. There do appear to be outliers within the "IMD AVERAGE SCORE" variable, however, displayed in **Figure 3**. **Figure 3** identifies two outliers: NHS Bradford City CCG and NHS Blackpool CCG, both located outside the upper limit. They will remain within the dataset as whilst they are extreme, they are not necessarily incorrect, reducing any misleading conclusions.

Population and Patients Seen for each NHS CCG

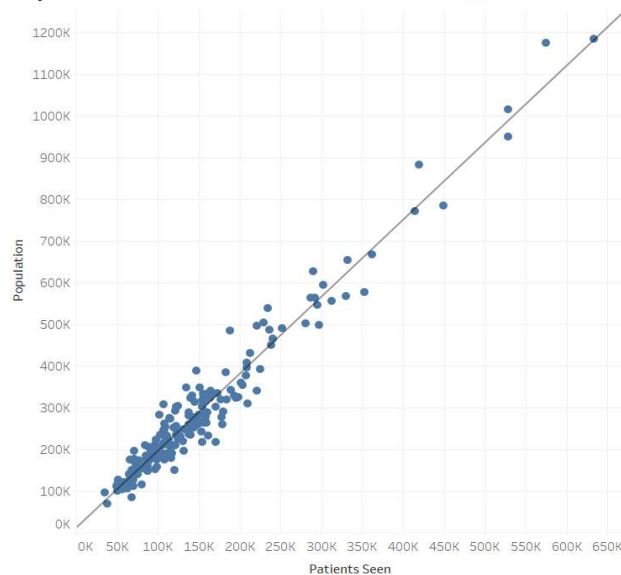


Figure 2: Scatter plot of Population by Patients Seen.

Distribution of Average IMD Scores

Comparing between NHS CCG

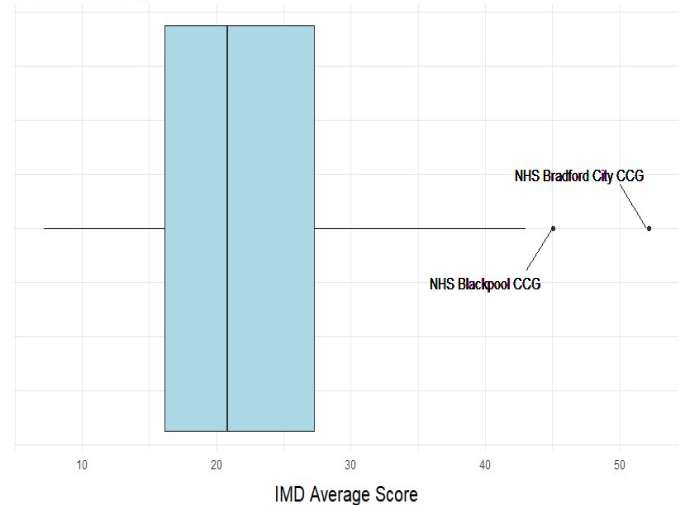


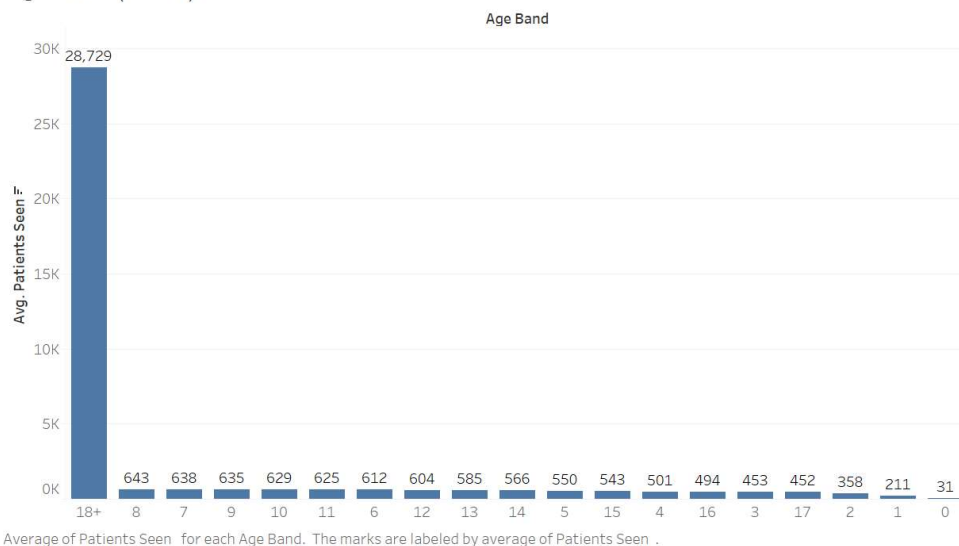
Figure 3: Box and whisker plot of Average IMD Score.

## DETAILED DATA ANALYSIS

### 1) Is there a relationship between the age and number of patients seen by a dentist over this period?

To examine the relationship between the number and age of patients treated, it is necessary to determine the average patients seen for each age band, within the 12 months. **Figure 4**, below, confirms that the majority (76%) of patients seen and treated are over 18 years old. However, in isolation, this analysis may be misleading as it does not tell the full story of the age band of patients who make the most frequent visits to the dentist.

Age band (Years) vs Patients Seen



Average of Patients Seen for each Age Band. The marks are labeled by average of Patients Seen .

Figure 4: Initial pattern between Age Band and Average Patients Seen

**Figure 5**, below, shows that 79% of the population of patients are 18 years and above. This aligns with our expectations; majority of the population is responsible for the highest number of visits to the dentist. Therefore, to get a fair analysis, it is necessary to assess the trend that exists between the age band and the percentage of patients seen in each age band's population.

Age band (Years) vs Population

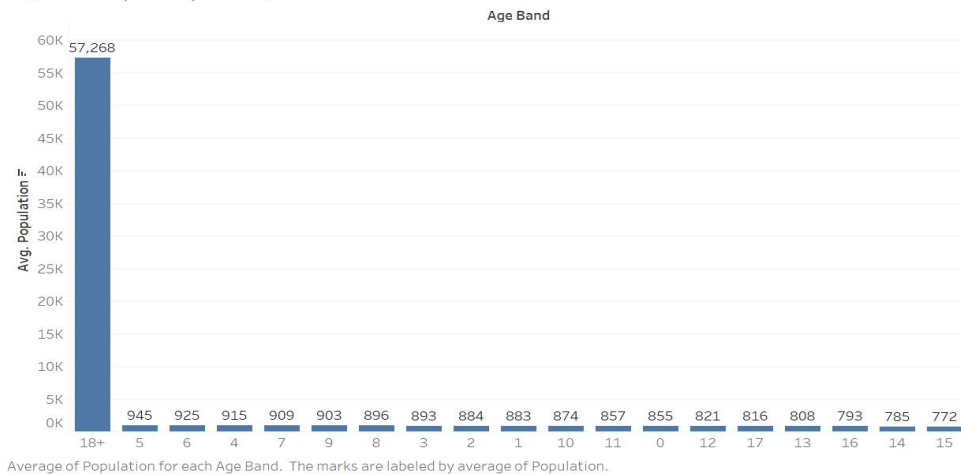


Figure 5: Trend between Age Band and Average Population

**Figure 6**, below, examines the visits to the dentist relative to the individual population of each age band. The results show that 12-year-olds (73.6%) visit the dentist most frequently. Meanwhile, as expected, 0-year-olds (3.7%) make the least visits to the dentist.

Age Band (Years) vs Percentage of Patients Seen in Population

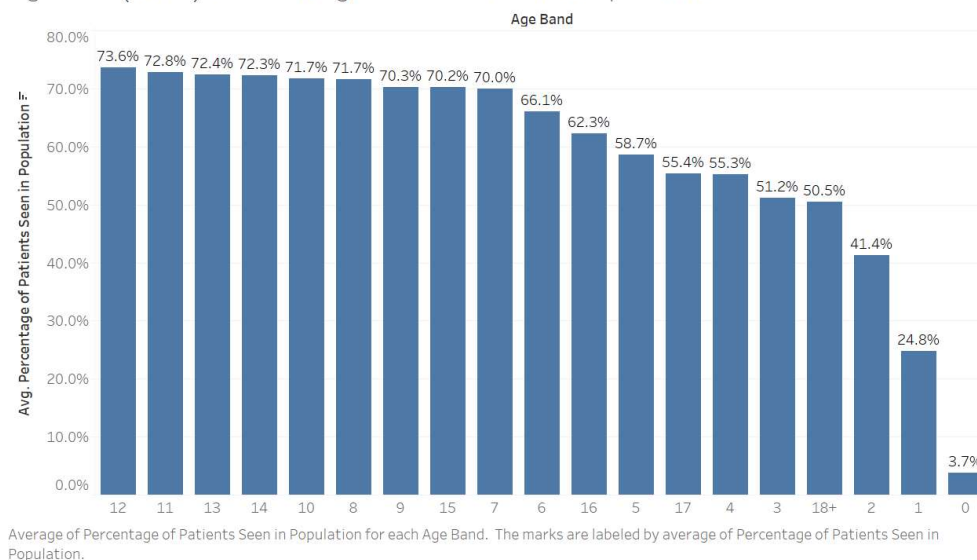


Figure 6: Trend between Age Band and Percentage of Patients Seen in Population

## 2) What is the effect of deprivation on this relationship?

The “Patient: Population Ratio” is the ratio of patients seen to the population of each area, with each CCG being represented as the percentage of people. Creating this new variable “Patient: Population Ratio”, reveals an association between it and the “IMD AVERAGE SCORE” variable. This can be seen in **Table 3** seen below, which shows a Pearson’s Coefficient Correlation value of 0.38.’

Table 3: Correlation Matrix using Pearson's Coefficient Correlation, rounded to 2 d.p.

|                           | Patients Seen |            |               |                   |                           |
|---------------------------|---------------|------------|---------------|-------------------|---------------------------|
| Patients Seen             | 1.00          | Population |               |                   |                           |
| Population                | 0.98          | 1.00       | Dentist Count |                   |                           |
| Dentist Count             | 0.95          | 0.95       | 1.00          | IMD Average Score |                           |
| IMD Average Score         | 0.03          | -0.04      | -0.05         | 1.00              | Patient: Population Ratio |
| Patient: Population Ratio | 0.14          | -0.06      | 0.03          | 0.38              | 1.00                      |

**Figure 7** represents the plot of the “IMD AVERAGE SCORE” variable against the “Patient: Population Ratio” variable.

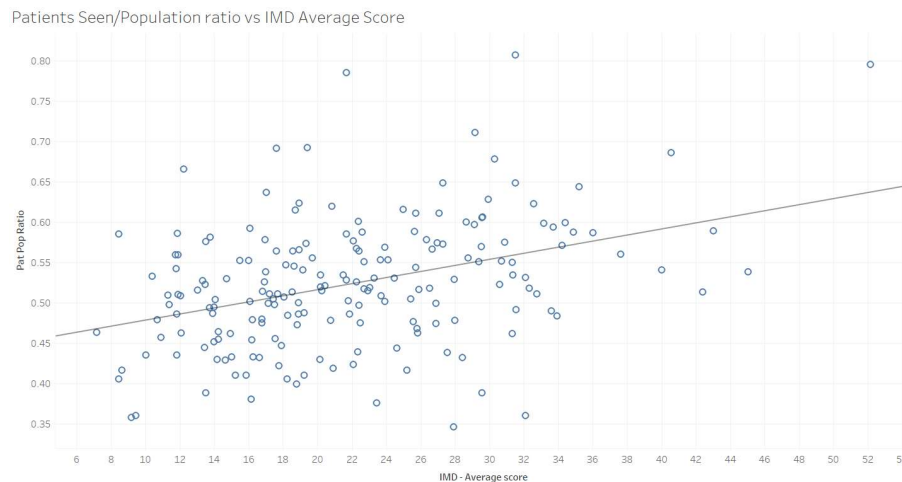


Figure 7: Scatter plot of Patients Seen and Population Ratio by IMD Average Score.

A trend can be seen from the data implying that with an increase in the IMD Average deprivation score, the frequency of patients seen in CCGs concerning their respective populations also shows an increase. **Figure 8** further shows the correlation of IMD average score with each age band respectively.

Correlation of IMD Average Score to each Age Band.

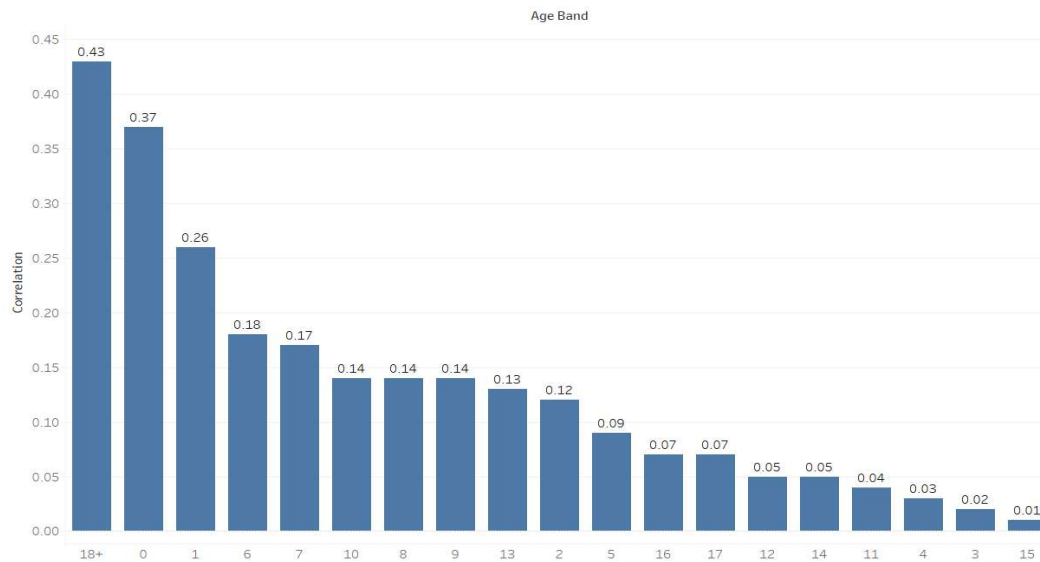


Figure 8: Bar plot of Varying Correlation Values for each Age Band.

The age bands “0” and “18+” have the highest correlation of 0.37 and 0.43 respectively. While age band “1” shows a correlation of 0.26 and the remaining age bands show correlations of less than 0.2. Consequently, the data implies that the age bands “0”, “18+” are the highest affected followed by the age band “1” while the remaining bands in the data set do not show a strong correlation.

### 3) A person aged 55 in 2019 will reach the State Pension age in 2031. What types of CCGs face the greatest shortage of dentists in 2031?

State pension age refers to the age at which one can start receiving regular payments from the government without working. In the UK, this is anyone over the age of 67. Whilst, in theory, dentists can continue to work after that age, it is safe to assume that majority of them will retire in 2031. Therefore, it is important to evaluate the CCGs mostly at risk of shortage of dentists.

To determine the types of CCGs that face the greatest shortage of dentists in 2031, it is necessary to get the dentist count for each CCG and dentist age group. Using this, the percentage of dentists aged 55+ years in each CCG can be calculated, therefore producing an analysis of the CCG types at risk of the greatest shortage of dentists by 2031.

Top 10: Highest Percentages of dentists aged 55+ (years) in each CCG

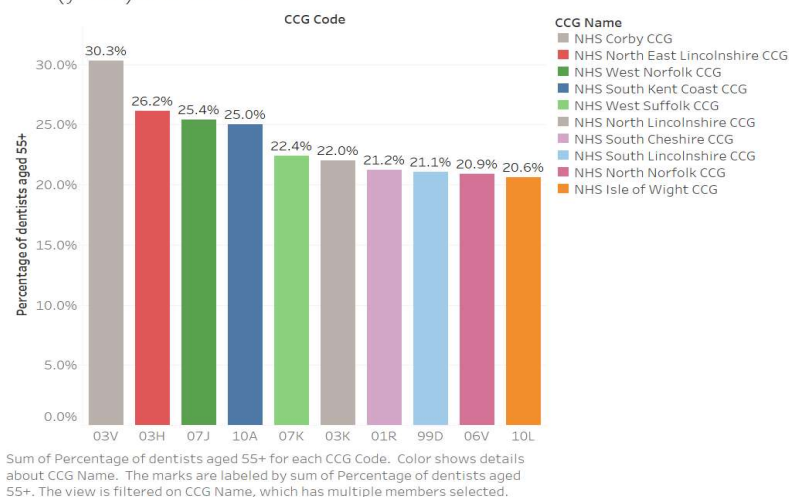
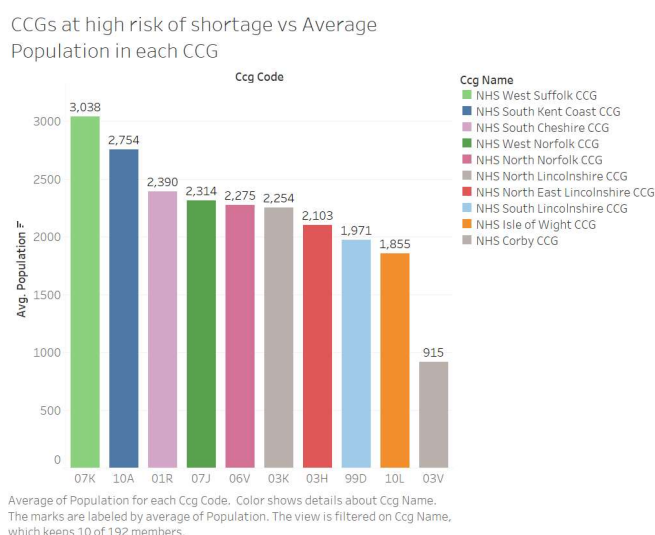


Figure 9: Top 10 CCGs facing the highest shortage risk by 2031.

**Figure 9** shows that NHS Corby CCG (03V) has 30.3% of dentists aged 55+ years, followed by NHS North East Lincolnshire CCG (03H), which has 26.3% of its dentist around that age. There is about a 6% difference between the 2<sup>nd</sup> and 10<sup>th</sup> spots in **Figure 9**. This means that most CCGs in **Figure 9** face around the same risk of shortage of dentists by 2031.

Further investigation is needed to determine what is peculiar among the CCGs that face a high risk of shortage of dentists. The population of individuals in those CCGs when compared to that of all others is seen to be below the average (50<sup>th</sup> percentile) as shown in **Figure 10** and **Table 4**. Therefore, the CCG types at risk of shortage of dentists by 2031 are those with a low population, relative to the other CCGs. It turns out that the CCG with the highest risk of shortage (NHS Corby) has the lowest population across all 192 CCGs as shown in the 0<sup>th</sup> percentile in **Table 4**. The fact that the CCG types at risk of shortage of dentists generally have a low population of individuals is good news because, due to the low population, there would not be an influx of people needing to go to the dentist unlike other CCGs with large populations.



*Figure 10: Relationship between CCGs and average population*

*Table 4: The percentiles of the population of individuals across all 192 CCGs*

| PERCENTILE | 0%  | 25%  | 50%  | 75%  | 100%  |
|------------|-----|------|------|------|-------|
| POPULATION | 915 | 2405 | 3238 | 4325 | 15599 |

## CONCLUSION

The goal of this report was to analyse NHS dentist data that spans for 12 months till 31<sup>st</sup> March 2019 and to answer the three research questions. By doing so, three interconnected conclusions can be drawn. Firstly, an analysis of the ages of patients visiting the dentist suggests that children between the ages of 9 and 14 were the most frequent visitors than any other age group. Secondly, deprivation was identified to be positively correlated with the frequency of adult patients visiting the dentist. This suggests that more deprived areas are found to have received more adult patients across the 12-month period than less deprived areas. The implications of this may mean that more deprived areas face higher demands for dentists, and consequently have a higher population to dentist ratio. The final conclusion identified that the type of CCGs that will have a shortage of dentists in the year 2031 are those that are in the low population areas. The worst affected CCGs are predicted to be NHS Corby and NHS North East Lincolnshire. Overall, this report finds evidence to support an association between deprivation and the demand for dental services. As a consequence, future government policy should take this into account, reducing the inequality between areas with different levels of deprivation.