Executive Summary

The National Health Service (NHS) and General Practitioners (GPs) have been under immense scrutiny, with appointment waiting times on the increase country-wide, and cancer survival rates low in comparison to other developed countries [1]. The NHS is claimed to be underfunded and stretched for resources ([2], [3]), but could perhaps be more effective with the resources available. In this report, appointment data collected over an 18 month period by the NHS [4], was used for the analysis and evaluation. The outcome data are presented from the following two questions, relating to the area of Coventry and Rugby, in Great Britain:

1. Have the weekly counts of Face-to-Face, Telephone and Home Visit appointments increased over the time period 05/03/2018 to 31/08/2019, and as a result, has the workload on GPs also increased over this time period?
2. Does the waiting time for an appointment have an effect on the percentage of appointments not attended?

The evidence collected for both questions included statistical tests and graphical plots, to assess whether measurable statistical differences could be demonstrated and visualised. The outcomes could then be used to determine whether it could be economically worthwhile to establish new strategies for doctor:patient interactions.

To take account of the variation by day of the week, weekly appointment counts were considered for the purposes of comparison. Seasonal variation was considered as a possible factor and as such were also considered for the models. The evidence gathered from the first question clearly demonstrated that Telephone appointments rose during the time period (highly statistically significant p-value of 0.00001). However, there was insufficient statistical evidence to suggest that the number of Home Visits and Face-to-Face appointments changed measurably over this time (p-values greater than 0.05). Telephone and Home Visits showed no statistically significant difference between seasons. Face-to-Face Appointments did show a statistically significant increase for the autumn season, though this could not be utilised in the analysis as there was only one autumn season in our dataset.

The second question was chosen to assess whether a better strategy to offering appointments could be determined, after it was shown that the annual cost of missed appointments was in excess of 200 million GB pounds [5]. The outcome of the analysis was a very clear 'Yes'. Again, both statistical tests and graphical representations were used which showed a clear positive association between a longer waiting time and an increased percentage of appointments not attended. Of the 21 different combinations tested, 16 showed a statistically significant increase in the percentage of 'did not attend' appointments linked with longer waiting times. The minimum time a patient had to wait for an appointment was less than half a day and the maximum was over 28 days. As the time between the initial contact between patient and doctors' surgery rose, so did the percentage of missed appointments.

To conclude, since the analyses undertaken demonstrated that numbers of Telephone appointments increased, whilst Home Visit and Face-to-Face appointments remained constant over the time between March 2018 and August 2019, it was therefore also shown that the workload for GPs had increased, albeit by only a small amount. There was also shown to be a clear association between appointment waiting time and the percentage of missed appointments, with evidence to suggest that the longer the wait, the less the per-day 'did not attend' appointments increased by.

When considering the reported lack of sufficient funding available for the NHS, and acknowledging the high cost associated with missed appointments, the results of these investigations clearly demonstrate the need for establishing new strategies. In particular practices must try to increase the number of same day appointments, for example by telephone and establish straight forward methods for patients to notify their GB practice should they no longer need or be able to attend their appointment, so that this could then be offered to other patients. A system for assigning these could be implemented to ensure as high a percentage of these cancelled appointments could be reassigned.

1. Demand on Coventry and Rugby GPs

Introduction

The question given by the client was: Has demand on Coventry and Rugby GPs increased over the recent past? This was refined based on online literature [6]. There were no data available on the number of GPs employed at each practice, making it impossible to calculate a per capita workload. Instead, the number of appointments made was considered to assess whether overall demand for GPs increased. This was split by appointment type, however 'Unknown' and 'Video' appointments were discarded from the analysis for the reasons of 'hard to quantify' and 'low appointment counts' respectively. Exploratory analysis undertaken demonstrated that there was a clear effect that week day had on the number of appointments. There were two approaches which could have been taken for this, either factoring week day into the model, or using the weekly count of the appointments. The latter approach was chosen, along with a specified time period, since this gave a clearer visualisation of the data and allowed for effect of number of working days to be taken into account. The seasonal variation was also taken into consideration as to the effect this might have on the demand for appointments. The question was reformulated to: Have the weekly count of Face-to-Face, Telephone and Home Visit appointments increased over the time period 01/03/2018 to 31/08/2019 in the Coventry and Rugby area?

Methods

For the purposes of these analyses all activities, including internal meetings and missed appointments, were considered as demand, since GPs have to be available for both. It was acknowledged that time regained from missed appointments was unlikely to be utilised effectively, since appointment time in the UK averages only 9.22 minutes [7].

In order to determine if time of year, given by 'Season' in the following table, had an effect on the median weekly appointment count, the data were also visualised graphically (see Graphics section):

Season	Face-to-Face	Telephone	Home Visit
Autumn	38791	4244	558.0
Spring	35542	4290	516.5
Summer	33854	4153	519.0
Winter	36603	4770	566.0

To classify which season a week fell in, each day in the week was given a season variable as defined online [8]. Then the mode season was taken as the season for that week: For example if three days were classified as summer and four as autumn, the season variable for that week would be autumn.

Model Formulation

In order to make the graph clear and easy to follow, it was decided to overlay a single model to a scatter plot. It was important therefore to select an optimum model. For this a generalised linear model (GLM) was chosen as it allows for the outcome (dependent) variable to have a range of error distribution models other than the normal via a link function. As the outcome variable was count data, a Poisson model was chosen to be the distribution family using the log link function. The models considered the response count variable and predictor variables: week number and Boolean (0 or 1) variables for each different season. The optimum model was selected using the Bayesian Information Criterion (BIC) from a few candidate models, denoted below for each appointment type (4 S.F):

Model Face-to-Face: $\mathbb{E}(y_i) = 0.0003w_i + 10.41sp_i + 10.39s_i + 10.56a_i + 10.40wi_i$
Model Telephone: $\mathbb{E}(y_i) = 0.0038w_i + 0.8237sp_i + 0.8184s_i + 0.8211a_i + 0.8222wi_i$
Model Home Visit: $\mathbb{E}(y_i) = 0.0011w_i + 6.202sp_i + 6.183s_i + 6.307a_i + 6.229wi_i$

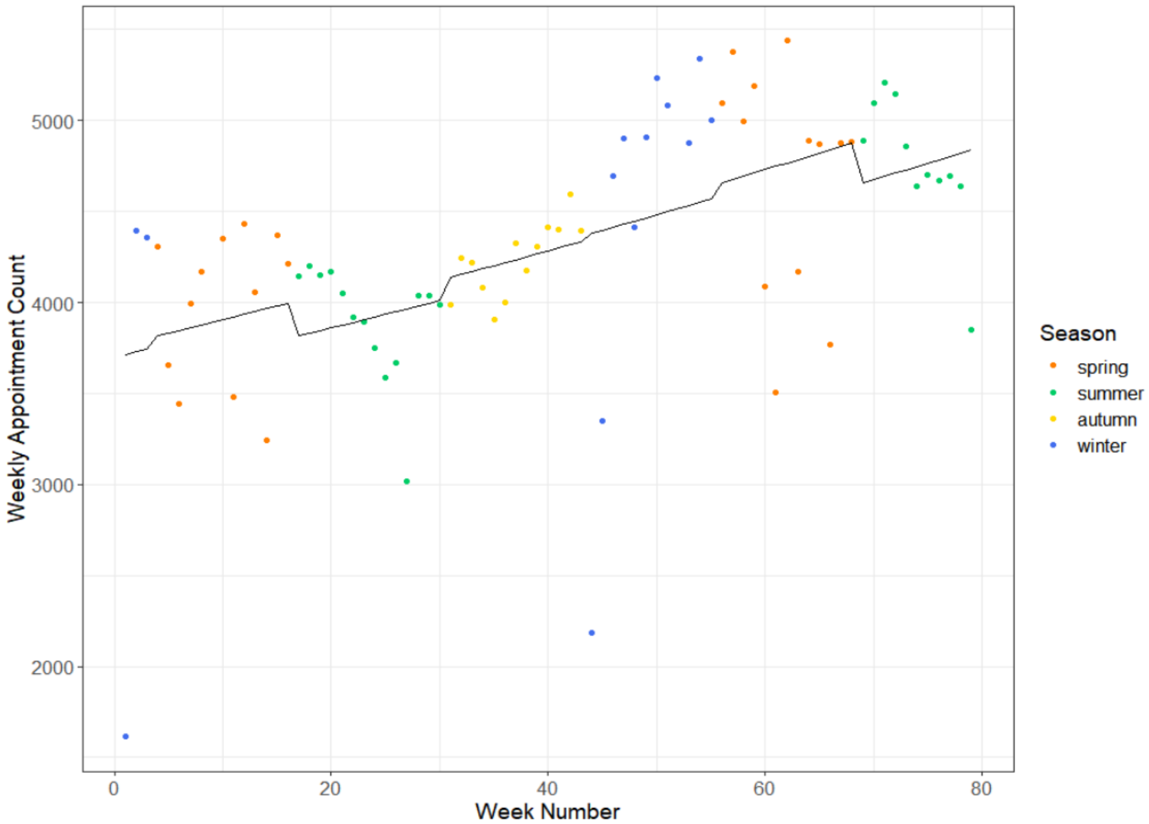
Where w_i = Week Number, season variables: sp_i = Spring Season, s_i = Summer Season, a_i = Autumn Season, wi_i = Winter Season and $\mathbb{E}(y_i)$ denotes the expected value for observation y_i.

The Face-to-Face model showed a slight hike in expected value associated with the autumn season, demonstrated by an increased parameter value for a_i in the model. Both Telephone and Home Visit models have expected values slightly less dependent on season as their parameter values are more constant across the seasons. All three show a positive trend between week number and expected count, denoted by the parameter value of w_i. Of the three models, Telephone appointments looked to have the greatest increase. These observations made from the models was assessed using statistical tests (see Statistical Tests section) to test whether the increase was of statistical significance.

Graphics

The aim of producing graphics was to visualise any obvious trends, backup the conclusion made from the models and provide insight into what statistical tests might be relevant. The graphic included in the text relates to the weekly Telephone appointments as this demonstrated the biggest increase based on week number as demonstrated by the w_i parameter value.

Figure 1: Telephone Weekly Appointment Count with Poisson Regression Line



This graph helps visualise the upwards trend of Telephone appointments and slight season variation discussed in the models section. Similar visualisations for Face-to-Face and Home Visit appointments can be found in the technical appendix [9]. They too reiterate the observations from the models.

Statistical Tests

The two observations to be tested are: Does season have a statistically significant effect on Weekly Appointment count? and Is there a statistically significant increase between week number and appointment count? This will show whether the number of appointments are increasing over time. To understand whether the observations seen from the models and visualisations are statistically significant, it was first important to test the assumptions of the data to see which test was most suitable. Throughout this entire section and the rest of the document, a p-value of 0.05 was used when checking for significance. When checking for a difference between the two years, a one-sided test was used, where a significant result would demonstrate that there was a statistically significant increase between the two compared years and as such there would be sufficient evidence to suggest that the workload for GPs had increased over the time period considered. This can be highlighted in the following hypotheses:

Null hypothesis (H0): There is no increase between the two years compared.
Alternative hypothesis (H1): The year 2019 data has a greater number of appointments than 2018.

To test whether season had a statistically significant effect either an ANOVA Test or a Kruskal-Wallis Test was used. To run an ANOVA Test, it is required that the groups have a homogeneous variance. A

Levene's Test was used to test this assumption, where a significant p-value demonstrates the variances are not the same and a Kruskal-Wallis Test should be used over an ANOVA. These tests were run on each appointment type and the results are summarised in the following table:

Appointment Type	Test	P-Value (4D.P)	Seasons Significant
Face-to-Face	ANOVA	0.0022	Yes
Telephone	Kruskal-Wallis	0.3634	No
Home Visit	ANOVA	0.0715	No

This table demonstrates that the season was significant for Face-to-Face appointments only. Therefore, Telephone and Home Visits data can be directly compared between data from years 2018 and 2019 as they contain a similar number of observations. Face-to-Face appointments should also be compared between seasons. Unfortunately, there were no data for Autumn 2019, which is the season statistically different from the rest and one that would be expected to be different.

To test a difference between the years, Telephone and Home Visit appointments were compared just between 2018 and 2019 data, whilst Face-to-Face appointments were also subsetted by season and compared between the seasons for both years. An underlying assumption to perform parametric tests is that the data come from a normal distribution. This can be checked by running a Shapiro test. A significant result indicates the data does not come from a normal distribution and a Wilcoxon Sign-Ranked Test should be used. If the test does not show statistical significance then a further F-Test is required to test whether the variances between both populations were equal. If the result of this test was statistically significant and the variances different a Welch's T-Test was applied, otherwise a Students T-Test was used. The outcomes from these tests can be summarised as follows, where H0 True (null hypothesis true) denotes there is no difference between the two years and H1 True (alternative hypothesis true) denotes there is a statistically significant difference:

Appointment Type	Test	Season	P-Value (5D.P)	Test Outcome
Telephone	Wilcoxon	NA	0.00001	H1 True
Home Visit	Wilcoxon	NA	0.81530	H0 True
Face-to-Face	Wilcoxon	NA	0.09381	H0 True
Face-to-Face	Wilcoxon	Spring	0.54517	H0 True
Face-to-Face	Wilcoxon	Summer	0.89698	H0 True
Face-to-Face	Wilcoxon	Winter	0.09375	H0 True

This table demonstrates that the only statistically significant increase between the two years occurred for Telephone appointments. It is worth observing that none of the comparisons followed a normal distribution and as such the Wilcoxon Sign-Ranked Test had to be used in each situation, which as a non-parametric test is less powerful than its parametric T-Test counterparts.

Critical Discussion

An optimum model, selected using the Bayesian Information Criterion demonstrated an increase in the number of Telephone appointments, potential increase for Home Visits and evidence of seasonal variation for Face-to-Face appointments during the time period evaluated. When these observations were tested using statistical tests, the number of Telephone appointments showed a statistically significant increase over time, whilst Home Visits showed no statistically significant difference over the time period considered. Face-to-Face appointments showed no statistically significant increase over time, but were seen to peak in the autumn season which was backed up with a statistically significant result.

The analysis was limited by the availability of full data collections for 18 months only. This led to there being just one autumn season and as such Face-to-Face appointments could not be tested by year and by autumn season. When analysing by week it only gives a relatively short amount of time in which to observe changes and few data points. Subtle longer term changes may become apparent in the future, if more data are collected over a longer time period. For example, it may be that instead of classifying the season by date a more suitable approach would be by looking at school holidays instead.

2. Missed Appointments

Introduction

Missed appointments cost the NHS over GBP 200 million a year [5]. In the light of this, the aim of the analysis was to assess whether appointment waiting time has an effect on this and thereby provide some insight into how these values could be reduced. The question was phrased as 'Does the waiting time for a GP appointment affect the percentage of appointments not attended?' In order to tie in with the first question, the area to be considered was again limited to Coventry and Rugby.

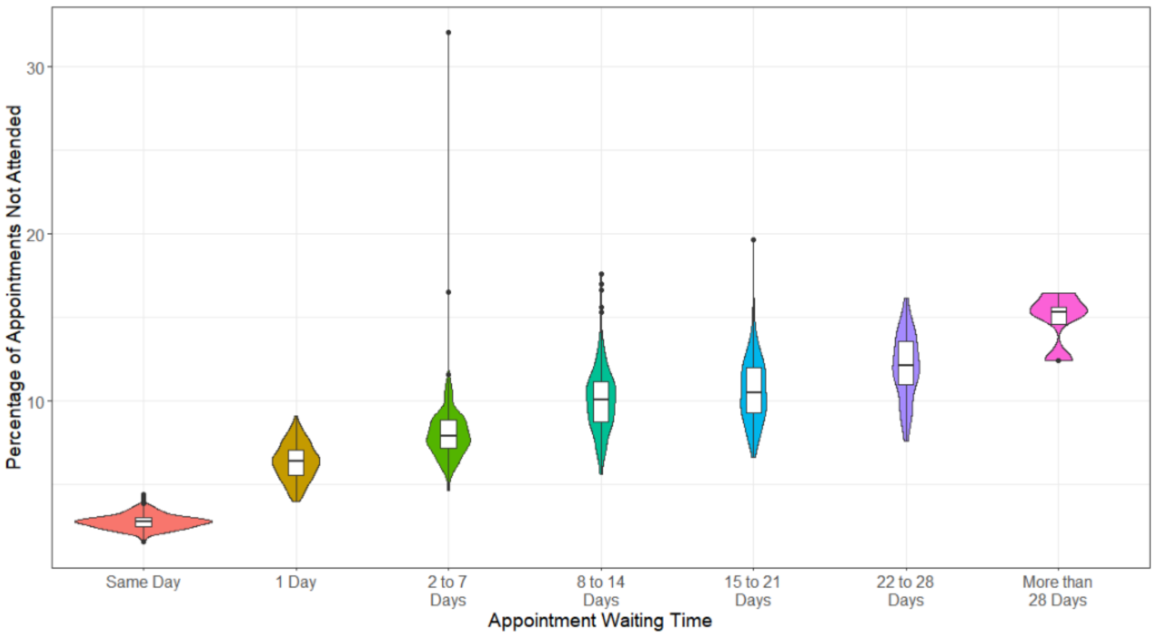
Analysis

To answer the question, only appointments directly related to GP:patient interactions were considered to be of interest. For this reason appointments with a status of 'unknown' or 'not provided' were discarded from the analysis. Those for which the waiting time was 'unknown' or there was a data issue were also discarded to ensure only relevant GP appointments were included in the analysis.

Some days had very few appointments due to clinics being shut and national holidays. When calculating percentage, the number of observations heavily influences outcome. Therefore a low number of appointments as a denominator could seriously skew the 'Did Not Attend (DNA)' percentage and lead to non-representative results. This led to the decision being made to discard all data point entries with less than or equal to 20 appointments for the given appointment waiting time category.

The following graph shows a boxplot inside a violin plot. This enables the distribution of the data points to be visualised from the violin plot as well as key quartile and anomalous point information from the boxplot.

Figure 2: Percentage of Appointments Not Attended vs Appointment Waiting Time



The plot clearly demonstrates a strong upwards trend between appointment waiting time and percentage of missed appointments. This was tested using statistical tests (see Statistical Tests section).

Another consideration was the per-day effect of increased appointment waiting time. The choice was made to use the median rather than the mean as this would limit the effect of outliers on the percentage of appointments not attended, since the data is positively skewed. This was then calculated by taking the median percentage not attended and dividing by the median number of days in the group + 1 (to allow for Same Day to be divided by 1 and not 0). To help understanding, these values are displayed as both increase per day and the median missed per 1000 of each time period in the table below. The number of observations in each group is also shown. Where relevant numbers are displayed to 2 decimal places.

N	Time Group	Missed per 1000	Increase per day per 1000	Increase (%) per day
249	Same Day	27.51	27.51	2.75
117	1 Day	63.66	31.83	3.18
249	2 to 7 Days	79.09	14.38	1.44
246	8 to 14 Days	100.76	8.40	0.84
193	15 to 21 Days	104.60	5.51	0.55
39	22 to 28 Days	121.39	4.67	0.47
4	More than 28 Days	153.33	4.65	0.46

This demonstrates the effect per day of extra wait decreasing the longer the appointment takes and emphasises the benefit of seeing patients as soon as possible.

Statistical Tests

To test the hypothesises generated from the plots, that number of appointments missed increases as time from booking to appointment increases, either an analysis of variance (ANOVA) Test (parametric) or Kruskal-Wallis Test (non-parametric) can be used. These tests detect whether at least one of the comparison groups comes from a different population than the rest. So, if the outcome from this test is significant, then a further multiple comparisons test can be used to see which combinations of the booking times are statistically different. To assess which test was more appropriate for our data, a Levene's Test was used. This assesses whether the different groups have a homogeneous variance, which is a required assumption for the ANOVA Test to hold. The outcome from this test had an F-value of 28.93 over 6 Degrees of Freedom and a very significant p-value $< 2.2 \times 10^{-16}$. This meant that the ANOVA Test could not be used and led to the implementation of a Kruskal-Wallis Test. The hypothesis for this test is formulated as follows:

H0: Percentage of appointments not attended is not affected by appointment waiting time.
H1: Percentage of appointments not attended is affected by at least one of the appointment waiting time categories.

The outcome of this test had a highly significant p-value of $< 2.2 \times 10^{-16}$, leading to the rejection of the null hypothesis and acceptance of the alternative hypothesis. Now that it has been statistically shown that the Kruskal-Wallis Test is significant, it is known that at least one of the appointment waiting time categories is different. To see which combinations are significantly different, a multiple comparison test was used [10]. Of the 21 possible comparisons, only five were not statistically significant in their difference: 22-28 days to $>$ 28 days; 8-14 days to 15-21 days; 8-14 days to $>$ 28 days; 15-21 days to 22-28 days and 15-21 days to $>$ 28 days. All other combinations were statistically significant in their difference, with the longer time period having a larger DNA percentage compared with the shorter one. This demonstrates that towards the latter end of the date range there was insufficient evidence to suggest that the results came from different populations. This is, in part due to the larger range of values covered in Figure 2. A reduced sample size of these data categories will also lower the statistical power of the tests (the ability of the statistical tests to detect a difference if there is one).

Critical Discussion

The results showed a statistically significant, positive relationship between longer booking to appointment times and an increased percentage of appointments where patients did not attend (DNA). Evidence did suggest that the increase was most noticeable at the beginning, emphasising that working on getting people seen quicker would be the best way to reduce missed appointments. Improving the ease in which people can cancel and change their appointments could also diminish the number of DNA appointments for those with longer waiting times.

The main limitation of this data analysis was the number of appointments in the 22 to 28 Days and More than 28 Day categories, with part of this being due to discarding data points with less than or equal to 20 observations. As such these had 39 and 4 suitable observations in respectively which reduces the precision of the interquartile range observed. This leads to the multiple comparisons test being less likely to demonstrate a significant result with these groups. If more data had been collected it is likely that these combinations would have also generated a statistically significant result due to the tighter confidence intervals calculated and associated higher statistical power.

References

[1] C. Lynch, "Measuring up: how does the uk compare internationally on cancer survival?," 2019. *Cancer Research UK*. [online] Available from: accessed 26/01/2020: https://scienceblog.cancerresearchuk.org/2019/09/11/measuring-up-how-does-the-uk-compare-internationally-on-cancer-survival/ (accessed 26/01/2020).

[2] S. Neville, L. Hughes, and B. Staton, "Sharp deterioration in NHS performance loads pressure on tories," 2019. *The Financial Times*. [online] Available from: "https://www.ft.com/content/a5585afe-06cf-11ea-9afa-d9e2401fa7ca" (accessed 26/01/2020).

[3] "Anon," 2019. *BMA*. [online] Available from: accessed 26/01/2020: "https://www.bma.org.uk/collective-voice/influence/key-negotiations/nhs-funding"(accessed 26/01/2020).

[4] Anon, "Nhs data," 2019. *Cancer Research UK*. [online] Available from: accessed 26/01/2020: https://digital.nhs.uk/data-and-information/publications/statistical/appointments-in-general-practice (accessed 26/01/2020).

[5] M. Bulman, "NHS England losing 216m pounds a year to missed GP appointments with one in 20 wasted, figures show," 2019. *The Independent*. [online] Available from: "https://www.independent.co.uk/news/uk/home-news/gp-appointment-missed-costing-nhs-england-waste-a8706706.html" (accessed 26/01/2020).

[6] Peng and Matsui, "The art of data science. a guide for anyone who works with data," 2017.

[7] E. Pickover, "NHS patients getting less time with their GPs compared to other developed countries," 2017. *The Independent*. [online] Available from: "https://www.independent.co.uk/news/health/nhs-gp-doctor-appointment-time-comparison-patients-latest-a8044681.html" (accessed 26/01/2020).

[8] Anon. *Time and Date* [online] Available from: accessed 26/01/2020: "https://www.timeanddate.com/calendar/seasons.html" (accessed 26/01/2020).

[9] S. Brownsey, "st344_individual," 2019. url: https://github.com/Brownsey/st344_individual.

[10] R. Documentation, "kruskalmc."