Causes of Scottish Cancelled Operations

Cameron J. Dobbie, Brian R. Daniel, Roan Murray

September 9, 2024

1 Overview

Having a planned operation cancelled has the capacity to change someone's life forever. This report investigates these cancellations and their reasons through the procurement, exploration and analysis of 5 datasets from the NHS Open Data website. We compare different Hospitals and Boards in hopes of finding outliers in the data which we then explore, and if necessary suggest government resources be provided to them to help solve their issues. We also look at the effect of COVID-19 on the cancellation rates and create projections to forecast when the NHS Cancellation rates would recover from the pandemic. Furthermore we identify spikes in the data and explore flu cases as an explanation for these spikes - with positive correlations found. Map-based visualizations assist in telling the story of cancellation rates throughout the report. Understanding and finding ways to improve the NHS Cancellation rates is of paramount importance as peoples lives and health are at stake. We explore the underlying story behind NHS cancelled operations and possible factors that effect these rates - in the hopes that this information could ultimately prove useful in lowering them.

Throughout this report we deem a p value < 0.05 to be statistically significant to match industry standards.

2 Introduction

Context and motivation The NHS is one of the most important government services and cancelled operations are incredibly disruptive to the lives of Scottish residents. In many cases they can be the difference between life and death. Therefore it is of extremely high interest to investigate and deepen our understanding of cancellation rates - and how to improve them.

This paper will analyse the cancellation rates of planned operations across various hospital and NHS boards in Scotland over the course of the last 9 years (May 2015 to January 2024) - as well as the various factors affecting these rates. COVID-19 and Influenza are both possible factors and we shall explain anomalies in the data by linking them to historical data on these large respiratory disease outbreaks. Furthermore we shall investigate which hospitals and boards are under performing and over performing - and explain the stories behind them.

Previous work Previous work in this area has indicated that COVID-19 has played a role in NHS Cancellations in vital areas such as emergency general surgery [4], indicating it could be an area of interest for our project. Furthermore it has been indicated that many different factors are impacting the NHS's ability to carry out operations, such as: winter pressures, failure of adequate preoperative assessment, staff shortages, access to operating theatres, equipment shortages, and critical care capacity. Hospital factors associated with a higher risk of cancellation were the presence of an emergency department and enhanced ward care areas. The cost in lost operating theatre time is as high as £400 million per year [5] adding a quantifiable monetary value on the huge cost to public health. Statistical releases from Public Health Scotland indicate that operations have not yet recovered to pre-COVID levels [11] investigating when they might recover would fall within the scope of our project.

Objectives Firstly we shall explore the given data - visualizing and providing statistics on the number of planned operations and their reasons for cancellations.

Secondly we seek to answer the following questions: 1. Which hospitals or healthcare boards have statistically significant cancellation rates, potential underlying reasons thereof and if they require special attention. 2. How COVID-19 impacted the NHS and it's cancellation rates - and when operations are projected to recover. 3. What interesting patterns exist in the data and can influenza cases be a possible explanation of them?

3 Data

Data provenance All five CSV data sets were taken from the Open-data NHS website under 3 different topics. First the cancelled planned operations [10], hospital codes [13] and also the [14]. The Data is released under the Open Government Licence which means that it is free for anyone to use [1] so long as any sources are cited.

Data description We used the following datasets with the following counts :

Name:	Data Count:
Cancellations By Board	1575
Cancellations Scotland	105
Cancellations By Hospital	4502
Hospital	276
Respiratory Scot	5070

Table 1: Dataset Names and Counts

With the following fields (Formatted for readability):

Cancellations Scotland:	Cancellations by board:	Cancellations by hospital:
Month	Month	Month
Total Operations	Total Operations	Total Operations
Total Cancelled	Total Cancelled	Total Cancelled
Cancelled By Patient Reason	Cancelled By Patient Reason	Cancelled By Patient Reason
Clinical Reason	Clinical Reason	Clinical Reason
OtherReason	OtherReason	OtherReason
Country	HBT	Hospital

Table 2: Table Fields

Hospitals:	Respiratory Scotland:
Hospital Code	Month
Hospital Name	Flu Cases
Post Code	

Table 3: Table Fields (Continued)

* Hospital contained the following fields which were not used: Address line 1, Address line 2, Address line 3, Address Line 4, Health Board, HSCP, Council Area, Immediate Zone, Data Zone

** Cancellation Scotland, Cancellations by board and Cancellations by hospital all contained multiple Quality Flag (QF) fields which have been removed.

Cancellation Scotland, Cancellations by board and Cancellations by hospital's month fields all ranged from May 2015 - January 2024 (105 Months)

Respiratory Scot's Week fields corresponded to October 2016 - March 2024 (90 Months)

Data processing All of our data sets needed to be cleaned. Our initial three data sets all had quality flag fields for fields related to number of operations, total cancelled operations and cancelled operations by various reasons. They were deemed not needed and deleted. The 'Country' field was dropped from the Cancellations Scotland dataset as all entries corresponded to Scotland and thus were redundant. In 2 of the given datasets and the respiratory cases dataset, the fields representing dates were stored as integers. These were all converted to date-time fields for ease of data manipulation. After this we left-joined our Cancellations by Hospital dataset with the Hospitals dataset on the field 'Hospital' in Cancellations by Hospital and 'HospitalCode' in Hospitals corresponding to a unique 5 digit identifier code in both datasets. We joined only the Hospital names and postcodes fields from the Hospitals dataset. This gave us a cleaned data set with extra information so that we could use postcodes and Geopy to get the longitude and latitude coordinates to create map visualisations using Folium.

There were several hospitals in our 'Cancellations by Hospital' data that did not correspond to any hospital in our Hospitals dataset. After some research it was found that these hospitals were in fact all either merged, destroyed or replaced [12]. However there were only a handful of these and so the postcodes and hospital names were found for them manually and added in.

The weekly respiratory cases dataset for Scotland included counts of different respiratory-related infections in addition to flu cases stratified by type and subtype. These categories were removed so that only case-counts of influenza classified as type A or type B were retained. The weekly data was then grouped so that rows represented the number of reported influenza cases in Scotland each month.

4 Exploration and analysis

The first thing we wanted to investigate was which hospitals and boards in Scotland had statistically significant cancellation rates. Throughout the entire paper the same colour-blind friendly colour scale is used, for consistency and to add to map visualisations.

We constructed figure 1 to compare NHS boards. 2 tailed Z-tests were performed on all boards however 'The Highland board' alone resulted in a statistically significant p-value < 0.05. Existing media shows that a highland MSP has been calling for a new Hospital to be built [7]. In 2023, NHS Highland announced investigations as to the suitability of a site for the new Hospital [3]- which upon inspection refers to Belford Hospital which will be mentioned in the paragraph below. NHS Highland seemed hopeful however as previously mentioned there have been protests of delays in recent weeks - and an immediate solution is not clear.

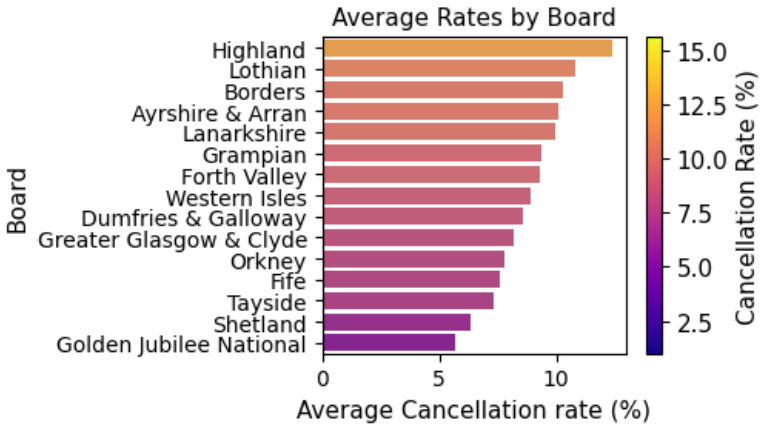


Figure 1: Bar Graph of the Average Cancellation Rate by Hospital

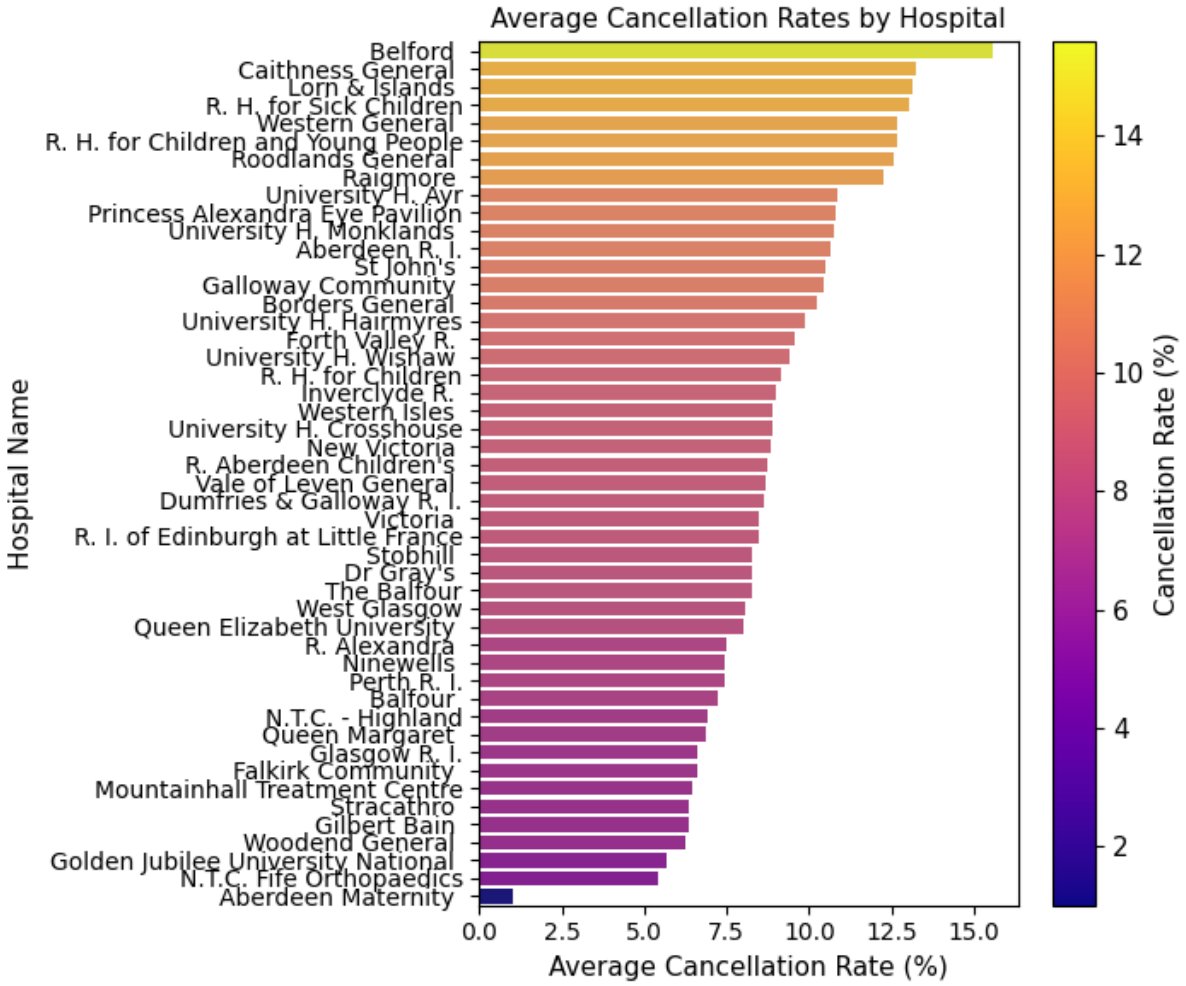


Figure 2: Bar Graph of the Average Cancellation Rate by Hospital

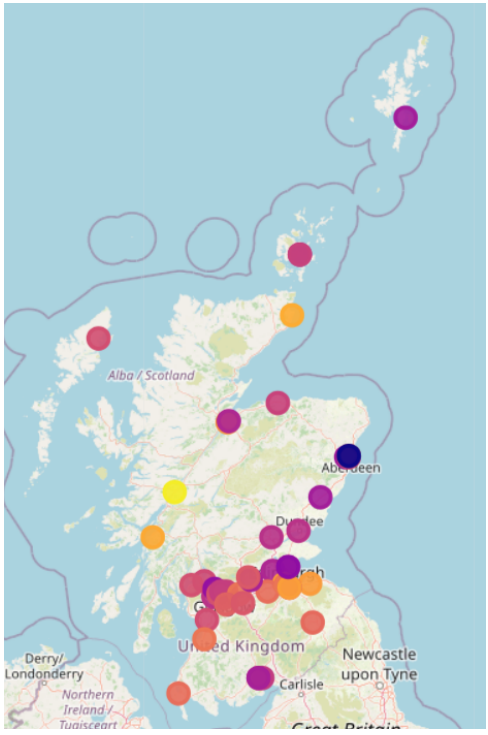


Figure 3: Map of Avg Cancellation Rate by Hospital

We then constructed Figure 2 and Figure 3 which compared all individual hospitals against each other and allowed us to visualise the cancellations on a map for another view of the data. 2 tailed Z-tests were performed on all hospitals and these were the only 2 that were statistically significant. Belford Hospital being the hospital with the worst cancellation rate of 15.58% which is nearly double the mean value of 8.97% and Aberdeen Maternity with a cancellation rate of 1.005% nearly 9 times less than the mean. These are statistically significant outliers in the data with Z-tests assuming a normal distribution resulting in p-values < 0.05 for both.

Aberdeen Maternity is a specialist hospital for expectant mothers to give birth in. The low cancellation rate is likely because the only planned operations are births and c-sections (operations that usually cannot be cancelled) It is reassuring that this critical service has not been disrupted while the rest of the NHS is under strain. It is one of only three maternity hospitals in Scotland. The other two are in not in the dataset which explains why it is the sole outlier.

The hospital with the largest cancellation rate is Belford Hospital. Which linking back to our comparison on the

Boards is also part of the worst performing NHS Board. The hospital is very out of date with locals saying "the hospital needed to be replaced with urgency"[8]. The NHS have delayed the plans to build a new hospital due to funding restraints and this has been leading locals to protest against the delays just a few weeks ago. [8] . With an incredibly high cancellation rate (15.58%), widespread local protests against the hospital and broad media attention it is clear that Belford hospital is in dire need of government resources.

It is apparent that there are points of strain in the NHS. While it is concerning that there are outliers in the negative direction - there is only 1 board and 1 hospital. Efforts should be made immediately to rectify this - however there do not seem to be multiple critical points. Given widespread concerns many would suggest that this is simply due to the system as a whole being under strain - making threshold for statistically significant outliers much higher.

Secondly we visualized the cancellation rates in Scotland by reason so we could identify any patterns in the data and other points of interest:

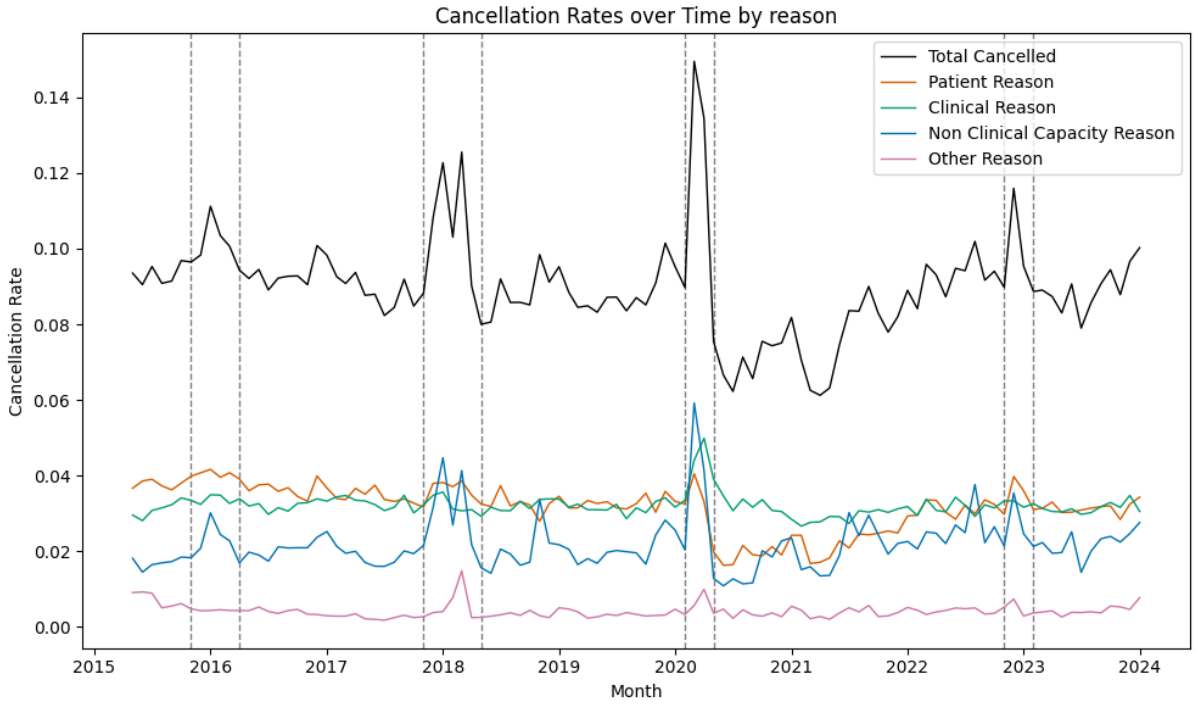


Figure 4: Cancellation rates

We visually identified 4 'spikes' in the data total cancellation rate. We conducted a 1 tailed Z test for each spike's peak assuming a normal distribution and calculated the following P-Values. The first spike on 1st of 2016 January had p-value ≈ 0.054, the 2nd spike on March 2018 had p-value < 0.05, the 3rd on March 2020 had p-value < 0.05, the 4th spike on December 2022 had p-value < 0.05. For the rest of the analysis we discarded first spike which was not statistically significant.

We hypothesized that the 2020 spike was due to the first COVID-19 lockdown as this corresponded to the biggest spike with all different reasons increasing, indicating widespread disruptions like those caused by a global pandemic. This aligns with reports from Public Health Scotland that stated the following: "In March 2020, of all planned operations: The decrease in number of planned operations and increase in Non-clinical/Capacity cancellations is related to the impact of the COVID-19 pandemic." [9]

To further investigate the effect of COVID-19 on the NHS we have split the data into 2 different sections. Pre-COVID (Up to February 2020) and post-COVID (from May 2020 onwards). We created a linear regression model to see how long it would take post-COVID levels to recover to their pre-COVID mean. The independent variable was *months* and the dependent variable was *count*. Months were

converted into integers for this regression. The model and it's resulting coefficients can be described by the following equations:

$$\hat{y} = count \text{ and } x = months$$

$$\hat{y} = \beta_1 x + \beta_0$$

$$\beta_1 = 0.00843 \text{ and } \beta_0 = -118898.256$$

With the mean of squared errors being calculated according to the following formula:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 11171475.023$$

Where n is the number of data points, y_i is the ith data point and $\hat{y}_i$ is the ith predicted value.

The regression is visualised in figure 5

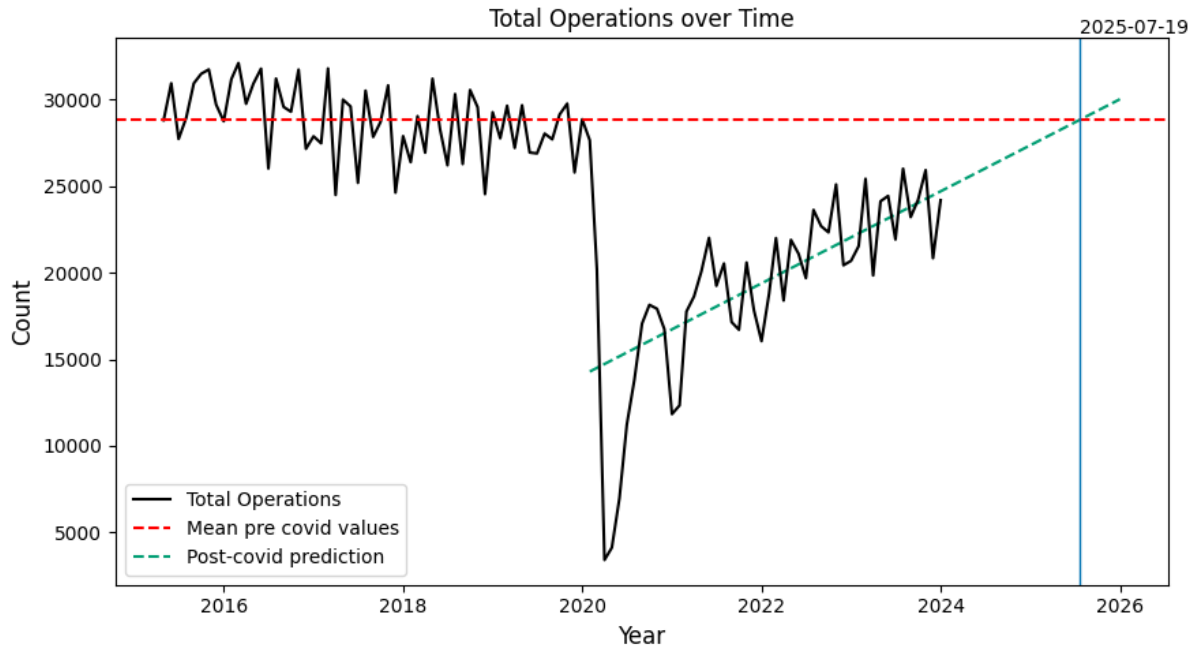
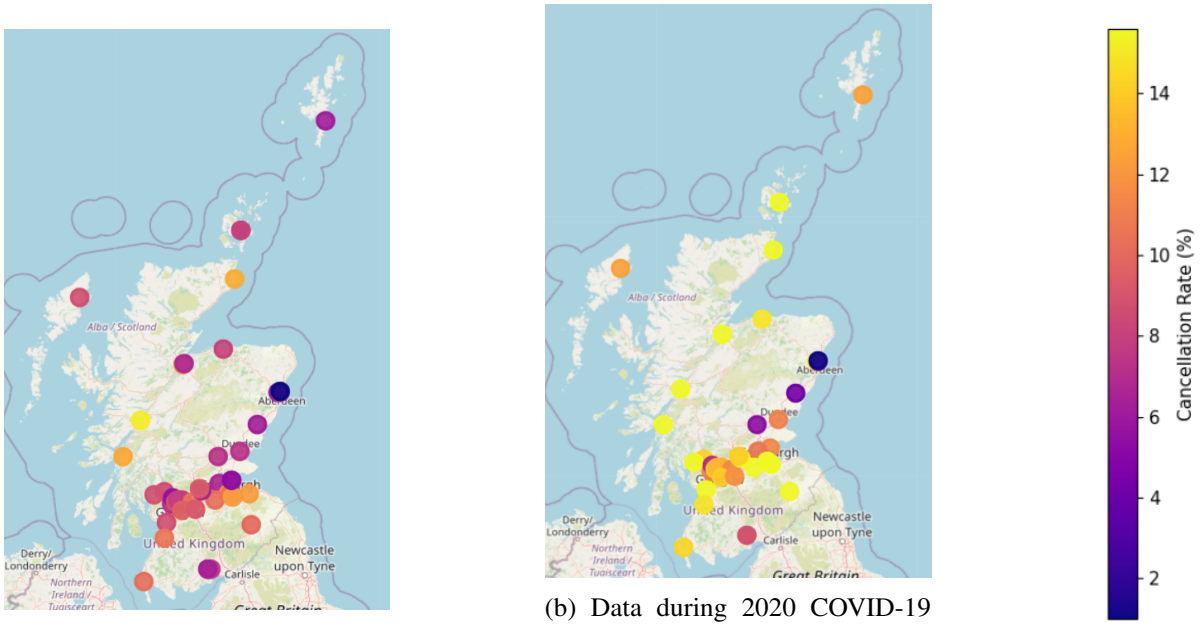


Figure 5: Linear regression to return to mean Total Operations

The pre-COVID mean was 28816.825 The model intersected this mean on the 19th of July 2025 meaning this projection estimates that the NHS will only fully recover it's number of operations over 5 years and 5 months from when COVID first caused a significant dip in operations. This is an incredibly large time period and it is clear that many lives would have been disrupted and lost due to these operations not taking place. Patients with terminal diseases can not wait over 5 years for treatment. Even if operations were not dealing with potentially fatal issues the loss of quality of life is unacceptable.

The average cancellation rate across the entire time frame of our data as well as just the COVID spike are shown in figure 6 and shows the change from the normal rate of pink, blue and purple being relatively low to the COVID map with far more orange and yellow indicating a significantly higher cancellation rate

The COVID-19 pandemic does not explain the other statistically significant spikes in figure 4 in , 2018 and 2022. The reason that most consistently spiked alongside total cancellation rates was Non-Clinical Capacity reasons which had Pearson's R of 0.84. As stated by the NHS Open data website this refers to: 'no beds available', 'staff not available', 'equipment not available/dirty' or 'theatre session overran therefore patient cancelled' [10].



(a) Average across all Months in Data

(b) Data during 2020 COVID-19 spike

Figure 6: Average vs COVID-19 spike maps

This posits the possibility of other significant events occurring during these periods—events pervasive enough to cause the cancellation of a substantial number of operations. After some investigation we found articles [6] [2] reporting flu spikes during the 2018 period as seen above in figure 4.

We hypothesized that flu might account for the other spikes in the data. To investigate this we obtained a flu cases data set from the NHS open data website and created the following visualization:

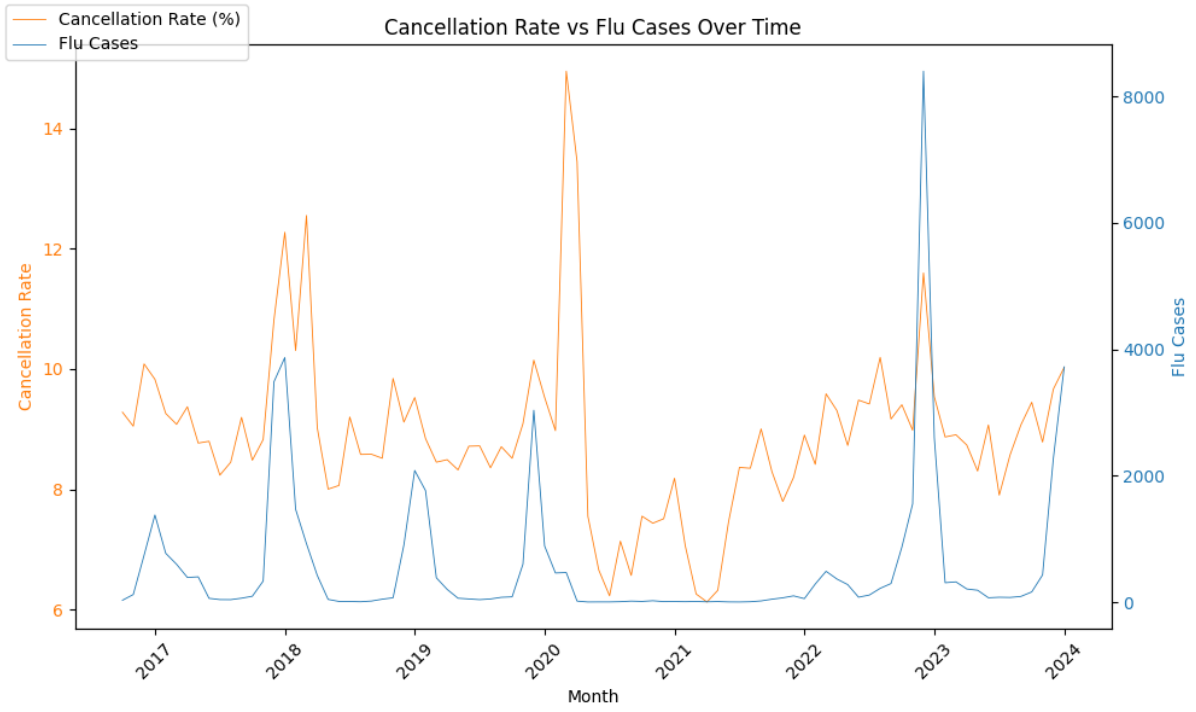
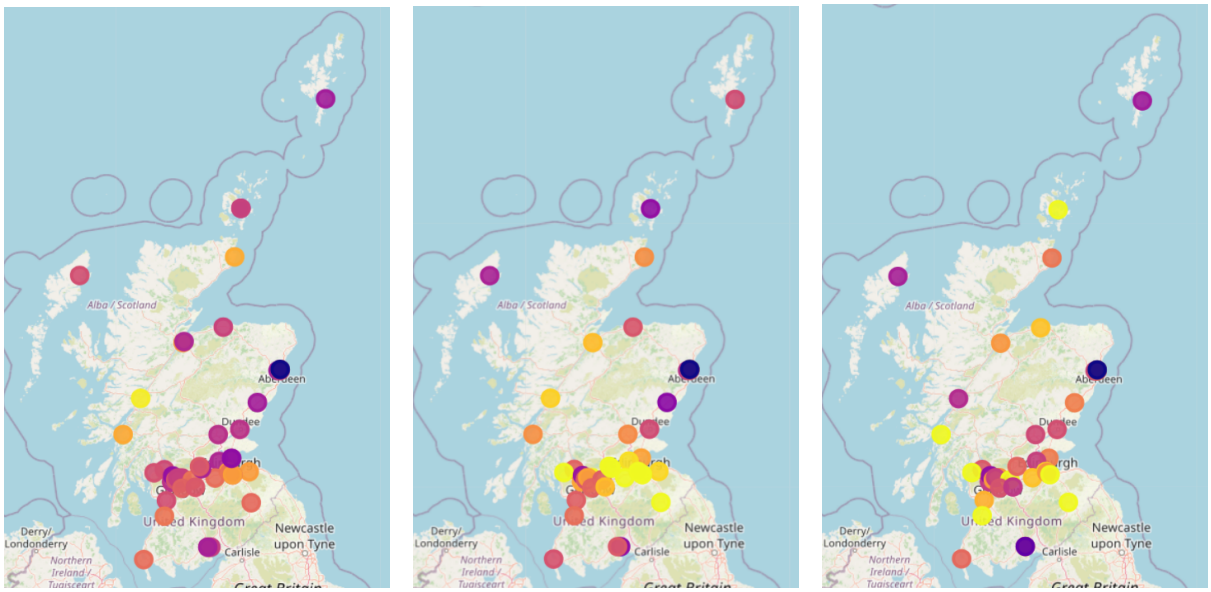


Figure 7: Cancellation Rate vs Flu Cases Over Time

To investigate this potential explanatory factor, we plotted the cancellation rate against flu cases to see if there were any obvious patterns in the data. As is evident in figure 7 a notable visual correlation

between flu cases and cancellation rates. Upon visual inspection of the data, it is evident that periods with higher flu activity correspond to spikes in the cancellation rate of operations. This visual correlation suggests a potential relationship between seasonal flu outbreaks and disruptions in the healthcare service.



(a) Average across all Months in Data (b) Data during 2018 flu spike (c) Data during 2023 flu spike

Figure 8: Average vs flu spike maps

To quantify this relationship, we conducted a Pearson correlation coefficient test on the flu cases and cancellation rate data. The analysis revealed a positive correlation ($r \approx 0.46$, $p < 0.00001$) between the number of flu cases and the cancellation rate of operations, however it is not large enough to be statistically significant. We assumed that the reason these results were so poor was because we included data from non-flu seasons.

We decided to do a map visualisation to investigate if there are any geographic trends in cancellation rates during peak flu times. We can see that when comparing 8b and 8c with 8a, it shows that cities are associated with higher cancellation rates during peak flu times, perhaps because the transmission rate is much higher in these areas.

To address this, we refined our analysis by focusing solely on flu seasons, defined as the winter months of December to January. Upon doing this, we found a strong relationship ($r \approx 0.78$, $p < 0.0005$) between flu cases and cancellation rates during peak flu seasons. These findings highlight the vulnerability of healthcare systems during peak flu periods and emphasize the need for strategies to mitigate these disruptions.

5 Discussion and conclusions

Summary of findings We identified 2 Hospitals as being statistically significant outliers with one maternity ward having an exceedingly low cancellation rate and The Belford Hospital having a high cancellation rate. This hospital with a high cancellation rate also had it's corresponding board (NHS Highland) as a statistically significant outlier. This brought our attention to cancellations being a regular problem in more rural areas. A brief investigation of existing media revealed many issues with the NHS Highland board and delayed to a new hospital being built to solve these issues. Clearly more resources are required for these areas.

Relating to our second question of how COVID-19 impacted the NHS, it is clear that this was a massive disruption. COVID reset operations to a very low level with our report projecting that 5 and a

half years since the first lockdown will be needed for recovery. It is clear that the NHS was not adequately prepared - and has not yet recovered.

Our investigation into the potential relationship between the number of flu cases and cancellation rates pointed to a substantial link between the two, specifically in the winter months. This could be due to an increase in flu patients in tandem with potential staff shortages due to flu. Our findings highlight the vulnerability of healthcare services during these times and suggest that preemptive measures are required such as reducing flu transmission, better patient management, and ensuring the availability of necessary resources.

Evaluation of own work: strengths and limitations Our team were prompted by our initial data visualisations to search for articles and data to explain the anomalies that were found. For example the rises in cancellation rates being linked to flu outbreaks. By doing this our paper follows a clear story of data exploration to discovery to then finding more data to explore. We also found very reputable sources for all our data which ensured the paper met a high ethical standard and none of our team had any conflicts of interest in regards to the research.

Some limitations of our study are that a lot of hospital data is confidential and so we have no right to use it. For example we would have wanted additional protected fields: sex, age of patient and type of operation. However we were not able to find data with these fields from reputable sources and with open licenses. With these fields we could have predicted which patients are most likely to cancel or have their operations cancelled and we could see if there is any bias in the system for example perhaps women aged 25-30 needing tonsils removed are more likely to get their operations cancelled then men of the same age and operation type or some other combination. We could have also found out if certain types of surgeries are more likely to get cancelled or if other types are less likely and theorised about why this might be.

Comparison with any other related work Our project investigates how COVID-19 effected cancellations. This is a widely studied area and there exists great overlap in the existing literature about COVID-19 and cancellations with Public Health Scotland commenting on this link [9]. Related work does mention that the NHS has not yet recovered from COVID however did not mention when that might happen [11] - whereas that does fall within the scope of our report. Related work does not comment on flu cases effecting cancellation rates - nor could we find any other reports about this. Seeing as though we've proven a correlation between flu cases and cancellation rates - further investigation into flu as a cause for cancellation rate strain would be appropriate - which we have not yet seen, although previous work does indicate that winter pressures pressure the NHS [5].

Improvements and extensions Adding onto the map visualisations we would have liked to have researched if there was a correlation between where a hospital was located (longitude and latitude) and its cancellation rate, or total operations. After doing this it would have been interesting to see if there was an optimal place to locate a new hospital such that the hospital would have the lowest possible cancellation rate however this would require a much more advanced regression model. It would have to use data of the population density of Scotland so that there would not just be a hospital up a Munro or in the Atlantic Ocean (where there would be the lowest population) and need to optimise for number of operations as well as cancellation rates as there is no point having a hospital which has a 0 cancellation rate but only 1 operation. As well as this, lots of other factors were deemed by the team to be significant to the model and so it was not feasible to complete in the time frame.

Further extensions could include finding more data about hospitals (ratio of staff to operations for example) to then create predictive models to further investigate factors leading to cancellation rates. Ultimately the NHS faces an enormous resource allocation problem - with limited money and staffing efficiency is incredibly important - therefore analyzing more factors leading to cancellations has the potential to quite literally save lives.

Relating to the question of the correlation between flu cases and cancellations - extensions on other winter factors could help explain the spikes. Given the regularity of these spikes it seems surprising that

the NHS takes such strain during these winter months. One could investigate why the NHS does not lower scheduled operations to then avoid cancellation rate spikes. Furthermore investigating the trade-offs of scheduling operations with high chances of being cancelled would be interesting.

References

[1] National archives. *Using Information under open government licence*. Last accessed 29 March 2024. URL: https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/.

[2] The Guardian. *Flu outbreak: UK deaths triple with GPs seeing major rise in patients*. URL: https://www.theguardian.com/society/2018/jan/18/flu-outbreak-gps-under-huge-pressure-as-deaths-soar-to-120.

[3] NHS Highland. *Ground investigations begin at proposed site for new hospital*. URL: https://www.nhshighland.scot.nhs.uk/news/2023/11/ground-investigations-begin-at-proposed-site-for-new-hospital/#:~:text=NHS%20Highland%20is%20pleased%20to,to%20last%20for%20three%20weeks..

[4] National Library of Medicine. *Changes in Emergency General Surgery During Covid-19 in Scotland: A Prospective Cohort Study*. Last accessed 29 March 2024. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7454130/.

[5] National Library of Medicine. *Counting the cost of cancelled surgery: a system wide approach is needed*. Last accessed 29 March 2024. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7118877/.

[6] National Library of Medicine. *Scotland facing worst flu outbreak for six years*. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1128740/#:~:text=The%20latest%20figures%20show%20that,to%20all%20but%20emergency%20cases..

[7] Philip Murray. *Raigmore operation cancellations crisis 'highlights need for new hospital in Inverness, says Highland MSP*. URL: https://www.northern-times.co.uk/news/operation-cancellations-highlight-need-for-new-hospital-in-310965/.

[8] Press and Journal. *Hundreds of locals take to Fort William streets in protest as new hospital faces more delays*. URL: https://www.pressandjournal.co.uk/fp/news/highlands-islands/6404626/fort-william-belford-protest-700-attend/.

[9] 31 March 2020 Public Health Scotland. "Cancelled planned operations". In: (). URL: https://publichealthscotland.scot/publications/cancelled-planned-operations/cancelled-planned-operations-month-ending-31-march-2020/.

[10] Public Health Scotland. *Cancelled Planned Operation Datasets*. Last accessed 29 March 2024. URL: https://www.opendata.nhs.scot/dataset/cancelled-planned-operations.

[11] Public Health Scotland. *Cancelled planned operations Month Ending 31 January 2024*. Last accessed 29 march 2024. URL: https://publichealthscotland.scot/publications/cancelled-planned-operations/cancelled-planned-operations-month-ending-31-january-2024/#:~:text=During%20the%20latest%20month%2C%2010,(9.5%25%2C%202%2C750).

[12] Public Health Scotland. *Changes in Hospital codes*. Last accessed 29 March 2024. URL: https://publichealthscotland.scot/publications/acute-hospital-activity-and-nhs-beds-information-quarterly/acute-hospital-activity-and-nhs-beds-information-quarterly-quarter-ending-30-june-2022/official-data-release-information/.

[13] Public Health Scotland. *NHS Scotland Hospital Codes*. Last accessed 29 March 2024. URL: https://www.opendata.nhs.scot/dataset/hospital-codes.

[14] Public Health Scotland. *Viral Respiratory Diseases (Including Influenza and COVID-19) Data in Scotland*. Last accessed 29 March 2024. URL: https://www.opendata.nhs.scot/dataset/viral-respiratory-diseases-including-influenza-and-covid-19-data-in-scotland.