A step closer to understanding missed appointments at the NHS in England

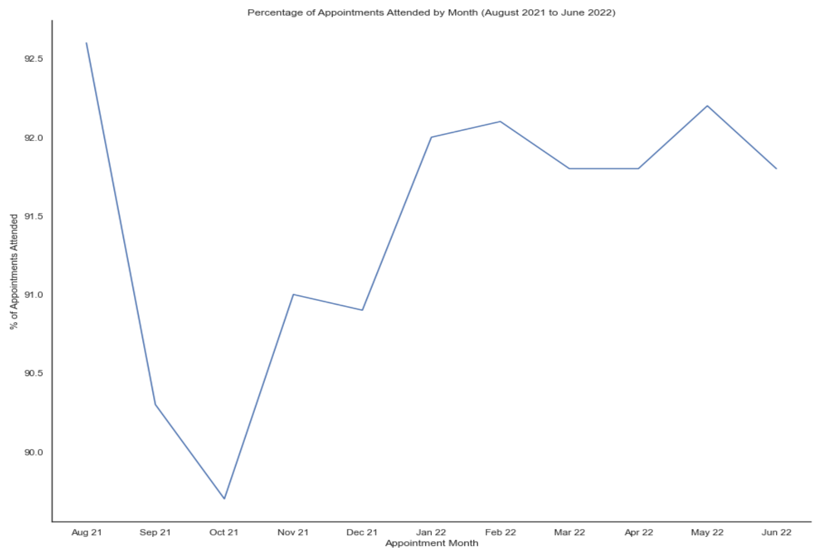
Context and purpose

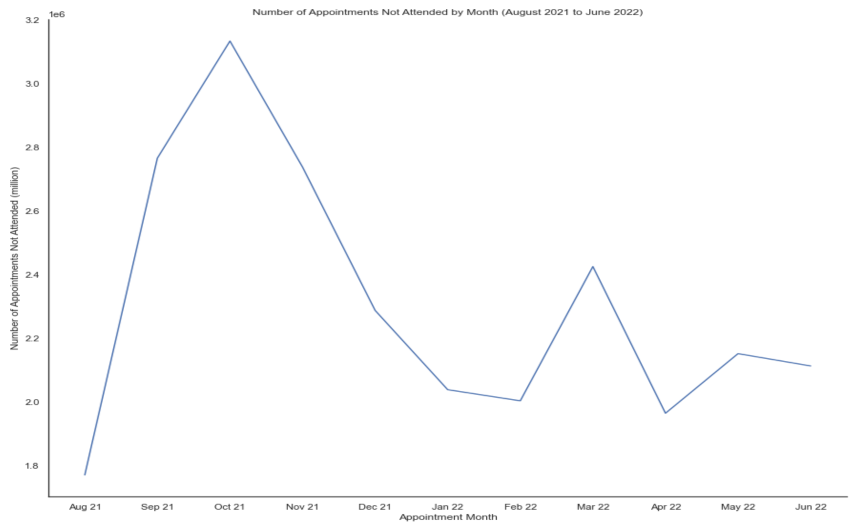
Reducing the number of appointments booked but not attended at the NHS in England is a worthwhile goal for the cost-effective operation of the publicly funded service, but also for social and public health reasons. The two graphs below show the size of the problem and opportunity. At a high level, the goal is to flatten and push the line down in the first graph showing the number of appointments not attended and to flatten and push the line up in the second graph which shows the percentage of appointments attended.

Over the eleven months represented in the two graphs on this page the mean and median percentage of appointments attended was approximately 91.5% and 92% respectively. Standard deviation is approximately 0.9%.

NHS England posed the following questions as key:

- Has there been adequate staff and capacity in the networks?
- What was the actual utilisation of resources?





A reformulation of these questions to make them more precise and answerable with the data we have could be:

> *What is the relationship between the number of appointments available at NHS England (approximately 1,200,000 per day) and the number of appointments historically recorded in our data?*

While the questions posed by NHS England are relevant, there probably other questions that must be addressed which may be even more directly related and therefore important for understanding and addressing the issue of missed appointments. Such a question could be:

> *Under what circumstances are appointments more or less likely to be missed?*

Analytical approach

We started with writing in our Jupyter Notebook our initial understanding of the challenge and opportunity relating to missed appointments at NHS England. Then we focused on getting familiar with the data publicly available from the NHS and working through an iterative approach updating our initial understanding of the challenge. What follows in this section is a high-level summary of our analytical approach. For more details, please see the Jupyter Notebook in our GitHub repo[1]. Also, please see the appendix to this document for details of the data files we used.

We started with importing the relevant Python libraries for our analysis: numpy, pandas, datetime, matplotlib and seaborn. All these are relevant and used in our analysis. We read the metadata text and loaded each of the data files as separate DataFrames and viewed the data, confirming and updating our initial understanding, including the unique values in each column. Where necessary we researched information online to enrich our understanding, for example regarding how the NHS is organised into Integrated Care Boards (ICBs) which is the basis for the locations dimension in our data. Please see the references at the end of this document for online sources we used to enrich our understanding and confirm our data is complete (for example all 42 ICBs of NHS England seem to be included in our data).

We checked for missing data using two methods on each data file:

1) Looked at the 'Non-Null Count' when applying the `.info()` method. For example,

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 137793 entries, 0 to 137792
Data columns (total 8 columns):
 #   Column                     Non-Null Count   Dtype
---  ------                     --------------   -----
 0   sub_icb_location_code      137793 non-null  object
 1   sub_icb_location_ons_code  137793 non-null  object
 2   sub_icb_location_name      137793 non-null  object
 3   icb_ons_code               137793 non-null  object
 4   region_ons_code            137793 non-null  object
 5   appointment_date           137793 non-null  object
 6   actual_duration            137793 non-null  object
 7   count_of_appointments      137793 non-null  int64
dtypes: int64(1), object(7)
memory usage: 8.4+ MB
```

2) Applied the `.dropna()` method and checked that no rows were dropped.

We paid special attention to the data quality issues noted in the metadata text and considered their possible implications and limitations for our analysis.

We used the logical operators to filter our data and the `.groupby()` method to combine multiple data into a single result. Most of our analysis involved looking at trends over time and we therefore used lineplots to visualise those trends. We also considered the effect of outliers in our visualisations and in some instances adjusted our graphs for the purposes of communicating our insights more effectively.

A key decision we made was regarding the period to use for our analysis. While we had data for 30 months from January 2020 to June 2022, we limited our analysis to the eleven months from August 2021 to June 2022. This decision was because the period prior to August 2021 was possibly too heavily affected by the Coronavirus lockdowns in England. England completed its phased exit from lockdown in July 2021.

Another key decision we made was regarding how much of our analysis would be focused on data aggregated by location. While such analysis may yield insights into factors that influence the rate of missed appointments based on differences across locations, we determined that at this stage our priority should be on trends over time and across non-location-specific factors such as service setting, context type, healthcare professional type, and appointment mode. The main reason for this decision is that in our opinion further analysis is needed to identify and account for the effect of differences in how relevant data is captured at different locations as this may have an impact on the reliability of comparisons.

Finally, before working with data from Twitter to identify the most popular hashtags relating to healthcare we considered the relevant ethical and code of conduct considerations.

[1] As we progressed with our analysis, we wrote about our approach in our Jupyter Notebook and in the GitHub repo README file. Please consider the Jupyter Notebook as the golden source – most up to date.

Visualisation and insights

Our initial analysis revealed insights and confirmed information we had from reading the metadata text regarding basic elements relevant to appointments at the NHS, such as the number and description of appointment modes, healthcare professionals, etc.. Please refer to the appendix and the Jupyter Notebook for details.

In addition to the two visualisations on page 1 of this report which show the rate at which appointments were attended and missed in the eleven months to 30 June 2022, we propose selecting the following for communicating with our stakeholders at the NHS to discuss and agree what further questions could be addressed through data analysis.

Utilisation by month

Our data analysis suggests no reason to start looking at increasing staff levels, given that in the eleven months to 30 June 2022 utilisation was between 66.3% and 84.5% (mean, 74.8%).

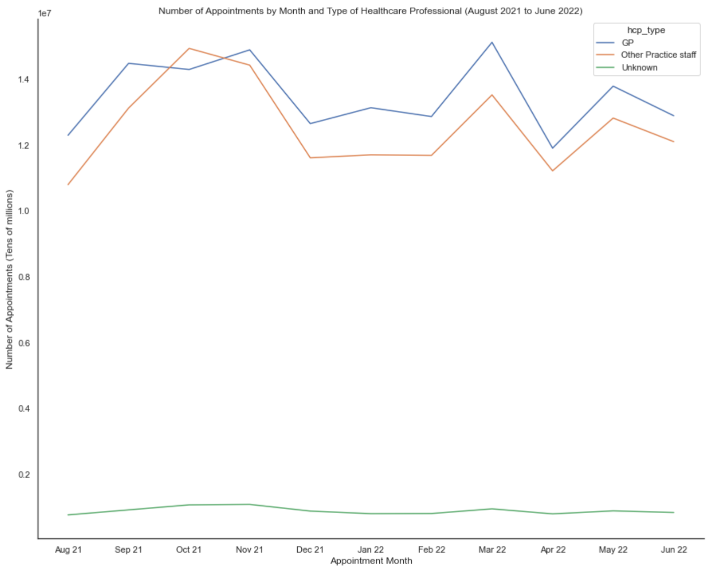
However, the answer might be different if we looked at utilisation by type of healthcare professional.

The graph for the number of appointments by month, as expected, has the same shape as the utilisation graph on the right.

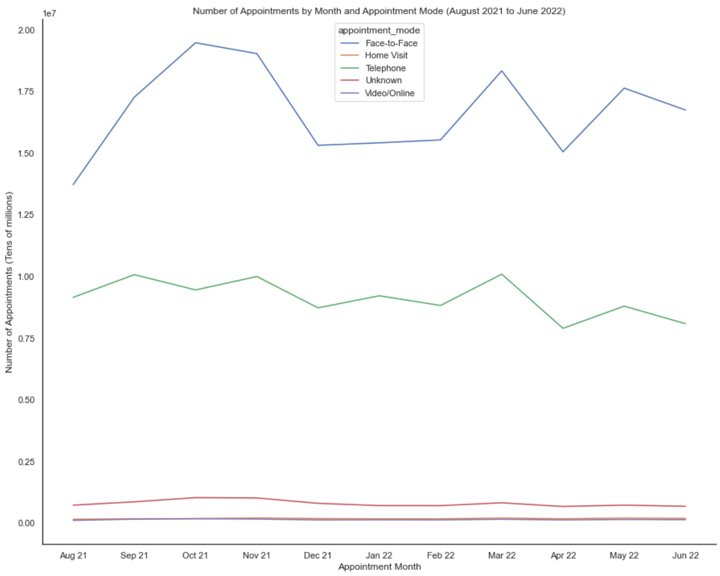
Appointments by month and type of healthcare professional

This graph shows the relative allocation of appointments between the two types of healthcare professionals.

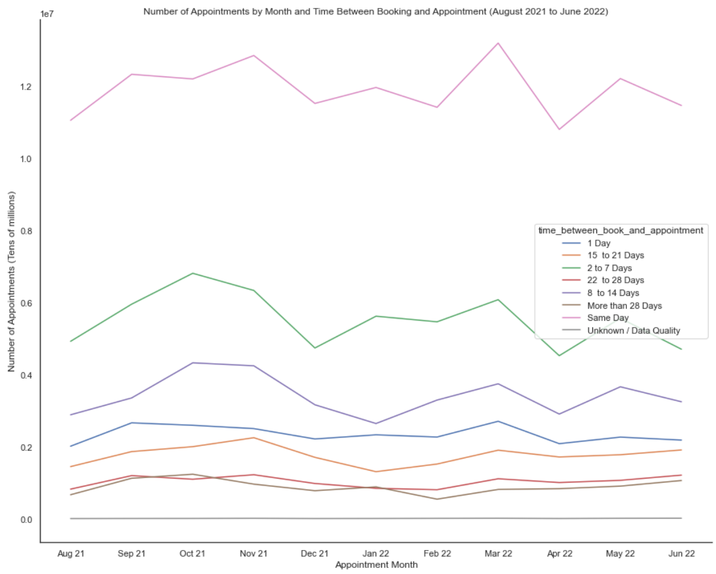
The trend over time, as expected, seems to match the trend for utilisation and appointments by month.



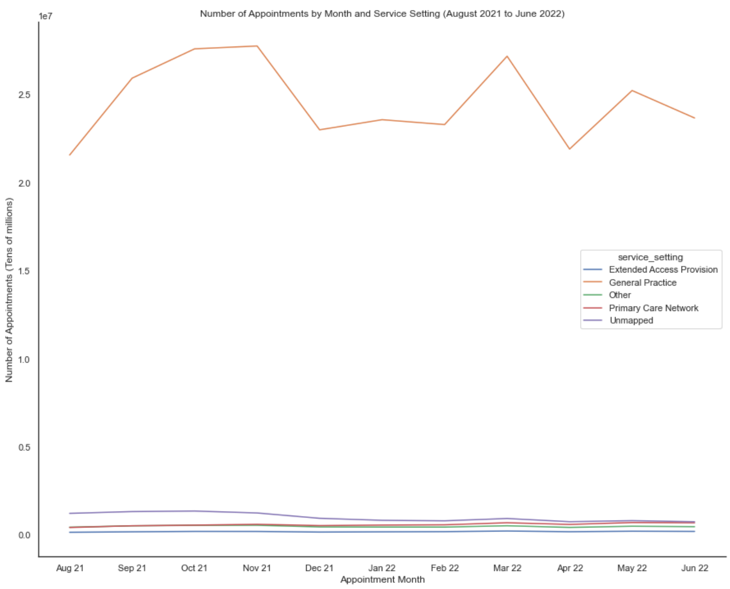
Appointments by month and mode
This graph when viewed together with the graphs on utilisation and total number of appointments by month shows that in the busiest months, utilisation and the number of appointments were driven up mainly by face-to-face appointments.



Appointments by month and time between booking and appointment
There doesn't seem to be a significant change over time in the distribution of appointments across the different categories. In addition, the movements in each category over time seem to be consistent with the movements in the other graphs above.



Appointments by month and setting
This graph shows that the GP setting is by far the most frequent. The pattern across time is, as expected, consistent with the pattern in the graphs above.



In addition to the above, it will be helpful to present to our stakeholders the barplot showing the frequency of Twitter hashtags, and the two boxplot graphs showing the variability around the median number of appointments by month and service setting.

Twitter data
Twitter data is an example of external data that can be insightful in understanding the behaviours and attitudes of the public towards the NHS and therefore enrich our understanding of issues and inform our recommendations and policy decisions. When using such data, it's important to be aware of its limitations, including the possibility for selection bias and polarised views that may not be representative of reality. Being aware of these limitations, Twitter and data from other similar external sources can be a powerful tool for understanding the people who use the NHS services.

Our analysis of Twitter hashtags relating to healthcare shows a neutral attitude. Further exploration, for example applying sentiment analysis not only on hashtags but also on the text within Tweets can reveal insightful trends and information about the public's reaction to events, such as policy changes as well as about the public's experience in their day-to-day interactions with the NHS services.

Patterns and predictions
In summary, our analysis shows the following patterns which may be worth validating first internally within our team, and then discussing with our stakeholders:

- NHS England appointments are missed on average at a rate of approximately 8%.
- One possible explanation for the spike in missed appointments in September and October 2021 could be the public's adjustment to the new norm after the end of Coronavirus lockdowns in England.
- Our data analysis does not show a need to increase the NHS appointments capacity given average utilisation at approximately 75%. However, further analysis may reveal pressure in particular locations, healthcare professionals, settings, and appointment modes.
- Appointments by month over the eleven months in our analysis show a consistent pattern across all the dimensions we used to group our data by.

In addition, in our presentation to our stakeholders we could consider recommending:

- Reformulation of the key questions and goals for further analysis.
- Further analysis using the current data – for example, to see the percentage of missed appointments by appointment mode, setting, and healthcare professional.
- Ways of working together in closer collaboration, perhaps adopting an agile / scrum framework to enable us to make faster interpretations of patterns we see in our data analysis and refine, change, and add to our analysis questions in real time.
- The design of AB testing to confirm or dismiss hypotheses over factors that may influence the likelihood for people to miss their appointments.
- Ways to overcome some of the data quality issues arising mainly from the fact that the data for policy decisions comes from systems that were designed for the day-to-day management of appointments.

Appendix

The data files we used in our analysis are:

metadata_nhs.txt	actual_duration.csv	appointments_regional.csv
national_categories.slsx	tweets.csv	

We'd be delighted to share a link to these files to anyone who might want to replicate and / or build on our analysis. As the first three data files overlap in their content, the table below might help understand what data is in each.

#	WWWWH	Name	Values	ad	ar	nc[2]
1	where	icb_ons_code	codes representing each of 42 icb locations	1	1	1
2	where	sub_icb_location_code	codes representing each of 106 sub-icb locations	1		
3	where	sub_icb_location_ons_code	codes representing each of 106 sub-icb locations	1		
4	where	sub_icb_location_name	names of each of 106 sub-icb locations	1		1
5	where	region_ons_code	codes representing each of 7 regions	1		
6	when	appointment_date	dates in ad are from 01.12.2020 to 30.06.2022 (seven months) dates in nc are from 01.08.2020 to 30.06.2022 (eleven months)	1		1
7	when	appointment_month	months in ar are from Jan 2020 to Jun 2022 (30 months) months in nc are from Aug 2021 to Jun 2022 (eleven months)		1	1
8	how_long	actual_duration	['1-5 Minutes', '6-10 Minutes', '11-15 Minutes', '16-20 Minutes', '21-30 Minutes', '31-60 Minutes', 'Unknown / Data Quality'] (7)	1		
9	who	hcp_type	[‘GP’, ‘Other Practice staff’, ‘Unknown’] (3)		1	
10	how	appointment_mode	['Face-to-Face', 'Home Visit', 'Telephone', 'Video/Online', 'Unknown'] (5)		1	
11	how_long	time_between_book_and_appointment	['Same Day', '1 Day', '2 to 7 Days', '8 to 14 Days', '15 to 21 Days', '22 to 28 Days', 'More than 28 Days', 'Unknown / Data Quality'] (8)		1	
12	how	service_setting	['General Practice', 'Extended Access Provision', 'Primary Care Network', 'Other', 'Unmapped'] (5)			1
13	what	context_type	['Care Related Encounter', 'Unmapped', 'Inconsistent Mapping'] (3)			1
14	what	national_category	['Patient contact during Care Home Round', 'Planned Clinics', 'Home Visit', 'General Consultation Acute', 'Structured Medication Review', 'Care Home Visit', 'Unmapped', 'Clinical Triage', 'Planned Clinical Procedure', 'Inconsistent Mapping', 'Care Home Needs Assessment & Personalised Care and Support Planning', 'General Consultation Routine', 'Service provided by organisation external to the practice', 'Unplanned Clinical Activity', 'Social Prescribing Service', 'Non-contractual chargeable work', 'Group Consultation and Group Education', 'Walk-in'] (18)			1
15	outcome_ variable	appointment status	['Attended', 'DNA', 'Unknown']		1	
16	how_many	count_of_appointments		1	1	1
				8	**7**	**8**

[2] ‘ad’, ‘ar’ and ‘nc’ are abbreviations for each of the first three data files.

References

https://digital.nhs.uk/services/organisation-data-service/integrated-care-boards

https://www.england.nhs.uk/integratedcare/integrated-care-in-your-area/

https://practiceindex.co.uk/gp/blog/news-prime-minister-candidate-wants-to-charge-no-show-patients/

https://digital.nhs.uk/data-and-information/publications/statistical/appointments-in-general-practice/appointments-in-general-practice-supporting-information#guide-to-data-files

https://www.businessnewsdaily.com/7488-twitter-for-business.html

https://developer.twitter.com/en/community/code-of-conduct#:~:text=Expected%20behavior,situation%20or%20someone%20in%20distress.

https://commonslibrary.parliament.uk/research-briefings/cbp-9068/