Regional variation of missed appointments in General Practice of Greater Manchester from 2019 to 2021

A dissertation submitted to The University of Manchester for the degree of MSc Bioinformatics and Systems Biology in the Faculty of Health

YEAR OF SUBMISSION:

2023

STUDENT ID:

11133575

SCHOOL:

UNIVERSITY OF MANCHESTER

CONTENTS:

Word count: 5,956

Table of Contents

List of Tables

List of Figures

ABSTRACT

Background:

Missed appointments have become an inevitable event placing a significant burden on healthcare by affecting the quality of primary care, and revenue and posing a risk to patient's health. I have investigated the regional variability of missed appointments in Greater Manchester for face-to-face appointments and other appointment types from 2019 to 2021.

Methods:

This was an observational study using data collected from 2019 to 2021 from the NHS database to analyse the rate of missed appointments in Greater Manchester for both face-to-face and other appointment types. Descriptive statistics and Time series box plots were used to provide a statistical summary of regional variability and visually represent this summary over time using data analyses software RStudio and Stata. Using Stata, Inferential analyses were conducted using Poisson regression to model the incidence rates (IRR) of missed appointments over time for both face-to-face and other appointment types.

Results:

The rate of missed appointments decreased from 2019 (5.39%) to 2021 (4.23%) for face-to-face appointments with a higher distribution of DNA rates observed in April to June 2019, compared to the same months in the following years. A similar trend was noticed in other appointment types where a wider dispersion was observed in months January, and May to August (2019). Although both types of appointments noticed a decrease in the incidence rate of missed appointments over time, other appointment types observed the most reduction in missed appointments compared to face-to-face appointments. A wide variability was observed in both of these appointment types, with months April to June particularly showed the lowest incidence rates of DNA cases in face-to-face and other appointment types, before a significant increase was observed between months October and December.

Conclusion:

My study indicates that there was a wide regional variability between 2019 to 2021 in Greater Manchester, with 2019 showing the highest level of DNAs, with potential reasons being health inequalities and deprivation. The findings highlighted that more patients attended other appointment types (telephone, video calls, online) compared to face-to-face appointments due to lack of transport issues and long waiting times. Overall, more research and strategies need to be revised to reduce missing appointments.

DECLARATION

I declare that the dissertation titled ‘ Regional variation of missed appointments in General Practice of Greater Manchester from 2019 to 2021 ‘ has been composed solely by myself, under the guidance of my supervisor Evan Kontopantelis. No portion of this work referred to in the dissertation has been submitted in support of another degree or qualification of this or any other university or other institute of learning.

INTELLECTUAL PROPERTY STATEMENT:

ACKNOWLEDGEMENTS

I would like to give my gratitude to my supervisor Evan Kontopantelis as the completion of this project would not have been possible without his guidance. He is an excellent supervisor who is very dedicated in his role. This has been a very interesting project and I have learnt valuable skills under Evan's guidance. Thank you.

INTRODUCTION:

Missed appointments known as Did Not Attend cases (DNA) are a common problem across the healthcare sector. These cases are reported when patients' who scheduled appointments, failed to attend without providing prior notice to their healthcare providers, (Marbouh et al, 2020). They pose an issue to General Practitioners (GP) led primary sectors as they result in a decreased clinical capacity, lower revenue and are a misuse of resources provided, (Alawadhi et al, 2021), (Marbouh et al, 2020), (Lee et al, 2019). Approximately 7.2 million GP appointments are missed, which indefinitely affects their workflow as 1.2 million GP hours are wasted and a loss of £216 million within the National Health Services (NHS) is experienced, (Margham et al, 2021). To limit this expenditure the NHS requests patients to simply cancel their appointments beforehand, (Parsons et al, 2023), (Martin et al, 2005), (Margham et al, 2021).

This comes at a time of constricted funding as an insufficient share of NHS expenditure is allocated to the GP sector with a reduction of 20% as of late, (Fisher et al, 2017). As a result, there's a significant shortage of GPs in primary care, (Owen et al, 2019). This is a challenging situation as there's a growing demand for GP services due to their crucial role as first consultants in primary care to diagnose and treat ailments, (Grol et al, 2018). These unattended cases act as an obstacle to achieving NHS's primary aim of reducing waiting time to less than 48 hours, to minimise the demand for quicker appointment allotments, (George et al, 2003), (Martin et al, 2005). Considering this, the high number of DNA cases amplifies the pressure on GPs by wasting appointments that could be allocated to patients presenting with more severe and urgent conditions, (Parsons et al, 2021).

According to research the probability of missing an appointment increases among patients who had already missed at least one appointment within 12 months, affecting both the GP and reception team alike, (Waller et al, 2000), (Kaplan-Lewis et al, 2013). The reception teams managing appointments are the most frustrated by missed appointments as they find it challenging to fit patients into sparse appointment slots, (Martin et al, 2005). In comparison, GPs are must less overset by these unattended cases and utilise this time to catch up on other works, (Margham et al, 2021). Rather, they are only disconcerted by the potential implications of these missed appointments. Continuous delays in the diagnosis of potential illnesses lead to deferrals in treatment and result in serious consequences regarding the patient's health, (Marbouh et al, 2020), (Kaplan-Lewis et al, 2013).

To reduce the rate of DNAs it is essential to understand the factors leading to this and how they influence attendance. Although recent studies characterise the impact of missed appointments in healthcare, there is still limited information regarding the patients who miss appointments, (Marbouh et al, 2020). However, current case studies have observed the most prevalent cases of missed appointments are known to be primarily associated with patients suspected of having multimorbidity, living in complex social settings, with unequal healthcare engagement, and socioeconomic deprivation, (McQueenie et al, 2019), (Marbouh et al, 2020), (Husain-Gambles et al, 2014), (Ellis et al, 2017). These individuals are likely to be ethnic minorities, with several studies reporting that some non-English speakers are more likely to miss appointments due to language barriers affecting their ability to cancel appointments, (Williamson et al, 2014), (Nguyen et al, 2011).

Other common factors in missed appointments include age, sex, transport difficulties, and staff-patient relationships which lead to miscommunication during appointment booking, (Nielsen et al, 2008), (Waller et al, 2000), (Martin et al, 2005). Other reasons were family and work conflicts, as well as long waiting times and fear of medical settings, (Martin et al, 2012), (Margham et al, 2021).
It's still unclear whether these factors are indefinitely associated with patients who miss appointments and how to address them, (Williamson et al, 2017). General practices have applied their strategies such as set number of reminders such as SMS messages and telephone calls, (Parsons et al, 2023), (Hashim et al, 2001). Despite this, patients from underprivileged backgrounds are unlikely to access the internet and technologies for contact. Other approaches discussed included charging patients for missing appointments, which disregards the socioeconomic complexities at play, (Parsons et al, 2023). There is little agreement as to what approach is best implemented.

This research aims to: (a) Investigate regional variability of missed appointments within Greater Manchester using their main Clinical Commissioning Groups (CCGs) over the years 2019 to 2021.
(b) Examine the differences in the rate of missed appointments for face-to-face and other types of appointments (c) Understand the socioeconomic complexities contributing to these missed appointments.
The CCG-level information utilised in this study was collected using data published in the NHS database. These databases were cleaned and analysed using descriptive statistics and inferential analyses such as Poisson regressing through data analysis software known as RStudio and Stata. The prevalence of low engagement in primary care can be addressed by studying regional variability in missed appointments, (Cashman et al, 2004). This could allow us to implement more effective interventions to improve the efficiency of the NHS and reduce health inequalities leading associated with missed appointments, (Cashman et al, 2004), (Mitchell et al, 2007).

METHODS:

1.1 Collecting Datasets

I investigated the level of missed appointments in GP appointments within the region of Greater Manchester. I collected datasets containing all months with relevant appointment data on missed appointments from years 2019 to 2022. To do this I used NHS Digital (https://digital.nhs.uk/data-and-information/publications/statistical/appointments-in-general-practice), an organised body within NHS England containing all medical records of different regions, (Morris, 2023), (Honeyman et al, 2016).

Only regional datasets were collected as they held information regarding the total number of appointments, appointment type, and attendance records for each month and year within different regions of England. Starting with April 2019, I downloaded the folder 'General Practice-April 2019:Regional' in ZIP format. I extracted this folder using the 'Extract All' option to access the information in this folder for cleaning. The regional ZIP folders for all years 2019, 2020, 2021, and 2022 were downloaded and extracted from its ZIP format. All timepoints beyond December 2022 were excluded as they were still incomplete, (Graham et al, 2009). I created four folders named '2019', '2020', '2021', and '2022' to store these regional datasets, so they were located in the same directory, to simplify the data cleaning process.

1.2 Clean and merge datasets

To ensure that the data was good quality, and errors don't affect the study's accuracy, I prepared the datasets for analysis by removing unnecessary columns and duplicated data, (Love et al, 2021), (Jr et al, 2021), (Van den Broeck et al, 2005). I downloaded RStudio (https://posit.co/download/rstudio-desktop/), a development software, and R 3.3.0+ (https://cran.rstudio.com/) to use the R programming language to clean the datasets, (Grömping, 2015)

I set the directory by loading datasets in the folders '2019', '2020, '2021', and '2022' into RStudio's environment, by selecting the 'Session', 'Set Working Directory' and 'Set Directory' tabs. To read these files, I installed the package 'readr' and ran this package into the console using the library function. I used the list.files function, as a character vector that lists all files from '2019' inside it, and merged these files into one dataset called 'D1', (Gruber, 2022), (Chan et al, 2018). This was done for all the files in '2020', '2021' and '2022' and the multimerger function used for merging high quantities of datasets, merged all these files into one dataset called 'FinalData' by setting auxiliary parameters to merge, (Cookson et al, 2011).

Package 'dplyr' was ran using the library function to access the subset function to remove unnecessary columns like 'CCG_ONS_CODE', 'STP_ONS', 'STPNM', 'REGION_ONS_CODE', 'REGION_NAME', 'TIME_BETWEEN_BOOK_AND_APPT' in D1 dataset, (Wickham et al, 2023), (Singh et al, 2019). Duplicated rows were removed using distinct function which only keeps unique rows in the dataframe. The write.csv command function saves and exports this database into a csv format in the directory where the folder 'Data' was located.

1.3 Organise the Dataset

I examined whether the level of DNAs varied across Greater Manchester over time by focusing on two types of GP appointments; Face-to-Face and Other appointment types. The subset function was used to differentiate 'FinalData' dataset into rows corresponding only to 'GP' and 'Face-to-Face' using columns 'hcptp' and 'apptmode' and named it 'GP/F2F'.

In the 'FinalData' database, I renamed the variables 'Home Visit', 'Telephone', 'Unknown', and 'Video/Online' into the variable 'Other'. I used the subset() function to only return rows corresponding to 'GP' and 'Other' variables.

Greater Manchester was the region of interest in this study, so I removed all other regions from GP/Other and GP/F2F datasets except ‘Bolton’, ‘Rochdale’, ‘Oldham’, ‘Wigan’, ‘Salford’, ‘Trafford’, ‘Manchester’, ‘Stockport’, ‘Tameside’, which were the CCG regions of Greater Manchester, (Dowle et al, 2019), (Wendt et al, 2022). I used the data.table and grep function to search for matches of characters in a given string, (Schork et al, 2022). I searched 'Bolton' along with other region names partially in GP/F2F, to retrieve all rows containing the full string name 'NHS Bolton CCG' and other region names. I applied the grep function to retrieve only rows with all 9 CCG regions in GP/Other using their partial names.

I examined the dataset to ensure that it had complete data, as otherwise statistical power of this study would be reduced, (Kang et al, 2013). To do this I searched for potential missing values in any of the timepoints using is.na function on GP/F2F and GP/Other datasets. I removed the year 2022 from both datasets using the operator '%in%' and dataset names as the data was incomplete compared to 2019, 2020, and 2021. To summarise DNA cases by their timepoints I aggregated the number of missed appointments in the 'countappt' column using summarise_at. I grouped and organised this aggregation by columns ccg_code, ccg_name, year, month, hcptp, and apptstatus using the group_by function.

I studied the number of DNA, Attended, and Unknown cases in depth by converting both 'GP/F2F' and 'GP/Other' databases from long format to wide format so that these variables were columns of their own with their related count data. The pivor_wider function using packages ‘usethis', 'devtools', and 'tidyr' tidied these datasets into this wide format individually, (Wickham, Bryan & Barrett, 2022), (Wickham et al, 2022), (Wendt et la, 2022), (Wickham et al, 2023). Packages ‘tidyr’ converted the datasets into wide formats, and packages ‘devtools’ and ‘usethis' simplified and automated repetitive tasks to do this conversion, (Garmendia et al, 2020).

1.4 Calculate Percentages

I investigated to see whether there were differences in the level of DNAs for different appointment types within the timepoints. To do this I calculated the percentages of DNAs for Face-to-Face and Other GP appointment types.

Using the cbind and rowSums function I combined and calculated the sum of every row in columns 'Attended', 'DNA', and 'Unknown' in GP/F2F dataset into a new column called 'Total'. I used the mutate function to find the percentages of DNAs in GP/F2F by dividing the DNA column by the Total column and calculating these percentages into a new Ayele column called 'DNA%'. I applied this mutate function and calculated the percentages of DNA into a 'DNA%' column for dataset GP/Other.

1.5 Descriptive Statistics

I examined the relationship between DNA, Attended, and Unknown cases from 2019 to 2021 within Greater Manchester, using summary statistics, (Kaur et al, 2018), (Kaliyadan et al, 2019). Using the tapply function, I calculated the summary statistics for the count and percentage data of these cases in GP/F2F and GP/Other databases, (Divisi et al, 2017). (Böhringer et al, 2013). Starting with GP/F2F dataset I used the tapply function along with columns 'year' and command 'summary' to get descriptive statistics such as mean, median, mode, range, and interquartile ranges. I applied the same process to GP/Other dataset to receive all the descriptive statistics for both count and percentage data.

Excel (https://www.microsoft.com/en-us/microsoft-365/excel) a spreadsheet software for data visualisation and analysis was used to present the descriptive statistics in a table format, (Al-Ubaydli et al, 2005). Years 2019, 2020, and 2021 were aligned vertically under the 'Year' column, and descriptive statistic values ‘Mean, ‘Standard deviation (std)’, ‘Minimum’, ‘Maximum’, ‘25th‘ and ‘75th‘ percentage quartile and ‘median’ were arranged horizontally under their column names, (Divisi et al, 2017).

1.6 Inferential analyses

I examined the outcome of this study which was the number of missed appointments within each timepoint. To do this I used the Poisson regression model to generate two models, one for GP and Face-to-Face and one for GP and Other appointment types, and both were adjusted for exposure which was time.

Stata (https://download.stata.com/download/), a statistical program for data analysis, was used to create these regression models using count data from GP/F2F and GP/Other datasets, (Boston et al, 1970). Both datasets were uploaded individually to stata using options 'File', 'Import', and 'text data(delimited, * .csv)' in this order. Both GP/F2F and GP/Other datasets adjusted to show time variable using the command gen time=time. I wanted to show all the months between 2019 to 2021, as 1 to 36, for an easier presentation so I replaced months between 2020 and 2021 to start from 12 and 24 using the replace function. I generated Poisson regression models for both GP/F2F and GP/Other datasets separately using the Poisson function, along with the variables ‘dna’, ‘time’, and ‘i.month’. The exposure was set to ‘total’ with irr stated in the function to get the incidence rate values of the measure of DNA cases.

1.7 Graphs

To visually represent the rate of DNAs in Face-to-Face and Other appointments clearly, I used box plots for both GP/F2F and GP/Other datasets. To do this I selected the following options 'Graphics', 'Box Plot' and adjusted several parameters. The 'Main' parameters orientation was changed to 'vertical' and 'DNA%' was chosen from the 'Variable' option, to display the percentage of DNAs in the boxplots for both appointment types. 'Group 1' and variable 'time' were selected from the 'Categories' parameter and the Y-axis parameter was changed in both datasets to 'Time' to show the time variable ranging from 1-36 for 2019 to 2021. The 'Title' parameter for GP/F2F was changed to 'Change in % of DNA cases over time for GP & Face-to-Face appointments', and 'Change in % of DNA cases over time for GP & Other appointments' for GP/Other dataset.

RESULTS:

2.1 Summary statistics:

Examining datasets from 2019 to 2021, I investigated the regional variability in missed appointments of face-to-face and other appointment types within Greater Manchester over these years. I used summary statistics for this because it's particularly useful in providing an understandable summary of the central tendency over time in Greater Manchester, to compare this regional variability. I analysed the variability in attended and unknown cases relative to any fluctuations in missed appointments over the years.

For face-to-face appointments, the year 2019 observed the highest rate of missed appointments with a mean of 5.39% (3,034 appointments, std: 0.77%) and a median of 5.3% (2,470 appointments). In contrast, this level decreased noticeably in 2020 (Table 1) with a mean of 3.98% (1,241 appointments, std: 1.28%) and a median of 3.94% (706 appointments), before observing a marginal increase in 2021 (mean:4.23%, median:4%). This suggests a wide variability in the rate of DNAs in Greater Manchester, due to the considerable decrease observed from 2019 to 2021.

Interestingly, the level of unknown cases observed a notable increase from 2019 onwards (mean: 2.76%, median: 2.67%, std: 0.82%), where the year 2020 experienced the highest level of unknown cases (mean: 4.49%, median: 3.72%, std: 2.63%), before dropping in 2021 (mean: 4.07%, median: 3.54%, std: 2.33%). This indicated that a number of these unknown cases could potentially be DNAs or Attended cases which failed to be recorded. This was supported by the rate of attended cases where from 2019 (mean: 91.85%, median: 92.08%) to 2020 (mean: 91.54%, median: 91.97%), there was a slight decrease in the rate of missed appointments. This further showed the effect of unknown cases on this trend, as the rate of DNAs decreased, the measure of attended cases were expected to increase proportionally to DNAs over these years but failed to do so. This disproportion between DNA and Attended cases in over time highlights the significance of the Unknown cases.

Other appointment types (videocalls, online, phone calls, house appointments) showed a wide variability in DNA rates from 2019 to 2021 with a decreasing trend observed over time.

The highest rate of DNAs was observed in 2019, with a mean of 3.43% (532 appointments, std: 1.2%) and a median of 3.14%. This rate decreased considerably in the following years, with 2021 having the lowest rate of missed appointments with a mean of 2.5% (1058 appointments, std: 0.63%) and a median of 2.36% (754 appointments). This trend in DNAs was proportional to the changes observed in the rate of attended appointments over time in Greater Manchester, where 2019 observed the highest rate of attendance (mean: 89.91%, median: 90.53%) before increasing notably in 2020 (mean: 93.67%, median: 94.46%).

The levels of unknown cases showed a significant effect on the proportionate relationship between DNA and attended cases. Face-to-Face appointments showed an increasing level of unknown cases, while other appointments observed an opposite trend with decreasing unknown cases. The highest level of unknown cases was observed in 2019 (mean: 6.76%, median: 6.58%), before decreasing substantially in 2020 (mean: 3.69%, median: 2.93%). This trend continued, with year 2021 exhibiting the lowest level of unknown cases (mean: 3.57%, median: 2.74%). This trend in unknown cases suggests a more accurate representation of the DNA and attended cases over time, potentially due to improved accuracy in appointment tracking.

Descriptive Statistics: GP and Face-to-Face appointments 2019-2021										
Count of cases						Percentage of cases				
	DNA					DNA%				
Year	mean	std	25%	50%	75%	mean	std	25%	50%	75%
2019	3034.44	1981.65	2136.75	2470	2892.25	5.39	0.77	4.91	5.30	5.65
2020	1240.83	1362.86	447.75	705.5	1662.75	3.98	1.28	2.97	3.94	4.80
2021	1326.29	1167.89	679.5	888.5	1353.25	4.23	1.48	3.49	4.00	4.74
	Attended					Attended%				
Year	mean	std	25%	50%	75%	mean	std	25%	50%	75%
2019	49328.63	22356.67	38851.75	43166	53753	91.85	1.04	91.45	92.08	92.53
2020	25230.44	19546.44	12728.5	18206.5	32851	91.54	2.77	89.97	91.97	93.39
2021	26969.77	16194.82	17412.5	21636.5	28265	91.70	3.20	90.5	92.09	93.93
	Unknown					Unknown%				
Year	mean	std	25%	50%	75%	mean	std	25%	50%	75%
2019	1558.73	1009.01	947	1282	1687.25	2.76	0.82	2.18	2.67	3.24
2020	1123.12	898.78	475.75	869	1617.75	4.49	2.63	2.50	3.72	5.90
2021	1285.1	1167.74	491.25	812.5	1699.5	4.07	2.33	2.46	3.54	4.81
For GP and Other types of appointments between 2019-2021:										
	DNA					DNA%				
year	mean	std	25%	50%	75%	mean	std	25%	50%	75%
2019	532.07	448.46	225.75	342.5	779.75	3.43	1.20	2.52	3.14	3.98
2020	859.62	625.97	499.5	705	952	2.65	0.72	2.17	2.52	3.06
2021	1058.19	701.94	580	754	1206.75	2.5	0.63	2.09	2.36	2.99
	Attended					Attended%				
year	mean	std	25%	50%	75%	mean	std	25%	50%	75%
2019	12305.4	6318.51	8124	9322	18161.75	89.81	2.13	88.81	90.53	91.26
2020	30095.71	16730.97	21502.25	27699	34642.75	93.67	2.34	91.70	94.46	95.38
2021	37573.66	16981.87	26343.25	32336	43880.75	93.94	1.90	93.18	94.13	95.61
	Unknown					Unknown%				
year	mean	std	25%	50%	75%	mean	std	25%	50%	75%
2019	909.81	544.72	562.25	653	1162.75	6.76	1.84	5.69	6.58	7.64
2020	1013.57	539.93	607.75	826	1370.25	3.69	1.97	2.07	2.93	5.52
2021	1299.35	631.52	785.25	1060	1770	3.57	1.95	2.20	2.74	4.12

Table 1 Descriptive Statistics (Mean, Median, std, 25%-75% IQR) for count and percentage data of missed appointments, attended appointments, and unknown appointments from 2019 to 2021

2.2 Time series analysis:

The research aimed to examine the seasonal distribution of missed appointments in both face-to-face and other appointment types over time. To achieve this, I used boxplots as they provide an intuitive and simple assessment of the data's symmetry, variance and outliers easily at first glance. This was so I could compare the changes in the distribution of DNAs between these two appointment types over time, using their approximate median values.

In 2019, the rate of DNA cases for face-to-face appointments were the highest observed, before notably decreasing from 2020 to 2021 (Table 1). Figure 1 illustrates the monthly scale of this decrease by examining the seasonal differences over time. The distribution of DNAs was higher in the months of April to June of 2019, compared to the same months in 2020 and 2021, which observed a considerably lower distribution. The year 2020 observed the lowest rate of missed appointments, particularly in May (month 17, median: 3.2%), compared to all other timepoints (Figure 1). Conversely, 2021 witnessed a slight increase in the rate of missed appointments from 2020. This was especially evident in the months September to November of 2021 which observed a higher median distribution of approximately 4.9%, 5% and 4.85% compared to 2020 which observed a lower distribution (4%, 4.01% and 3.8%) in the same months.

The rate of DNA cases showed a wide variability from 2019 to 2021, as evidenced by the considerably higher median distribution in January 2019 (Month 1, median: 5.1%), compared to January of 2020 (Month 13, median: 3.8%) and 2021 (Month 25, median: 3.4%). This wide variability was supported by the high dispersion seen in 2020 and 2021 between August and January (months 20 to 25), which observed a significantly large difference in their interquartile ranges (IQR). In contrast, the same months in 2019 observed no significant dispersion, rather a more stable trend of missed appointments was seen before a wide variability in the following years. Along with this, an extreme outlier of 16% was observed in September (month 33), potentially due to the naturally wide variability in the rate of DNAs in Greater Manchester.

I next analysed the distribution in other types of GP appointments to see if there was a decreased rate of missed appointments, compared to what was seen in face-to-face appointments in Greater Manchester.

My findings show that 2019 experienced the highest level of DNAs across all years analysed in this short time series (Table 1). Specifically, the distribution of DNAs was considerably higher between the months August and September 2019 (median: 3.8%). Contrastingly, the distribution of DNAs was considerably lower in 2020 (median: 2.5% and 2.6%) and 2021 (median: 2.4% and 2.3%) across the same months, with year 2021 exhibiting the lowest level of DNA rates. My results also revealed a wide variability in the level of DNAs observed over time, as evidenced by the wide dispersion seen in months January, May and August of 2019, where there was a significantly large distance between the IQRs. A few extreme outliers were also noticed in March and April 2019 (Months 2 and 3), with an estimated median distribution of 7.5% and 7.6% of DNAs.

Overall, face-to-face appointments exhibited a more extreme outlier distribution compared to other appointment types, which suggests that unusual data entries or events affected the data range. The level of dispersion in face-to-face appointments was exceedingly lower in 2019, compared to other appointment types, implying there was a wider variability in other appointment types.

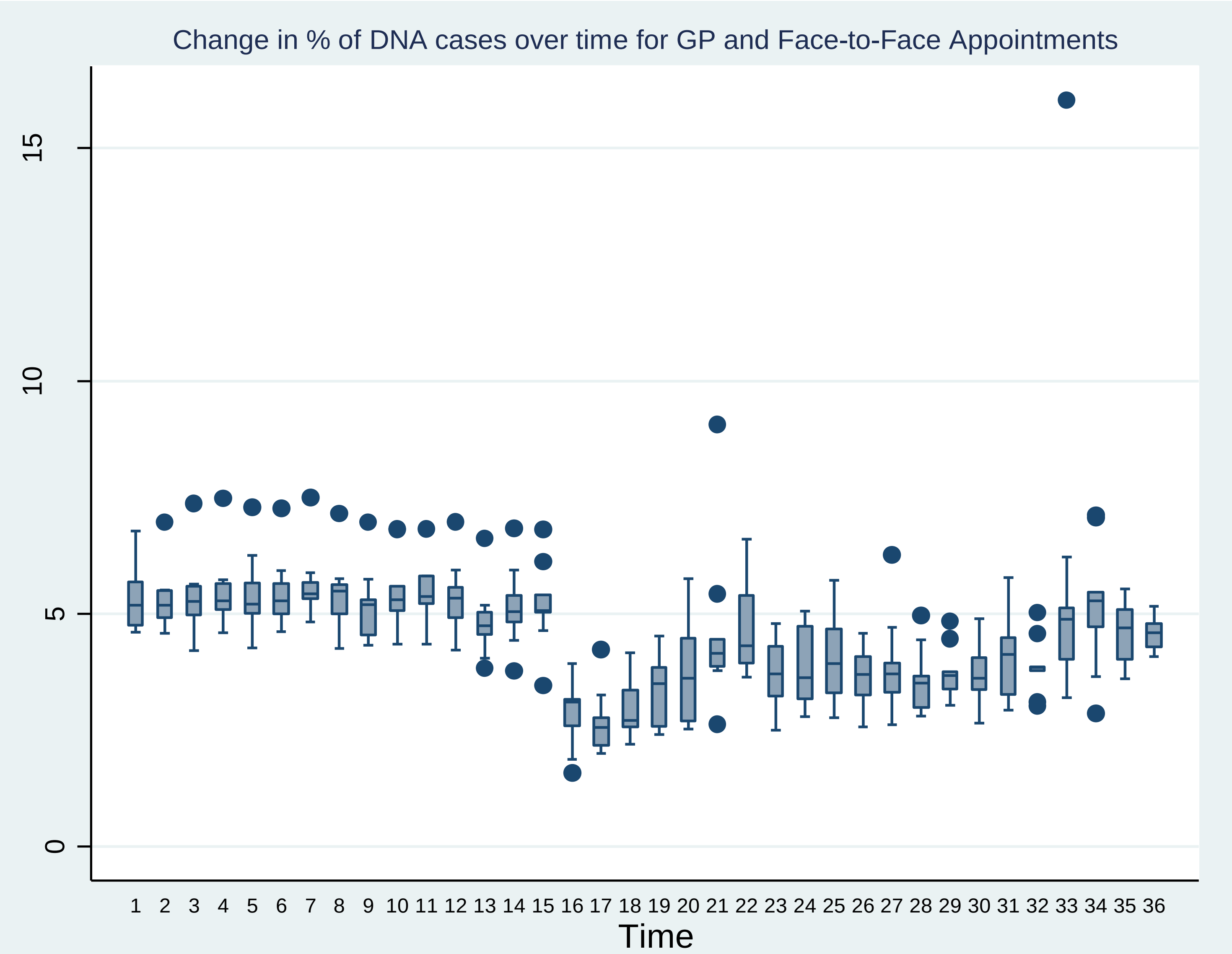

Figure 1 Box plots, showing the distribution of percentage of missed Face-to-Face appointments with general practitioners over time.

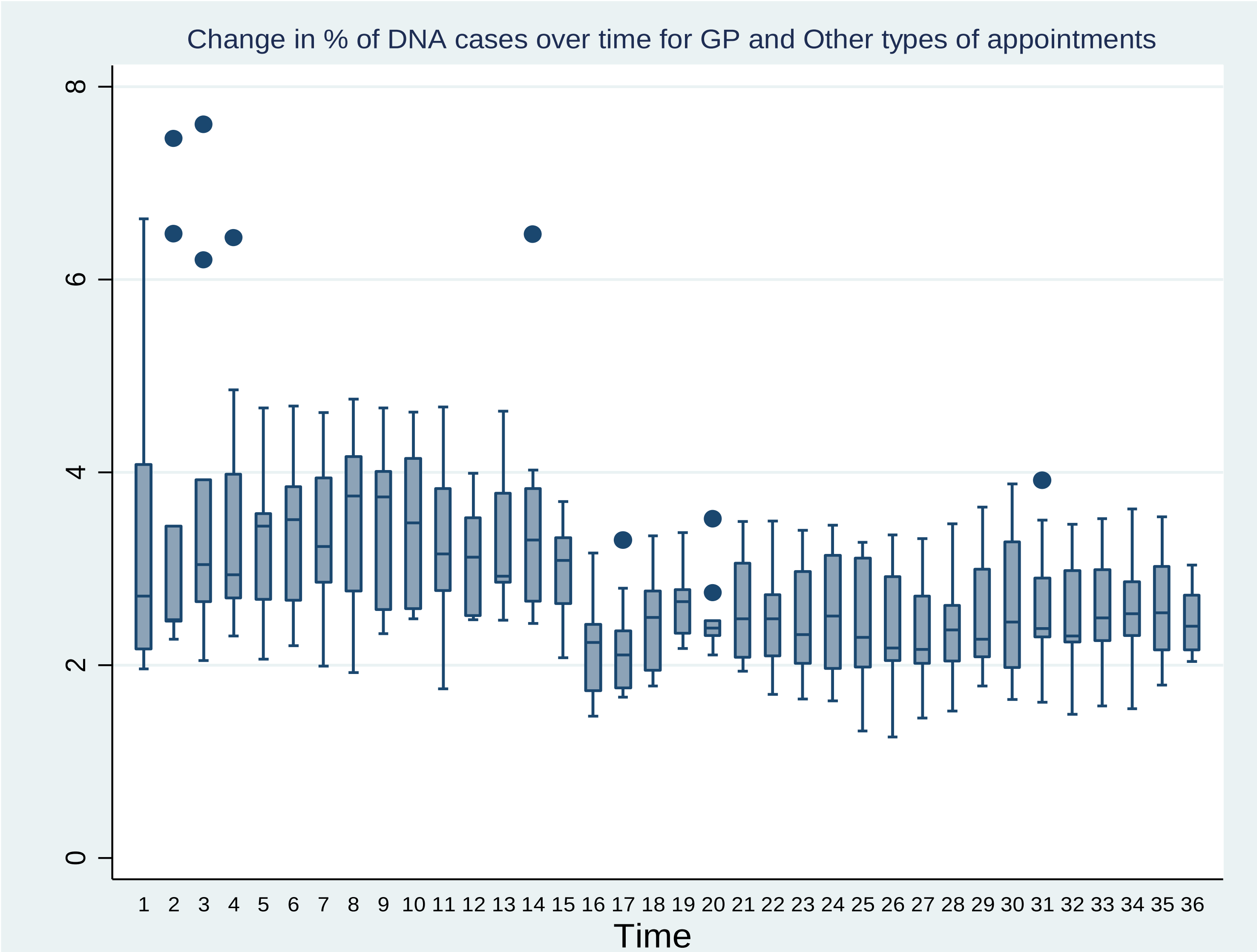

Figure 2 Box plots, showing the distribution of percentage of missed Other appointment types with GPs over time.

2.3 Statistical analyses:

The incident rate of missed appointments over time was analysed using the incident rate ratio (IRR) calculated in Poisson regression modelling, to determine the rate of DNA cases in Greater Manchester. I chose to use Poisson regression because it provides a great baseline for this type of count data. My sample size of interest was 9 CCG regions in Greater Manchester which was a relatively small sample size. Poisson regression was the most efficient model to estimate IRRs for count data with smaller sample sizes.

The study showed a gradual decrease in the missed face-to-face appointments over time approximately by 1.06% (IRR: 1.02), and strongly supported by the 95% confidence interval (CI 1.03% TO 1.08%). Using the reference group month (January) as a comparison, the months (indicator variables) were analysed to examine the changes in the rate of DNAs over the timescale of 2019 to 2021 (Table 2).

A wide variability in the rate of missed appointments was observed over time (Figure 1), where months February and March (months 2 and 3) exhibited a slight increase in DNA cases by 2.7% (IRR: 1.02) and 4% (IRR: 1.04), compared to the month (January). However, months April, May and June (months 4, 5, and 6) observed the lowest incidence rate in missed appointments, with a reduction of 4.8% (IRR: 0.95), 3.94% (IRR: 0.96), and 3.32% (0.97), respectfully. The 95% confidence intervals of 3.66% to 6.08% (CI 0.94 - 0.96) supported these results. The results in table 2 noticed an increase in the incidence rate of DNAs from July onwards with months October, November and December (months 10, 11 and 12) observing the highest incidence rate of missed appointments over time. These months expressed an increase in the number of DNAs by 18.4% (IRR: 1.18, CI 1.17- 1.2), 9.9% (IRR:9.9%, CI 1.08- 1.11) and 10.6% (IRR: 1.11, CI 1.09 – 1.11) respectively. The results revealed that there was a significant increase in the rate of DNAs from months April to June and October to December, which highlights the wide variability observed in this study.

I wanted to investigate whether more people attended the other types of appointments than the face-to-face appointments in Greater Manchester. I analysed the instance rate of other appointment types to examine whether there were any substantial differences in the measure of missed appointments between face-to-face and other appointments over time. No formal statistical comparisons were made between these groups, rather an overall estimation of the differences between the incidence rate ratios (IRRs), due to an overlap between the confidence intervals (CI).

Results from Table 3 show that other appointment types experienced a drop in the rate of missed appointments by 1.38% at each timepoint, supported by the 95% confidence interval (CL 0.9857 – 0.9865). Other appointment types noticed a wide variability in the rate of DNAs, particularly between months February and July. Initially months February and March experienced an increased in missed appointments by 3.2% (IRR: 1.0326) and 1.7% (IRR: 1.0169). However, this incidence rate drastically decreased from month March onwards, with months April, May and June exhibiting the lowest rate of DNAs, with a reduction of 8.1%, 7.3% , and 0.7% respectively. There was a significant increase in DNA rates from month June onwards, with July exhibiting an increase by 2.3% (IRR: 1.023). November showed the highest increase in missed appointments by 4.3% (IRR: 1.0432), compared to all other months, evidencing the overall wide variability observed in this study.

This research found that both groups experienced a decrease in the IRR of missed appointments for each month in the timeseries. Both groups experienced a similar trend where there was a reduction in the missed appointments between month April and June (Table 2). However, other appointment types experienced the most reduction in missed appointments, with an estimated difference of 3.3% and 3.4% in reduction from face-to-face appointments. Both groups experienced an increase in the incidence rate of DNA cases in month November. The wide variability observed between both groups over time suggests that there potentially other factors contributing to increase in DNA cases.

dna	IRR	Std.Err.	z	P>|z|	[95% Conf. interval]	
time	0.989424	0.00013	-80.84	0	0.989169	0.989679
month						
2	1.027928	0.005891	4.81	0	1.016446	1.03954
3	1.040873	0.00605	6.89	0	1.029082	1.052798
4	0.951208	0.006183	-7.7	0	0.939166	0.963405
5	0.9606	0.0062	-6.23	0	0.948525	0.972829
6	0.96681	0.006185	-5.28	0	0.954763	0.97901
7	1.0324	0.006316	5.21	0	1.020094	1.044854
8	1.027721	0.006552	4.29	0	1.01496	1.040642
9	1.22884	0.007161	35.36	0	1.214885	1.242955
10	1.184155	0.006881	29.09	0	1.170745	1.197719
11	1.099394	0.006662	15.64	0	1.086415	1.112528
12	1.106046	0.006951	16.04	0	1.092505	1.119754
_cons	0.056176	0.000234	-691.99	0	0.05572	0.056636

Table 2 Poisson Regression Model showing the Incidence Rates (IRR) of missed face-to-face appointments over time.

dna	IRR	Std.Err.	z	P>|z|	[95% Conf. interval]	
time	0.986161	0.000212	-64.72	0	0.985744	0.986577
Month						
2	1.032605	0.010748	3.08	0.002	1.011753	1.053888
3	1.016941	0.009874	1.73	0.084	0.997771	1.03648
4	0.918824	0.009252	-8.41	0	0.900867	0.937138
5	0.926644	0.009436	-7.48	0	0.908333	0.945325
6	0.993158	0.009774	-0.7	0.485	0.974184	1.012501
7	1.023324	0.010024	2.35	0.019	1.003865	1.04316
8	1.008946	0.010233	0.88	0.38	0.989089	1.029202
9	1.033243	0.010179	3.32	0.001	1.013483	1.053387
10	1.026545	0.010206	2.64	0.008	1.006734	1.046745
11	1.043292	0.010391	4.26	0	1.023124	1.063857
12	1.024925	0.010529	2.4	0.017	1.004494	1.045771
_cons	0.03862	0.000311	404.03	0	0.038015	0.039235

Table 3 Poisson Regression Model showing the Incidence Rates (IRR) of missed appointments for other appointment types over time.

DISCUSSION:

Inevitable events such as missed appointments affected the performance and revenue of primary care significantly, (Glowacka et al, 2009). In this study, I investigated the regional variability of missed appointments in Greater Manchester alone over the years 2019 to 2021, with two appointment types considered; Face-to-Face and Other types of appointments.

This study found an overall wide variability in the rate of missed appointments over the years 2019 to 2021 with 2019 showing the highest measure of DNAs before decreasing notably in 2020 in face-to-face appointments (Table 1). The median distribution of DNAs in 2020 was the lowest when compared to 2019 and 2021 (Figure 1). This variability was supported by outliers in September 2021 as well as the incidence rate ratio where months April to June had the lowest incidence rate of missed appointments before increasing significantly in months October to December (Table 2).

This high measure of DNA in 2019 correlated more with ethnic communities (22%) living in the most deprived conditions experiencing declining health and life expectancies, (Raleigh et al, 2021). The year 2020 observed the lowest measure of missed appointments despite outliers between 2020 to 2021 (Figure 1) and lower life expectancy as well as high COVID-19 cases observed, (Woolf et al, 2022), (Crimmins et al, 2015). These factors point to increased missed appointments, especially in ethnically diverse and deprived populations due to possible health inequalities, and fear of infection, (Pan et al, 2020), (Ayele et al, 2022). The year 2020 should have experienced a higher measure of DNAs but did not because of the increased disproportionate level of unknown cases (Table 1), which failed to be identified as DNA or attended.

There was a considerable decrease in the rate of missed appointments between 2019 to 2021 (Table 1) in other appointment types, where the highest median distribution in September 2019, declined significantly in September 2021, (Figure 2). More people attended other appointment types (Online, Home visits, Video calls, Telephone) compared to missed appointments in Greater Manchester. Studies have pointed to the rise of these online appointment options during the COVID-19 pandemic as the reason for this decrease, (Newbound et al, 2017). Convenience, transport, and no long waiting times were the notable reasons for this, with a study confirming that transport issues (29.9%) were significant barriers to attending face-to-face appointments, (Ofei-Dodoo et al, 2019), (Allan-Blitz et al, 2022).

3.1 Literature:

In context, most existing research analyses deeper into the prevalence of missed appointments as their primary objective, rather than focusing on the surface-level observational variability in this study. Few studies investigated regional variability in Greater Manchester alone, but various literature published regarding the variability of DNAs in different regions of the UK and non-western countries were used for comparison.

The findings from this study were in line with an existing study investigating nine primary clinics in Sheffield over two years, which observed a wide variability in DNAs (5.7% to 17%), (Waller et al, 2000). Similarly, a retrospective study investigating four practices in Leeds over 12-months noticed a similar range in variability (4.2% to 11.8%), (Neal et al, 2001). My research noted a similar range in the variability of missed appointments to these Sheffield and Leeds clinical studies which support the validity of my findings. All these studies also observed a similar correlation where populations of ethnic minorities and those living in greater deprivation are key factors in DNA cases, (Neal et al, 2001), (Waller et al, 2000), (Williamson et al, 2014).

When comparing my findings to other countries, the measure of DNA cases in the US was slightly similar in range, though a significantly wider variability (5% to 50%) was observed, (George et al, 2003). This was most likely because the study investigated all regions in the USA, compared to the smaller population size of Greater Manchester, (George et al, 2003). A more prominent reason for this wider variability observed in the US, was due to the lack of universal healthcare, as opposed to the UK, (Light et al, 2003). The USA also noted a substantially higher socioeconomic, gender, and racial depravity, as well as poverty, which all played significant factors as to why there was a wider variability in DNAs in America, (Light et al, 2003), (Crowley et al, 2020), (Singh et al, 2017), (Wang et al, 2019).

Comparing this research to non-western countries, there was a significantly wider difference in the rate of DNAs. Higher variability of DNAs was noted in an outpatient study in a Singapore clinic (21% to 39%) from 2000 to 2004, and an outpatient clinic in Saudi Arabia (40.7% to 63.5%) observed this wider variability in different parts of the day, (Lee et al, 2005), (Kasem et al, 2015). Factors such as high deprivation, differences in healthcare structures were a reason for these wide differences between the UK and Singapore, and Saudi Arabia, (AlOmar et al, 2018), (Alawadhi et al, 2021).

3.2 Strengths & Limitations:

This study used national data collected from NHS, to provide detailed information regarding the total appointments, CCG locations, and attendance of appointments, (Pastorino et al, 2019). This Explanatory data analysis method such as Descriptive summary statistics, Time series Box Plots, and Poisson regression models provided detailed information regarding the regional variability of DNA cases.
This data could be beneficial to future works in evaluating the best methods of solving healthcare inequalities in Greater Manchester. This data was well collected, with quality and authenticity considered before publishment to the NHS database and aligned with the main research objectives of this study, (Paradis et al, 2016), (McAfee et al, 2012).

Though this study provides an overall detailed cross-sectional analysis of regional variability of DNA cases in Greater Manchester, several methodological limitations are to be considered. A small sample size of only 9 CCG regions was used in this study, with CCG region 'Bury' missing from the datasets during collection. As a result, this subgroup population in Bury was underrepresented during this investigation and indefinitely reduced the validity and statistical power of this study regarding regional variability, (Marino et al, 2021), (Mack et al, 2018). Another limitation included the prominent level of missing data between 2019 to 2021 during data collection. These were recalled as unknown cases (Table 1) which failed to be recorded as DNA or Attended due to information loss or system errors, so a great percentage of DNAs were unaccounted for, (Kwak et al, 2017). This study also lacked a detailed investigation of COVID-19 and its effects on the rate of DNAs in Greater Manchester, considering there was a peak in infections from 2020 to 2021, (Carvalho et al, 2021), (Sterne et al, 2009). It was only discussed briefly on an assumption basis, rather than with certainty. Considering all of this, the implications of these findings must be approached with caution as a lot of these factors were unaccounted for when investigating the variability of DNAs in Greater Manchester.

3.3 Future Works:

Future research should consider both practice level and personal level factors to study the correlation in patient ages, gender, ethnicity, and deprivation along with total number of patients and GPs to variability in missed appointments, (Haneuse et al, 2011), (Parsons et al, 2023). This is to provide deeper understanding of DNA cases in different regions and revise current resolutions along with future strategies to address inequalities associated with DNA cases.

ONS-level data and the new ICS system will be applied to remove constraints placed on accessing certain information to acquire better-quality data for analysis, (Gongora-Salazar et al, 2022). To compare regional variabilities across a wider scale, more regions and a longer timescale will be investigated, rather than being limited to Greater Manchester alone. COVID-19 and other common illnesses indefinitely affected the year 2020 and onwards and will be researched in-depth to determine its impact on the attendance records within different regions, (Parsons et la, 2021).

CONCLUSION:

This study documented an overall wide variability in missed appointments in Greater Manchester from 2019 to 2021 for both face-to-face and other appointment types. Though the rate of missed appointments decreased from 2019 to 2021, there were some high incidence rates of DNAs along with some extreme outliers in certain months.

Surface level assumptions were addressed regarding these outliers and high IRRs, including low life expectancy, deprivation, and ethnically diverse populations all correlated to DNAs. The year 2020 particularly experienced the lowest rate of DNAs compared to 2019 and 2021, which raised questions considering 2020 experienced a lower life expectancy and increased COVID-19 cases. A reason identified was the increased rate of unknown cases that failed to be identified and could have increased the rate of DNAs in 2020.

More people attended other appointment types in comparison to face-to-face appointments, due to convenience and the lack of practical barriers such as transport, and waiting times, often seen in face-to-face appointments. The implications of these findings must be addressed with caution as there was no substantial quantitative evidence correlating the relationship between all these assumptions and DNA cases in Greater Manchester. A subpopulation in Manchester called Bury was underrepresented along with COVID-19 which was not a prominent factor in this study. This leads to further questioning the weight and effect of these factors regarding the rate of DNAs in Greater Manchester over time, had they been accounted for. Future works will analyse the impact of COVID-19 and factors such as deprivation, age, ethnicity, and their effect on DNA cases in Greater Manchester.

A clear understanding developed regarding the variability in Greater Manchester for face-to-face and other appointment types and the present inequalities causing missed appointments. Ultimately more research and strategies must be revisited to reduce missed appointments.

REFERENCES:

AlOmar, R.S., Parslow, R.C. & Law, G.R. (2018). Development of two socioeconomic indices for Saudi Arabia. BMC Public Health 18, 791. https://doi.org/10.1186/s12889-018-5723-z

Allan-Blitz, L.-T. et al. (2022). A pilot study: The impact of clinic-provided transportation on missed clinic visits and system costs among teenage mother–child dyads, Nature News. Nature Publishing Group. Available at: https://www.nature.com/articles/s41599-022-01342-x#citeas

Alawadhi, A., Palin, V., & van Staa, T. (2021). Prevalence and factors associated with missed hospital appointments: a retrospective review of multiple clinics at Royal Hospital, Sultanate of Oman. BMJ open, 11(8), e046596. https://doi.org/10.1136/bmjopen-2020-046596

Al-Ubaydli, M. (2005) Using Microsoft Excel, The BMJ. British Medical Journal Publishing Group. Available at: https://doi.org/10.1136/bmj.329.7481.s7

Ayele, T. A., Alamneh, T. S., Shibru, H., Sisay, M. M., Yilma, T. M., Melak, M. F., Bisetegn, T. A., Belachew, T., Haile, M., Zeru, T., Asres, M. S., & Shitu, K. (2022). Effect of COVID-19 pandemic on missed medical appointment among adults with chronic disease conditions in Northwest Ethiopia. PloS one, 17(10), e0274190. https://doi.org/10.1371/journal.pone.0274190

Böhringer, S. (2013) Dynamic parallelization of R functions, The R Journal. Available at: https://journal.r-project.org/articles/RJ-2013-029/

Boston RC, Sumner AE. STATA: a statistical analysis system for examining biomedical data. Adv Exp Med Biol. 2003;537:353-69. doi: 10.1007/978-1-4419-9019-8_23. PMID: 14995047.

Carvalho, T., Krammer, F., & Iwasaki, A. (2021). The first 12 months of COVID-19: a timeline of immunological insights. Nature reviews. Immunology, 21(4), 245–256. https://doi.org/10.1038/s41577-021-00522-1

Cashman, S. B., Savageau, J. A., Lemay, C. A., & Ferguson, W. (2004). Patient health status and appointment keeping in an urban community health center. Journal of health care for the poor and underserved, 15(3), 474–488. https://doi.org/10.1353/hpu.2004.0037

Chan B. K. C. (2018). Data Analysis Using R Programming. *Advances in experimental medicine and biology*, *1082*, 47–122. https://doi.org/10.1007/978-3-319-93791-5_2

Cookson, T. (2011) Merging multiple data files into one data frame: R-bloggers, R. Available at: https://www.r-bloggers.com/2011/04/merging-multiple-data-files-into-one-data-frame/

Crimmins E. M. (2015). Lifespan and Healthspan: Past, Present, and Promise. The Gerontologist, 55(6), 901–911. https://doi.org/10.1093/geront/gnv130

Crowley, R. et al. (2020) “Envisioning a better U.S. Health Care System for all: Coverage and cost of care,” Annals of Internal Medicine, 172(2_Supplement). Available at: https://doi.org/10.7326/m19-2415.

Divisi, D., Di Leonardo, G., Zaccagna, G., & Crisci, R. (2017). Basic statistics with Microsoft Excel: a review. Journal of thoracic disease, 9(6), 1734–1740. https://doi.org/10.21037/jtd.2017.05.81

Dowle, M. et al. (2019) Data.table: Extension of data.frame version 1.9.3 from R-forge, version 1.9.3 from R-Forge. Available at: https://rdrr.io/rforge/data.table/

Ellis, D. A., McQueenie, R., McConnachie, A., Wilson, P., & Williamson, A. E. (2017). Demographic and practice factors predicting repeated non-attendance in primary care: a national retrospective cohort analysis. The Lancet Public Health, 2(12), e551-e559.

Fisher, R. F., Croxson, C. H., Ashdown, H. F., & Hobbs, F. R. (2017). GP views on strategies to cope with increasing workload: a qualitative interview study. The British journal of general practice : the journal of the Royal College of General Practitioners, 67(655), e148–e156. https://doi.org/10.3399/bjgp17X688861

Garmendia, A. (2020) R for life sciences. Chapter 7, dplyr and tidyr: tidyverse packages to manage data. UPV - Instituto Agroforestal Mediterráneo (IAM). Available at: http://personales.upv.es/algarsal/R-tutorials/07_Tutorial-7_R-dplyr-tidyr.html

George, A., & Rubin, G. (2003). Non-attendance in general practice: a systematic review and its implications for access to primary health care. Family practice, 20(2), 178–184. https://doi.org/10.1093/fampra/20.2.178

Glowacka, K. J., Henry, R. M., & May, J. H. (2009). A hybrid data mining/simulation approach for modelling outpatient no-shows in clinic scheduling. Journal of the Operational Research Society, 60, 1056-1068.

Gongora-Salazar, P., Glogowska, M., Fitzpatrick, R., Perera, R., & Tsiachristas, A. (2022). Commissioning [Integrated] Care in England: An Analysis of the Current Decision Context. International journal of integrated care, 22(4), 3. https://doi.org/10.5334/ijic.6693

Graham J. W. (2009). Missing data analysis: making it work in the real world. *Annual review of psychology*, *60*, 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Grol, S.M. et al. (2018) The role of the general practitioner in multidisciplinary teams: A qualitative study in elderly care - BMC primary care, BioMed Central. BioMed Central. Available at: https://doi.org/10.1186/s12875-018-0726-5 (Accessed: March 25, 2023).

Grömping, U. (2015). Using R and RStudio for Data Management, Statistical Analysis and Graphics (2nd Edition). Journal of Statistical Software, Book Reviews, 68(4), 1–7. https://doi.org/10.18637/jss.v068.b04

Gruber, J.B. (2022) Basic Usage. Available at: https://cran.r-project.org/web/packages/LexisNexisTools/vignettes/demo.html

Haneuse, S., & Bartell, S. (2011). Designs for the combination of group- and individual-level data. Epidemiology (Cambridge, Mass.), 22(3), 382–389. https://doi.org/10.1097/EDE.0b013e3182125cff

Honeyman, M., Dunn, P. & McKenna, H. (2016) A digital NHS? An introduction to the digital agenda and plans for implementation. London: The Kings Fund.

Hashim, M. J., Franks, P., & Fiscella, K. (2001). Effectiveness of telephone reminders in improving rate of appointments kept at an outpatient clinic: a randomized controlled trial. The Journal of the American Board of Family Practice, 14(3), 193–196.

Husain-Gambles, M., Neal, R. D., Dempsey, O., Lawlor, D. A., & Hodgson, J. (2004). Missed appointments in primary care: questionnaire and focus group study of health professionals. British Journal of General Practice, 54(499), 108-113.

Jr., J.F.H. et al. (2021) Partial least squares structural equation modelling (PLS-SEM) using R, SpringerLink. Springer International Publishing. Available at: https://link.springer.com/book/10.1007/978-3-030-80519-7

Kaliyadan, F., & Kulkarni, V. (2019). Types of Variables, Descriptive Statistics, and Sample Size. Indian dermatology online journal, 10(1), 82–86. https://doi.org/10.4103/idoj.IDOJ_468_18

Kang H. (2013). The prevention and handling of the missing data. Korean journal of anaesthesiology, 64(5), 402–406. https://doi.org/10.4097/kjae.2013.64.5.402

Kaplan-Lewis, E., & Percac-Lima, S. (2013). No-show to primary care appointments: why patients do not come. Journal of primary care & community health, 4(4), 251-255.

Kasem, A.A., Althobaiti, T.S., Al-Jeaid, D.A.S. et al. (2015). An analysis of causes behind missed scheduled appointments at outpatient ENT clinics. Egypt J Otolaryngol 31, 1–3. https://doi.org/10.4103/1012-5574.152701

Kaur, P., Stoltzfus, J. and Yellapu, V. (2019) Descriptive statistics . International Journal of Academic Medicine. Available at: https://www.ijam-web.org/text.asp?2018/4/1/60/230853

Kwak, S. K., & Kim, J. H. (2017). Statistical data preparation: management of missing values and outliers. Korean journal of anesthesiology, 70(4), 407–411. https://doi.org/10.4097/kjae.2017.70.4.407

Lee, R.R.S. et al. (2019) “Factors affecting follow-up non-attendance in patients with type 2 diabetes mellitus and hypertension: A systematic review,” Singapore Medical Journal, 60(5), pp. 216–223. Available at: https://doi.org/10.11622/smedj.2019042.

Lee, V.J., Earnest, A., Chen, M.I. et al. (2005). Predictors of failed attendances in a multi-specialty outpatient centre using electronic databases. BMC Health Serv Res 5, 51. https://doi.org/10.1186/1472-6963-5-51

Light D. W. (2003). Universal health care: lessons from the British experience. American journal of public health, 93(1), 25–30. https://doi.org/10.2105/ajph.93.1.25

Love, S. B., Yorke-Edwards, V., Diaz-Montana, C., Murray, M. L., Masters, L., Gabriel, M., Joffe, N., & Sydes, M. R. (2021). Making a distinction between data cleaning and central monitoring in clinical trials. Clinical trials (London, England), 18(3), 386–388. https://doi.org/10.1177/1740774520976617

Mack, C., Su, Z. and Westreich, D. (2018) Managing Missing Patient Data in Patient Registries. White Paper, addendum to Registries for Evaluating Patient Outcomes: A User's Guide, Third Edition. Agency for Healthcare Research and Quality. Available at: https://www.ncbi.nlm.nih.gov/books/NBK493611/

Marbouh, D., Khaleel, I., Al Shanqiti, K., Al Tamimi, M., Simsekler, M. C. E., Ellahham, S., Alibazoglu, D., & Alibazoglu, H. (2020). Evaluating the Impact of Patient No-Shows on Service Quality. Risk management and healthcare policy, 13, 509–517. https://doi.org/10.2147/RMHP.S232114

Margham, T., Williams, C., Steadman, J., & Hull, S. (2021). Reducing missed appointments in general practice: Evaluation of a quality improvement programme in East London. *The British Journal of General Practice*, *71*(702), e31. https://doi.org/10.3399/bjgp20X713909

Marino, M., Lucas, J., Latour, E., & Heintzman, J. D. (2021). Missing data in primary care research: importance, implications and approaches. Family practice, 38(2), 200–203. https://doi.org/10.1093/fampra/cmaa134

Martin, S. J., Bassi, S., & Dunbar-Rees, R. (2012). Commitments, norms and custard creams–a social influence approach to reducing did not attends (DNAs). Journal of the Royal Society of Medicine, 105(3), 101-104.

Martin, C., Perfect, T., & Mantle, G. (2005). Non-attendance in primary care: the views of patients and practices on its causes, impact and solutions. Family practice, 22(6), 638-643.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. Harvard business review, 90(10), 60-68.

McQueenie, R., Ellis, D. A., McConnachie, A., Wilson, P., & Williamson, A. E. (2019). Morbidity, mortality and missed appointments in healthcare: a national retrospective data linkage study. BMC medicine, 17(1), 2. https://doi.org/10.1186/s12916-018-1234-0

Mitchell, A. J., & Selmes, T. (2007). A comparative survey of missed initial and follow-up appointments to psychiatric specialties in the United kingdom. Psychiatric services (Washington, D.C.), 58(6), 868–871. https://doi.org/10.1176/ps.2007.58.6.868

Morris, L.(2023). Health Education England finalises merger with NHS England. National Health Executive. Available at: https://www.nationalhealthexecutive.com/articles/health-education-england-finalises-merger-nhs-england

Morrissey, K., Spooner, F., Salter, J., & Shaddick, G. (2021). Area level deprivation and monthly COVID-19 cases: The impact of government policy in England. Social science & medicine (1982), 289, 114413. https://doi.org/10.1016/j.socscimed.2021.114413

Neal, R. D., Lawlor, D. A., Allgar, V., Colledge, M., Ali, S., Hassey, A., Portz, C., & Wilson, A. (2001). Missed appointments in general practice: retrospective data analysis from four practices. The British journal of general practice : the journal of the Royal College of General Practitioners, 51(471), 830–832.

Newbould, J., Abel, G., Ball, S., Corbett, J., Elliott, M., Exley, J., Martin, A., Saunders, C., Wilson, E., Winpenny, E. and Yang, M. (2017). Evaluation of telephone first approach to demand management in English general practice: observational study. bmj, 358.

Nguyen, D. L., Dejesus, R. S., & Wieland, M. L. (2011). Missed appointments in resident continuity clinic: patient characteristics and health care outcomes. Journal of graduate medical education, 3(3), 350–355. https://doi.org/10.4300/JGME-D-10-00199.1

Nielsen, K. M., Faergeman, O., Foldspang, A., & Larsen, M. L. (2008). Cardiac rehabilitation: health characteristics and socio-economic status among those who do not attend. European journal of public health, 18(5), 479–483. https://doi.org/10.1093/eurpub/ckn060

Ofei-Dodoo, S., Kellerman, R., Hartpence, C., Mills, K., & Manlove, E. (2019). Why Patients Miss Scheduled Outpatient Appointments at Urban Academic Residency Clinics: A Qualitative Evaluation. Kansas journal of medicine, 12(3), 57–61.

Owen, K., Hopkins, T., Shortland, T., & Dale, J. (2019). GP retention in the UK: a worsening crisis. Findings from a cross-sectional survey. BMJ open, 9(2), e026048. https://doi.org/10.1136/bmjopen-2018-026048

Pan, D., Sze, S., Minhas, J. S., Bangash, M. N., Pareek, N., Divall, P., Williams, C. M., Oggioni, M. R., Squire, I. B., Nellums, L. B., Hanif, W., Khunti, K., & Pareek, M. (2020). The impact of ethnicity on clinical outcomes in COVID-19: A systematic review. EClinicalMedicine, 23, 100404. https://doi.org/10.1016/j.eclinm.2020.100404

Paradis, E., O'Brien, B., Nimmon, L., Bandiera, G., & Martimianakis, M. A. (2016). Design: Selection of Data Collection Methods. Journal of graduate medical education, 8(2), 263–264. https://doi.org/10.4300/JGME-D-16-00098.1

Parsons, J., Abel, G., Mounce, L. T., & Atherton, H. (2023). The changing face of missed appointments. The British journal of general practice : the journal of the Royal College of General Practitioners, 73(728), 134–135. https://doi.org/10.3399/bjgp23X732249

Parsons, J., Bryce, C. and Atherton, H. (2021) Which patients miss appointments with General Practice and the reasons why: A systematic review, British Journal of General Practice. British Journal of General Practice. Available at: https://doi.org/10.3399/BJGP.2020.1017 (Accessed: March 25, 2023).

Pastorino, R., De Vito, C., Migliara, G., Glocker, K., Binenbaum, I., Ricciardi, W., & Boccia, S. (2019). Benefits and challenges of Big Data in healthcare: an overview of the European initiatives. European journal of public health, 29(Supplement_3), 23–27. https://doi.org/10.1093/eurpub/ckz168

Raleigh, V. and Holmes, J. (2021) The health of people from ethnic minority groups in England, The King's Fund. Available at: https://www.kingsfund.org.uk/publications/health-people-ethnic-minority-groups-england

Schork, J. (2022) Grep & grepl R functions (3 examples): Regexpr, Gregexpr & Regexec, Statistics Globe. Available at: https://statisticsglobe.com/grep-grepl-r-function-example

Singh, G. K., Daus, G. P., Allender, M., Ramey, C. T., Martin, E. K., Perry, C., Reyes, A. A. L., & Vedamuthu, I. P. (2017). Social Determinants of Health in the United States: Addressing Major Health Inequality Trends for the Nation, 1935-2016. International journal of MCH and AIDS, 6(2), 139–164. https://doi.org/10.21106/ijma.236

Singh, G & Soman, B. (2019). Data Transformation using dplyr package in R. 10.13140/RG.2.2.10397.46565.

Sterne, J.A.C. et al. (2009) Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls, The BMJ. British Medical Journal Publishing Group. Available at: https://www.bmj.com/content/338/bmj.b2393

Van den Broeck, J., Cunningham, S. A., Eeckels, R., & Herbst, K. (2005). Data cleaning: detecting, diagnosing, and editing data abnormalities. PLoS medicine, 2(10), e267. https://doi.org/10.1371/journal.pmed.0020267

Waller, J., & Hodgkin, P. (2000). Defaulters in general practice: who are they and what can be done about them?. Family practice, 17(3), 252–253. https://doi.org/10.1093/fampra/17.3.252

Wang, Z. and Sun, J. (2019) “Explaining the poverty difference between the US and the UK: A Shapley Income-distribution decomposition approach,” Applied Economics Letters, 27(17), pp. 1438–1441. Available at: https://doi.org/10.1080/13504851.2019.1688238.

Wendt, C. J., & Anderson, G. B. (2022). Ten simple rules for finding and selecting R packages. PLoS computational biology, 18(3), e1009884. https://doi.org/10.1371/journal.pcbi.1009884

Wickham, H., Bryan, J. and Barrett, M. (2022) Automate package and project setup [R package usethis version 2.1.6], The Comprehensive R Archive Network. Comprehensive R Archive Network (CRAN). Available at: https://cran.r-project.org/web/packages/usethis/index.html

Wickham, H. et al. (2023) A grammar of data manipulation . Comprehensive R Archive Network (CRAN). Available at: https://cran.r-project.org/web/packages/dplyr/index.html

Wickham, H. et al. (2022) Tools to make developing R packages easier [R package devtools version 2.4.5], The Comprehensive R Archive Network. Comprehensive R Archive Network (CRAN). Available at: https://cran.r-project.org/web/packages/devtools/index.html

Williamson, A. E., Mullen, K., & Wilson, P. (2014). Understanding "revolving door" patients in general practice: a qualitative study. BMC family practice, 15, 33. https://doi.org/10.1186/1471-2296-15-33

Williamson, A. E., Ellis, D. A., Wilson, P., McQueenie, R., & McConnachie, A. (2017). Understanding repeated non-attendance in health services: a pilot analysis of administrative data and full study protocol for a national retrospective cohort. BMJ open, 7(2), e014120.

Williamson, D. F., Parker, R. A., & Kendrick, J. S. (1989). The box plot: a simple visual method to interpret data. Annals of internal medicine, 110(11), 916–921. https://doi.org/10.7326/0003-4819-110-11-916

Woolf, S. H., Masters, R. K., & Aron, L. Y. (2022). Changes in Life Expectancy Between 2019 and 2020 in the US and 21 Peer Countries. JAMA network open, 5(4), e227067. https://doi.org/10.1001/jamanetworkopen.2022.7067