Modelling demand in elective care in the National Health Service in England

Amber Sian Lewis-Roberts

Professional Economist BSc and Apprenticeship Level 6

School of Economics

University of Kent, July 2024

Abstract

This report investigates demand for elective health care in England provided by the National Health Service (NHS), in response to changes in average waiting time for treatment. It uses regression analysis to estimate the relationship between demand and waiting time, presenting the results as measures of elasticities. Using average waiting time as a proxy for price, demand theory suggests that as the average waiting time increases, the demand for health care should fall. However, despite increases in average waiting time to 14.4 weeks, the total elective care waiting list has continued to grow (7.6 million in November 23, NHS England, 2019). This suggests that the demand for health care is inelastic in response to changes in waiting time. The data for this report covers 117 general and acute hospital trusts over 92 months from NHS England and NHS Digital monthly statistical reports and uses 10 dependent and independent variables selected from literature. The data was split into three periods - pre-pandemic, pandemic, and post-pandemic – to reduce the impact of the COVID-19 pandemic on results and to see whether any factors had changed over time. Statistical testing guided the modelling. The pre-pandemic and post-pandemic models are least squares dummy models with fixed effects and time dummies with statistically insignificant variables removed. No model was identified for pandemic data. The pre-pandemic results show that if the waiting time increased by 1%, the demand for health care would fall by 0.073%. For post-pandemic, if the waiting time increased by 1%, the demand for health care would fall by 0.146%. This shows that demand is inelastic with respect to average waiting times and is consistent with the hypothesis and literature. Statistical testing showed that the models are not reliable, so the coefficients must be interpreted with caution and the models are inappropriate for policy decisions.

Acknowledgements

Thank you to the University of Kent tutors, academic advisors, apprenticeship advisors and administrators for their support and guidance over the last four years. Thank you to my fellow apprentices – particularly those in DHSC past and present – for sharing in the ups and downs of the past four years and being a constant source of support and friendship. Thank you to the Department of Health and Social Care – line managers, my teams, grade managers, friends, and colleagues – for all their support and encouragement during my studies allowing me to develop professionally while completing the degree apprenticeship.

Finally, and most importantly, thank you to my husband Tim for his unwavering support, belief and encouragement through this programme and everything it entailed. I would not have got to the end without him.

INTRODUCTION

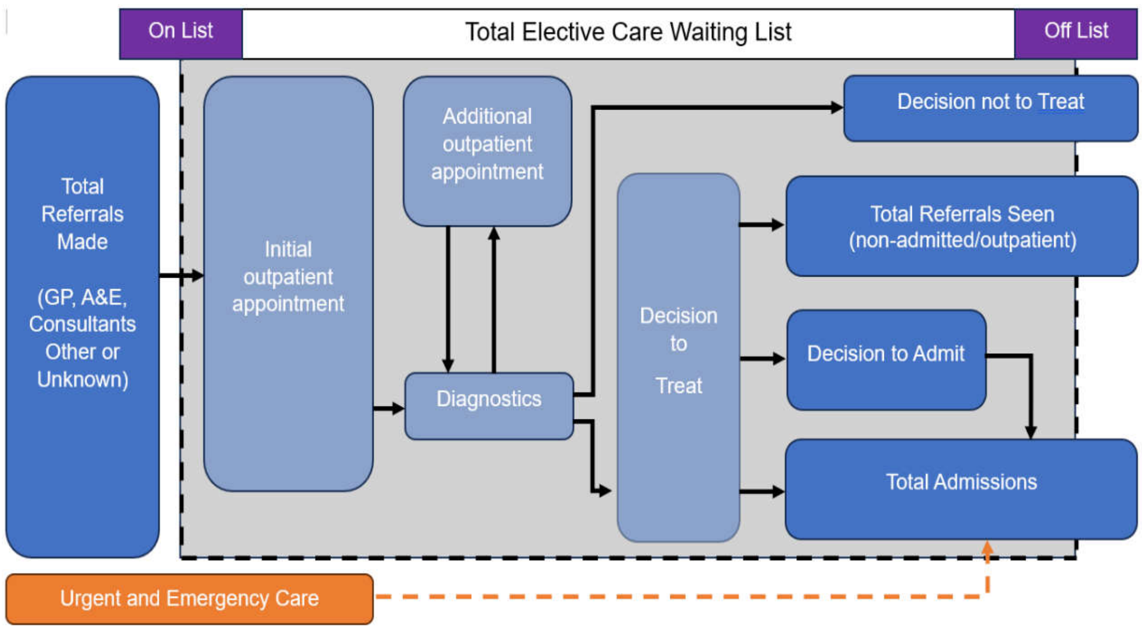
This report investigates whether the responsiveness (elasticity) of the demand for elective health care in England, as provided by the National Health Service (NHS), has remained inelastic despite increasing waiting times for treatment. It does this using regression analysis to model the relationship of demand and supply variables to assess the responsiveness of demand to changes in waiting time for treatment. The results are presented as measures of elasticities and discussed in the context of economic theory and literature.

i) Background

Elective care is defined as planned care – unlike emergency care which is unplanned – usually involving specialist care or surgery following a referral from a General Practitioner (GP) or other health care professional (Nuffield Trust, 2019). Patients waiting to receive elective treatment make up the total waiting list in England.

The diagram below shows this report defines the patient journey within the elective care system for the modelling of demand (loosely based on Findlay, 2021). The grey area is the total waiting list, and the darker blue boxes will be used as variables within the model.

Figure 1: Visual representation of the elective care system.



An individual patient receives a referral for an initial appointment with a view to obtaining a consultant-led diagnosis and treatment decision. The referral adds them to the total waiting list. Where a decision not to treat is made, they are removed from the waiting list, otherwise they are either seen as an outpatient, or admitted for treatment (day case or overnight). Once they receive their first treatment, they are removed from the total waiting list. Some patients receive a decision to admit but may not have date for treatment. They remain on the list until their first treatment. Included in the diagram are emergency admissions - they are included in total hospital admissions and share resources with elective care.

Since April 2016, the total waiting list grew from over 3,600,000 with an average waiting time of 6.6 weeks to over 7,600,000 in November 2023 with an average waiting time of 14.4 weeks (NHS England, 2019).

This indicates demand for health care is increasing and not being met efficiently, raising questions for policy makers around how to address this. Thus, it is necessary to understand how demand responds to changes to different factors as prescribed by economic theory, to provide suggested areas for policymakers to consider.

ii) Economic Theory

Demand and supply are fundamental concepts in economics used to analyse and predict participant behaviour to determine the price of goods or services within a market.

Generally, demand refers to the quantity that consumers are willing and able to pay for, with demand theory stating that (all other things held constant) as price increases or consumer income falls, demand for the good falls (Mankiw, 2018). This is also applicable to supply – a supplier will provide what they are willing/able to sell, so if prices are higher, a supplier will try to sell more for more profit.

The price charged for goods/services is key to determining demand and supply. Assuming a 'perfect' market with price as the only factor, prices would be fixed where the amount of the goods/services supplied equals the amount that is demanded by consumers (Mankiw, 2018) – the market is in equilibrium.

Economists measure the responsiveness of demand relative to price (or other) using elasticities (Mankiw, 2018). An elasticity less than 1, indicates demand is less responsive to changes in other factors (inelastic). An elasticity greater than one indicates it is more responsive to changes (elastic).

Applying this to health/health care, Morris et al (2012) states that health is not a typical good – it is not tangible and cannot (usually) be traded – but can be obtained through the consumption of health care. Demand for health care is therefore the derived demand for health (Santana et al, 2021), and health care can be thought of as a normal good. Improved health is what the market wants, but health care is what is needed to help achieve this. Arguably the market does not want to consume health care (not pleasurable), but consumers are willing to pay for it in their pursuit of health (Morris et al, 2012).

Health care also differs due to an imbalance of information between participants - consumers (patients) tend to have an economic agent (GP) guiding their consumption decision which can generate demand (Rosen, 2014).

From a macroeconomic perspective, there is evidence in literature to show that health has a positive effect on economic growth – the better the health of the population, the higher the rate of economic growth (Ridhwan, 2022). Thus, health is desirable for consumption and investment to individuals and population (Santana et al, 2021).

Applying demand theory to health care, as the price of health care increases, demand for health care should fall. However, in England, health care is 'free' at the point of entry, so from a consumer perspective, price and income are not considerations and alternatives are sought.

Following literature (discussed below), this report uses the expected average waiting time for treatment to represent price in the regression modelling.

Demand theory postulates that as the waiting time increases, demand for health care falls. However, as the total waiting list continues to grow, this implies the demand for health care is not responsive to average waiting time (it is inelastic) and other factors are influential. This has been observed in literature, and this report hypothesizes that this is still the case, but using more recent data, also seeks to identify any other statistically significant factors to changes in the demand for health care.

iii) Literature

- **Previous modelling of demand – issues and common features**

Earlier studies such as Blundell and Windmeijer (2000) and Gravelle et al (2003) define demand as utilized healthcare, using data for patients seen to measure demand. This essentially means that data that could be defined as the supply of health care was being used for demand. However, following greater data availability, Martin et al (2007) was able to use patient utility to model demand as the data could be separated into both inpatient and outpatients, referrals, and admissions. This meant that demand and supply could be modelled separately which Martin et al (2007) argues produces more accurate models.

Whether modelled together or separately, the literature generally agrees that supply variables are important considerations when determining demand (Blundell and Windmeijer 2000, Gravelle et al 2003 and Martin et al 2007 and Oliveria 2007). The exception is Samsudin (2010) who looked to identify the determinants of demand for health with emphasis on the influence of personal characteristics, socio-economic factors, supply, and health variables. Supply variables were not included within their models however they indicate that this may be the result of data availability rather than assuming they are not relevant.

Blundell and Windmeijer (2000), Gravelle et al (2003) and Martin et al (2007) all use waiting time as a replacement for price in demand theory, which is also used in this report. However, Blundell and Windmeijer (2000) use waiting time as the dependent variable when analysing health care utilisation which may indicate endogeneity issues between the independent variables (meaning they are not truly independent). They addressed this by employing lags to the dependent variable. Interestingly, Oliveria (2007) did not include waiting time as a variable but had included costs of alternative treatments instead. However, this study modelled utilisation of elective care within the Portuguese health care system rather than England, so this appears to be a closer approximation of price.

The issues of endogeneity are also expressed in Gravelle et al (2003) in their study of utilisation of different admission types for small geographical areas in England, Oliveria (2007) and Samsudin (2010). They all suggest using a two-stage model to deal with this but are unclear on how they decided which variables to use as instruments. Gravelle et al (2003) went further to state the identification of instrumented variables was difficult with almost infinite combinations and so employed fixed effects to their models instead. More variables and difficulties in determining endogeneity increases the chances of autocorrelation between the variables and biased estimates in this report. Statistical testing will be used to guide the model specification in this report.

Brailsford et al (2004) argues for the inclusion of emergency care data when modelling the elective care system as the two systems share resources. They find that problems in emergency care are closely linked to high volumes of demand and pressure in elective care. This study was limited to one hospital and modelled emergency care, but the importance of including emergency data is also supported by Gravelle et al (2003). Brailsford et al (2004) finds that pressures on staffing and bed occupancy were influential supply factors when modelling demand for emergency services and should be considered with elective care too. This fits with the other studies advocating for the inclusion of supply variables mentioned above.

All the papers reviewed used smaller samples than this report is attempting. They either use specific geographical areas (Oliveria 2007), specific hospitals (Brailsford, 2004), treatment types (Martin et al 2007), or types of admissions (Gravelle et al, 2007). This contrasts with what is being attempted in this report, as data has been collected for elective care across England. Usually, larger data samples provide more robust estimates but with panel data, the possibility of heteroskedasticity increases as the potential for variance between the residuals is increased. Robust standard errors will be used in this report to try and mitigate this.

Using smaller samples introduces debate between the papers about the inclusion of patient characteristics. Looking at individual hospitals, wards, or treatments, particularly in terms of utilisation, using patient characteristics would seem sensible to understand who accessed the services. Oliveria (2007) and Samsudin (2010) hold the opposite views with the former advocating for geographical factors as driving forces, and the latter stating that socio-economic characteristics had little impact. These studies cover larger geographical areas than others though, but this report will follow suit and not include patient characteristics.

- **Results from literature**

Using waiting time as a proxy for price as an independent variable, Gravelle et al (2003) and Martin et al (2007) provided results of the responsiveness of demand. Gravelle et al (2003) estimates that the basic models with fixed effects show that a 1 unit increase in waiting time leads to a fall of 0.094 units in health care utilisation (acute wards in this case). However, they emphasis that these models while providing an intuitive interpretation, were driven by the RESET test and R-squared values of the regression, are mis-specified.

Martin et al (2007) developed separate demand and supply models of two elective treatment types using quarterly data from 200 hospitals over seven years between 1995 and 2002. They argued combining supply and demand led to misleading estimations of the variable coefficients, impacting interpretations. The data was split between inpatients and outpatients, and two model types were presented – Ordinary Least Squares (OLS) and Seemingly Unrelated Regression (SUR) with static and dynamic effects. The static demand models (no lagged dependent variable) for routine surgery for inpatients produce elasticities of -0.198 (OLS) and -0.069 (SUR), and for outpatients at -0.034 (OLS) and -0.063 (SUR), showing that demand is inelastic. Using SUR models was preferred as OLS models appeared to over-estimate the coefficients. For simplicity, this report will not model inpatient and outpatient demand separately.

- **Contribution and expectations of this report**

This report will add to the body of literature by following the example of Martin et al (2007) in separating the data for demand and supply but using more recent data to look at demand between April 2016 and November 2023. This report will differ from previous studies as it attempts to look at all elective care within England.

Recommendations from literature for including supply and emergency care variables will be followed within the modelling, along with quality or performance indicators to reduce omitted variable bias. All the papers discussed the complex nature of demand and supply, so this paper will attempt to model the two together by including supplied health care data to start with. If this approach does not produce robust results, the approach of Martin et al (2007) in separating demand and supply models will be followed but with combined inpatient and outpatient data.

The expectation is that demand in health care has remained inelastic in response to waiting times, however how inelastic remains to be seen, and it is unclear whether other factors also appear statistically significant.

DATA, VARIABLES AND QUALITY

i) Data

The data for this project was collated from several monthly statistical publications from NHS England, and NHS Digital. These monthly publications were downloaded, filtered by organisation code or hospital trust as appropriate, and the relevant variables copied to a master workbook. Each variable was copied to a different page in the workbook to avoid mismatching trusts data, and so it could be cleaned correctly before being combined into a summary spreadsheet with pages for each period.

The data consists of 117 general and acute hospital trusts in England from April 2016 to November 2023 to create a panel set and was analysed as three time periods.

April 2016 to February 2020 – pre-pandemic.

March 2020 to February 2022 – pandemic.

March 2022 to November 2023 – post-pandemic.

Splitting into three periods meant that the COVID-19 pandemic could be covered separately and would not disproportionately impact the post-pandemic period.

ii) Variables

Literature and available published data informed variable selection, and they can be split into demand, price and supply as required by economic theory. Table 1 lists the variables, model notations, definitions, calculation methods and data sources.

Table 1: Variables covering 117 hospital trusts between April 2016 and November 2023

VARIABLE	SHORTHAND	DESCRIPTION	SOURCE
Total Referrals Made	TOTAL_DEMAND	The number of patients that have received referrals to secondary care – i.e. added to the waiting list. Sum of GP (specific acute) and Other Referrals Made (specific acute) together	April 2016 to May 2020 **Statistics (2019)** Monthly Hospital Activity. June 2020 to November 2023 **NHS England (n.d.c)** - Monthly Outpatient Referrals Data
Decision to Admit	DTA	The number of patients receiving notification that they will be admitted for treatment.	Consultant-led Referral to Treatment Waiting Times - **NHS England (2019)**
Total Patients Seen	TOTAL_SUPPLY	The total number of inpatients and outpatients seen per month. (The amount of health care supplied). Sum of inpatient admissions and outpatients seen.	April 2016 to May 2020 **Statistics (2019)** Monthly Hospital Activity. June 2020 to November 2023 **NHS Digital. (n.d.a).** Hospital Episode Statistics
Average Median Wait Time (days)	AVG_WAIT	The expected average time in days spent on the waiting list.	Consultant-led Referral to Treatment Waiting Times - **NHS England (2019)**
Staff sickness absence rate	ABS_RATE	The proportion of NHS hospital trust staff absent due to sickness.	NHS Sickness Absence Rates - **NHS Digital (2022a)**
Deaths in Surgery	DEATHS_RATE	The proportion of deaths within 30 days of being discharged from surgery including elective care procedures.	Summary Hospital-Level Mortality Indicator - **NHS Digital (n.d.b)**

Full Time Equivalent Staff	FTE_STAFF	The number of full-time equivalent NHS hospital staff.	NHS Workforce Statistics - **NHS Digital (2022b)**
Bed Occupancy Rate - Day	OCC_DAY_RATE	The proportion of day beds occupied - used by patients not requiring an overnight stay.	Bed Availability and Occupancy Data – Day only - **NHS England (n.d.a)**
Bed Occupancy Rate - Overnight	OCC_NIGHT_RATE	The proportion of overnight beds occupied - used by patients staying overnight.	Bed Availability and Occupancy Data – Overnight - **NHS (2023)**
Day Case Rate	DAY_CASE_RATE	The proportion of patients admitted for day treatment only out of total admissions.	April 2016 to May 2020 **Statistics (2019)** Monthly Hospital Activity. June 2020 to November 2023 **NHS Digital. (n.d.a)**. Hospital Episode Statistics
Emergency Admission Rate	EM_ADM_RATE	The proportion of patients admitted from emergency care from total admissions. Calculated from total admissions and total non-elective admissions	April 2016 to May 2020 **Statistics (2019)** Monthly Hospital Activity. June 2020 to November 2023 **NHS Digital. (n.d.a)**. Hospital Episode Statistics

Additional variables were removed as the models developed and are not included in the table. For reference, this includes the total number of patients on the elective waiting list (TOTAL _PATHWAYS, NHS England 2019), total number of outpatients seen (TOTAL_SEEN, Statistics (2019) and NHS Digital (n.d.a)), and total number of patients admitted (inpatients) (NO_ADM, Statistics (2019) and NHS Digital (n.d.a)).

Other variables from literature were considered, but due to data quality or self-reported bias they have not been included.

Data includes all non-specialised hospital trusts (one or more hospitals), thus capturing demand for health care and supply of health care throughout England over ninety-two months. Population characteristics relevant to individual trusts or over more time have not been included representing a departure from literature. (Blundell & Windmeijer 2000, Gravelle et al 2003, Brailsford et al 2004).

iii) Dependent and Independent

The initial dependent variable was TOTAL_PATHWAYS (all existing and new referrals), representing the measurable total demand for health care (Gravelle 2003, and Martin et al 2007). However, as this project developed, the modelling was split into demand and supply (Martin et al, 2007), with the health care demanded (TOTAL_DEMAND) and the health care supplied (TOTAL_SUPPLY) as the dependent variables.

Following demand theory, the "average waiting time in days that a patient is expected to be on the waiting list" (AVG_WAIT) represents the price mechanism in the regression (Martin et al 2007, Gravelle et al 2003).

The remaining independent variables act as controls to the market which could affect both the demand for and supply of health care and have been guided by literature discussed above. They could influence a patients' decision to access health care, the practitioners (economic agents) referring them, or constraints to the supply of health care supplied. These controls are important to reduce omitted estimate bias and improve the robustness of the models by capturing their relationship with demand and supply.

DAY_CASE_RATE was not included in the initial regression models due to suspected overlap, however it was subsequently added as a further control variable.

iv) Data Quality

Issues of data quality needed to be addressed to ensure the robustness of the model.

- **Monthly v Quarterly**

Bed occupancy and part of the deaths in surgery data was published as quarterly rather than monthly. For these variables, the quarterly data was assumed to apply to all months within the quarter, thereby creating a monthly series to match the other variables. It is acknowledged that this may have slightly skewed the results.

- **Mergers**

Between April 2016 and November 2023, several hospital trusts merged to form either a 'new' trust with a new name and organisation code, used an existing code and new name, or kept one of the original codes and names. To deal with the mergers, the final list of trusts was taken from the November 2023 data, and all mergers were assumed to have applied prior to April 2016. Trusts were reconciled using the organisation codes with the November 2023 names and codes applied. Original codes and names were removed.

For whole numbers, the figures for each trust were added together prior to the merger date. For percentages or averages, the figures for each trust were added then divided by the number of trusts involved in the merger.

- **Specialty trusts**

The publications included NHS trusts that provide specialist services, such as community, mental health, cancer treatments, orthopedics, or children's services. These trusts have been removed to avoid skewing results as they do not function like general and acute (non-specialised) trusts.

- **Pandemic Data**

During the pandemic, hospital trusts had to be organized differently due to infection control procedures in place to limit the spread of COVID-19. This meant that hospital capacity and bed occupancy could not be measured in the same way, and can't be compared with previous years (NHS, n.d.). Thus, the variables for bed occupancy have been removed from the pandemic model.

- **Missing Data**

There are months where trusts do not report one or more variables creating an unbalanced panel. Due to the size of this panel, it seemed unlikely that the missing values would affect the overall models and no attempt has been made to estimate or substitute those points. The modelling software also takes account of missing variables.

MODELLING

Regression analysis was performed on each period using econometric techniques in GRETL statistical software following the same methodology. Robustness tests guided the model selection, particularly specification (RESET), heteroskedasticity (White's), autocorrelation (Wooldridge), normality tests, testing for omitted variables (Wald's test) and where applicable, the Hausman test for consistency/bias. Where the p-values of the tests were less than 0.05, the null hypothesis was rejected. All periods started with the assumption that the relationship between the variables could be explained using Ordinary Least Squares (OLS) linear regression with the variables in their level forms.

Frequency distributions of the variables found them all to be skewed, so each variable was transformed into log form. The final models are analysed as log-log regressions and the coefficients can be read as elasticities.

i) Specifications

The initial modelling assumed that demand and supply could be captured within one model using TOTAL_PATHWAYS as the dependent variable. Several model specifications were attempted, but they lacked robustness, and no preferred models could be identified. An overview of the issues is covered under Methodology – part 1 below.

Following Martin et al 2007, demand and supply were split into two models and considered separately, but with two simplifications:

Inpatient and outpatient data was combined (as seen in Gravelle 2003).

Fewer variables were included - no patient characteristics or geographical considerations.

Although both demand and supply were modelled, only the specifications for demand and issues encountered are discussed in Methodology - part 2 below.

The following final models have been identified.

Pre-pandemic:

$$\log(TOTAL\_DEMAND) = \alpha + \beta_1 \log(AVG_{WAIT})_{it} + \ \beta_2 \log\,(FTE\_STAFF)_{it} + \beta_3 \log\,(EM\_ADM\_RATE)_{it} + \beta_4 S_t + \ u_i$$

Post-pandemic:

$$\log(TOTAL\_DEMAND) = \ \alpha + \beta_1 \log(AVG\_WAIT)_{it} + \ \beta_2 \log\,(FTE\_STAFF)_{it} + \beta_3 S_t + \ u_i$$

Both models are Least Squares Dummy Variable specifications, where trust differences were held fixed, and with unidentified time dependent variables ($\beta_n S_t$). As discussed later, a pandemic model specification could not be identified.

ii) Methodology – Part 1

OLS models in log-log forms gave high R-squared results with most variables indicating they were statistically significant for all periods. However, robustness testing of the models indicated that none were of an adequate specification, tall had heteroskedasticity, autocorrelation, and that none had errors distributed normally despite logs of the variables taken.

Least Squares Dummy Variable (LSDV) models with both fixed and random effects were considered to account for potential omitted variable bias and address issues of heteroskedasticity and misspecification in the OLS model.

Fixed Effects (FE) models attempted to address any omitted variable bias and account for unobserved differences between the trusts that did not change over time but are correlated with other variables within the model. Although they exhibited high R-squared values, heteroskedasticity and autocorrelation remained. The Wald test confirmed that there were time dependent variables missing from the models. Removing any insignificant variables had no effect on heteroskedasticity or autocorrelation and affected the significance of the remaining variables raising questions of endogeneity.

A Random Effects (RE) model was compared against the FE model. This model differs from the FE model as it assumes that the unobserved differences between the trusts are not correlated with the independent variables. It uses the Hausman test to assess whether it is preferred over the FE model and indicate bias in the model. The null hypotheses of estimates being consistent were rejected. The estimates were not consistent and there was bias in the models. Removing statistically insignificant variables did not improve the Hausman test.

FE models with time dummies were preferred, but both specifications were rejected due to bias in the estimates, and poor robustness testing. This is not unexpected as was consistent with literature (Blundell & Windmeijer 2000 and Gravelle et al 2003). The bias suggests there is endogeneity within the models.

To deal with the bias, two stage least squares (2SLS) regressions were constructed. The 2SLS substitutes the endogenous variables with fixed estimates thus removing the correlation with the error term and increasing the reliability of the remaining coefficients by reducing the likelihood of overestimated. However, there are issues identifying which direction the endogeneity flows and which variables should be used as instruments due to how demand and supply are linked in health care (Gravelle, 2003).

For simplicity, the following variables were assumed to be exogenous (not affected by other variables in the model): TOTAL_MADE, DTA, ABS_RATE, DEATHS_RATE, OCC_DAY_RATE, OCC_NIGHT_RATE and EM_ADM_RATE. The remaining variables were assumed to be endogenous and used as instruments within the regression. Lags to DTA were attempted as patients can be admitted in a later calendar month after receiving an admittance decision.

The 2SLS had mixed results. The Sargan test indicated the pre-pandemic model was not over-identified, but the others were. The Cragg-Donald test indicated that all three periods contained weak instruments within the models (test statistic of less than 10), and the Hausman test of OLS estimates being consistent could not be rejected for any of the models. This suggested that 2SLS was not the appropriate specification. Due to the difficulties of correctly identifying the appropriate instruments per literature, for simplicity, 2SLS was rejected.

As total demand for health care would rely on existing demand as well as new demand, an autoregressive model was tested by lagging the dependent variable in the models. This can also help with issues of autocorrelation and misspecification. Keele and Kelly (2005), argue that lagged dependent variables can be used when the system being modelled is considered dynamic – or rather they can be used where past information also informs the future.

Including lagged variables improved test results for autocorrelation and misspecification, particularly for the pre-pandemic data. However, the magnitudes of the coefficients are tiny, with most less than 0.001. With all variables in log form, the coefficients represent elasticities between the dependent and independent variables, but due to their size, interpretation was difficult. This model type could be useful for forecasting future demand as used in Shah et al 2024, but that is beyond the scope of this project. This model type was rejected, though acknowledged that it satisfied more robustness checks than other specifications.

iii) Methodology – part 2

The modelling switched to the approach of Martin et al (2007) of modelling demand and supply separately. The dependent variable for demand changed to TOTAL_DEMAND (monthly referrals). The independent variables remained the same. An additional independent variable was added – proportion of elective admissions dealt with the same day (DAY_CASE_RATE) – to further account for differences in the types of admissions.

The specification choices were again guided by robustness testing, and all regressions were run using robust standard errors to attempt to correct for heteroskedasticity.

- **Pre-pandemic**

The initial model contained all independent variables producing an R-squared of 81.4%, with all variables appearing statistically significant. The magnitudes of the coefficients were more consistent with literature with the smallest at -0.47 for ABS_RATE and the largest at 0.68 for FTE_STAFF. However, the RESET test had a p-value of less than 0.05 indicating the specification was not adequate. No further testing was undertaken.

When time and trust dummies were added as per indications seen in previous models, the R-squared increased to 98.3%, though only three variables (AVG_WAIT, FTE_STAFF and EM_ADM_RATE) remained statistically significant. The RESET test had a p-value of 0.51 indicating the specification is adequate. The Wald test was run against the insignificant variables, the time dummies, and the trust dummies. This showed that the insignificant variables had no explanatory power within the model (p-value greater than 0.05), but that the time and trust dummies did. Re-running the model with the insignificant variables removed led to a reduced R-squared and slight increase in the standard errors of the regression, however the RESET test p-value increased to 0.999 suggesting this is a stronger model specification.

A FE model was run holding the differences between trusts constant over time. This increased the significance of the AVG_WAIT variable, increased the R-squared back to 98.3%, and reduced the standard errors of both the variables and the overall regressions. The Wald test confirmed that the time dummies still had explanatory power within the model. Heteroskedasticity and autocorrelation are present in the model, and the residuals are not distributed normally. Using lagged variables as proposed by Keele and Kelly (2005) were included to address autocorrelation but did not resolve the issue and were removed.

Although heteroskedasticity and autocorrelation are present in the model, and the residuals are not distributed normally, an FE model with time dummies is preferred for this data set.

- **Pandemic**

The same approach was followed for the pandemic data excluding the bed occupancy variables.

The initial OLS model specification was inadequate, producing RESET test p-values of below 0.05. No further testing was conducted. Adding time and trust dummies to the model increased the R-squared to 96.5%, but the RESET test p-value was still less than 0.05.

Only two variables appear to be statistically significant (AVG_WAIT and ABS_RATE). The Wald test confirmed that the remaining variables appeared to have no explanatory power within the model, but the time and trust dummies did. However, removing the insignificant variables from the OLS model did not produce an adequate specification according to the RESET test. It appears that OLS models are not suitable for this data, which may be reflective of the different pressures on the health care system during the pandemic.

Both FE and RE models were attempted, with the FE model yielding the same R-squared and standard error for the regression as the OLS model with time dummies included. The RE model was run and initially showed three variables as significant (AVG_WAIT, FTE_STAFF and ABS_RATE). The Wald test confirmed that the remaining variables had no explanatory power and could be excluded from the model. The Hausman test produced a high chi-squared statistic and low p-value indicating that the GLS estimates are not consistent and there is bias and endogeneity within the model.

Due to the difficulties in addressing endogeneity discussed earlier, for simplicity, no further modelling was attempted, and a preferred model has not been identified.

- **Post-pandemic**

Bed occupancy variables have been reinstated.

The first OLS model yielded an R-squared of 75.9% for the regression, but the RESET test reported a p-value of less than 0.05. Adding time and trust dummy variables increased the R-squared value to 98.5%, and the RESET test had a p-value of 0.71 indicating that the specification is adequate. However, only two variables appeared statistically significant, AVG_WAIT and FTE_STAFF. The Wald test was run against insignificant variables, and with a p-value of 0.28, it can be inferred that they have no explanatory power within the model. The Wald test was also run against time and trust dummies, finding that they did have explanatory power.

Removing the insignificant variables reduced the R-squared to 98.3%, but the RESET test p-value increased to 0.98. However, the magnitudes of the coefficient's seem over-inflated. AVG_WAIT records a coefficient of -0.199, which is consistent with the findings of Martin et al 2007 when modelling demand using an OLS model. FTE_STAFF though, records a coefficient of 1.619 which seems overestimated. While possible that this is an impact of the health care system recovering after the pandemic (more people presenting to their GP to receive a referral, or GPs being more accessible after infection control measures were lifted), it seems unrealistic.

To account for any omitted variable bias, the FE and RE models were tested again. The FE model reported the higher R-squared value of 98.5%, while maintaining a low standard error for the regression at 0.07. AVG_WAIT and FTE_STAFF remained statistically significant after the insignificant identified variables had been removed, however their coefficients had changed. AVG_WAIT was slightly lower at -0.146, but FTE_STAFF had fallen to 0.471, which appears more sensible for interpretation. The Wald test confirmed that there are likely missing time dependent variables in the model, and the time dummies should remain. Testing confirmed the presence of heteroskedasticity and autocorrelation.

The RE model indicated that the same variables as OLS and FE models were statistically significant, however, the Hausman test statistic reported a p-value of less than 0.05, indicating that there is bias in the model, and that FE would be preferred. Removing the insignificant variables did not improve the Hausman test statistic, and thus the FE model is the preferred model.

The FE model with time dummies is the preferred model, though it is acknowledged that heteroskedasticity and autocorrelation are present.

RESULTS AND INTERPRETATION

Literature suggested there would be difficulties with the model specifications, and this has been evident throughout this project. The final model specifications for pre-pandemic and post-pandemic data were the Least Squares Dummy Variables models, holding trust differences constant, and including time dummies.

No model specification was identified for pandemic data.

i) Results

Table 2 below sets out the results of the preferred models identified for each of the three periods described in this report. Due to the number of time dummy variables, the coefficients for these have not been included, but the Wald Omitted Variables test results to confirm their explanatory power within the models are reported. Any variables that have been omitted from the final models have also been excluded.

Table 2: Statistical results of the final models

VARIABLE	PRE-PANDEMIC	PANDEMIC	POST-PANDEMIC
L_AVG_WAIT	-0.073** (0.036)		-0.146*** (0.044)
L_DTA			
L_FTE_STAFF	0.217* (0.122)		0.471* (0.241)
L_ABS_RATE			
L_DEATHS_RATE			
L_OCC_DAY_RATE			
L_OCC_NIGHT_RATE			
L_EM_ADM_RATE	-0.184** (0.102)		
L_DAY_CASE_RATE			
TIME DUMMIES	Wald Test p-value: **5.51e-99**		Wald Test p-value: **4.48e-78**
R-SQUARED **“Goodness-of-fit”**	98.3% (0.983)		98.5% (0.985)
STANDARD ERROR	0.068		(0.067)
Variable Significance: * = p<0.1, ** = p<0.05 and *** = p<0.01			

It was not possible to identify a preferred model for the pandemic period and so the coefficients have been excluded. This is likely due to several factors:

The health care system had to be organised differently both in terms of resources (staff, beds, equipment, wards), and for infection control purposes (NHS, n.d.).

People’s health seeking behaviours changed in response to the pandemic e.g. some couldn’t access their GP or chose to not seek health care at that time (SAGE, 2021).

Data quality – reorganization of health care may have affected how, when or if data was collected.

While this means it isn't possible to make any comparisons for this period, it does show that splitting the data to separate the pandemic was a good approach to avoid the remaining results being skewed.

ii) Interpretation

All variables have been logged, meaning that the coefficients can be interpreted as elasticities.

For **pre-pandemic**, three variables identified as statistically significant, AVG_WAIT, FTE_STAFF and EM_ADM_RATE. If the average waiting time increased by 1%, the coefficient indicates that the demand for health care as measured by referrals made would fall by 0.073%. This is a small % and shows that demand for health care in England is very inelastic and shows little response to changes in the average waiting time. This estimate is close to that of Gravelle et al (2003) of 0.094. However, unlike Gravelle et al (2003), this model split demand and supply to give a more accurate estimate (Martin et al 2007). It is acknowledged that this chosen specification is likely overestimating the coefficient.

For staff, if the number of full-time staff increased by 1%, the demand for health care would increase by 0.217%. While this seems to be heading in the wrong direction, it may be an indication of how accessible health care is. As the percentage is less than 1, the variable is considered inelastic, however, it suggests that more staff could generate more patient referrals, meaning demand has increased (Rosen, 2014). There is no indication in the model where those additional staff members are within the hospital system. If the staff levels increase for consultants and A&E teams which are both alternative sources of referrals, it is feasible that more staff will review more patients and may then refer them to the elective care system for treatment (Brailsford et al, 2004). This could be an example of supplier-led demand (Rosen, 2014), particularly as patients rely on health care professionals when making decisions about their care. Thus, the variable makes statistical and economic sense.

The emergency admission rate also showed as statistically significant and indicated that a 1% increase in emergency admission rates would lead to a reduction in the referrals made of 0.184%. With the percentage being less than 1, this variable is again inelastic. The direction of the coefficient is reasonable as emergency care will share the same resources as elective care (Brailsford et al, 2004). This variable also makes statistical and economic sense.

Post-pandemic, the coefficient for AVG_WAIT indicates that if the average waiting time increased by 1%, the demand for health care would fall by 0.146%. This shows that demand post-pandemic remains inelastic, although slightly more responsive to changes in waiting time than before the pandemic. This is closer to the estimates given in Martin et al 2007 (between 0.198 and 0.034), though acknowledged that it could be overestimated due to selected model specification.

Full time staff was again statistically significant, but the coefficient had increased to 0.471, meaning the total number of referrals is more responsive to an increase in full-time staff than prior to the pandemic. An increase of 1% in full-time staff would lead to an increase in demand of around 0.471%. This can again be interpreted in terms of accessibility to health care. Survey data showed that over the course of the pandemic, many people chose not to seek health care through fear of overwhelming the NHS or struggled to get access to their GP for referrals (SAGE, 2021). It is estimated that around 10 million people did not seek health care during the pandemic period (Shah et al 2024). To meet an increase in demand, you need more resources including staff. It is therefore feasible that increasing staff increases the accessibility of health care, and thus would increase the demand for health care. In the supply side of the equation, it would also be expected that increasing staff would lead to an increase in the number of patients seen. For the total waiting list, an increase in staff should reduce the overall list as more patients can be seen at any one time.

Emergency admissions are not statistically significant in the post-pandemic model, which is surprising as if an estimated 10 million people did not seek health care, it is possible they may enter the system as emergency admissions.

Overall, the results confirm the previous literature expectations and hypothesis that demand for elective care in England has remained inelastic in response to changes in average waiting times. It further indicates that full-time staff and emergency admissions are statistically significant when modelling demand and from the size of the coefficients may be more influential.

iii) Limitations

The models have some positives, such as a high R-squared score implying a 'good-fit' to the data, and low standard errors for the variables and models. The panel data was large which usually yields more accurate results, and the final model specifications did satisfy the RESET test indicating the specifications were adequate. However, as discussed throughout the report, there are limitations to the results.

- The models lack robustness – despite efforts to address heteroskedasticity using robust standard errors and autocorrelation using lagged dependent variables, testing indicated the presence of both in all models. Normality issues also persisted despite logging the variables to address this.
- The coefficients are likely overestimated due to inadequate model specifications (Martin et al 2007).
- The models use fixed effects to address omitted variable bias in relation to differences with trusts, but the Wald test indicated that there were time varying factors missing from the model which likely impact the results.
- The data set was constructed by the author, which is a positive in that the data was chosen specifically for this report and quality issues could be controlled for. However, the Wald tests indicated that there are missing variables from the model, so the data set could be improved. Further consideration should also be given as to whether the metrics for the current variables are the most appropriate – would growth rates have been better?
- It is likely there are trust dependent variables also missing from the model, but the statistical software detected multi-collinearity when including trust dummies in the final model, and this could not be tested for.
- Endogeneity was highlighted in several research papers discussed in the literature review, and it is highly likely it has remained in the final models. Attempts to utilize Two-Stage models to address this were also unsuccessful, although it was also clear in literature that identifying the variables to be instrumented was difficult (Gravelle et al 2003).
- The panel set is large, involving 117 trusts for 21-47 months at a time across 7-9 independent variables, but it is possible data sets over longer periods would be more

accurate, particularly in relation to post-pandemic data, as this may still be influenced by the pandemic.

- Modelling demand on a national scale may have been inappropriate, and a better approach would have been to consider smaller geographical areas or treatment areas. This may have enabled patient characteristics to have been incorporated reducing omitted variables bias.

The limitations demonstrate the difficulties of modelling the health care system in England, and further evidence that the health care does not behave like a typical good as described by Morris et al 2012. Overall, the difficulties in identifying suitable model specifications show the complexity of the relationships between variables. Despite having some positive traits, the specifications are not ideal at present and require further development to be useful to policy.

CONCLUSION AND FUTURE DISCUSSION

This report reinforced difficulties of modelling demand for health care discussed in the literature - overestimation, endogeneity, poor specifications. The results support the original hypothesis and literature findings (Martin et al 2007, and Gravelle et al 2003) that demand for elective care is inelastic with respect to changes in average waiting time. However, the models lack the robustness needed to be reliable, having failed several statistical test discussed in the limitations section above.

As the models are unreliable, there are no realistic policy recommendations at present, although the results do support the assertion made by Brailsford et al (2004) of the importance of considering emergency care when investigating elective care.

They also support the approach of Martin et al (2007) of splitting demand and supply data to give more robust results.

The results tentatively suggest that staff numbers also play a role given that this was statistically significant in both identified models. The reports of high staff vacancies particularly in nursing suggests this is not far-fetched (BMA, 2024), and Brailsford et al (2004) indicated staff as a resource pressure point.

However, there are several options for further research:

- Breaking the data into specific treatment functions, by region, Integrated Care Board or even trusts, may produce more statistically robust results, and allow population characteristics to be incorporated as per some literature.
- Investigating which variables could be considered truly exogenous or endogenous, as well as more time spent considering the appropriate data lags to some variables may also prove advantageous.
- Considering how this could link with forecasting future demand.
- Looking at interactions with emergency care, primary care, and social care.
- Adding more variables such as bed discharge data to identify more unobserved effects.

Regardless of the limitations, the report shows that the economics of health care is a complex (and interesting) issue, with plenty of options available for further research.

REFERENCES

Blundell, R. and Windmeijer, F. (2000). Identifying demand for health resources using waiting times information. Health Economics, 9(6), pp.465–474. doi: https://doi.org/10.1002/1099-1050(200009)9:6%3C465::aid-hec525%3E3.0.co;2-h.

BMA (2024) NHS Medical Staffing Data Analysis, British Medical Association. Available at: https://www.bma.org.uk/advice-and-support/nhs-delivery-and-workforce/workforce/nhs-medical-staffing-data-analysis (Accessed: 19 May 2024).

Brailsford, S.C., Lattimer, V.A., Tarnaras, P. and Turnbull, J.C. (2004). Emergency and on-demand health care: modelling a large complex system. Journal of the Operational Research Society, 55(1), pp.34–42. doi: https://doi.org/10.1057/palgrave.jors.2601667.

Cox, A., Chicksand, D. and Ireland, P. (2005). Overcoming demand management problems: The scope for improving reactive and proactive supply management in the uk health service. Journal of Public Procurement, 5(1), pp.1–22. doi: https://doi.org/10.1108/jopp-05-01-2005-b001

Findlay, R. (2021) Using referral-to-treatment (RTT) data to plan the Covid Recovery, Gooroo Blog. Available at: https://blog.gooroo.co.uk/2021/04/using-referral-to-treatment-rtt-data-to-plan-the-covid-recovery/ (Accessed: 10 May 2024).

Gravelle, H., Sutton, M., Morris, S., Windmeijer, F., Leyland, A., Dibben, C. and Muirhead, M. (2003). Modelling supply and demand influences on the use of health care: implications for deriving a needs-based capitation formula. Health Economics, 12(12), pp.985–1004. doi: https://doi.org/10.1002/hec.830.

Keele, L. and Kelly, N.J. (2006). Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables. Political Analysis, 14(2), pp.186–205. doi: https://doi.org/10.1093/pan/mpj006

Mankiw, N.G. (2018) Principles of Economics. Boston: Cengage.

Martin, S., Rice, N., Jacobs, R. and Smith, P. (2007). The market for elective surgery: Joint estimation of supply and demand. Journal of Health Economics, 26(2), pp.263–285. doi: https://doi.org/10.1016/j.jhealeco.2006.08.006

Morris, J. (2019) Elective care | Nuffield Trust, Nuffield Trust. Available at: https://www.nuffieldtrust.org.uk/news-item/elective-care [Accessed: 10 May 2024].

Morris, S. (2012). Economic analysis in health care. Chichester: Wiley

NHS Digital. (n.d.a). Provisional Monthly Hospital Episode Statistics for Admitted Patient Care, Outpatient and Accident and Emergency data. [online] Available at: https://digital.nhs.uk/data-and-information/publications/statistical/provisional-monthly-hospital-episode-statistics-for-admitted-patient-care-outpatient-and-accident-and-emergency-data [accessed 1 March 2023]

NHS Digital. (n.d.b). Summary Hospital-level Mortality Indicator (SHMI) - Deaths associated with hospitalisation. [online] Available at: https://digital.nhs.uk/data-and-information/publications/statistical/shmi [accessed 30 March 2024]

NHS England (n.d.a). Statistics» Bed Availability and Occupancy Data – Day only. [online] Available at: https://www.england.nhs.uk/statistics/statistical-work-areas/bed-availability-and-occupancy/bed-data-day-only/ [accessed 11 March 2024]

NHS England (n.d.b). Statistics» Bed Availability and Occupancy – KH03. [online] Available at: https://www.england.nhs.uk/statistics/statistical-work-areas/bed-availability-and-occupancy/bed-availability-and-occupancy-kh03/ [accessed 11 March 2024]

NHS England (n.d.c). Statistics» Monthly Outpatient Referrals Data. [online] Available at: https://www.england.nhs.uk/statistics/statistical-work-areas/outpatient-referrals/mrr-data/ [accessed 30 March 2024]

NHS England (2019). Statistics» Consultant-led Referral to Treatment Waiting Times [online] England.nhs.uk. Available at: https://www.england.nhs.uk/statistics/statistical-work-areas/rtt-waiting-times/ [accessed 1 March 2024]

NHS Digital (2022). NHS Workforce Statistics - NHS Digital. [online] NHS Digital. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/nhs-workforce-statistics [accessed 30 March 2024]