The Relationship between Health Expenditures and Diagnostic Waiting Times with Special Attention to Cancer

THESIS

MSc Health Economics, Policy & Law

J.M. Groen

Research Report

Rotterdam, 23rd of June, 2021

Student ID | 581198

Supervisor | C.S. Brindley

Word count | 19,955 (excl. references and appendices)

Abstract

Background Over the past decade, essential parts of the NHS England are experiencing the worst performance against waiting time targets since they were set in 2000. In light of the current Covid-19 pandemic, disrupting the provision of health care, these developments have become increasingly alarming. This study empirically investigates the effect of health spending on diagnostic waiting time performance. A panel data set was analysed covering a period of six years at the level of Clinical Commissioning Groups in England. A special interest goes to cancer, for which timely diagnosis and treatment are particularly essential in maximising health gains. This paper aims to contribute to the understanding of the channels through which the long waiting times for diagnostics in England can be reduced.

Methods Data on diagnostic waiting times have been derived from the NHS Statistics on Cancer Waiting Times and Diagnostic Waiting Times & Activity. Two cancer-specific waiting time targets were examine measuring the percentage of patients seen within two weeks after urgent GP referral (two-week-wait targets). Additionally, 15 key diagnostic waiting time targets (not specific to cancer) were studied representing the percentage of patients receiving a diagnostic test within six weeks after referral (six-week-wait targets). The dataset was analysed using the First Differences Instrumental Variables estimation technique. Health spending was instrumented using the distance from target index. The analytic sample contained 1,098 observations on 188 CCGs over a six year period (UK financial years 2014 tot 2019). Regional variation was examined and robustness checks were performed to assess the sensitivity of the results to alternative scenarios.

Results Both two-week-wait targets reported no significant effect of health spending. For eight key diagnostics a significant effect of spending was found on six-week-wait target performance. All endoscopy key diagnostic presented a positive coefficient on health spending. A 1% increase in spending correlated with an increase in six-week-wait target performance of 0.83 %-points ($p = .014$) for Colonoscopy, 0.70 %-points ($p = .012$) for Cystoscopy, 0.76 %-points ($p = .021$) for Flexible Sigmoidoscopy, and 0.82 %-points ($p = .010$) for Gastroscopy. Positive effects of spending were also found for key diagnostics Peripheral Neurophysiology and MRI. A 1% increase in health spending was associated with an increase in target performance of 0.68 %-points ($p = .005$) for Peripheral Neurophysiology, and 0.12 %-points ($p = .012$) for MRI. A negative effect of health spending was found for key diagnostic Audiology Assessments. A 1% increase in spending correlated with 0.34 %-points ($p = .002$) decrease in six-week-wait target performance. No significant effects were observed for the seven remaining key diagnostic waiting time targets. The findings were found to vary across regions.

Discussion This study has been the first to empirically investigate the link between health expenditures and diagnostic waiting times. Using the FDIV estimation technique allows for causal interpretation of the estimates. In line with previous work on waiting times for elective surgery, health spending was found to positively impact six-week-wait target performance for most key diagnostics. The largest effects were found for the endoscopy key diagnostics which are commonly used to diagnose intestinal cancers. Especially these diagnostics can maximise health gains by reducing long waiting times. Even though this study cannot uncover the exact channels through which health spending affects waiting times, it substantiates the importance of further research on how demand and supply characteristics can reduce excessive waiting times. Future work is encouraged to take a closer look at these diagnostic waiting times and build on the lessons learned in this study. It may be especially interesting to consider for future studies to consider regional variation and investigate the efficiency of health spending.

Table of contents

List of abbreviations

AA	AUDIOLOGY ASSESSMENTS (SIX-WEEK-WAIT TARGET)
BE	BARIUM ENEMA (SIX-WEEK-WAIT TARGET)
CCG	CLINICAL COMMISSIONING GROUPS
COL	COLONOSCOPY (SIX-WEEK-WAIT TARGET)
CT	COMPUTED TOMOGRAPHY (SIX-WEEK-WAIT TARGET)
CWT	CANCER WAITING TIMES
CYS	CYSTOSCOPY (SIX-WEEK-WAIT TARGET)
DEXA	DUAL ENERGY X-RAY ABSORPTIOMETRY (SIX-WEEK-WAIT TARGET)
DFT	DISTANCE FROM TARGET
DRG	DIAGNOSIS RELATED GROUP
DWT	DIAGNOSTIC WAITING TIMES
ECG	ECHOCARDIOGRAPHY (SIX-WEEK-WAIT TARGET)
EP	ELECTROPHYSIOLOGY (SIX-WEEK-WAIT TARGET)
FD	FIRST DIFFERENCE
FD IV	FIRST DIFFERENCE INSTRUMENTAL VARIABLES
FE	FIXED EFFECTS
FFS	FEE-FOR-SERVICE
FS	FLEXIBLE SIGMOIDOSCOPY (SIX-WEEK-WAIT TARGET)
GAS	GASTROSCOPY (SIX-WEEK-WAIT TARGET)
GDP	GROSS DOMESTIC PRODUCT
GP	GENERAL PRACTITIONER/PRACTICE
HMO	HEALTH MAINTENANCE ORGANISATION
ICS	INTEGRATED CARE SYSTEM
IMD	INDEX OF MULTIPLE DEPRIVATION
IoD	INDEX OF DEPRIVATION
IV	INSTRUMENTAL VARIABLE
LSDV	LEAST SQUARE DUMMY VARIABLE
MRI	MAGNETIC RESONANCE IMAGING (SIX-WEEK-WAIT TARGET)
NHS	NATIONAL HEALTH SERVICE
NOUS	NON OBSTETRIC ULTRASOUND (SIX-WEEK-WAIT TARGET)
OECD	ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT
OLS	ORDINARY LEAST SQUARES
OOP	OUT-OF-POCKET
PMI	PRIVATE MEDICAL INSURANCE
PN	PERIPHERAL NEUROPHYSIOLOGY (SIX-WEEK-WAIT TARGET)
POLS	POOLED ORDINARY LEAST SQUARES
RE	RANDOM EFFECTS
RP	RESPIRATORY PHYSIOLOGY (SIX-WEEK-WAIT TARGET)
SWW	SIX-WEEK-WAIT TARGET
TWW	TWO-WEEK-WAIT TARGET
TWWBS	TWO-WEEK-WAIT BREAST SYMPTOMS TARGET
UK	UNITED KINGDOM
URO	URODYNAMICS (SIX-WEEK-WAIT TARGET)
YLLs	YEARS OF LIFE LOST
ZCM	ZERO CONDITIONAL MEAN

List of tables & figures

TABLES

FIGURES

APPENDIX

Introduction

CONTEXT

For the past decades, long waiting times have been a major policy concern within the National Health Service (NHS) England. Since the 1990s, waiting times have dominated public and political debate in the UK. Starting in 2000, targets were set with the aim to guarantee maximum waiting times for health services (Boyle, 2011). These targets significantly reduced waiting times during the first decade of the 21st century (Propper, Sutton, Whitnall & Windmeijer, 2007). However, essential parts of the NHS are experiencing the worst performance against these waiting time targets since they were set. This persistent downward trend reflects the wider pressures the health and social care system endures. Long waiting times are a symptom of more people needing treatment than the NHS has the available capacity. This reflects a decade of significantly lower than average funding growth for the NHS and workforce shortages, coupled with growing and changing population health needs (Boyle, 2011; Thorlby et al., 2019). Additionally, these pressures are exacerbated by cuts to social care and public health budgets (Thorlby et al., 2019; Crawford & Emmerson, 2012). There is a widespread concern that capacity (concerning both clinical space and workforce) is not keeping up with current demand and ultimately affects patients (Brown et al., 2014).

For diseases like cancer in particular, where timely diagnosis and treatment are essential, this downward trend in waiting time performance is extra alarming. Timely diagnosis may allow cancer to be identified at a treatable stage and prevent complications (Hamilton, 2020). This reasoning has even underpinned the design of the UK health care delivery for decades; aiming to diagnose at least 75% of cancers at stage I or II (i.e. potentially curable) by 2028, up from approximately 53% in 2018 (Hamilton, 2020). Screening programmes for cancer only identify 10% of adult cancer cases, leaving the remainder of patients to present with symptoms. Most of these patients visit their GP and are then offered an urgent referral using the two-week-wait (TWW) system, which calls for specialist input within 14 days (Hamilton, 2020). Since 2012, a increasingly downward trend is observed for two-week-wait target performance. As of 2018, performance has even dropped below the operational standard of 93% of patients seen within the target (Nuffield Trust, 2021).

Another important pathway is the six-week-wait (SWW) Diagnostic Waiting Time target. Timely diagnosis is important to patients and essential in improving health outcomes. The SWW target was introduced for 15 key diagnostics[1]. The operational standard is set at 99% of patients who should receive a diagnostic test within six weeks after referral. Since 2014, the percentage of people waiting for more than six weeks has been increasing. From 2018 onwards, target performance scores are declining even more rapidly (Nuffield Trust, 2021b). The six-week-wait targets encompass imaging, physiology, and

[1] 1. Magnetic Resonance Imaging 2. Computed Tomography 3. Non-Obstetric Ultrasound 4. Barium Enema 5. Dual Energy X-ray Absorptiometry 6. Audiology Assessments 7. Echocardiography 8. Electrophysiology 9. Peripheral Neurophysiology 10. Respiratory Physiology 11. Urodynamics 12. Colonoscopy 13. Flexi Sigmoidoscopy 14. Cystoscopy 15. Gastroscopy.

endoscopy diagnostics which are used to identify a wide range of health conditions. Some of these key diagnostics are important for cancer diagnosis as well. Not all patients present with symptoms that result in an urgent TWW referral by their GP. The remaining share of cancers is likely to be identified within the six-week-wait diagnostic pathway (NHS England, 2015).

As such, these waiting time trends are distressing. Yet, in light of the current Covid-19 pandemic, their significance becomes even more explicit. The pandemic crucially adds to the existing pressures on capacities and health needs. In the first months after the coronavirus outbreak, TWW performance decreased to 90% and 77%, respectively, for GP urgent referral when cancer is suspected and in case of breast symptomatic (when cancer is not initially suspected). At the end of 2020, these scores even dropped to 88% and 70%, respectively (Nuffield Trust, 2021). Also, SWW target performance declined drastically, reaching 42% on average in May 2020. After a substantial recovery, but still far from reaching the operational standard, SWW performance was measured at 71% in January 2021 (Nuffield Trust, 2021b).

The sharp increase in cancer diagnostic waiting times has far-reaching health consequences, which are highlighted by several modelling studies on cancer deaths resulting from the delays due to Covid-19 in England. Sud et al. (2020) found that delays in the TWW pathway over a three-month lockdown period would result in 181 - 542 additional lives lost and 3,316 - 9,948 YLLs for respectively a 25 - 75% backlog in TWW-referrals[2]. Also, a delay in additional diagnostic capacity with provision spread across months 3 - 8 after lockdown would result in 401 – 1,231 additional lives lost and 7,332 - 22,635 YLLs under respectively a 25 - 75% backlog scenario[3]. A two-month delay in TWW investigatory referrals resulted in an estimated loss of 0.0 - 0.7 life-years per referred patient, depending on age and tumour type (Sud et al., 2020). The authors argue that a prompt provision of additional capacity to address the backlog of diagnostics will minimise deaths as a result of diagnostic delays that could add to those predicted due to expected presentational delays (Sud et al., 2020). Another study conducted by Maringe et al. (2020) estimated an increase of 7.9 - 9.6% in the number of deaths due to breast cancer up to five years after diagnosis compared to pre-pandemic figures. For colorectal cancer, they estimated a 15.3 - 16.6% increase; for lung cancer a 4.8 - 5.3% increase; and for oesophageal cancer, 5.8 - 6.0% increase up to 5 years after diagnosis. They estimated the total additional YLLs across these cancers to be around 59,000 – 63,000 years.

RELEVANCE & RESEARCH QUESTION

The two modelling studies mentioned above are just a selection of the expanding evidence on the potential health losses as a result of diagnostic delays. The two studies present different figures reflecting their different methods, cancer sites and assumptions. However, Hamilton (2020) argues that perhaps a precise figure is not needed, the loss of health (and life) as a consequence of delays in diagnostics is

[2] 25% backlog in referrals: 181 additional lives and 3,316 life-years lost. 50% backlog in referrals: 361 additional lives and 6,632 life-years lost. 75% backlog in referrals: 542 additional lives and 9,948 life-years lost.

[3] 25% backlog diagnostics: 401 additional lives and 7,332 life-years lost. 50% backlog diagnostics: 811 additional lives and 14,873 life-years lost. 75% backlog diagnostics: 1,231 additional lives and 22,635 life-years lost.

substantial whatever the method used. Hamilton (2020) suggests that what matters most now is the recovery plan. But what should that recovery plan look like?

Even prior to the Covid-19 pandemic, many authors and policy makers debated that the Government should increase investments in cancer services to ensure the NHS could meet the rising demand and cancer outcomes (Brown et al., 2014). Additionally, investments are considered to be particularly crucial in diagnostic services, where rising demand was already starting to outstrip the resources available in 2014 (Brown et al., 2014). Further, Thorlby et al. (2019) argue that recovering waiting time target performance requires substantial initial investments to stabilise staff numbers, increase the workforce in key services, and increase equipment and facilities.

In sum, investments in capacity and workforce are widely considered to help reduce waiting times and are regarded as a crucial short-term plan of action. Keeping in mind that resources were already scarce before the pandemic, the economic consequences of the pandemic make tackling the excessive pre-existing and extra waiting times even more challenging. Ultimately, this all boils down to the question of 'how to spend the money' to reduce waiting times. This substantiates the importance of understanding how diagnostic waiting times, and especially those for cancer, are impacted by demand and supply characteristics.

An important first step in this comprehension is to determine if and, if so, what kind of relationship exists between health spending (reflecting supply and demand-side determinants) and diagnostic waiting times. Estimating these waiting time elasticities can inform further research on the channels through which supply and demand determinants impact waiting times. To date, the effect of health spending has been investigated by Siciliani and Hurst (2003 & 2003b) for elective surgery waiting times. They found a substantial negative correlation indicating that higher health expenditures are associated with lower waiting times. Their comprehensive OECD Working Papers and proposed theories serve as a theoretical fundament for this study. It is expected to find a similar relationship for diagnostic waiting times. The research question for this study is:

How are health expenditures related to diagnostic waiting time performance in England from 2014 to 2019?

STRUCTURE OF THE PAPER

This paper continues by outlining the historic context of waiting times in the NHS England and discussing the organisation of the English health care system. Subsequently, the theoretical framework and conceptual model are specified. Next, the empirical strategy is discussed in which the selection of the most appropriate estimation technique is explained and relevant assumptions are considered. Thereafter, the data and analytic sample are described. In the results section, the descriptive statistics and regression results are presented. Then, the results are discussed and hypotheses which may help explain the findings are proposed. At last, the limitations of this study are addressed after which some concluding remarks and implications for future research are made.

Institutional background

HISTORIC CONTEXT

During the 1990s, the Conservative government instituted market-oriented reforms on the supply side of the UK health care market. One of the primary aims was reducing long waiting times for elective care (Propper, 1995; Cutler, 2002). Waiting times decreased, but not by as much as was desired by the government nor the public. In 1997, the average waiting time for elective care was still around 23 weeks and maximum waiting times of over 18 months still existed. Partly due to concerns about NHS waiting time performance, a new Labour administration was voted in. The new government ended the market reforms and adopted a target-based waiting time policy in 2000 (Propper et al., 2007). Waiting times were published and used as a basis for sanctions (e.g. stricter supervision) and rewards (e.g. freedom to keep certain surpluses).

Propper et al. (2007) examined whether this target policy significantly reduced waiting times. They used the natural experiment arising from policy differences between England and Scotland and found that waiting time targets indeed significantly reduced waiting times in England. Studies by Hauck and Street (2006) and Dimakou et al. (2009) led to similar conclusions. Thus far, waiting time targets seemed to do their job. Then why have waiting times risen again?

The renewed rise in waiting times from 2010 onwards reflects an imbalance between allocated resources and growing and changing population health needs (Brown et al., 2014; Thorlby et al., 2019). An ageing population combined with a wider range of available treatments due to technological innovations, cause the demand to increase. At the same time, the NHS supply side has been facing a much lower than average funding growth in addition to workforce shortages. Cuts to social care and public health budgets exacerbate these pressures by making it harder to keep people healthy outside of hospitals (Wyattt, 2019; Thorlby et al., 2019). Moreover, Brown et al. (2014) argue that deficits in supply are not only caused by practical capacity issues. The organisational efficiency of health care is crucial as well. Loss of national and local leadership and infrastructure, fragmentation of commissioning across the patient pathway and variation in the roles and responsibilities of organisations may lead to supply inefficiencies. Altogether, the mechanisms behind the renewed rise in waiting times are vast and not straightforward. In any way, they have led to one of Boris Johnson's first promises, when he became Prime Minister in 2019, to be fixing long waiting times (Prime Minister's Office, 2019). It seems like waiting times in England have turned full circle.

Thorlby et al. (2019) indicate several priorities for the Government. In the short term, initial investments are required to stabilise and increase staff numbers, and increase capacity in terms of equipment and facilities. In the long term, enhancing preventive health care and finding ways to provide health services more efficiently are vital. However, these long-term solutions require physicians and health care managers to have the time and skills to redesign the provision of services.

Ultimately, the existing challenges have become even more intricate due to the Covid-19 pandemic. Between January and September 2020 there were 4 million fewer referrals to outpatients than in the same period the year before (Thorlby, Fraser & Gardner, 2020). The number of patients referred by their GP for suspected cancer dropped by 60% in April 2020 compared to April 2019 (Thorlby et al., 2020). By the end of 2020, considerable progress was made in restarting routine hospital services. However, pre-pandemic levels have not yet been reached, not to mention the extra activity needed to address the backlog due to the pandemic. Waiting time target performance has reached the lowest level since targets were set, with more patients experiencing long delays in diagnosis and treatment (Gardner & Fraser, 2021).

THE ENGLISH HEALTH CARE SYSTEM & CLINICAL COMMISSIONING GROUPS

In England, all residents are entitled to free public health care through the NHS. The NHS Constitution encompasses all rights related to NHS care and includes the waiting time targets set for various health services (Department of Health & Social Care, 2012). The responsibility for health legislation and general policy rests with Parliament, the Secretary of State for Health, and the Department of Health. Day-to-day responsibility (i.a. managing the NHS budget) lays with a government agency called NHS England.

In 2016, the UK spent 9.8% of its GDP on health care. Public expenditures – mainly related to the NHS – made up almost 80% of this amount (Office for National Statistics, 2018). NHS funding mainly comes from general taxes (80%). The remaining 20% comes from national insurance, which is a payroll tax paid by employees and employers. The NHS also receives income from co-payments and patients covered by private health insurance using NHS services. Approximately 10.5% of the population holds voluntary private medical insurance (PMI) to obtain faster access to elective care in the private sector (Thorlby, 2020). For public services, the degree of cost-sharing is very limited. Health care is free at the point of use for primary care visits, specialist consultation, and outpatient and inpatient hospital services. Only for certain services, such as travel vaccinations and outpatient prescription medicines, out-of-pocket (OOP) payments apply.

NHS England allocates shares of the national budget to 191 Clinical Commissioning Groups (CCGs). CCGs were established as part of the Health and Social Care Act in 2012 and allocate health care budgets at the local level. NHS providers contract with local CCGs to provide health services. Over the years, CCGs have been merging from initially 211 in 2012 to 191 in April 2019. Especially from 2018 onwards, mergers have taken place. The reason underlying these mergers is to move towards so-called Integrated Care Systems (ICSs). The NHS Long Term plan (NHS England, 2019) specifies that there would typically be one CCG per ICS to allow for better collaboration in the allocation and provision of health care (Das-Thompson, 2019).

Public hospitals are mainly reimbursed at nationally determined diagnosis-related group (DRG) rates, which include the costs of medical staff. These DRG payments make up about 60% of hospitals' income. The remainder comes from activities such as mental health care, education and research funds (Thorlby, 2020). Nearly all specialists employed in NHS hospitals are paid on a salary basis. Salary rates

are set nationally as part of a contract between the Department of Health and the British Medical Association. Specialists are free to (additionally) engage in private practice. In 2006, approximately 55% of physicians performed private work. This proportion is declining as the earnings gap between public and private practice narrows (NHS Digital, 2018). Concerning primary care, most GPs are private contractors (60%) or are employed on a salary basis (22%). The payment for private contractors consists for 60% out of capitation for essential services, 15% fee-for-service (FFS) payments for optional services, and around 10% performance-related payments (Thorlby, 2020).

Every three years, NHS budgets are set nationally. Since 2010, the allocation of funds by the central government has grown much slower. The long-term historical growth rate was on average 4% in real terms between 1950 and 2010 (Crawford and Emmerson, 2012). The average spending growth rate between 2010 and 2015 was 1.2% (Nuffield Trust, 2018). The mismatch between funding, demand and the cost of providing services has led to NHS hospitals and other providers accumulating an underlying deficit of £4.3 billion (Nuffield Trust, 2018).

Theoretical framework

This section starts by explaining waiting times as rationing mechanism in national health systems. Subsequently, the conceptual model by Siciliani and Hurst (2003/2003b) on the determinants of waiting times is discussed. Along with theories and evidence on the dynamic nature of waiting times, this model serves as the theoretical foundation for this study.

WAITING TIMES AS RATIONING MECHANISM

In the presence of excess demand, waiting times can help to balance the supply and demand for health services (Lindsay & Feigenbaum, 1984; Martin & Smith, 1999). In traditional markets, prices are used to ration goods and services and reconcile demand and supply. However, in the NHS England, people face zero or low co-payments due to public and or private insurance. Subsequently, there is a limited reduction in demand due to prices. Instead, waiting times are used to reconcile demand and supply (Siciliani & Iversen, 2012).

According to Siciliani and Iversen (2012), longer waiting times may reduce demand for at least two reasons. Firstly, when waiting is long, patients are more willing to go for private treatment. Secondly, patients may opt for an alternative treatment (e.g. a pharmaceutical alternative). On the other hand, longer waiting times may increase the supply of health services. Out of altruism, providers may be willing to exert greater effort when the wait is long. Moreover, waiting times are used as targets. Providers face financial or non-financial (e.g. more regulation) sanctions when they exceed waiting time targets. Also, hospital managers may face a higher risk of losing their when waiting times are high (Lindsay & Feigenbaum, 1984; Martin & Smith, 1999; Propper et al., 2007).

A certain degree of waiting may also be preferable in terms of equity. Patients are supposed to wait their turn, irrespective of their ability to pay or other non-clinical characteristics. However, some authors raised concerns about excessive waiting times creating inequalities by income or educational level. Only patients who can afford to pay out of pocket or hold private health insurance (often linked to employment status), may opt for private treatment when waiting is long. If this shift in demand to private facilities is insufficient to substantially reduce waiting times (e.g. when waiting times are excessive) patients without the resources to opt for private treatment are stuck on the waiting list. Subsequently, they may face inequalities in patient experience and or health outcomes (Siciliani, 2015; OECD, 2020).

To conclude, in health systems like the NHS, waiting times are to some extent desired to harmonise demand and supply. Therefore, optimum waiting times will not be zero in a system where patients face no direct prices. However, the substantial increase in waiting times over the past decade seems to be exceeding the goal of rationing and may harm patient experience and health outcomes.

CONCEPTUAL MODEL

Fig. 1 presents the conceptual model as proposed by Siciliani and Hurst (2003/2003b) on the determinants of waiting lists and waiting times. In short, waiting times are impacted by demand factors affecting the inflow to the waiting lists and supply factors affecting the outflow.

To start with, the model illustrates that the demand for health services is determined by the population's health status and the state of medical technology which influences the range of available treatments and the expectations of patients. At the same time, physicians play a key role in managing demand. The thresholds set by physicians to refer and add patients to the waiting list impacts the demand. The severity of these referral thresholds may be influenced by the method of physician-remuneration. Physicians paid on a fee-for-service (FFS) basis may offer patients fast access (i.e. maintain short queues) compared to salaried physicians (Iversen & Luras, 2002). On the other hand, allowing salaried physicians to work both in the public and private sector, can encourage some doctors to increase the waiting times in the public sector to stimulate the demand for their private health services (DeCoster et al., 2002; Morga & Xavier, 2001). Furthermore, financial incentives for patients (e.g. the extent of cost-sharing, private health insurance and price of private treatment) are also expected to influence demand.

The supply for health services is considered to depend on both public and private capacity combined with the level of productivity. Martin and Smith (1999) provide evidence for a negative correlation of the number of available beds and waiting times (elasticity = -0.242). Lindsay and Feigenbaum (1984) found a similar negative correlation for both the number of available beds and physicians. Productivity is expected to be impacted by the payment method for physicians and hospitals. Physicians paid on a FFS basis are likely to deliver more procedures than salaried physicians. For example, Gosden et al. (2001) found that, compared to capitation, FFS remuneration resulted in a higher quantity of primary care services. Over the past decades, many studies found similar results indicating that the payment method for physicians impacts the demand of health services (Delattre & Dormont, 2003; Dijk et al., 2013; Seyedin et al., 2020; Seyedin et al., 2021). For hospital remuneration, activity-based funding (such as DRG) was found to encourage a higher productivity compared to fixed budgets (Biorn, Hagen, Iversen & Magnussen 2002; Clemmesen & Hansen, 2003). Also, Mot (2003) found that replacing FFS payments by fixed budget payments for specialists in the hospital reduced the average admission rate and led to longer waiting times for surgery in the Netherlands.

It is likely that there are dynamic effects from waiting times to the quantities demanded and supplied in health care. In private markets these quantities respond to price. In the health sector however, instead of prices, waiting times are used to ration health services. All else being equal, waiting times may decrease when longer waiting times encourage patients to pay for private treatment or take out private health insurance (Besley et al., 1998). In addition, long waiting times may reduce demand by discouraging physicians from making referrals and adding patients to the waiting list (Iversen, 1997). Supply may increase due to longer waiting times because money may 'follow the queue' (Gravelle, Smith & Xavier, 2003; Iversen, 1993). Also, longer waiting times stimulate to reduce unused capacity (Cooper,

1981). Similarly, long waiting times can serve as a signal for private providers to expand activity. The existence of these dynamic effects may help to achieve equilibrium waiting times.

These dynamic effects have been widely studied (mainly with UK data). Siciliani and Iversen (2012) describe the demand and supply functions as:

$$[1] \qquad Y_i^D = \alpha_0 + \alpha_1 w_i + \alpha_2 x_i^D + \alpha_3 z_i + e_i^D$$

$$[2] \qquad Y_i^S = \beta_0 + \beta_1 w_i + \beta_2 x_i^S + \beta_3 z_i + e_i^S$$

where Y_i^D and Y_i^S denote respectively the (log of) demand and supply of health care in area i, and w is the (log of) waiting time. They expect that, $\alpha_1 < 0$ (longer waits associated with lower demand), and $\beta_1 > 0$ (longer waits associated with more supply). Vector x_i^D contains variables which affect demand (e.g. proportion of elderly in the area). Vector x_i^S contains variables which affect supply (e.g. number of doctors). Vector z_i contains variables which affect both the supply and demand of services. For instance, a hospital with higher quality may simultaneously have a higher demand and lower supply if quality is costly and can be traded-off with quantity.

Martin and Smith (2003) provide cross-sectional evidence for equations [1] and [2]. They found that the demand-elasticity in NHS England is between 0 and -0.2; a 1% increase in waiting times decreases demand at most by 0.2%. Supply was found to be elastic; between 2.1 – 5.9 depending on the year. The inelasticity of demand is important from a policy perspective. It indicates that investments to expand supply will have a substantial effect in reducing waiting times because the supply-increase will only be offset to a small extent by an increase in demand (Siciliani & Iversen, 2012). Gravelle et al. (2003) found a similar effect for demand-elasticity by using panel-data methods. Additionally, Fabbri and Monfardini (2009) came to the same conclusion using survey data in contrast to the other studies which used administrative data.

Assuming that demand and supply are in equilibrium $Y_i^D = Y_i^S = Y_i$, Siciliani and Iversen (2012) suggest that waiting times can be written directly as a function of demand and supply shifters:

$$[3] \qquad w_i = \gamma_0 + \gamma_1 x_i^D + \gamma_2 x_i^S + \gamma_3 z_i + e_i$$

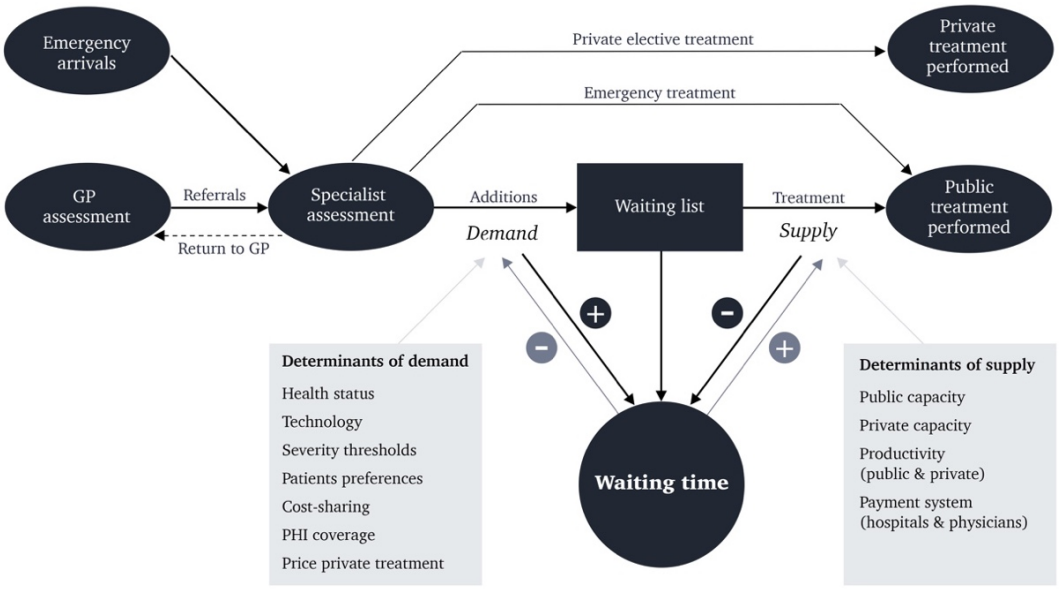
Estimating equation [3] empirically has the benefit that no information on activity is necessary and dynamic effects between waiting times and demand or supply are resolved. However, for variables affecting demand and supply simultaneously (z_i) the estimated coefficient gives the net effect. It is not possible to disentangle to what level the effect is due to demand or supply.

For this study, equation [3] is rewritten to study the effect of health spending (reflecting supply and demand characteristics) on diagnostic waiting time target performance:

$$[4] \qquad \mathrm{Y}_{it} = \eta_0 + \eta_1 Spending_{it} + \eta_2 x_{it1} + \cdots + \eta_k x_{itk} + \alpha_i + \varepsilon_{it}$$

where Y_{it} is the waiting time target performance across CCG i in year t. The expression $Spending_{it}$ contains the explanatory variable of interest measuring health spending per capita varying across CCG and over time. Control variables affecting both health spending and waiting times (e.g. proportion of elderly) are indicated by x_{itk}. The unobserved heterogeneity constant over time is indicated by α_i, and the idiosyncratic error is captured by ε_{it}. Equation [4] presents the general model of interest. The exact specification of this model is further elaborated in the empirical strategy. Siciliani and Hurst (2003/2003b) found a negative correlation between health expenditure and waiting times for elective surgery. This study expects to find a similar relation for diagnostic waiting times.

Figure 1. Conceptual model by Siciliani and Hurst (2003/2003b) on the determinants of waiting lists and waiting times.



Data

Secondary panel data are used at the CCG level in England covering a period of six years from 2014 to 2019. These years capture the UK financial years (FY) running from April through March[4]. All variables used in this study vary at the CCG-year level. Data is analysed for the period of FY2014 – FY2019, since not all data was available prior to FY2014 (see section on analytic sample below). Moreover, more recent data was available for FY2020. However, the outbreak of Covid-19 in March 2020 has heavily disrupted health care and increased waiting times. Therefore, the observations for FY2020 are considered incomparable to the years before and are excluded from analysis.

DEPENDENT VARIABLES: WAITING TIME TARGET PERFORMANCE

The performance on several diagnostic waiting time targets (which are set in the NHS Constitution) is used to study the effect of health spending on waiting times. Two types of targets are considered; the two-week-wait (TWW) cancer waiting time targets and six-week-wait (SWW) diagnostic waiting time targets. Data are derived respectively from NHS Cancer Waiting Times (CWT) Statistics (NHS England, 2020b) and NHS Diagnostic Waiting Times & Activity (DWT) Statistics (NHS England, 2020c). Table 1 provides an overview of all waiting time targets studied.

The TWW targets measure the percentage of patients seen by a specialist consultant within 14 days after GP referral. A distinction is made when cancer is suspected (TWW) and when patients present with breast symptoms but cancer is not initially suspected (TWWBS). This TWWBS target was introduced because only half of diagnosed breast cancers were coming through the standard TWW route. As a result, a significant proportion of patients was not benefitting from the faster pathway by the TWW (NHS Improvement, 2009). The operational standard for both two-week-wait targets is set at 93%. At least 93% of patients should be seen within two weeks after GP referral. The clock starts when the receipt of referral is made by the GP (day zero). The clock ends when the patient is first seen by either a specialist consultant or receives a diagnostic (NHS England, 2020b).

The SWW targets are not specified for cancer specifically but concern general waiting times for 15 key diagnostics. An overview of the specific procedures included within each key diagnostic is found in Table A11 in the appendix. The SWW targets measure the percentage of patients receiving a diagnostic within six weeks after referral. The operational standard for all six-week-wait targets is set at 99%. At least 99% of patients should receive a diagnostic within six weeks after referral. The clock starts when a referral is made by a physician. All referral routes (e.g. referral by a GP, hospital-based clinician, etc.) and all settings (e.g. outpatient clinic, inpatient ward, etc.) are included. The clock stops when the patient receives the procedure. When a patient cancels or misses an appointment the clock is set to zero. Patients are excluded when the diagnostic procedure is part of a treatment plan. Also, test carried out as part of

[4] For example, the financial year of 2014 (FY2014) runs from April 2014 through March 2015.

national screening programmes are excluded, but diagnostic tests triggered by abnormal screening results are included (NHS England, 2015). Some key diagnostics include procedures which are commonly used to diagnose cancer. For example, endoscopy diagnostics are used to diagnose intestinal cancers. Other key diagnostics include procedures which are rarely or never used in cancer diagnostics (e.g. Audiology Assessments, Respiratory Physiology). Table 1 further specifies for what purpose each key diagnostic is used.

All waiting time statistics are monthly reported by commissioners. Monthly submissions are sent to the a database hosted by NHS Digital (NHS England, 2020b). Submissions and data are closely monitored and validated by NHS England to identify any large errors in the data. However, it remains difficult to identify if providers are correctly following the guidelines on when a clock can be restarted or stopped. However, as the information should be used within the organisation (e.g. hospital) for operational delivery, it is likely that the returns from individual organisations are signed off at director level and provide an accurate reflection of the situation within that organisation (NHS England, 2015b).

Monthly waiting times are expected to be serially correlated. For example, long waiting times in March affect subsequent waiting times in April. Therefore, annual waiting time target performance scores were computed using geometric mean calculation which allows to account for these compounding effects.

Table 1. Overview of the dependent variables.

Two-week-wait targets	Percentage of patients seen by a specialist consultant within two weeks after GP referral.
TWW	*In case of suspected cancer.*
TWWBS	*In case of breast symptoms when cancer is not initially suspected.*
Six-week-wait targets	Percentage of patients receiving a diagnostic test within six weeks after physician referral.
Imaging	
Magnetic Resonance Imaging	*Used to image the whole body, head, spine, chest, abdomen, pelvis, heart, bone, liver, gallbladder, pancreas, kidneys, angiography (vascular system). Has many different diagnostic purposes as well as cancer diagnosis. MRI is more time consuming than CT, but yields more detailed images.*
Computed Tomography	*Used to image whole body, head, spine, sinuses, chest, abdomen, pelvis, calcium scoring, angiography, cerebral vessels, bone, colon, pulmonary arteries, kidneys. Used for many different diagnostic purposes as well as cancer diagnosis. CT is less time consuming than MRI, but yields less detailed images.*
Non-Obstetric Ultrasound	*Used to examine soft tissue and fluid filled organs: thyroid, abdomen, pelvis, scrotum, kidneys, bladder, bone, retina, coronary artery, female genital tract, oesophagus, upper gastrointestinal tract, liver, bile duct, pancreas, peritoneum. It can detect abnormalities such as tumours. Obstetric ultrasounds (i.e. ultrasounds on the reproductive tract of pregnant women) are excluded.*
Dual Energy X-ray Absorptiometry	*Used to examine bone density (calcium content) to detect bone conditions. Not used for cancer diagnosis (only as part of treatment plan, which is excluded in SWW target).*
Barium Enema	*Used to examine colon and rectum. Used to detect cancer, non-cancerous growths (polyps), inflammation, ulcers, and other diseases.*
Physiology	
Audiology Assessments	*Used to assess hearing, balance or suitability for hearing aid or cochlear implant. Not used for cancer diagnosis.*
Echocardiography	*Used to examine structural and functional abnormalities of the heart. Used to diagnose heart failure, valve diseases, thrombus, infective vegetation. Could also be used to diagnose heart tumours. However, cardiac tumours are extremely rare.*
Electrophysiology	*Used to examine the heart's electrical conduction system. Used to diagnose abnormal heart rhythms. Not used for cancer diagnosis.*
Peripheral Neurophysiology	*Used to examine nerve conduction and electrical activity of the muscle. Mostly used to diagnose neuro-muscular diseases. Not used for cancer diagnosis.*
Respiratory Physiology	*Used to diagnose sleep-breathing problems (sleep studies). Not used for cancer diagnosis.*
Urodynamics	*Used to examine disfunctions of the bladder, sphincters, and urethra. Not used for cancer diagnosis.*
Endoscopy	
Colonoscopy	*Used to examine the large intestine (colon) and rectum. Used to diagnose causes of abdominal pain, intestinal problems, polyps, and colon cancer.*
Cystoscopy	*Used to diagnose bladder diseases (stones, inflammation) and bladder cancer.*
Flexible Sigmoidoscopy	*Used to examine the rectum and lower colon. Used to diagnose swollen tissue, ulcers, polyps, and cancer. Less invasive than colonoscopy.*
Gastroscopy	*Used to examine the upper gastrointestinal tract (oesophagus, stomach and first part of small intestine). Used to diagnose swallowing problems, abdominal pain, stomach ulcers, reflux diseases, cancer.*

Note: Operational standard of two-week-wait targets is set at 93%. Operational standard of six-week-wait targets is set at 99%.

EXPLANATORY VARIABLE: HEALTH SPENDING PER CAPITA

Annual health spending statistics at the CCG level from FY2014 to FY2019 are derived from NHS England's Budget Allocations Statistics (NHS England, 2020d). Based on an analysis of the reported health spending available up to 2014, health expenditures are found to be closely correlated with budget allocations. Therefore, this study uses budget allocations as proxy for health spending to explore the effect on waiting times.

NHS England allocates budgets to each CCG consisting of three funding streams: core services, primary care, and specialised services. This study considers the core services funding stream of which approximately two-thirds is assigned to hospital-based care. The remaining share is assigned to community services and mental health care (NHS England, 2020). The primary care funding stream is excluded because of its distal link to hospital waiting times (e.g. through improved prevention, early detection, and management of chronic conditions). The specialised services funding stream is excluded as well because it relates to uncommon conditions, for which there are few providers and the costs are very high.

Each CCG's budget allocation is based on a weighted capitation formula consisting of four main elements: a target allocation, baseline, distance from target and pace of change (NHS England, 2019). Target allocations estimate the relative need and unavoidable costs for health services between CCGs. For example, CCGs with a higher percentage of elderly have a relatively higher demand for health services. As an illustration of unavoidable cost differences between CCGs, variation in input prices across regions impacts the costs of health services as well. Funding stream shares are set nationally. Of the total NHS budget, 75% is allocated to core services, 8% to primary care, and 17% to specialised services. To determine the share allocated to each funding stream per CCG, weighted populations are calculated per funding stream (funding stream target share). To calculate the target allocation, this funding stream target share is multiplied with the national budget per capita and several indices to control for differences in relative need and unavoidable costs:

$$[5] \quad Target\ allocation = national\ budget\ per\ capita * \ funding\ stream\ target\ share * age\ index * \ additional\ needs\ index * \ input\ price\ index$$

The age index reflect the demographic profile of a CCG. The additional needs index captures local deprivation and other factors likely to influence the need for health services. The input price index adjusts for unavoidable differences in the costs of delivering health services including remoteness and input prices (e.g. staff, equipment, buildings). A CCG's final allocation per capita is derived by multiplying the target allocations with a distance from target (DFT) index:

$$[6] \quad Allocation\ per\ capita = target\ allocation * \ distance\ from\ target\ index$$

This involves comparing the target allocation with the CCG's previous year's allocation (baseline). Based on this difference, the DFT index is constructed in combination with a pace of change (POC) parameter. This POC parameter is set nationally and sets a minimum growth in allocations for core CCG allocations and a higher growth in allocations for CCGs furthest under target. The aim of the POC policy is to move CCGs towards their target allocations over time and smooth potentially large fluctuations in funding as they may affect the provision of health services. Those further below (above) target will see the largest annual increase (decrease) in allocations.

In conclusion, this study uses the final core services allocation per CCG as a proxy for annual hospital spending. The allocations are measured per capita to allow comparisons across CCGs with different population sizes. The price effects of inflation are corrected by using the Consumer Price Index (CPI) for health services to transform allocations into current prices (Statista, 2021).

INSTRUMENTAL VARIABLE: DISTANCE FROM TARGET INDEX

The distance from target (DFT) index is used to instrument health spending in this study. Statistics on the DFT index are derived from NHS England Budget Allocations Statistics (NHS England, 2020d). To keep in line with the data on health spending (proxied by core services budget allocations), this study uses the DFT index for the core services funding stream.

The DFT index determines a CCG's final allocation per capita in combination with the target allocation. The DFT index is calculated annually based on two elements; a pace of change (POC) parameter which is set nationally, and the percentage difference between a CCG's per capita actual (baseline) and target allocation:

$$[7]\ DFT\ index\ = difference\ baseline\ and\ target\ allocation * pace\ of\ change$$

Baseline below target allocation suggests that a CCG has a higher need than the current budget is able to cover and growth is required. In practice, a DFT index value of one indicates that a CCG's baseline is equal to its target allocation, which results in no real growth. A DFT index greater (less) than one means a CCG's baseline is below (above) targets and thus its allocation needs to increase (decrease).

The aim of the DFT index is to align current allocations with target allocations while avoiding sharp shocks to budgets. To dampen this effect, actual allocations are based on applying growth to the baseline allocation to move towards the target, rather than immediately changing the target figure. This level of growth is set nationally according the POC policy which indicates how quickly CCGs should move towards their target allocations. This POC is informed by policy considerations such as the maximum decrease and maximum increase that can be implemented without disproportionately affecting the availability, quality and efficiency of health services.

For the DFT index to be a relevant and valid instrument it should fulfil two requirements; it should strongly impact health spending, and the DFT index should be exogenous (i.e. not directly impact waiting time target performance). In the empirical strategy, these assumptions are further discussed.

COVARIATES

Table 2 presents the set of control variables used in this study. A distinction is made between time-varying and time-constant covariates. Time-varying variables are included in models across all estimation techniques. Summary statistics on these covariates are reported in Table 3. Time-constant variables are only included in models estimated by Pooled OLS or Random Effects regression. Summary statistics on these variables are reported in Table A1 in the appendix. For all estimations, the average annual number of people registered with a GP was used as weight.

This study controls for several time-varying characteristics. Data on the age distribution across CCGs (share of persons in a certain age group) are obtained from the NHS Outcomes Tool (NHS England, 2018). Socio-economic control variables include the Indices of Deprivation sourced from the English IMD Indices (National Statistics, 2015), measuring the level of deprivation across certain socio-economic areas (income, employment, crime, housing, living environment, education, employment). Including the Market Forces Factor (MFF) index allowed to control for regional differences in input prices. Data on the MFF index was derived from NHS England Budget Allocations Statistics (NHS England, 2020).

Time-constant covariates include the share of population with chronic health problems, variables for the ethnicity composition of the CCG population, share of population born in non-European countries, five variables on employment and economic activity, and three variables on socio-economic household characteristics (lone person, lone parent, no car). Also, nine regional dummies were included to control for time-fixed regional differences. All data on time-constant control variables was obtained from Nomis census data (Nomis, 2011).

Table 2. Overview of covariates.

Time-varying covariates	Population age groups	% of population aged 0-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80+
	IoD - income	Proportion of population experiencing deprivation relating to low income
	IoD - employment	Proportion of working age population in an area involuntarily excluded from the labour market.
	IoD - education	Lack of attainment and skills in the local population.
	IoD - health	Risk of premature death and the impairment of quality of life through poor physical/mental health.
	IoD - crime	Measures the risk of personal and material victimisation at local level.
	IoD – barriers to housing/services	Measures the physical and financial accessibility of housing and local services.
	IoD – living environment	Measures the quality of both the 'indoor' and 'outdoor' local environment.
	IMD - Index of multiple deprivation	Overall measure of multiple deprivation by people living in an area. The seven indices of deprivation are combined and weighted to calculate the Index of Multiple Deprivation.
	Market Forces Factor index	An estimate of unavoidable cost differences between health care providers, based on their geographical location. MFF reflects that unit staff, land and building input costs are higher in some parts of the country (e.g. in London) than in others.

Table 2 continued. Overview of covariates.

Time-constant covariates	Health problem	% of population with a long-term health problem or disability (aged 16-64 and all ages).
	Provides unpaid care	% of population who provides unpaid care.
	Country of birth non-EU	% of population whose country of birth is non-EU.
	Ethnicity	% of population with a certain ethnicity (Asian, Black, Mixed, Other, White).
	Tenure	% of households who owns, privately rents, or socially rents a house.
	Economically active	% of population employed (including students and temporarily unemployed).
	Economically inactive	% of population retired, long-term health problems, and permanently unemployed.
	Unemployed	% of population temporarily unemployed.
	Occupation type	% of population aged 16-74 with professional, agricultural or no qualifications.
	Lone person 65+	% of households with lone person over 65 years of age.
	Lone parent	% of households with lone parent.
	No cars or vans	% of households with no cars or vans
	Region dummies	Nine region dummies.

ANALYTIC SAMPLE

Fig. 2 displays the selection process of the analytic sample. The DFT index and MFF index statistics were only initiated as of FY2014. Including observations for FY2013 would complicate a direct comparison of the IV and non-IV estimation. In addition, adding observations for FY2013 is unlikely to add much power changing the results. The sample remains of a sufficiently large size, which should reveal effects if they exist. Therefore, observations for FY2013 were excluded from analysis.

Furthermore, observations for CCG codes 13Q and X24 were excluded. 13Q is the commissioner code for the NHS England-commissioned armed forces activity. This is the commissioner organisation for the eligible armed forces population. This unit is not included because no data on demographics and budget allocations available. Even if data would be acquired, this CCG described mostly men and is not bound to a specific geographic region, which makes it incomparable to other CCGs. For similar reasons, observations for X24 were excluded. This unit comprised the aggregate of NHS England.

Due to some missing values the final model was estimated for 1,098 observations across 188 CCGs over a time span of 6 years. The panel data set is almost balanced; 181 CCGs were observed for the maximum of 6 years. The remaining CCGs were observed for 2 years. No CCGs were excluded in order to generate a balanced panel. Because the panel is not severely unbalanced, estimates are likely to be very similar to the estimates for a balanced panel (Baltagi, 2005).

Figure 2. Selection process analytic sample.

Empirical strategy

This section starts by describing the study design and general model of interest after which the selection of the most appropriate estimation technique is discussed. Secondly, assumptions underlying the model's validity and reliability are evaluated. Statistical methods are informed by, inter alia, Wooldridge (2013), Angrist and Pischke (2015), Angrist (2009), and Baltagi (2005).

STUDY DESIGN

This retrospective study empirically investigates the relationship between health spending and waiting time target performance scores for diagnostics using panel data. 188 CCGs across England are followed over six years from FY2014 to FY2019[5]. Panel data estimation techniques using multiple linear regression methods are used to explore the relationship of interest for various waiting time targets as described in the data section. Statistical analyses are performed using Stata 16.1 (StataCorp, 2019) statistical software. The following general panel data model for CCG i in year t was considered to determine the most appropriate estimation technique:

$$[8] \qquad \mathrm{Y_{it}} = \beta_0 + \beta_1 \ln(spending)_{it} + \beta_2 \ln(x)_{it}^{V} + \beta_3 \ln(x)_{t}^{C} + \mu_{it}\,, \qquad \text{where } \mu_{it} = \alpha_i + \varepsilon_{it}$$

Here, $\mathrm{Y_{it}}$ denotes the waiting time target performance of interest. The expression $spending_{it}$ captures the log of health spending per capita. Vector $(x)_{it}^{V}$ measures CCG-level covariates in log varying over time and vector $(x)_{i}^{C}$ represents time-constant covariates in log. The composite error term μ_{it} is composed by time-invariant CCG-level unobserved heterogeneity α_i and a CCG-specific idiosyncratic error ε_{it}.

POOLED ORDINARY LEAST SQUARES ESTIMATION

Firstly, pooled ordinary least squares (OLS) is used to study the association between health spending per capita and waiting time target performance. Time-invariant unobserved heterogeneity α_i at the CCG-level is treated as part of the error term. Thus, for pooled OLS estimation the error term is the composite error term μ_{it}. The residuals in equation [8] were found to be heteroskedastic by employing the Breusch-Pagan test (1797). Therefore, heteroskedastic-robust standard errors are used clustered at the CCG level.

When OLS assumptions are fulfilled, pooled OLS estimation is the best linear unbiased estimator (BLUE). However, in observational panel data, observations are not independently distributed over time. Therefore, the error term is likely to be serially correlated:

$$[9] \qquad Corr(\mu_{it}, \mu_{it-1} | \mathbf{X_i}) \neq 0$$

[5] Each financial year (FY) runs from April through March. For example, FY2014 comprises the period starting in April 2014 through March 2015.

where $\mathbf{X_i}$ denotes all independent variables. Even when assuming that the idiosyncratic error ε_{it} is not serially correlated, the unobserved heterogeneity α_i is common to all observations of the same unit:

$$[10.1] \quad \mu_{it} = \alpha_I + \varepsilon_{it}$$

$$[10.2] \quad \mu_{it-1} = \alpha_i + \varepsilon_{it-1}$$

As a result, time-fixed unobserved CCG characteristics may be correlated with the dependent variable and health spending per capita (violation of the zero conditional mean assumption), which causes the estimates to be biased. This was substantiated by the Cumby-Huizinga test (1992), which rejected the null hypothesis that the errors are serially uncorrelated ($p < .001$). Therefore, OLS estimates cannot be interpreted as causal. The results of the Cumby-Huizinga test can be found in the appendix in Table A12.

The endogeneity issues arising from the time-invariant unobserved heterogeneity can be eliminated by using other panel data estimation techniques. Random Effects (RE), Fixed Effects (FE), and First Differences (FD) methods control for the time-invariant unobserved heterogeneity. In the following section, the choice between these estimation techniques is explained.

CHOICE OF ESTIMATION TECHNIQUE

Random Effects vs Fixed Effects/First Differences

Methods like FE and FD estimation use the within CCG variation of the data to account for serial correlation[6]. The RE method quasi-demeans the data by using both the between and within variation. When the unobserved heterogeneity α_i is uncorrelated with the independent variables, methods which only use the within variation are inefficient. In that case, RE estimation would be more efficient. Additionally, RE has the advantage of providing estimates for time-constant characteristics.

However, it is likely that the link between health spending and waiting time target performance is impacted by time-constant unobserved factors (e.g. unrecognized socio-economic characteristics). To test this, RE estimates are compared to FE estimates to determine if there are significant differences in time-varying coefficients. In that case, time-invariant unobserved factors impact the relationship of interest and RE estimates are biased. The Hausman[7] test (1978) was employed which rejected the null hypothesis of consistent RE coefficients on time-varying variables for all outcome measures with the exception of Audiology Assessments ($p = .0896$). The output of the Hausman test can be found in the appendix in Table A13. Using FE or FD estimation is considered to be more appropriate to analyse the data set.

[6] Fixed Effects estimation eliminates α_i by demeaning the data. First Difference estimation eliminates α_i by differentiating the data.

[7] The Hausman test was performed using a user-written Stata command [xtoverid] in order to conduct the Hausman test using RE and FE with robust standard errors to account for heteroskedasticity.

Fixed Effects vs First Differences

The choice between the FE and FD approach hinges on the relative efficiency of the estimators (i.e. the structure of serial correlation of the idiosyncratic error ε_{it}). The estimates of both methods are unbiased under the same assumptions, but if the idiosyncratic error ε_{it} is serially correlated, FD is more efficient and gives valid standard errors. If the idiosyncratic error ε_{it} is serially uncorrelated, FE is more appropriate. The Inoue & Solo (2006) LM-test was used to test for serial correlation. This test the null hypothesis of no autocorrelation of any order against two alternative hypothesis; autocorrelation of some order, and autocorrelation up to order 1. The results can be found in Table A14 in de appendix. For all dependent variables except Audiology Assessment ($p = .061$) the null hypothesis was rejected at 5% significance. As a result, the FD approach was considered to be most appropriate.

FIRST DIFFERENCES ESTIMATION

The FD method uses first-difference transformation of the data to eliminate time-invariant unobserved heterogeneity from the panel data model. The model in first differences for CCG i in year t takes the following form:

$$[11] \quad \Delta \mathrm{Y}_{it} = \beta_0 + \delta_2 \Delta d2_t + \cdots + \delta_T \Delta dT_t + \beta_1 \Delta \ln (spending)_{it} + \beta_2 \Delta \ln(x)_{it1} + \cdots + \beta_k \Delta \ln (x)_{itk} + \Delta \varepsilon_{it}, \quad t = 1, 2, \ldots, 6$$

Where Y_{it} denotes the waiting time target performance of interest. The expression $spending_{it}$ captures the log of health spending per capita. Year-fixed effects are captured by $\delta_T \Delta dT_t$ for T=2, 3, ..., 6. Expression x_{itk} consists of the log of control variables varying over time and across CCG. The error term contains the idiosyncratic error ε_{it}. The residuals in equation [11] were found to be heteroskedastic by employing the Breusch-Pagan test (1797). Therefore, heteroskedastic-robust standard errors are used clustered at the CCG level.

Even though the FD approach allows to control for time-invariant unobserved characteristics α_i, endogeneity issues due to unobserved time-varying factors ε_{it} remain. These issues may cause the estimates to be biased, which makes causal inference inappropriate. Therefore, the relationship between health spending and waiting time target performance is additionally analysed using a First Differences Instrumental Variables (FDIV) approach.

FIRST DIFFERENCES INSTRUMENTAL VARIABLES ESTIMATION

The FD approach controls for several time-varying variables that may be correlated with waiting time target performance and health spending. However, the coefficient of health spending may be biased due to omitted (time-varying) variables that are correlated with the dependent variable and health spending (e.g. the number of physicians at the CCG level or practice variation). An additional source of endogeneity may be related to the use of CCG budget allocations as a proxy for health spending. The true (unknown)

value for health spending at the CCG level, may be considered as the sum of the budget allocation value and an error term varying at the CCG level m_{it}:

$$[12] \qquad health\ spending_{it} = budget\ allocation_{it} + m_{it}$$

where i indicates the CCG, and t the year. Expression m_{it} indicates an (idiosyncratic) error term that varies over time and across CCG. Even though budget allocations and health spending were found to be closely correlated up to 2014, there might be a slight difference in reality which is represented by m_{it}. The elimination of time-invariant heterogeneity from the regression model by using the FD approach reduces the probability that the measurement error term m_{it} is correlated with the error term of equation [11]. However, it is still be possible that the error term in differences $\Delta\varepsilon_{it}$ is correlated with the measurement error in differences Δm_{it}, which may cause a bias to the FD estimates on health spending.

In addition to endogeneity issues resulting from measurement error and or omitted time-varying factors, reverse causality may cause health spending to suffer from endogeneity as well. Longer waiting times at the CCG level could result in prolonged suffering and health losses for patients. This results in additional health needs for which the target allocations will be adjusted.

Instrumenting health spending using a variable uncorrelated with ε_{it} can solve these endogeneity issues due to omitted (time-varying) factors, measurement error and or reverse causality. This study uses the DFT index as an instrumental variable. The data section contains a detailed description of the DFT index. In short, combined with the target allocation, the DFT index determines a CCG's final allocation per capita [6]. The DFT index is annually calculated based on a pace of change (POC) parameter which is set nationally and the difference between a CCG's actual and target allocation. The model estimated by the FDIV approach is identical to equation [11]. However, the DFT instrument is used to disentangle the effects of changes in health spending on waiting time target performance:

$$[13] \qquad \Delta\ln(health\ spending)_{it} = \rho_1 + \rho_2\Delta\ln(DFT)_{it} + \rho_3\Delta\ln(\mathrm{x})_{it1} + \cdots + \rho_k\Delta\ln(\mathrm{x})_{itk} + \Delta\varepsilon_{it}$$

For the DFT index to be a valid instrument, it should meet two criteria. Firstly, it should strongly predict health spending, which is expected because the final allocations per capita are calculated by multiplying the target allocation with the DFT index [6]. This can be assessed by an examination of the significance of the excluded instrument in the first-stage Instrumental Variable (IV) regression. Secondly, the DFT index should be exogenous. The instrument should not be correlated with any other determinants of the dependent variable with the exception of health spending. This exclusion restriction consists of two parts. The instrument should be randomly assigned (i.e. independent of potential outcomes, conditional on covariates) and the instrument should have no effect on the dependent variable other than through the first-stage channel. When the exclusion restriction is violated, the DFT index correlates with the standard errors, which causes the estimates to be biased. Because the exclusion restriction involves the covariance between the instrument and the unobserved idiosyncratic error ε_{it},

this assumption generally cannot be statistically tested. Alternatively, economic reasoning must substantiate instrument exogeneity.

In this study, the differences in health spending due to the DFT index should independently vary from CCG characteristics. In other words, CCGs should have no influence on the DFT index. As mentioned in the data section, the DFT index is determined by the difference between target and baseline allocation in combination with the POC parameter [7]. As discussed, target allocations are likely to be affected by omitted (time-varying) variables and reverse causality. For example, practice variation may impact health care utilisation at the CCG level which is a component of the need index (NHS England, 2019). However, the extent to which unobserved CCG characteristics may influence the difference between baseline and target allocation is considered to be limited. The target allocation is calculated based on a well-defined, complicated formula which accounts for demographic differences, differences in need, and differences in unavoidable costs [5]. Therefore, it can be argued that it is difficult for CCGs to influence the direction and magnitude of the difference between baseline and target allocation. Assuming that unobserved (time-varying) factors influencing the target allocation (e.g. practice variation) are stable over time, these factors influence current and target allocations similarly. This limits the degree of control by CCGs on the difference between current and target allocations. Also, the DFT index is determined by the POC policy [8] which is set nationally and affects the DFT index completely outside the control of CCGs. On the basis that CCGs have limited control over the difference between baseline and target allocation combined with no control over the POC policy, it is plausible that the DFT index instrument fulfils the exclusion restriction.

Furthermore, the DFT instrument can be used to assess whether health spending is, as expected, an endogenous regressor (i.e. impacted by time-varying omitted variables and or reverse causality). The *C-test* (also: 'difference-in-Sargan' test) is used to test the null hypothesis that the regressor is exogenous. Additionally, considerable differences between the FD and FDIV coefficients on health spending could also indicate effects of time-varying unobserved factors. If health spending is found to be exogenous, FD estimation is considered to be more efficient than the FDIV approach. If health spending is indeed endogenous, FDIV estimation should be used to account for omitted (time-varying) variables.

VALIDITY & RELIABILITY

For the results to be valid and reliable, several assumptions need to be fulfilled. The first four OLS assumptions (1 – 4) are demanded to obtain unbiased estimates. In order for the estimates to be efficient, two extra conditions (5 – 6) have to be satisfied. A brief evaluation of these assumptions is provided.

1. Linearity

Informed by existing literature (Siciliani and Hurst, 2003b; Siciliani and Iversen, 2012), equation [11] assumes a linear relationship between waiting time target performance and the log of health spending, and the included time-varying control variables. To assess linearity, the residuals were plotted against the fitted values of the dependent variable. These residual plots can be found in the appendix in Table A16.

2. Random sample

There is a random sampling of observations. The analytic sample contains all existing CCGs in the period of analysis. Therefore, the sample encompasses the full target population. This reduces concerns about the expected value of the estimators being equal to the population parameters.

3. No (perfect) multicollinearity

Each explanatory variable changes over time for at least some CCGs and no (perfect) linear relationships exist among the explanatory variables. Because panel data is used with heterogenous units (CCGs) present, multicollinearity is not a large concern. Pearson's correlation matrixes were examined for the independent variables. A few variables reported a high degree of multicollinearity. After careful consideration, while keeping the comparability of the models across estimation techniques in mind, three time-constant variables were removed from the model used for OLS estimation (share of people aged 16-64 with long term health problems, share of house owners, share of population economically inactive).

4. Zero conditional mean (strict exogeneity)

For each time period, the expected value of the idiosyncratic error ε_{it} given the independent variables $\mathbf{X_i}$ in all time periods and the unobserved effect α_i is zero [12]:

$$[14] \qquad \mathrm{E}(\varepsilon_{it} \,|\, \mathbf{X_i}, \alpha_i) = 0.$$

FD estimates cannot be interpreted as causal when the strict exogeneity assumption is violated. As discussed, using an instrument to generate exogenous variation in health spending can solve issues of endogeneity related to unobserved (time-varying) factors. Additionally, if the functional form of the econometric model misses non-linearities or interaction terms, the estimates may be biased. Previous

work by Siciliani and Iversen (2012) proposes a log-log configuration to investigate the relationship of interest. The Ramsey/Pesaran-Taylor RESET test (1999) was employed for three functional form specifications; linear, linear-log, log-log. For most dependent variables (11 out of 17) the RESET test did not reject the null hypothesis ($E(Y_{it}|\mathbf{X_i})$ is linear in $\mathbf{X_i}$) using a linear-log configuration[8]. The results of the RESET test can be found in Table A17 in the appendix. No literature was found suggesting to add certain polynomials or interactions to the model of interest. To make comparison across dependent variables possible, the same linear-log configuration was used for all outcome measures.

5. Homoskedasticity

Conditional on the independent variables $\mathbf{X_i}$, the variance of the errors is constant [15]. The Breusch-Pagan test (1979) was employed to assess the variance of the errors. For all dependent variables, the null hypothesis of the error variances being equal was rejected. To account for the heteroskedasticity of the errors, heteroskedastic-robust standard errors were used in further analyses. Robust standard errors are clustered at the CCG level.

$$[15] \qquad \mathrm{Var}(\Delta\varepsilon_{it}|\mathbf{X_i}) = \sigma^2, \qquad t = 2,..,6$$

6. Normality

Conditional on $\mathbf{X_i}$, the $\Delta\varepsilon_{it}$ are independent and identically distributed normal random variables. The Shapiro-Wilk test (1965) was used to assess the normality of the distribution of the error term. The output of the Shapiro-Wilk test can be found in Table A18 in the appendix. For all dependent variables, the null hypothesis of the errors being normally distributed was rejected. However, due to the analytic sample being reasonably large (N=1,098), $\hat{\beta}_1$ approximately follows a normal distribution centered at β_1. Therefore, the errors are assumed to be normally distributed. Additionally, Kernel Density plots were obtained, which indicate a close to normal distribution. These plots are found in Table A19 in the appendix.

[8] RESET test passed: Two-Week-Wait, Dual Energy X-ray Absorptiometry scan, Magnetic Imaging Resonance, Non-Obstetric Ultrasound, Respiratory Physiology, Urodynamics, Electrophysiology, Colonoscopy, Cystoscopy, Gastroscopy, Flexible Sigmoidoscopy. RESET test failed: Two-Week-Wait Breast Symptoms, Computed Tomography, Echocardiography, Peripheral Neurophysiology, Audiology Assessments, Barium Enema.

Empirical results

DESCRIPTIVE STATISTICS

Before discussing the main results of the econometric analyses, descriptive statistics are explored. Sample characteristics, trends in waiting time performance and absolute patient volumes are presented. The analysed period covers FY2014 to FY2019. In order to obtain a broader picture of waiting times and patient volumes over time, the presented trend figures in this paper comprise FY2013 to FY2020 (April 2013 – March 2021). The data from March 2020 onwards (FY2020) reveals a dramatic decline in waiting time target performance due to the Covid-19 pandemic. Because of the large impact of the Covid-19 pandemic disrupting health care, these observations are incomparable to previous years. Therefore, these observations have been excluded from statistical analysis.

Table 3 summarizes the characteristics of the analytic sample. Descriptives on time-constant covariates (only relevant for POLS estimation) are found in Table A1 in the appendix. The analytic sample includes 1,098 observations on 188 CCGs. On average, only two waiting time performance targets meet their operational standard; two-week-wait target when cancer is suspected (TWW) and the six-week-wait target for key diagnostic Non-Obstetric Ultrasound (NOUS). The six-week-wait target for Urodynamics (URO) and two-week-wait target in case of breast symptoms (TWWBS) present the lowest average performance scores. Furthermore, the targets on key diagnostics Barium Enema (BE) and Electrophysiology (EP) reveal missing values (BE: N=114, EP: N=380). Additional analysis indicated that for these diagnostics a substantial number of observations scored exactly 100% (BE: N=830, EP: N=718). A further investigation of the monthly commissioner statistics revealed that the absolute patient volumes for these key diagnostics are substantially lower compared to the other key diagnostics. This explains the high amount of missing values; if there are no patients to be seen, no target performance score can be obtained. Moreover, it can be argued that a 100% waiting time performance score for these diagnostics is more readily achieved than for high-volume diagnostics. Because of the little variance in these performance scores, econometric analyses are not likely to yield valid and reliable results. Therefore, the six-week-wait targets for BE and EP were excluded from further analyses.

Fig. 3 illustrates the trend in waiting time performance for both two-week-wait cancer waiting time targets over time. On average, TWW performance decreased with 3.18 %-points[9] from FY2014 to FY2019. Within that same period, TWWBS decreased more drastically with 10.53 %-points[10] on average. Since 2018, the operational standard of 93% is not being met for both two-week-wait targets. Moreover, Fig. 3 demonstrates the sudden drop in waiting time performance as a consequence of the Covid-19 outbreak at the end of March 2020. After a short period of recovery during the summer of 2020, performance further declined in the autumn of 2020.

[9] Average TWW target performance in FY2014 = 94.27% (sd. 2.14). Average TWW target performance in FY2019 = 91.09% (sd. 4.42). The average number of people registered with GP over period 2014-2019 per CCG is used as weight.

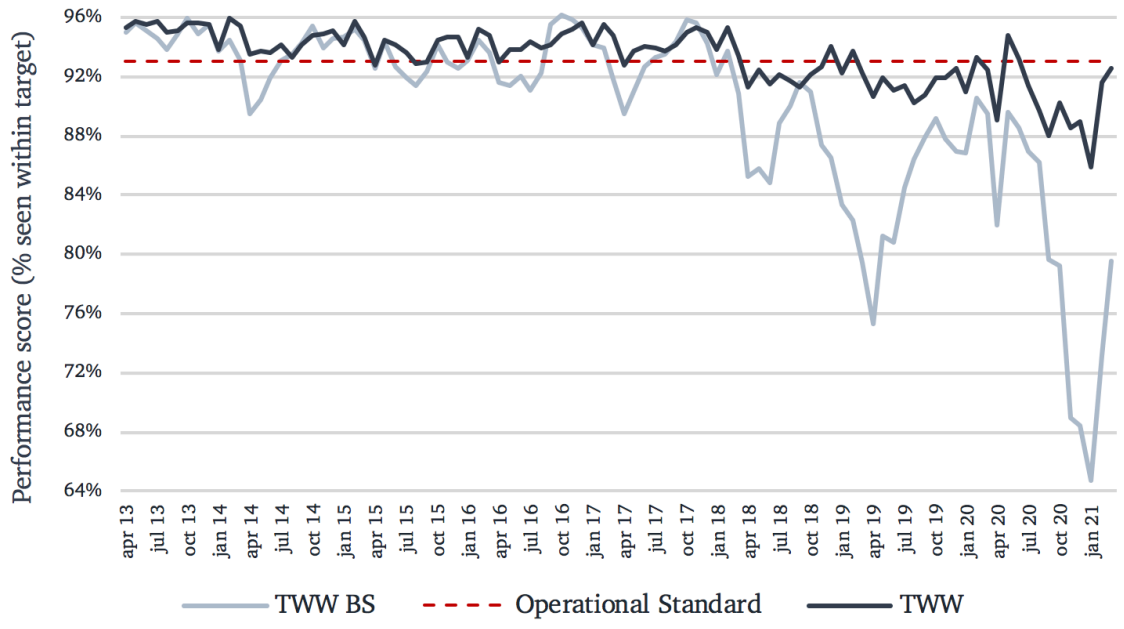
[10] Average TWWBS target performance in FY2014 = 93.03% (sd. 5.25). Average TWWBS target performance in FY2019 = 82.50% (sd. 17.38). The average number of people registered with GP over period 2014-2019 per CCG is used as weight.

Table 3. Sample characteristics (N = 1,098) (number of units (CCGs) = 188).

			Mean	Std dev	Min	Max
Dependent	***Two-week-wait targets***					
(within target %)	Two-week-wait (TWW)		93.51	3.42	73.68	98.86
	Two-week-wait Breast Symptomatic (TWWBS)		90.19	10.76	16.69	100
	Six-week-wait targets					
	Computerised Tomography (CT)		98.90	1.84	82.47	100
	Dual Energy X-ray Absorptiometry (DEXA)		98.99	3.04	59.06	100
	Magnetic Resonance Imaging (MRI)		98.38	2.16	77.49	99.98
	Non-Obstetric Ultrasound (NOUS)		99.03	1.77	80.54	100
	Barium Enema (BE)		99.45	1.91	83.60	100
	Audiology Assessments (AA)		98.20	2.87	75.42	100
	Echocardiography (ECG)		96.92	5.11	51.11	100
	Electrophysiology (EP)		98.02	7.51	21.18	100
	Peripheral Neurophysiology (PN)		97.17	4.62	48.38	100
	Respiratory Physiology (RP)		95.02	7.35	38.11	100
	Urodynamics (URO)		89.15	9.88	44.09	100
	Colonoscopy (COL)		93.96	8.28	44.02	100
	Cystoscopy (CYS)		92.92	7.08	50.06	100
	Flexible Sigmoidoscopy (FS)		94.96	7.13	42.55	100
	Gastroscopy (GAS)		95.13	6.50	48.01	100
	Total aggregate		96.06	2.79	78.74	99.77
Explanatory	Hospital spending per capita (current prices £)		1171.40	119.24	895.21	1577.11
Instrument	Distance from target index		1.00	0.05	0.88	1.34
Covariates	Market forces factor index		1.00	0.07	0.93	1.16
(time-varying)	Average number of people registered with GP		275,752.00	139,636.60	75,326.00	931,122.00
	Aged (% population)	0-9	11.71	1.40	7.52	18.07
		10-19	11.09	0.90	7.58	16.57
		20-29	13.47	3.11	8.84	26.54
		30-39	14.06	3.31	8.77	26.25
		40-49	13.77	1.05	11.05	17.41
		50-59	13.20	1.45	7.01	16.08
		60-69	10.49	2.24	4.07	16.33
		70-79	7.56	2.13	2.19	14.22
		80+	4.66	1.33	1.46	8.39
	Index of deprivation	Barriers	22.19	6.48	7.48	49.31
		Crime	2.02	0.44	0.81	3.27
		Health	1.99	0.58	0.58	3.64
		Education	21.47	8.34	3.53	56.88
		Employment	0.11	0.04	0.05	0.24
		Environment	22.16	9.56	4.52	59.52
		Income	0.14	0.05	0.05	0.31
		Multiple	21.67	7.86	7.18	52.14

Note: Outcome measures report waiting time target performance scores. These characteristics of the analytic sample cover the period of analysis (T=6): FY2014 to FY2019. The operational standard for TWW targets is 93%. The operational standard for SWW targets is 99%. The average number of people registered with GP over period 2014-2019 per CCG is used as weight.

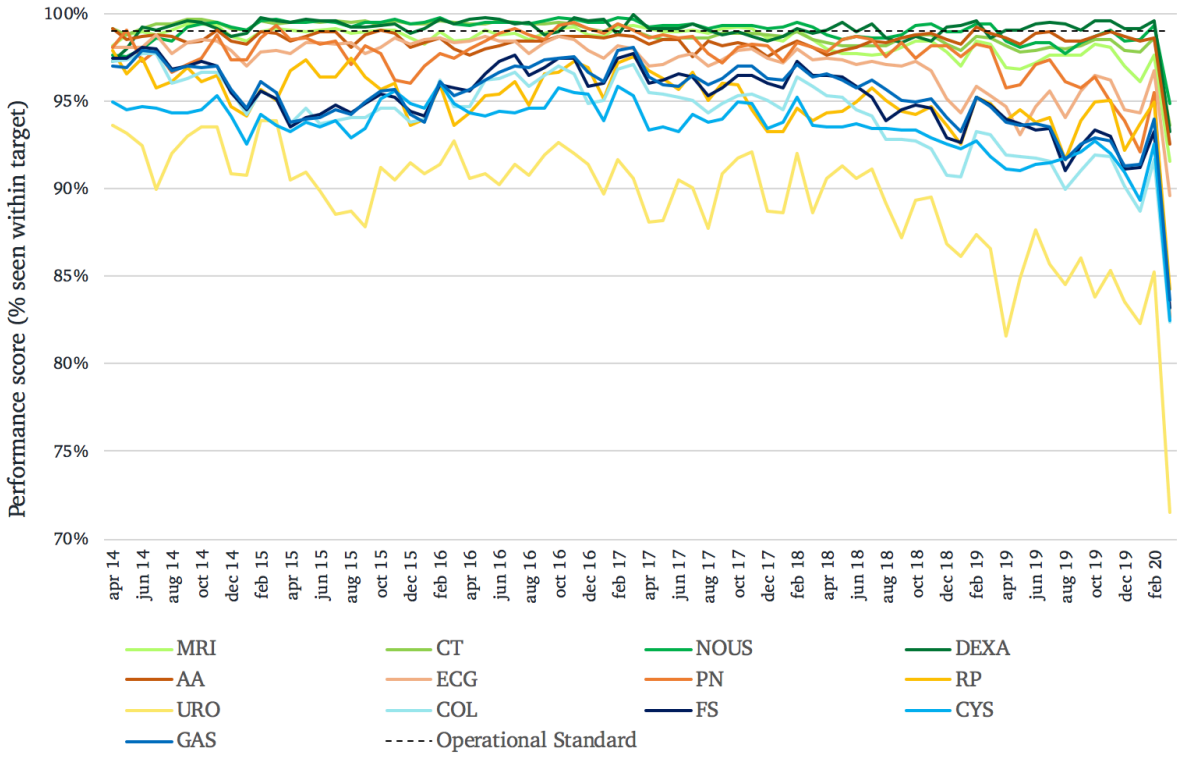
Figure 3. Trend in two-week-wait target performance for the Cancer Waiting Time Targets over time (FY2013 – FY2020).



NB: The period of analysis consists of FY2014 to FY2019. Trends are presented for FY2013 to FY2020 to present a broader picture of performance over time and to demonstrate the impact of the Covid-19 pandemic as of March 2020.

Fig. 4 displays the trend in waiting time performance for each six-week-wait key diagnostic over time (FY2014 – FY2019). On average, the total six-week-wait target performance is 3.31 %-points lower in FY2019 compared to FY2014[11]. Key diagnostic Urodynamics and diagnostics related to endoscopy (COL, CYS, FS, GAS) present the largest declines and lowest performance scores over time. Imaging diagnostics (CT, DEXA, MRI, NOUS) and Audiology Assessments (AA) demonstrate the highest performance scores and lowest declines over time. Fig. 5 presents the waiting time performance for the six-week-wait targets aggregated into three diagnostic categories (imaging, physiology, endoscopy) and an overall aggregate. It reveals the six-week-wait targets for endoscopy and physiology diagnostics substantially decline over time. Imaging diagnostic waiting time performance also decreases, but relatively less pronounced. In Fig. 6 six-week-wait target performance per diagnostic has been plotted up to March 2021. All diagnostics show a dramatic fall in waiting time performance as a result of the Covid-19 pandemic. On average, the total six-week-wait performance dropped from 93.90% in FY2019 to 58.80% in FY2020. Up to March 2021 waiting time performance scores have been recovering, but are still far from reaching pre-Covid rates.

Figure 4. Trend in six-week-wait target performance for 13 key diagnostics[12] over time (FY2014 – FY2019).

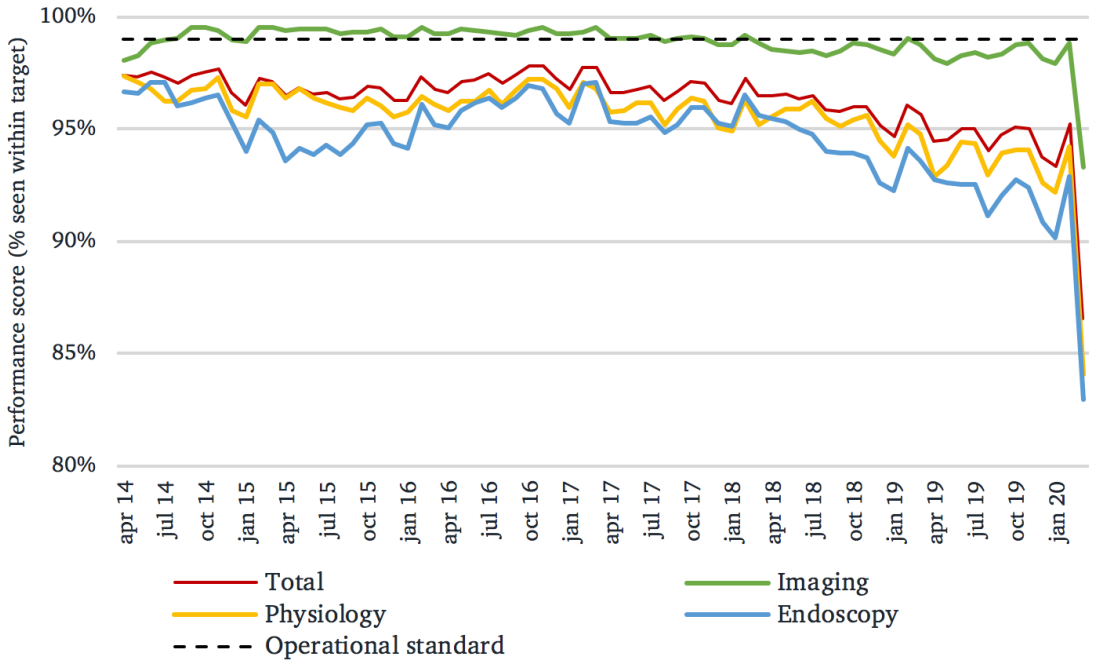


NB: This figure covers only the period of analysis for FY2014 to FY2019. Fig. 6 presents these trends for FY2013 to FY2020 to present a broader picture of performance over time and to demonstrate the impact of the Covid-19 pandemic as of March 2020. Fig. 4 is presented separately since it more clearly depicts the differences in performance between key diagnostics.

[11] Average six-week-wait target performance in FY2014 = 96.99% (sd. 1.97). Average six-week-wait target performance in FY2019 = 93.66% (sd. 3.75). The average number of people registered with GP over period 2014-2019 per CCG is used as weight.

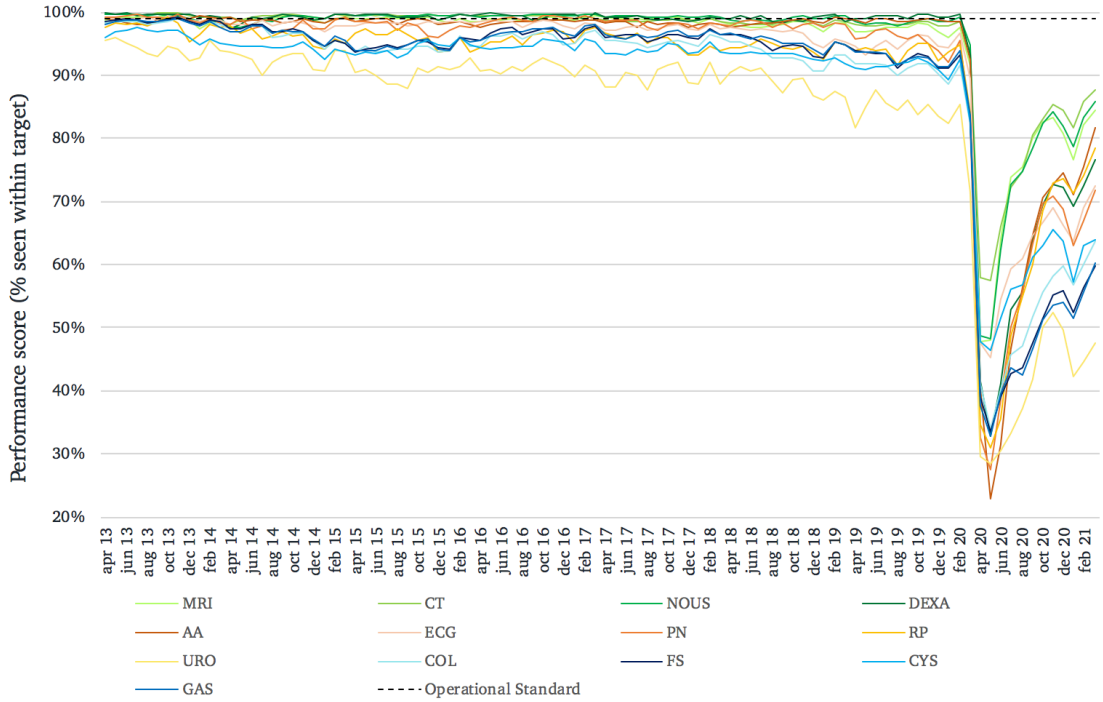
[12] **MRI**: Magnetic Resonance Imaging, **CT**: Computed Tomography, **NOUS**: Non-Obstetric Ultrasound, **DEXA**: Dual Energy X-ray Absorptiometry, **AA**: Audiology Assessments, **ECG**: Echocardiography, **PN**: Peripheral Neurophysiology, **RP**: Respiratory Physiology, **URO**: Urodynamics, **COL**: Colonoscopy, **FS**: Flexi Sigmoidoscopy, **CYS**: Cystoscopy, **GAS**: Gastroscopy.

Figure 5. Trend in six-week-wait target performance per key diagnostic category and the total aggregate over time (FY2014 – FY2019).



NB: This figure covers the period of analysis for FY2014 to FY2019. The imaging category contains the aggregate for key diagnostics: MRI, CT, DEXA, NOUS (BE excluded). The physiology category contains the aggregate for key diagnostics: AA, ECG, PN, RP, URO (EP excluded). The endoscopy category contains the aggregate for key diagnostics: COL, CYS, FS, GAS. The total aggregate score contains these 13 key diagnostics (BE and EP excluded).

Figure 6. Trend in six-week-wait target performance for 13 key diagnostics[13] over time (FY2013 – FY2020).



NB: This figure covers FY2013 to FY2020 to present a broader picture of performance over time and to demonstrate the impact of the Covid-19 pandemic as of March 2020. Fig. 4 presents these trends for just the period of analysis (FY2014 to FY2019) which more clearly presents the differences in six-week-wait performance between key diagnostics.

[13] **MRI**: Magnetic Resonance Imaging, **CT**: Computed Tomography, **NOUS**: Non-Obstetric Ultrasound, **DEXA**: Dual Energy X-ray Absorptiometry, **AA**: Audiology Assessments, **ECG**: Echocardiography, **PN**: Peripheral Neurophysiology, **RP**: Respiratory Physiology, **URO**: Urodynamics, **COL**: Colonoscopy, **FS**: Flexi Sigmoidoscopy, **CYS**: Cystoscopy, **GAS**: Gastroscopy.

Absolute patient volumes and patient shares across the waiting time targets over time have been examined in Fig. 7.1 – Fig. 11. All figures contain data from FY2013 up to FY2020 to present a broader picture of the developments over time and the consequences of the Covid-19 pandemic. Please note that the period of analysis encompasses FY2014 to FY2019 (i.e. reaches from April 2014 – March 2020).

Fig. 7.1 displays that the absolute patient volumes for the TWW cancer waiting time target has increased from approximately 120,000 patients in FY2014 to 160,000 patients in FY2019. The relative share of patients seen within standard (i.e. two weeks) is decreasing over time (Fig. 7.2). Fig. 8.1 reveals the absolute patient volumes for the TWWBS target are decreasing over time. In FY2014 approximately 18,000 patients were referred, whereas in FY2019 approximately 10,000 patients. Simultaneously, the relative share of patients not seen within standard (two weeks) increases substantially over this period (Fig. 8.2).

Figure 7.1. Total patient volumes (in thousands) over time for the two-week-wait (TWW) target.

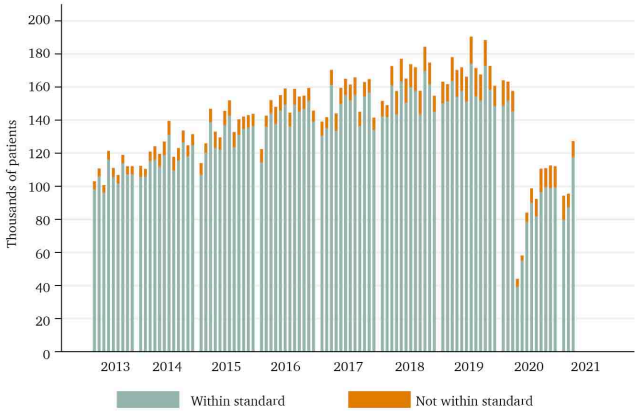


Figure 7.2. Patient shares (%) over time meeting the two-week-wait (TWW) target.

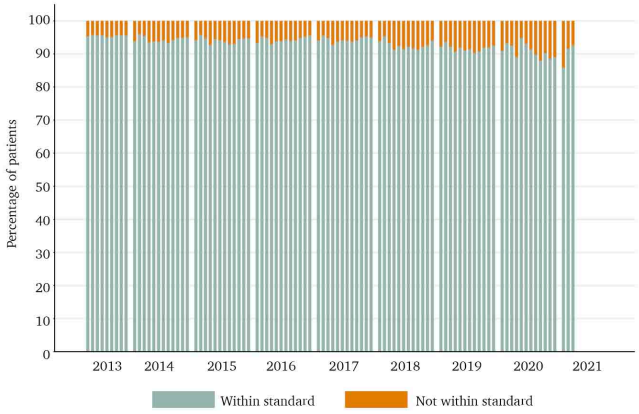


Figure 8.1. Total patient volumes (in thousands) over time for the two-week-wait in case of breast symptoms (TWWBS) target.

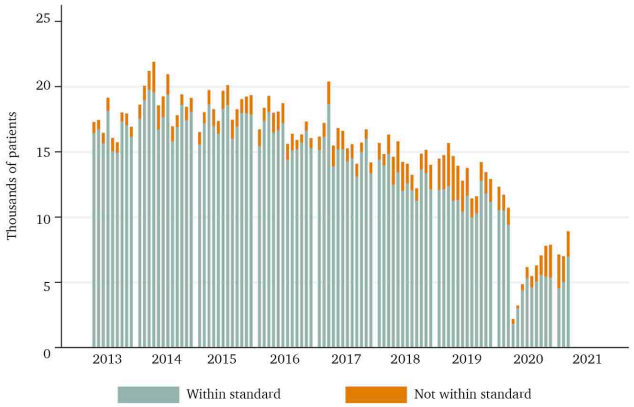
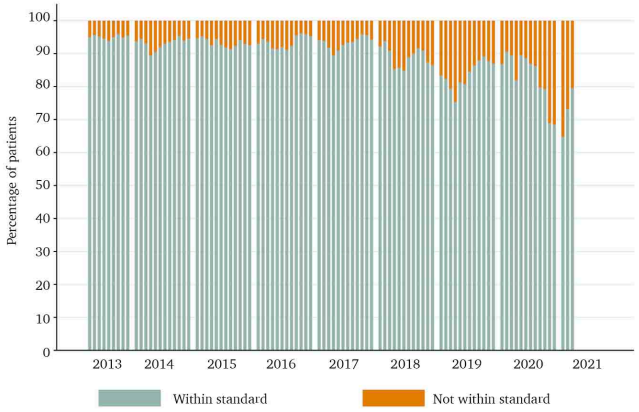


Figure 8.2. Patient shares (%) over time meeting the two-week-wait in case of breast symptoms (TWWBS) target.



Similar developments are observed for the six-week-wait Diagnostic Waiting Time targets. As shown in Fig. 9.1, the absolute volume of total patients referred via the six-week-wait pathway increases substantially over time. Approximately 800,000 patients were referred in FY2014 compared to over 1 million patients in FY2019. The relative share of patients exceeding the six week threshold is expanding over the period of analysis (Fig. 9.2).

Additionally, absolute patient volumes and relative shares of patients were plotted for each key diagnostic (Fig. 10.1 and 10.2 respectively). As absolute patient volumes are increasing over time, the relative shares of patients per diagnostic remain similar (Fig. 10.2). In addition, Table A9 in the appendix reports the average annual growth rates per key diagnostic. Most diagnostic show similar growth rates (average = 107.80%). Audiology Assessments (AA) reports a lower growth rate of 102.35%. Finally, Fig. 10.3 displays the relative shares of patients per diagnostic category. Imaging diagnostics includes the largest share of patients (approximately 70%), physiology diagnostics has a share of approximately 20% of patients and endoscopy diagnostics is the smallest group of patients with a share of 10%.

Figure 9.1. Total patient volumes (in thousands) over time for the six-week-wait target.

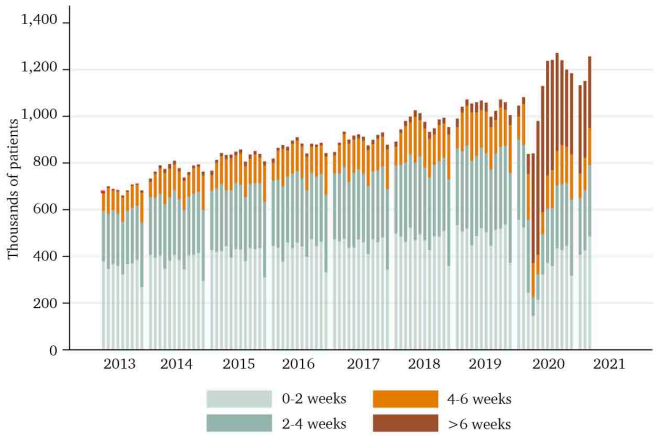


Figure 9.2. Patient shares (%) over time meeting the six-week-wait target.

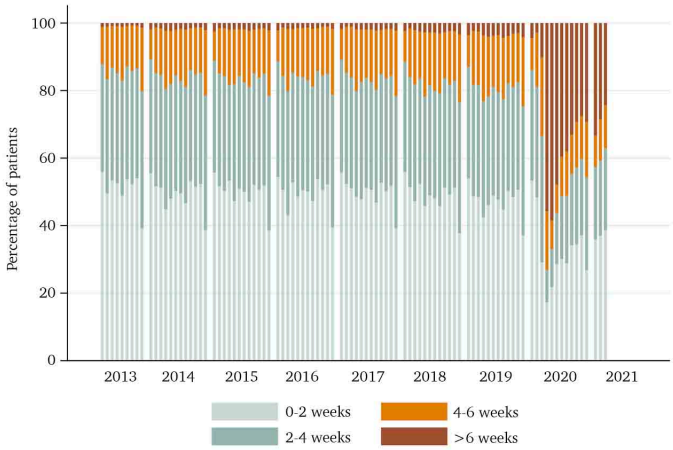


Figure 10.1. Patient volumes (in thousands) over time per key six-week-wait diagnostic.

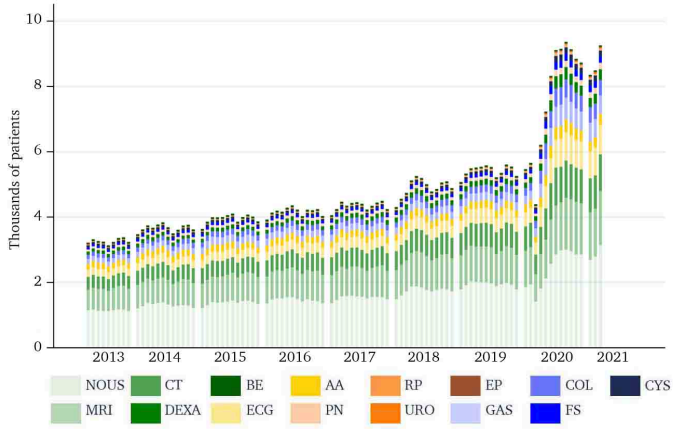


Figure 10.2. Patient shares (%) over time per key six-week-wait diagnostic.

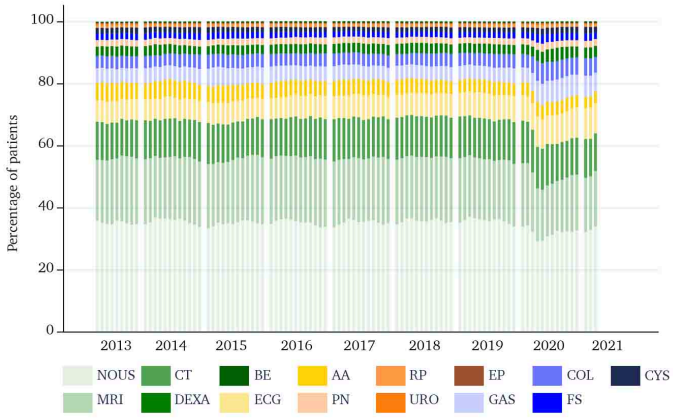
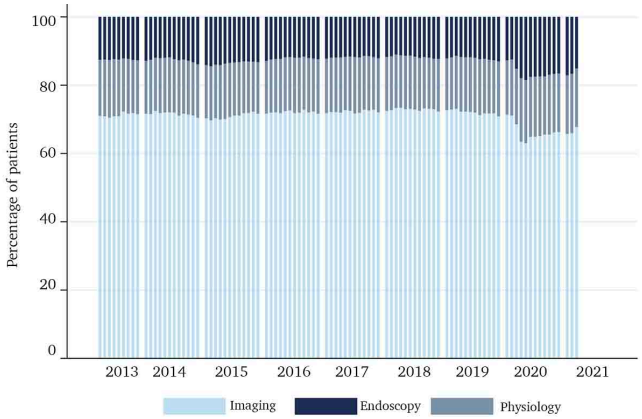


Figure 11. Patient shares (%) over time per six-week-wait diagnostic category.



MAIN RESULTS

Table 4 summarizes the regression results of the effect of increases in log health spending per capita on each outcome measure. The complete regression results including coefficients on control variables can be found in the appendix (Table A2 and A3). Column 1 of Table 4 presents the results for each model using POLS estimation. These regressions include time-constant covariates, year fixed effects and region controls as outlined in the data section. Standard errors are clustered at the CCG level. Health spending has a significant positive effect on waiting time performance for TWWBS, CT, ECG, PN, and CYS. The POLS estimates cannot be interpreted as causal, since health spending and waiting times may be correlated with unobserved time-constant CCG characteristics. Also, the POLS estimators are inefficient due to serial correlation in the error term, which results in invalid standard errors impacting the significance and confidence intervals. When controlling for autocorrelation of the errors and time-invariant unobserved CCG characteristics using the FD estimator (column 2), most significant effects weaken (lower level of significance) or disappear (no significance). Only PN and CYS remain significant. Moreover, COL and FS become significant at a 10% significance level. These significant FD estimation results present a positive correlation between health spending and waiting time performance. As discussed in the empirical strategy, these estimates are likely to be biased due to omitted time-varying CCG characteristics impacting both the explanatory variables and outcome measure.

To account for these endogeneity issues, health spending was instrumented using the DFT index. The first-stage *F*-statistic substantially larger than 10 (*F*-stat. = 130.68) indicates that the instrument has strong predictive power for health spending in the first stage. The Kleibergen-Paap rk LM test confirmed this by rejecting the null hypothesis (Chi-sq(1) = 49.2, $p<.001$) that the instrument has insufficient explanatory power to predict the endogenous variable in the model for identification of the parameters. Second-stage estimates from two-stage least squares estimation using the FDIV approach are shown in column 3. For most outcome measures with a significant coefficient on health spending the *C*-test of endogenous regressors led to a rejection of the null hypothesis of exogeneity of health spending at the 5% significance level[14]. This indicates that health spending is indeed an endogenous regressor, which is likely to be impacted by omitted (time-varying) variables. Therefore, using a FDIV approach is desirable to address these endogeneity issues. Remaining outcome measures, with the exception of Urodynamics, presented *p*-values < .10 for the *C*-statistic[15]. The results of the *C*-test can be found in Table A15 in the appendix. Even though the *C*-test's null hypothesis was not rejected at 5% significance for all outcome measures, the coefficients on health spending become considerably larger compared to column 2. These results are in line with the argument that unobserved time-varying factors may cause the estimates of health spending in column 2 to be biased. For example, changes in hospital efficiency over the years at CCG level may impact both health spending and waiting time performance. Accordingly, the results of the FDIV estimation technique are considered to be the most appropriate and thus are further discussed in this paper.

[14] Results C-test of endogenous regressors: MRI: $p = .008$, AA: $p = .005$, PN: $p = .049$, CYS: $p = .038$

[15] Results C-test of endogenous regressors: URO: $p = .149$, COL: $p = .093$, FS: $p = .095$, GAS: $p = .061$

Considering the FDIV estimates (column 3), 8 out of 15 outcome measures present a statistically significant coefficient on health spending. Most of these (n=6) present a positive effect of health spending. Following the models' linear-log functional form, point estimates can be interpreted as 1% increase in health spending is associated with $\frac{\beta}{100}$ %-points change in the dependent variable, *ceteris paribus*. For example, a 1% increase in health spending per capita is associated with 0.83 %-points increase in Colonoscopy target performance ($\beta = 82.77$, $p = .014$). Thus, more spending leads to more patients receiving a Colonoscopy within six weeks (target) after referral. Similar effects are observed for the other endoscopy measures (Cystoscopy: $\beta = 70.45$, $p = .012$; Flexible Sigmoidoscopy: $\beta = 75.96$, $p = .021$; Gastroscopy: $\beta = 81.73$, $p = .010$). The effect for MRI shows the same positive direction, but its coefficient is nearly six to seven times smaller compared to the effects for the endoscopy diagnostics; a 1% increase in health spending correlates with 0.12 %-points increase in MRI target performance ($\beta = 11.85$, $p = .012$). These findings are in line with the results of Siciliani and Hurst (2003b), reporting that higher health spending is associated with lower waiting times. The results for physiology diagnostics, show varied effects in terms of size, direction and significance level. Peripheral Neurophysiology presents a positive and sizeable effect ($\beta = 67.62$, $p = .005$) close to the results for the endoscopy diagnostics. Yet, the other physiology diagnostics display an unexpected direction for the effect of health spending. For Audiology Assessments, a 1% increase in health spending is significantly (at the 1% level) correlated with 0.34 %-points decrease in waiting time performance ($\beta = -34.12$, $p = .002$). In contrast, Urodynamics presents a negative correlation ($\beta = -74.96$, $p = .074$) at a lower significance level (10%).

Since there are many different outcome measures considered in this study (n = 15), the outcomes reporting a significant effect of health spending are further explored. However, additional analyses are carried out for all 15 outcomes measures and corresponding results can be found in the appendix (Tables A4-A9).

Table 4. Overview of POLS, FD and FDIV regression results for each dependent variable estimating the effect of log health spending per capita on waiting time target performance.

	1		2		3	
	POLS		**FD**		**FDIV**	
	N=1098	*p*	N=914	*p*	N=914	*p*
Two Week Wait targets						
Two-Week-Wait (TWW)	4.53 (4.07)	.267	-2.86 (5.08)	.574	7.58 (9.25)	.412
Two-Week-Wait Breast Symptoms (TWWBS)	21.12** (9.70)	.031	-15.83 (9.6)	.101	4.65 (24.34)	.849
Six Week Wait targets						
Imaging						
Computerised Tomography (CT)	4.22** (1.76)	.017	-1.06 (3.06)	.730	4.11 (5.70)	.471
Dual Energy X-ray Absorptiometry (DEXA)	-3.34 (2.76)	.228	-7.63 (6.39)	.234	-2.09 (17.37)	.904
Magnetic Resonance Imaging (MRI)	2.55 (1.95)	.192	0.30 (2.18)	.889	11.85** (4.70)	.012
Non-obstetric Ultrasound (NOUS)	2.30 (1.58)	.147	-2.78 (3.47)	.424	-9.01 (9.07)	.320
Physiology						
Audiology Assessments (AA)	-5.80 (3.63)	.112	-9.55 (6.64)	.152	-34.12*** (11.24)	.002
Echocardiography (ECG)	8.35** (3.94)	.035	-14.89 (10.16)	.145	-6.88 (16.45)	.676
Peripheral Neurophysiology (PN)	7.15* (4.13)	.085	25.37* (14.14)	.074	67.62*** (24.05)	.005
Respiratory Physiology (RP)	4.89 (7.14)	.494	-21.66 (15.09)	.153	-35.91 (31.70)	.257
Urodynamics (URO)	6.29 (10.23)	.539	-29.57 (30.96)	.341	-74.96* (41.96)	.074
Endoscopy						
Colonoscopy (COL)	9.11 (7.87)	.249	29.14* (17.59)	.099	82.77** (33.56)	.014
Cystoscopy (CYS)	18.73** (8.54)	.030	29.63* (17.09)	.085	70.45** (28.17)	.012
Flexible Sigmoidoscopy (FS)	3.14 (6.63)	.636	30.10* (18.02)	.097	75.96** (32.92)	.021
Gastroscopy (GAS)	2.30 (6.42)	.721	24.77 (17.19)	.151	81.73** (31.77)	.010
First stage results:						
Δ ln distance from target index					0.51*** (0.05)	.000
F-stat excluded instrument					130.68	

Note: All regressions include a constant term and year dummies (for 2015-2018). Complete regressions results including coefficients on the control variables can be found in the appendix (Table A2 and A3). The average number of people registered with GP over period 2014-2019 is used as weight. Robust standard errors clustered at CCG level are in the parentheses. Significant at: *10%, **5%, ***1% level.

HETEROGENEITY OF THE RESULTS

The results presented above indicate that for several key diagnostics there is a significant effect of health spending on six-week-wait target performance. These effects are likely to vary across regions. For example, regions with a higher population density may have larger hospitals. In turn, these larger hospitals can allow for economies or diseconomies of scale (i.e. efficiency or inefficiency due to size). Also, regions may differ in socioeconomic characteristics influencing the demand for health services. These regional differences in supply and demand characteristics may impact the link between health spending and waiting time performance.

To examine this regional variation, the analyses were disaggregated for different regions across England. To maintain large enough samples, the nine regions available in the data set (Fig. 12.1) were comprised into four main regions; North, Midlands & East, London, and South (Fig. 12.2). Table 5 presents the results for these disaggregated FDIV analyses. For most outcome measures, the effect of health spending becomes insignificant. This may partly be explained by the fact that the sample sizes per region are smaller than the original sample, resulting in larger standard errors, which makes it more difficult to pick up a significant effect. Alternatively, the absence of significant effects may simply indicate that effects of health spending on waiting time performance do not exist or are not pronounced enough in those regions. In general, the effects for region Midlands & East seem to drive the overall results for most outcome measures. Moreover, region South and region North report significant effects of health

spending for some outcomes. For the region of London, none of the outcome measure reports a significant effect of health spending.

Additionally, regional variation was examined for the outcome measures reporting non-significant effects of health spending in the original model (Table A4 in the appendix). Fo two-week-wait target performance a positive effect of health spending was found for region Midlands & East significant at the 5% level ($\beta = 55.46$, $p = .044$). A 1% increase in health spending per capita is associated with 0.55 %-points increase in TWW target performance in this region, *ceteris paribus*. The other outcome measures did not report significant effects across regions.

Figure 12.1. Nine regions across England used to compose four larger regions (Fig. 12.2) to examine regional variation of the results.

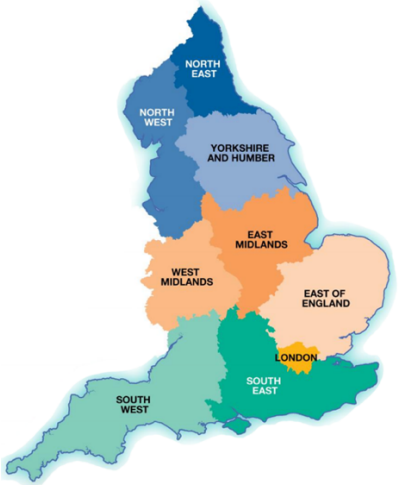


Figure 12.2 The four regions across England used to examine regional variation of the results.

Table 5. FDIV regression results of log health spending per capita on six-week-wait target performance across four geographical regions.

	All regions	**North**	**Midlands & East**	**London**	**South**
	N=914	N=295	N=274	N=160	N=182
	Units=187	Units=59	Units=58	Units=32	Units=38
Magnetic Resonance Imaging (MRI)	11.85** (4.7)	12.27 (10.99)	14.68 (10.03)	-6.95 (5.07)	10.58 (12.24)
Audiology Assessments (AA)	-34.12*** (11.24)	-65.4** (29.89)	-3.52 (16.9)	-24.01 (16.62)	-62.49* (34.68)
Peripheral Neurophysiology (PN)	67.62*** (24.06)	23.52 (48.44)	94.63 (65.22)	32.98 (36.76)	68.86* (41.44)
Urodynamics (URO)	-74.96* (41.96)	76.25 (111.57)	-208.18* (114.15)	8.16 (47.55)	-50.61 (83.51)
Colonoscopy (COL)	82.77** (33.56)	35.53 (74.19)	238.14** (102.82)	32.11 (27.39)	-19.2 (40.58)
Cystoscopy (CYS)	70.45** (28.17)	-54.89 (115.22)	89.24* (47.41)	25.79 (46.41)	73.2* (43.63)
Flexible Sigmoidoscopy (FS)	75.96** (32.92)	126.56 (96.91)	156.31** (69.89)	63.61 (39.6)	-54.91 (51.47)
Gastroscopy (GAS)	81.73** (31.77)	22.96 (55.80)	185.77** (85.12)	31.49 (31.79)	15.43 (59.82)

Note: All regressions include a constant term and include year dummies (for 2015-2018). Results for outcome measures reporting a non-significant effect of log health spending per capita can be found in the appendix (Table A4). The average number of people registered with GP over period 2014-2019 is used as weight. Robust standard errors clustered at CCG level are in the parentheses. Significant at: *10%, **5%, ***1% level.

ROBUSTNESS CHECKS

Exclusion of outliers

To test whether the results are driven by outliers, observations were dropped for which target performance was less than 70% or was exactly 100%. Estimates may somewhat changes as less observations result in larger standard errors. However, assuming the effects found in the original model exist in reality, coefficients should remain in the same direction and significant. Table 6 compares the results for the original model and the model in which outliers were excluded. Almost all effects remain similar sized, in the same direction, and significant at the 5% level. Only the effect for the six-week-wait target performance for Urodynamics (URO) changes substantially from being significant at the 10% level, to a *p*-value of almost 0.3. Hence, the effect of health spending on URO target performance is not robust to excluding outliers.

Table 6. FDIV regression results of log health spending per capita on six-week-wait target performance comparing the model excluding outliers in waiting time performance scores (scores <70% and of exactly 100% are excluded).

	Original model (N=914)	*p*	Model excl. outliers	*p*	N
Magnetic Resonance Imaging (MRI)	11.85 (4.70)	.012	11.85 (4.70)	.012	914
Audiology Assessments (AA)	-34.12 (11.24)	.002	-34.18 (12.21)	.005	805
Peripheral Neurophysiology (PN)	67.62 (24.06)	.005	70.79 (28.68)	.014	694
Urodynamics (URO)	-74.96 (41.96)	.074	-33.08 (31.83)	.299	729
Colonoscopy (COL)	82.77 (33.56)	.014	49.57 (24.68)	.045	857
Cystoscopy (CYS)	70.45 (28.17)	.012	50.34 (23.54)	.032	856
Flexible Sigmoidoscopy (FS)	75.96 (32.92)	.021	53.42 (29.73)	.072	828
Gastroscopy (GAS)	81.73 (31.77)	.010	56.81 (25.73)	.027	885

Note: FDIV regressions include a constant term and year dummies (for 2015-2018). Results for outcome measures reporting a non-significant effect of log health spending per capita can be found in the appendix (Table A5). The average number of people registered with GP over period 2013-2019 is used as weight. Robust standard errors clustered at CCG level are in the parentheses.

Alternative calculation performance scores

As described in the data section, the geometric mean was used to calculate average annual waiting performance scores to account for possible serial correlation in monthly waiting time statistics. The robustness of the results to different computation methods of the outcome measure was assessed. The models were re-estimated for two alternative calculation method; the arithmetic mean and the mean based on the total number of patients seen within target per year divided by the total number seen per year ('year mean'). Table 7 presents that the effects remain similar and significant across the computation methods. The 'year mean' model shows relatively lower significance levels of health spending. This can be explained by larger standard errors due to the less specific computation method of target performance. Still, most effects remain significant at the 5% level. Only for URO the effect becomes considerably smaller and less significant, yet remains significant at the 10% level.

Table 7. FDIV regression results of log health spending per capita on six-week-wait target performance comparing different calculation methods of target performance.

	Original model (N=914)	*p*	Arithmetic mean	*p*	*Year mean*	*p*
Magnetic Resonance Imaging (MRI)	11.85 (4.70)	.012	11.80 (4.64)	.011	11.03 (4.68)	.018
Audiology Assessments (AA)	-34.12 (11.24)	.002	-34.37 (10.97)	.002	-34.64 (12.25)	.005
Peripheral Neurophysiology (PN)	67.62 (24.06)	.005	66.65 (23.19)	.004	69.37 (23.55)	.003
Urodynamics (URO)	-74.96 (41.96)	.074	-75.57 (40.86)	.064	-60.30 (35.63)	.091
Colonoscopy (COL)	82.77 (33.56)	.014	77.32 (31.09)	.013	86.41 (36.08)	.017
Cystoscopy (CYS)	70.45 (28.17)	.012	67.28 (26.56)	.011	74.06 (29.48)	.012
Flexible Sigmoidoscopy (FS)	75.96 (32.92)	.021	73.97 (30.49)	.015	77.74 (37.26)	.037
Gastroscopy (GAS)	81.73 (31.77)	.010	78.2 (29.56)	.008	84.01 (35.64)	.018

Note: FDIV regressions include a constant term and year dummies (for 2015-2018). Results for outcome measures reporting a non-significant effect of log health spending per capita can be found in the appendix (Table A6). The average number of people registered with GP over period 2014-2019 is used as weight. Robust standard errors clustered at CCG level are in the parentheses. Number of observations for all measures is N=914.

Alternative Index of Deprivation measure

The Indices of Deprivation (IoD) together constitute – weighted with different strengths – a score of multiple deprivation; the Index of Multiple Deprivation (IMD) (National Statistics, 2015). Because of multicollinearity, the IMD was not included in the original model. Yet, the results should be robust to substituting the separate IoDs with the aggregate IMD score. If not, this could indicate possible measurement errors in one of the IoDs, causing the estimates to be biased. However, Table 8 shows that the results for each model including the IMD remain very similar to the original model.

Table 8. FDIV regression results of log health spending per capita on six-week-wait target performance comparing the model using an alternative aggregated measure of deprivation; the Index of Multiple Deprivation (IMD).

	Original model	***p***	**Model with IMD**	***p***
Magnetic Resonance Imaging (MRI)	11.85 (4.7)	.012	11.73 (4.54)	.010
Audiology Assessments (AA)	-34.12 (11.24)	.002	-32.5 (10.83)	.003
Peripheral Neurophysiology (PN)	67.62 (24.06)	.005	61.08 (22.35)	.006
Urodynamics (URO)	-74.96 (41.96)	.074	-75.30 (40.75)	.065
Colonoscopy (COL)	82.77 (33.56)	.014	75.53 (32.00)	.018
Cystoscopy (CYS)	70.45 (28.17)	.012	63.53 (26.72)	.017
Flexible Sigmoidoscopy (FS)	75.96 (32.92)	.021	70.66 (31.15)	.023
Gastroscopy (GAS)	81.73 (31.77)	.010	74.99 (30.58)	.014

Note: FDIV regressions include a constant term and year dummies (for 2015-2018). Results for outcome measures reporting a non-significant effect of log health spending per capita can be found in the appendix (Table A7). The average number of people registered with GP over period 2014-2019 is used as weight. Robust standard errors clustered at CCG level are in the parentheses. Number of observations for all measures is N=914.

Balanced panel

The findings were assessed to be robust to excluding CCGs with incomplete observations on each time period (i.e. making the panel balanced). Table 9 presents that coefficients for the balanced panel remain very similar to the unbalanced (original) panel. This substantiates that the attrition is happening at at random and does not impact the results. The degree of attrition is limited; the balanced panel contains 9 observations and 7 CCGs less than the original panel.

Table 9. FDIV regression results of log health spending per capita on six-week-wait target performance comparing the original model estimated for an unbalanced panel and a balanced panel.

	Unbalanced panel N=914 Units=188	***p***	**Balanced panel N=905 Units=181**	***p***
Magnetic Resonance Imaging (MRI)	11.85 (4.7)	.012	12.14 (4.7)	.010
Audiology Assessments (AA)	-34.12 (11.24)	.002	-33.90 (11.24)	.003
Peripheral Neurophysiology (PN)	67.62 (24.06)	.005	67.48 (23.98)	.005
Urodynamics (URO)	-74.96 (41.96)	.074	-75.76 (41.90)	.071
Colonoscopy (COL)	82.77 (33.56)	.014	82.66 (33.65)	.014
Cystoscopy (CYS)	70.45 (28.17)	.012	69.94 (28.16)	.013
Flexible Sigmoidoscopy (FS)	75.96 (32.92)	.021	76.35 (33.19)	.021
Gastroscopy (GAS)	81.73 (31.77)	.010	81.54 (31.88)	.011

Note: FDIV regressions include a constant term and year dummies (for 2015-2018). Results for outcome measures reporting a non-significant effect of log health spending per capita can be found in the appendix (Table A8). The average number of people registered with GP over period 2014-2019 is used as weight. Robust standard errors clustered at CCG level are in the parentheses.

Discussion

SUMMARY OF FINDINGS

This study finds significant and robust effects of health spending per capita on six-week-wait target performance for 8 key diagnostics. Most of these results are in line with the findings of Siciliani and Hurst (2003b) reporting that a higher level of health spending is systematically associated with lower waiting times for elective surgery. A relatively sizeable and positive effect is reported for all four endoscopy key diagnostics; increased health spending is associated with a higher percentage of patients seen within six weeks. With regards to physiology diagnostics, Peripheral Neurophysiology (PN) reports a similarly sized positive effect. Unexpectedly, a negative correlation was found for Audiology Assessments (AA) and Urodynamics (URO), indicating that higher spending is associated with a lower six-week-wait target performance. However, the effect for URO was not robust to re-estimating the empirical model while excluding outliers in performance scores. Hence, caution is demanded when drawing conclusions based on this result. Furthermore, MRI is the only imaging key diagnostic for which six-week-wait target performance is significantly impacted by health spending, which also presents a positive correlation. Yet, its point estimate is nearly six to seven times smaller compared to those of the endoscopy diagnostics and the PN diagnostic. The findings vary across four geographical regions in England. Mainly region Midlands & East seems to drive the overall results. For some outcome measures, regions South and North help explain the effects found in the full model. Region London reports no significant effect of health spending for any of the outcome measures.

This section continues by proposing potential mechanisms which may help to understand the findings using the conceptual model by Siciliani and Hurst (2003/2003b). These hypotheses may inform future studies investigating the link between health spending and waiting times.

POTENTIAL MECHANISMS

Non-significant effects of health spending

Not all outcome measures report a significant effect of health spending. This paragraph explores potential explanations for the absence of significant effects. These considerations may help inform the design of future research.

Firstly, increases in health spending per capita at the CCG level may not reflect supply factors particularly attributing to certain outcome measures. As elaborated in the data section, health spending is proxied by CCG core services allocations per capita. These allocations cover services related to acute and general hospital care, mental health, prescribing, community, and maternity. In other words, this measure for health spending does not exclusively cover the spending on, for instance, ECG procedures. Therefore, there could be some sort of ‘mismatch’ between the measure of health spending and the measure of waiting time performance, possibly allowing true effects to go unobserved in this study.

Secondly, in some cases the outcome measure may not capture the effect – if there is one - properly. In this study, waiting times are studied as target performance scores capturing the percentage of patients seen within a certain amount of time. If in reality an increase in spending reduces waiting times, but not by enough for patients to fall within target (two-week-wait or six-week-wait), this effect of health spending is not picked up. For example, if on average patients wait 7 weeks for an ECG and health spending reduces the wait by 2 days, this is not reflected by the ECG performance score for the percentage of patients seen within six weeks.

Furthermore, the two-week wait (TWW) target may have specific problems for capturing an effect. The TWW measures the percentage of patients waiting less than two weeks after an urgent GP referral to any kind of specialist consultation. Thus, the TWW covers a broad range of waiting times relating to all sorts of specialisms within the hospital. If there are effects of higher spending in certain specialist departments, these may not be pronounced enough to be revealed in the overall TWW target. Notably, when the analysis was disaggregated for four geographical regions, a significant effect ($\beta = 55.46$) at the 5% level was found for the TWW target for region Midlands & East (Table A4). This indicates that for this specific region, the effect of health spending on TWW target performance is pronounced enough to be detected.

An alternative explanation for not finding significant effects may relate to a limited variance in target performance over time. This may explain why most imaging diagnostics report no significant effect of spending considering that they show the least decrease in six-week-wait target performance over time (Fig. 4) and the highest average target performance score (98.95%). This suggests that the variance in target performance of these diagnostics (CT, DEXA, NOUS) may not be pronounced enough to reveal a significant effect of spending. In line with this proposed theory is that, within the imaging category, MRI demonstrates the largest decline in performance while reporting a significant effect of spending.

Thus far, it is primarily explored that in reality there are effects of health spending but they somehow do not significantly show up in the analyses. However, it could also be argued that these effects do not exist in reality. More specifically, supply side investments may not translate into better waiting time performance. Many authors propose that it is not only about spending the money, but how the money is spent (Appleby, 2005; Allder, Walley & Silvester, 2011; Taylor, 2014). For instance, the efficiency of health spending may be negatively impacted by how commissioners invest in capacity. Temporary increases in capacity are essential as a short-term strategy to meet targets, but are often wasteful and expensive in the long term, which results in unsustainable waiting time reductions (Appleby, 2005). Moreover, a better control of capacity and demand variation, a CCG's organisational commitment & persistence in reducing waiting times, and hospitals' production efficiency are considered to be crucial for health spending to translate into better waiting time performance (Appleby, 2005; Allder et al., 2011; Taylor, 2014). These organisational and institutional factors serve as an interesting starting point for further empirical research and could contribute to the understanding of how health spending can reduce long waiting times.

This section continues by discussing the significant effects of health spending per capita on six-week-wait target performance of various key diagnostics.

Relative price of private treatment

A first hypothesis concerns the relative price of private treatment for patients. Notably, the effect of health spending is significant, positive and largest for all four endoscopy diagnostics (COL, CYS, FS, GAS). This suggests that particularly these diagnostics are sensitive to demand and or supply characteristics. At the same time, of all six-week-wait target key diagnostics considered in this study, endoscopy diagnostics are relatively the most costly for patients in the private health sector. On average, endoscopy procedures cost around £1,800 in private facilities, whereas imaging and physiology procedures cost on average around £550 and £390 respectively (see Table A10 in the appendix sourced from Private Health care UK, 2020). This relatively high price of private treatment may cause the demand for endoscopy diagnostics in public health care to be inelastic. Facing high out-of-pocket (OOP) costs, patients are less likely to opt for private treatment when waiting is long (Siciliani and Hurst, 2003; Siciliani and Iversen, 2012). Subsequently, it can be argued that, when demand is inelastic, these diagnostics are more susceptible to supply side interventions. This is supported by Shenbagaraj et al. (2017) who provide evidence that the largest challenges in meeting endoscopy waiting times are capacity, staffing issues and unplanned demand. Then, if increased health spending per capita reflects more funding into capacity, this may explain the positive association with waiting time performance for these endoscopy diagnostics.

Moreover, this hypothesis may also help explain the effect found for MRI six-week-wait target performance. Following the endoscopy diagnostics, MRI is the most expensive procedure (of those examined in this study) in private health care (average price = £1,075). The relative private price hypothesis could also clarify why a rather unexpected effect of health spending was found for Audiology Assessments (AA) target performance. A 1% increase in spending is associated with 0.34 %-points decrease in AA six-week-wait target performance, *ceteris paribus*. On average, AA in private facilities are relatively inexpensive (£135). Thus, OOP costs are relatively low in absolute terms. Patients may prefer to pay rather than to wait when AA waiting times are high. As a result, assuming spending is constant, waiting times for AA in public health care will stagnate or decrease. This is substantiated by the relatively constant six-week-wait target performance for AA over time. Over time, AA target performance decreased with 0.32 %-points from 98.45% in FY2014 to 98.13% in FY2019. This decrease is more than ten times as small compared to the average decrease in key diagnostic target performance over that same period (from 96.99% in FY2014 to 93.66% in FY2019). Additionally, compared to the average of 107.80%, AA has the lowest average annual growth rate in terms of patient volumes on over the period of analysis (102.35%). All average annual growth rates can be found in the appendix in Table A9. A visual presentation is found in Fig. 10.1

Relative patient volumes

A second suggested mechanism considers the relative patient volumes per key diagnostic. The endoscopy key diagnostics encompass the lowest patients volumes (10%) compared to the imaging (70%) and physiology (20%) key diagnostics (Fig. 11). Perhaps these smaller-scaled services are more susceptible to supply side investments as they can yield relatively greater benefits from extra capacity. Keeping in mind that outcome measures capture the percentage of patients seen within six weeks, then – by example – one extra cystoscopist can treat a higher share of the total number of cystoscopy patients[16] compared to the share one extra radiologist can treat of the total number of patients needing a CT[17]. Moreover, larger-scaled services may face scale inefficiencies. For example, small inefficiencies in administrative tasks may add up to substantial losses in productivity (Taylor, 2014). Also, it can be argued that it is more difficult for larger hospital departments to allocate funding efficiently. In other words, it may be harder for larger departments to pinpoint how to spend the money effectively compared to smaller-sized departments within the hospital.

This patient volume hypothesis assumes diminishing marginal returns to scale. As a diagnostic becomes larger in terms of patient volumes, one extra unit of capacity (reflected by health spending) may yield relatively less and less benefits in supply and subsequently waiting time performance. In line with this proposed mechanism is the similar effect found for Peripheral Neurophysiology (PN), which is another low-volume key diagnostic[18]. Also, this volume-mechanism may explain why the effect for MRI (relatively high volume[19]) is considerably smaller than the health spending coefficients for the endoscopy procedures and PN. Because of its higher patient-volume, it may take higher investments to obtain similar benefits or the beneficial effect of funding may be dampened due to inefficiencies in the production of health services.

Investments in equipment and technological innovations

Another hypothesis which may help explain the positive correlation of health spending and MRI six-week-wait target performance considers investments into equipment and technological innovations. In addition to investments to increase the workforce, some key diagnostics can also greatly benefit from more funding into (better) equipment. This may be especially relevant to the MRI key diagnostic. Of all (imaging) diagnostics studied, MRI is on average the most-time consuming procedure[20]. Investing in new and innovative MRI units can considerably speed up the process and ensure treating more patients in less time. Additionally, older MRI equipment has a high risk of failure and breakdowns which may also cause delays (European Society of Radiology, 2014). Moreover, the UK has the lowest number of MRI units per capita among comparable Western-European countries (e.g. UK has less than a third compared to Germany) (The Health Foundation, 2019). This indicates that there is much room to improve MRI

[16] Cystoscopy makes up around 2% of total patients in the SWW pathway.

[17] CT makes up around 10% of total patients in the SWW pathway.

[18] Peripheral Neurophysiology makes up around 2% of total patients in the SWW pathway.

[19] MRI makes up around 20% of total patients in the SWW pathway.

[20] On average: MRI scans take 45-90 minutes, Non-Obstetric Ultrasounds take 30 minutes, CT & DEXA scans take 10-20 minutes.

waiting times through investments in new and better MRI scanners. This notion has been underlined by many authors calling for substantially more funding into MRI units over the past years (European Society of Radiology, 2014; Royal College of Radiologists, 2017/2019; Graves et al., 2017; The Health Foundation, 2019; NHS England, 2021). However, investing in extra and better MRI equipment is costly. The cost of MRI machines can range from £0.6 to £1.9 million, not to mention the additional costs of constructing MRI suites (Keefer, 2019). This may explain why a relatively small positive effect of health spending on six-week-wait target performance for MRI was found. Investments in capacity through equipment and technological innovations may have beneficial effects on waiting times, but their high costs may dampen the effect.

Relative potential health losses from delayed diagnostics

Overall, it seems that diagnostics with a large potential health loss from delayed diagnosis report a significant and positive effect of health spending on six-week-wait target performance. For example, endoscopy diagnostics are commonly used to diagnose intestinal cancers, whereas Respiratory Physiology is generally used to diagnose sleep disorders. Both of these types of conditions can have serious health consequences for patients. However, without devaluing the way patients experience and are affected by their condition in either of these cases, it could be argued that the diagnosis of intestinal cancers is more pressing; potential health losses in terms of morbidity and mortality could be higher from a delayed diagnosis. This raises the question whether investments are perhaps allocated more efficiently (i.e. translate better into reduced waiting times) for such diagnostics. For example, larger potential health losses from delayed diagnosis may increase the scrutiny about failing to meet waiting time targets. In that case, physicians, hospital managers and or commissioners might exert even greater effort to use resources as effectively as possible. However, it is difficult to establish the precise health consequences of delayed diagnosis of each of the key diagnostics as they do not solely target on type of health condition. For instance, MRI scans are used to diagnose brain tumours (large potential health loss of delayed diagnosis), but also to diagnose fluid accumulation as a result of a bone fracture (relatively small potential health loss of delayed diagnosis). In any way, this hypothesis builds on the argument that reducing waiting times is not just about spending the money, but about how to spend the money most effectively. It can be interesting for future research to investigate this allocative efficiency of health spending.

Regional variation

The findings were found to differ across four geographical regions in England. Region Midlands & East mainly seemed to drive the overall results. This regional variation may indicate important differences in demand and supply characteristics impacting waiting times across regions. For example, efficiency of health spending may differ across regions; regions with a higher population density (e.g. London) may have larger hospitals, which may result in inefficiencies of scale causing investments to not effectively translate into reduced waiting times. Future studies may consider differences across regions when further investigating the relationship between spending and waiting times.

LIMITATIONS

This study faces some limitations which future work might address. First, health spending is proxied by CCG core services allocations. Although health expenditures and allocations were found to be highly correlated, there may still be differences from actual spending. Additionally, besides general hospital spending (which was of particular interest in this study) the CCG core services allocations include other components, such as acute hospital care, maternity care, mental health services, community services, and prescribing of medicines. Therefore, increases in CCG core services allocations may reflect increases in these other components as well, distorting the effect on waiting times for diagnostic hospital services. Moreover, this aggregated level of spending may allow for effects to go unobserved. Ideally, future research is able to address these issues and find more specific measures for health spending relating to distinct waiting time outcome measures.

Secondly, the dependent variables studied may have complicated capturing the effect of health spending properly. Waiting times were analysed as performance scores measuring the percentage of patients seen within a certain amount of time (two weeks or six weeks). For example, if the reduction in waiting times associated with spending is not enough for patients to fall within these target time frames, the beneficial effect of spending goes unrecognized. Therefore, it is recommended that future work considers mean waiting times in number of days instead when investigating the link between spending and waiting times.

At last, even though this study contributes to understanding how diagnostic waiting times may be impacted, it does not provide information on how the exact channels through which spending affects waiting times. As elaborated above, some hypotheses come to mind which may help explain the findings. However, future research is needed to better comprehend how demand and supply factors affect waiting times. Nevertheless, with these caveats in mind, the key findings still imply several significant and substantial effects which deserve further investigation.

CONCLUSION

This study has been the first to empirically investigate the link between health expenditures and waiting times for diagnostics. A special interest went out to diagnostic waiting times relating to cancer, as timely diagnosis for cancer is particularly important. Using the FDIV approach allows for causal interpretation of the estimates. Although no significant effects of health spending were found for both cancer-specific diagnostic waiting time targets (two-week-wait targets), a positive association was found for all six-week-wait targets related to the endoscopy key diagnostics. These endoscopy diagnostics are commonly used to diagnose intestinal cancers and may therefore particularly benefit from reducing long waiting times. There is growing data suggesting that delays in diagnosis are associated with a more advanced stage of cancer and it has been well established that prognosis is stage dependent (Neal, 2007; Tørring et al., 2017). This association has been widely recognised for rectal cancer (Ramos et al., 2007). Also, some studies found the same for colorectal cancer (Ramos et al., 2007; Neal, 2007). For oesophageal cancer, delays result in worse short-term outcomes since doubling times are shorter compared to colorectal

cancer (Grotenhuis et al., 2010). Therefore, delays may be more significant in this group. Furthermore, a delay in the diagnosis of bladder cancer has been found to increase the mortality risk independent of the tumour grade and disease stage (Liedberg et al., 2002; Hollenbeck et al., 2010). All in all, there has been much literature providing evidence that diagnostic delays of intestinal cancers matter for health outcomes (Neal, 2007). This becomes even more significant since the UK reports poorer cancer outcomes compared to many Western European countries (Berrino et al., 2007). In addition, long endoscopy waiting times will not only affect cancer care, but also many other conditions for which a delayed diagnosis may result in worse patient experience and considerable health losses (Lay et al., 1997).

The current Covid-19 pandemic highlights the negative health consequences of diagnostic delays. A recent study by Peery et al. (2018) calculated the impact of a 6-month suspension of elective colonoscopy. They estimated a delayed diagnosis of more than 2,800 colorectal cancers and 22,000 high-grade adenomatous polyps with malignant potential. The 6-month mortality rate for those (of the adenomatous polyps) eventually diagnosed with colorectal cancer was estimated to increase by 6.5% (Pita-Fernández et al., 2016). Considering that by March 2021 six-week-wait target performance for the endoscopy key diagnostics has only recovered up to approximately 50% (Fig. 6) adds to the concerns about the potential health losses due to diagnostic delays.

In conclusion, even though this study cannot uncover the exact channels through which health spending affects waiting times, the significant effects found for health spending substantiate the importance of further research on how demand and supply factors can reduce waiting times and maximise health gains. Future work is encouraged to take a closer look at the diagnostic waiting times examined in this study and to address the limitations mentioned above. Furthermore, future studies may consider regional differences since the effect of health spending was found to vary across geographical regions. At last, it may be especially interesting to investigate the efficiency of health spending ('how the money is spent') to identify mechanisms to reduce long waiting times.

References

Allder, S., Silvester, K. & Walley, P. (2010). Managing capacity and demand across the patient journey. *Clinical Medicine, Journal of the Royal College of Physicians of London.* Royal College of Physicians.

Allder, S., Walley, P. & Silvester, K. (2011). Is follow-up capacity the current NHS bottleneck? *Clinical Medicine, Journal of the Royal College of Physicians of London.* Royal College of Physicians.

Appleby, S. (2005). Cutting NHS waiting times: identifying strategies for sustainable reductions. King's Fund Publications.

Baltagi, B.H. (2005). Econometric analysis of panel data. Chapter 9. Third edition. John Wiley & Sons, Ltd. West Sussex, England.

Berrino F, De Angelis R, Sant M, Rosso S, Lasota M, Coedergh JW, Santaquilani M. Survival for eight major cancers and all cancers combined for European adults diagnosed in 1995–1999: results of the EUROCARE-4 study. Lancet Oncol. 2007;8:773–783.

Besley, T. et al. (1998). Private and public health insurance in the UK. European Economic Review. Vol.42: 491-497.

Biorn, E., Hagen, T. Iversen, T. & Magnussen, J. (2002). The effect of activity-based financing on hospital efficiency. Working Paper 8, Health Economic Research Programme University of Oslo.

Boyle, S. (2011). United Kingdom (England): health system review. Health Systems in Transition. Vol. 13(1).

Brown, H., Ellins, J., Kearney, J., Singh, K., Jackson, O., Krell, H. et al. (2014). Measuring up? The health of NHS cancer services. Health Services Management Centre: University of Birmingham & ICF GHK Consulting. Commissioned by Cancer Research UK.

Clemmesen, F. & Hansen, M. (2003). Erfaringerne med meraktiviteitsfinansiering af sygehuse. Samfundsokonomen. Vol.3: 11-16.

Crawford, R. & Emmerson, C. (2012). NHS and social care funding: the outlook to 2021/22. Nuffield Trust & Institute for Fiscal Studies.

Coomber, N. (2014). Delivering Cancer Waiting Times: a good practice guide. NHS Interim Management and Support.

Cooper, R.B. (1981). Basic queueing theory. North Holland Inc, New York.

Cutler, D. (2002). Equality, Efficiency, and Market Fundamentals: the Dynamics of International Medical Care Reform. Journal of Economic Literature 40, 881-906.

Das-Thomson, J. (2019). Merging CCGs to align commissioning. Pharmafield.
URL: https://pharmafield.co.uk/health care/merging-ccgs-to-align-commissioning/

DeCoster, C., Carriere, K.C., Peterson, S., Walld, R. & MacWilliam, L. (2002). Waiting times for surgical procedures. Medical Care. Vol.37(6): 187-205.

Delattre, E. and Dormont, B. (2003). Fixed Fees and Physician-Induced Demand: A Panel Data Study on French Physicians. Health Economics, 12, 741-754.

Department of Health & Social Care (2012). NHS Constitution for England. UK Government.
URL: https://www.gov.uk/government/publications/the-nhs-constitution-for-england/the-nhs-constitution-for-england

Dijk, C.E. van, Berg, B. van den, Verheij, R.A., Spreeuwenberg, P., Groenewegen, P.P., Bakker, D.H. de. Moral Hazard and supplier-induced demand: empirical evidence in general practice. Health Economics: 2013, (22(3), 340-352.

Dimakou, S., Parkin, D., Devlin, N. & Appleby, J. (2009) Identifying the impact of government targets on waiting times in the NHS. Health Care Manag Sci, 12: 1-10.

Downing A, Aravani A, Macleod U, et al. Early mortality from colorectal cancer in England: a retrospective observational study of the factors associated with death in the first year after diagnosis. Br J Cancer 2013;108:681–5.

Ellis, A. & Fry, R. (2010). Regional health inequalities in England. Regional Trends 42: 2010 edition.

European Society of Radiology (2014). Renewal of radiological equipment. Insights Imaging. Vol 5: 543-546.

Gardner, T. & Fraser, C. (2021). Longer waits, missing patients and catching up. The Health Foundation.
URL: https://www.health.org.uk/news-and-comment/charts-and-infographics/how-is-elective-care-coping-with-the-continuing-impact-of-covid-19

Gosden, T, Forland, F. Kristiansen, I.S., Sutton, M., Leese, B., Giuffrida, A. et al. (2001). Impact of payment method on behaviour of primary care physicians: a systematic review. Journal of Health Services Policy. Vol. 6: 44-55.

Gravelle, H., Smith, P.C. & Xavier, A. (2003). Waiting lists and waiting times: a model of the market for elective surgery. Oxford Economic Papers. Vol. 55(1): 81-103.

Graves, M., Malcolm, P., Lipton, A., Scurr, E., Horne, D., Eve, J. & Hall, A. (2017). Magnetic resonance imaging (MRI) equipment, operations and planning in the NHS. Report from the Clinical Imaging Board. Clinical Imaging Board.

Grotenhuis BA, van Hagen P, Wijnhoven BPL, et al. Delay in diagnostic workup and treatment of esophageal cancer. J Gastrointest Surg 2010;14:476–83.

Hamilton, W. (2020). Cancer diagnostic delay in the COVID-19 era: what happens next? The Lancet. Vol. 21:1000-1002.

Hauck, K. and Street, A. (2006). Do targets matter? A comparison of English and Welsh national health properties. Clayton, Australia: Monash University. York, England: University of York.

Harrison, A. & Appleby, J. (2009). Reducing waiting times for hospital treatment: lessons from the English NHS. Journal of Health Services Research & Policy. Vol. 14(3): 168-173.

Hobbs, R., Bankhead, C., Mukhtar, T., Stevens, S., Perera-Salazar, R., Holt, T. et al. (2016). Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007-2014. The Lancet. Vol. 387(10035): 2323-2330.

Hollenbeck, B.K., Dunn, R.L., Ye, Z., Hollingsworth, J.M. Skolarus, T.A., Kim, S.P., Montie, J.E. et al. (2010). Delays in diagnosis and bladder cancer mortality. American Cancer Society. Vol. 116: 5235-5242.

Health Foundation (2019). Lack of investment in NHS infrastructure is undermining patient care. The Health Foundation. URL: https://www.health.org.uk/news-and-comment/news/lack-of-investment-in-nhs-infrastructure-is-undermining-patient-care

Iacobucci, G. (2020). Covid-19: NHS leaders braced for longer waiting times as service deals with fallout. BMJ: London. Vol 370.

Iversen, T. (1993). A theory of hospital waiting lists. Journal of Health Economics. Vol.12: 55-71.

Iversen, T. & Luras, H. (2002). Waiting times as a competitive device: an example from general medical practice. International Journal of Health Care Finance and Economics. Vol.2: 189-204.

Keefer, A. (2019). How much do MRI machines cost? eHow Health.
URL: https://www.ehow.co.uk/about_4731161_much-do-mri-machines-cost.html

Lay CS, Tsai YT, Teg CY, et al. Endoscopic variceal ligation in prophylaxis of first variceal bleeding in cirrhotic patients with high-risk esophageal varices. Hepatology 1997;25:1346–50.

Liedberg, F., Anderson, H., Mansson, A. & Mansson, W. (2002). Diagnostic delay and prognosis in invasive bladder cancer.

Lindsay, C.M. & Feigenbaum, B. (1984). Rationing by waiting lists. American Economic Review. Vol.74: 404-17.

Maringe, C., Spicer, J., Morris, M., Purushotham, A., Nolte, E., Sullivan, R. et al. (2020). The impact of the COVID-19 pandemic on cancer deaths due to delays in diagnosis in England: a national, population-based, modelling study. Lancet Oncol. Vol. 21: 1023-34.

Martin, S. & Smith, P.C. (1999). Rationing by waiting lists: an empirical investigation. Journal of Public Economics. Vol.71: 141-64.

Martin, S. & Smith, P.C. (2003). Using panel methods to model waiting times for National Health Services surgery. Journal of the Royal Statistical Society. Vol.166: 1-19.

Morga, A. & Xavier, A. (2001). Hospital specialists' private practice and its impact on the number of NHS patients treated and on the delay for elective treatment. Discussion papers in Economics. The University of York, n.2001/01.

Mot, E.S. (2003). Paying the medical specialist: the eternal puzzle: experiments in the Netherlands. PhD thesis, Amsterdam.

National Statistics (2015). The English Indices of Deprivation 2015. Department for Communities and Local Government. Statistical Release, September 2015.

Neal, R.D. (2009). Do diagnostic delays in cancer matter? Br. J. Cancer. Vol. 101(2): S9-S12.

NHS Digital (2018). NHS workforce statistics. Author: Workforce and Facilities, NHS Digital.
URL: https://files.digital.nhs.uk/pdf/j/r/nhs_workforce_statistics__december_2017_bulletin.pdf

NHS England (2015). Diagnostic Waiting Times and Activity. Guidance on completing the diagnostic waiting times & activity monthly data collection. Prepared by Analytical Service (Operations).
URL: https://nhsenglandfilestore.s3.amazonaws.com/stats/DM01-guidance-v-5.32.doc

NHS England (2015b). Monthly Diagnostic Waiting Times and activity. Data quality and Methodology Statement. Prepared by Analytical Service (Operations). URL: https://nhsenglandfilestore.s3.amazonaws.com/stats/Diagnostics-Data-Quality-Statement.pdf

NHS England (2016). General Practice forward view. NHS England Publications.
URL: https://www.england.nhs.uk/wp-content/uploads/2016/04/gpfv.pdf

NHS England (2018). CCG outcomes tool. (NB. 2019 value estimated based on national population growth rate for England (0.5535%). Population values missing for CCGs discontinued before 2018 or established in 2019. New ONS estimates available in Oct. '20.
URL: https://www.england.nhs.uk/ccg-out-tool/

NHS England (2019) The NHS Long Term Plan. NHS England Publications.
URL: https://www.longtermplan.nhs.uk/wp-content/uploads/2019/08/nhs-long-term-plan-version-1.2.pdf

NHS England (2020). Clinical Commissioning Groups (CCGs). NHS England Statistics.
URL: https://www.england.nhs.uk/commissioning/who-commissions-nhs-services/ccgs/

NHS England (2020b). National Cancer Waiting Times monitoring dataset guidance – version 11.0. NHS England Statistics.

NHS England (2020c). Diagnostic Waiting Times and Activity. NHS England Statistics.
URL: https://www.england.nhs.uk/statistics/statistical-work-areas/diagnostics-waiting-times-and-activity/

NHS England (2020d). Allocations. NHS Statistics. URL: https://www.england.nhs.uk/allocations/

NHS England (2021). Diagnostic imaging network commercial structure and operational governance guide. NHS England Publications. URL: https://www.england.nhs.uk/wp-content/uploads/2021/04/B0030-Operational-governance-guide-April-2021.pdf

NHS England & Improvement (2019). Technical guide to allocation formulae and pace of change. NHS England Publications.

NHS Improvement (2009). Going further on cancer waits: the symptomatic breast two week wait standard: a guide to support implementation. NHS Improvement.

Nomis (2011). 2011 Census [webpage]. Nomis Official Labour Market Statistics. URL: https://www.nomisweb.co.uk/sources/census_2011

Nuffield Trust (2018). Budget 2018: what it means for health and social care. Nuffield Trust, The Health Foundation & The King's Fund. URL: https://www.nuffieldtrust.org.uk/resource/budget-2018-what-it-means-for-health-and-social-care

Nuffield Trust (2021). Cancer Waiting Times. Nuffield Trust & The Health Foundation, QualityWatch. URL: https://www.nuffieldtrust.org.uk/resource/cancer-waiting-time-targets#background

Nuffield Trust (2021b). Diagnostic Test Waiting Times. Nuffield Trust & The Health Foundation, QualityWatch. URL: https://www.nuffieldtrust.org.uk/resource/diagnostic-test-waiting-times#background

Nuti, S. & Vainieri, M. (2012). Managing waiting times in diagnostic medical imaging. BMJ Open: 2(6).

OECD (2020). Waiting times for health services: next in line. OECD Health Policy Studies.

Office for National Statistics (2018). UK Health Accounts: 2016. Office for National Statistics. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/health caresystem/bulletins/ukhealthaccounts/2016

Peak, S. (2018). Delivering endoscopy services to meet the demands of today and the challenges of tomorrow. Vanguard Health care Solutions.

Peery AF, Crockett SD, Murphy CC, et al. Burden and cost of gastrointestinal, liver, and pancreatic diseases in the United States: update 2018. Gastroenterology 2019;156:254–72.

Pita-Fernández S, González-Sáez L, López-Calviño B, et al. Effect of diagnostic delay on survival in patients with colorectal cancer: a retrospective cohort study. BMC Cancer 2016;16:664.

Prime Minister's Office (2019). Boris Johnson's first speech as Prime Minister: 24 July 2019. UK Government. URL: https://www.gov.uk/government/speeches/boris-johnsons-first-speech-as-prime-minister-24-july-2019

Private Health care UK (2020). Cost of private health care. Retrieved: 16 June 2021. URL: https://www.privatehealth.co.uk/costs/

Propper C. (1995). Agency and incentives in the NHS internal market, Social Science and Medicine 40, 1683-1690.

Propper C., Sutton M., Whitnall C. & Windmeijer F. (2007). Did 'targets and terror ' reduce waiting times in England for hospital care? Working Paper No. 07/179 published in The B.E. Journal of Economic Analysis & Policy 8, (1), Article 5 (2008).

Ramos, M., Esteva, M., Cabeza, E., Campillo, C., Llobera, J. & Aguiló, A. (2007). Relationship of diagnostic and therapeutic delay with survival in colorectal cancer: a review. Eur J Cancer. Vol. 43(17): 2467-78.

Ransom, S.B. et al. (1996). The effect of capitated and fee-for-service remuneration on physician decision making in gynaecology. Obstetrics and Gynaecology. Vol.5: 707-710.

Royal College of Radiologists (2017). NHS must do more to future-proof its MRI capacity, say imaging experts. Clinical Imaging Board. URL: https://www.rcr.ac.uk/posts/government-pledges-%C2%A3200m-new-imaging-scanners

Royal College of Radiologists (2019). Government pledges 200m for new imaging scanners. Clinical Imaging Board. URL: https://www.rcr.ac.uk/posts/nhs-must-do-more-future-proof-its-mri-capacity-say-imaging-experts

Seyedin, H., Afshari, M., Isfahani, P., Rakhshan, A., Hasanzadeh, E. & Taherimirghaed, M. (2020). Main factors leading to supplier-induced demand in Iran: a comprehensive review. Health Scope. Vol. 9(2).

Seyedin, H., Afshari, M., Isfahani, P., Hasanzadeh, E., Radinmanesh, M. & Corani Bahador, R. (2021) The main factors of supplier-induced demand in health care: a qualitative study. J Educ Health Promot. Vol. 10: 49.

Shenbagaraj, L., Thomas-Gibson, S., Stebbing, J., Broughton, R., Dron, M., Johnston, D. et al. (2017). Endoscopy in 2017: a national survey of practice in the UK. Frontline Gastroenterology: 10(1): 7-15.

Siciliani, L. (2015). Waiting times: evidence of social inequalities in access for care. Data and Measures in Health Services. Health Services Research. Springer, Boston, MA: 1-17.

Siciliani, L. & Hurst, J. (2003). Tackling excessive waiting times for elective surgery: a comparison of policies in twelve OECD countries. OECD Health Working Papers No. 6.

Siciliani, L. & Hurst, J. (2003b). Explaining waiting times variations for elective surgery across OECD countries OECD Health Working Papers No. 7.

Siciliani, L. & Iversen, T. Waiting times and waiting lists. The Elgar Companion to Health Economics. Chapter 24.

Siu, A.L. et al. (1988). Use of the hospital in a randomised trial of prepaid care. JAMA. Vol.259: 1343-1346.

StataCorp (2019) Stata Statistical Software: Release 16. College Station, Texas: StataCorp LLC.

Statista (2021). Consumer price index (CPI) of medical and paramedical services annually in the United Kingdom (UK) from 2003 to 2020. Statista 2021.
URL: https://www.statista.com/statistics/286517/medical-services-consumer-price-index-cpi-annual-average-in-uk/

Stevens, S. & Pritchard, A (2020). Letter to NHS Trusts: important and urgent – next steps on NHS response to COVID-19. NHS England & Improvement.

Sud, A., Torr, B., Jones, M.E., Broggio, J., Scott, S., Loveday, C. et al. (2020). Effect of delays in the 2-week-wait cancer referral pathway during the COVID-19 pandemic on cancer survival in the UK: a modelling study. Lancet Oncol. Vol. 21: 1035-44.

Taylor, J. (2014). Seven ways to no delays. NHS Improving Quality.
URL: https://www.england.nhs.uk/improvement-hub/wp-content/uploads/sites/44/2017/11/Seven-Ways-to-No-Delay.pdf

Thorlby, R. (2020). The English health care system. International Profiles of Health Care Systems (2020), 59-69.

Thorlby, R., Fraser, C. & Gardner, T. (2020). Non-COVID-19 NHS care during the pandemic: activity trends for key NHS services in England. The Health Foundation.

Thorlby R., Gardner T. & Turton C. (2019) NHS Performance and Waiting Times: Priorities for the Next Government. The Health Foundation.

Tørring ML, Murchie P, Hamilton W, et al. Evidence of advanced stage colorectal cancer with longer diagnostic intervals: a pooled analysis of seven primary care cohorts comprising 11 720 patients in five countries. Br J Cancer 2017;117:888–97

Valori, R.M., Barton, R. & Johnston, D.K. (2009). The English National Endoscopy Quality Assurance Programme: quality of care improves as waits decline. Gastrointestinal Endoscopy. Vol. 69(5).

Wilson, S.E. & Longmire W.P. (1978). Does method of surgeon payment affect surgical care? Journal of Surgical Research. Vol.24: 457-468.

Wooldrigde, J. (2013). Introductory Econometrics: a modern approach. 5th edition. Chapters 13-15. South-Western, Cengage Learning.

Wyatt, S. (2019). Waiting times and attendance durations at English accident and emergency departments. The Strategy Unit. NHS Midlands and Lancashire, Commissioning Support Unit.

Appendices

Table A1. Sample characteristics for time-constant covariates included in Pooled OLS estimation (N=1098).

	Variable		Obs	Mean	Std	Min	Max
Covariates	Health problem (% population)	Aged 16-64	1098	12.76	2.63	7.63	20.51
Time-constant		All ages	1098	17.61	3.08	11.20	25.60
	Provides unpaid care (% population)		1098	10.22	1.29	6.50	13.00
	Country of birth non-EU (% population)		1098	9.77	9.71	1.10	42.40
	Ethnicity (% population)	Asian	1098	7.90	9.76	0.50	60.90
		Black	1098	3.69	5.87	0.10	27.20
		Mixed	1098	2.27	1.64	0.50	7.60
		Other	1098	1.09	1.53	0.10	10.60
		White (base)	1098	85.27	16.20	27.80	98.80
	Tenure (% households)	Owner	1098	63.03	11.22	24.20	82.70
		Private rent	1098	17.06	5.86	9.20	40.10
		Social rent	1098	17.76	6.93	6.20	43.70
	Unemployed and economically active (% population)		1098	6.34	2.06	3.42	15.61
	Economically active including employed (% population)		1098	70.00	3.26	55.55	77.93
	Occupation (% population aged 16-64)	Professional	1098	17.26	4.25	9.02	31.86
		Agricultural	1098	0.81	0.98	0.03	5.42
		No qualifications	1098	14.81	4.05	6.14	29.53
	Lone person aged >65 (% households)		1098	12.25	2.09	6.00	17.76
	Lone parent (% households)		1098	10.66	2.25	6.83	19.04
	No cars or vans (% households)		1098	26.19	11.78	10.50	64.80
	Region dummies	North East	57				
		North West	189				
		Yorkshire Humber	132				
		East Midlands	116				
		West Midlands	126				
		East England	114				
		South East	210				
		South West	64				
		London (base)	192				

Note: Outcome measures report waiting time target performance scores. The operational standard for Two Week Wait targets is 93%. The operational standard for Six Week Wait targets is 99%. The average number of people registered with GP over period 2014-2019 per CCG is used as weight.

Table A2. Complete POLS, FD & FDIV regression results for two-week-wait target performance.

	TWW (cancer suspected)			TWWBS (cancer not initially suspected)		
	POLS	**FD**	**FDIV**	**POLS**	**FD**	**FDIV**
ln hospital spending per capita (£)	4.53 (4.07)	-2.86 (5.08)	7.58 (9.25)	21.12** (9.70)	-15.83 (9.60)	4.65 (24.34)
ln market forces factor index	-15.29 (11.27)	-62.82 (78.01)	-74.24 (81.73)	-72.95** (29.05)	-152.51 (126.15)	-174.92 (123.65)
ln % population aged						
0-9	3.91 (3.900)	-8.15 (9.18)	-8.66 (8.96)	-6.44 (12.36)	-66.39** (29.75)	-67.40** (29.35)
10-19	-13.17 (5.93)	-4.28 (12.31)	-3.84 (12.13)	-33.05** (14.59)	-15.04 (30.19)	-14.18 (29.75)
20-29	1.67 (2.71)	24.46** (11.67)	25.23** (11.46)	-0.48 (7.94)	34.77 (27.84)	36.26 (27.06)
30-39	-7.82 (4.80)	29.16* (17.19)	28.96* (17.05)	-3.71 (13.93)	23.75 (37.32)	23.35 (36.85)
40-49	12.80** (4.91)	29.38** (13.45)	29.68** (13.25)	34.59*** (12.24)	26.44 (34.54)	27.03 (33.92)
50-59	-0.74 (5.20)	12.85 (13.12)	14.24 (13.04)	-11.69 (13.89)	-85.81* (44.63)	-83.09* (43.29)
60-69	-4.36 (4.29)	22.07** (10.04)	22.98** (9.81)	5.64 (10.93)	68.92** (31.30)	70.70** (31.38)
70-79	-1.42 (3.38)	19.10** (9.48)	18.77** (9.43)	-16.73* (8.96)	52.10* (28.16)	51.46* (27.88)
ln Index of Deprivation						
Barriers to housing/services	-0.88 (1.13)	-2.15 (2.61)	-2.71 (2.67)	-2.36 (2.96)	2.85 (10.57)	1.75 (10.87)
Crime	0.20 (1.60)	-2.24 (2.80)	-2.35 (2.73)	-0.79 (5.16)	-9.52 (10.29)	-9.73 (10.05)
Education	1.02 (1.33)	3.22 (5.14)	2.95 (5.14)	6.96** (3.52)	1.82 (16.28)	1.28 (16.12)
Employment	-5.64 (4.91)	0.52 (14.96)	1.72 (14.85)	-7.25 (12.98)	-47.04 (43.08)	-44.70 (43.07)
Income	-4.23 (4.73)	-2.36 (13.83)	-2.73 (13.52)	-27.11** (12.60)	-48.60 (49.02)	-49.31 (48.42)
Living environment	0.14 (0.76)	1.98 (1.52)	1.37 (1.51)	-1.23 (2.01)	1.70 (5.04)	0.49 (4.71)
Time-fixed variables & region dummies	Yes	No	No	Yes	No	No
Year 2015	-0.10 (0.28)	1.11* (0.57)	1.06* (0.59)	-0.06 (0.63)	4.42** (1.97)	4.32** (1.95)
Year 2016	-0.17 (0.58)	1.44** (0.62)	0.86 (0.84)	-0.93 (1.20)	4.83** (1.89)	3.69 (2.26)
Year 2017	-0.30 (0.77)	0.75 (0.51)	0.78 (0.50)	-0.73 (1.63)	2.53** (1.45)	2.59* (1.45)
Year 2018	-2.10** (1.05)	-1.09** (0.49)	-1.08** (0.48)	-7.98** (2.48)	-5.79*** (1.63)	-5.77*** (1.60)
Year 2019	-2.82** (1.36)	(omitted)	(omitted)	-9.03** (3.27)	(omitted)	(omitted)
Constant	47.18 (38.57)	-1.17 (0.78)	-1.01 (0.80)	-87.43 (101.34)	-6.45** (2.52)	-6.14** (2.54)
First stage results:						
ln distance from target index			.509***			.509***
			(.045)			(.045)
F-stat excluded instrument			130.68			130.68
Obs	1098	914	914	1098	914	914
R^2	.232	.077	.075	.299	.111	.110

Note: All regressions include a constant term. POLS regressions include time-constant covariates and 8 region dummies not depicted in this table. All regressions include year dummies. The average number of people registered with GP over period 2014-2019 is used as weight. Robust standard errors clustered at CCG level are in the parentheses. Significant at: *10%, **5%, ***1% level.

Table A3. Complete POLS, FD & FDIV regression results for six-week-wait target performance of 13 key diagnostics.

	Magnetic Resonance Imaging (Imaging)			Audiology Assessments (Physiology)		
	POLS	**FD**	**FDIV**	**POLS**	**FD**	**FDIV**
ln hospital spending per capita (£)	2.55 (1.95)	0.30 (2.18)	11.85** (4.70)	-5.80 (3.63)	-9.55 (6.64)	-34.12*** (11.24)
ln market forces factor index	0.85 (7.32)	18.17 (24.00)	5.54 (24.99)	8.00 (9.17)	-455.77** (205.63)	-428.88** (197.66)
ln % population aged						
0-9	1.44 (3.77)	0.45 (4.86)	-0.12 (4.91)	7.76** (3.56)	0.27 (7.84)	1.49 (7.94)
10-19	-3.71 (3.63)	7.05 (7.19)	7.54 (7.24)	-5.43 (4.44)	-13.34 (10.23)	-14.37 (10.31)
20-29	-0.01 (1.77)	15.25** (5.93)	16.09*** (5.87)	0.72 (1.89)	1.75 (13.00)	-0.05 (12.89)
30-39	-1.66 (3.55)	14.25 (8.72)	14.03 (8.56)	-9.65** (4.19)	-1.51 (16.76)	-1.04 (16.46)
40-49	8.43*** (3.17)	15.22** (6.10)	15.55** (6.06)	6.47** (2.73)	9.37 (9.00)	8.66 (9.07)
50-59	-2.47 (3.48)	15.37** (6.76)	16.90*** (6.36)	-5.46 (3.62)	5.99 (11.89)	2.72 (12.02)
60-69	-3.17 (2.55)	0.28 (4.88)	1.28 (4.86)	2.77 (2.32)	-7.52 (10.68)	-9.66 (10.76)
70-79	3.91** (1.73)	11.06** (5.05)	10.79** (4.97)	-0.66 (2.40)	-3.59 (9.48)	-2.82 (9.30)
ln Index of Deprivation						
Barriers to housing/services	0.21 (0.63)	1.20 (1.74)	0.58 (1.73)	-0.61 (0.74)	1.30 (2.03)	2.62 (2.21)
Crime	-0.16 (1.15)	1.021 (1.75)	0.91 (1.70)	-1.03 (1.05)	1.43 (2.32)	1.68 (2.34)
Education	0.71 (1.07)	1.33 (4.16)	1.03 (4.05)	0.18 (1.22)	5.77 (4.92)	6.42 (4.95)
Employment	-0.52 (4.07)	-8.91 (8.53)	-7.58 (8.27)	-6.18* (3.69)	-12.00 (11.00)	-14.82 (10.99)
Income	-1.08 (2.87)	7.43 (9.64)	7.03 (9.35)	7.19* (3.65)	3.40 (10.55)	4.26 (10.68)
Living environment	-0.60 (0.43)	0.54 (0.93)	-0.14 (0.97)	0.28 (0.57)	0.39 (1.37)	1.83 (1.62)
Time-fixed variables & region dummies	Yes	No	No	Yes	No	No
Year 2015	-0.06 (0.17)	1.06*** (0.33)	1.00*** (0.32)	0.22 (0.23)	0.22 (0.42)	0.34 (0.41)
Year 2016	0.06 (0.32)	1.33*** (0.35)	0.69* (0.38)	0.44 (0.37)	0.23 (0.58)	1.60** (0.65)
Year 2017	0.10 (0.46)	1.01*** (0.28)	1.04*** (0.28)	0.29 (0.55)	-0.29 (0.45)	-0.36 (0.45)
Year 2018	-0.72 (0.62)	0.17 (0.33)	0.18 (0.33)	1.37* (0.71)	0.98*** (0.36)	0.96*** (0.35)
Year 2019	-1.72** (0.83)	(omitted)	(omitted)	1.50 (0.94)	(omitted)	(omitted)
Constant	92.20*** (20.90)	-1.47*** (0.42)	-1.30*** (0.41)	136.35*** (33.56)	-0.60 (0.47)	-0.98** (0.49)
First stage results:						
ln distance from target index			.509***			.509***
			(.045)			(.045)
F-stat excluded instrument			130.68			130.68
Obs	1098	914	914	1098	914	914
R^2	.242	.052	.047	.158	.024	.014
	Peripheral Neurophysiology (Physiology)			**Urodynamics (Physiology)**		
	POLS	**FD**	**FDIV**	**POLS**	**FD**	**FDIV**
ln hospital spending per capita (£)	7.15* (4.13)	25.37* (14.14)	67.62*** (24.05)	6.29 (10.23)	-29.57 (30.96)	-74.96* (41.96)
ln market forces factor index	-36.12*** (13.71)	-187.56 (132.10)	-233.77* (138.18)	32.00 (32.19)	16.19 (379.61)	65.84 (370.59)
ln % population aged						
0-9	-4.32 (5.34)	4.37 (17.48)	2.28 (16.92)	-0.41 (13.77)	-17.28 (37.63)	-15.03 (36.75)
10-19	0.21 (6.35)	-1.91 (18.72)	-0.14 (17.99)	-4.90 (15.70)	82.28 (41.20)	80.37* (41.19)
20-29	-1.08 (2.74)	-38.54* (18.66)	-35.46* (18.10)	8.88 (7.66)	66.93 (40.98)	63.62 (41.14)
30-39	-5.41 (5.25)	-36.01 (22.19)	-36.83* (22.14)	21.59 (15.14)	72.70 (46.33)	73.58 (46.10)
40-49	2.23 (5.20)	-9.69 (17.86)	-8.47 (17.90)	18.97 (13.10)	28.79 (35.67)	27.48 (35.35)
50-59	-7.30 (6.10)	6.16 (24.18)	11.77 (25.07)	6.50 (14.34)	145.11*** (40.17)	139.08*** (39.61)
60-69	0.51 (4.32)	-42.95** (18.14)	-39.27** (17.54)	7.63 (12.75)	4.30 (36.82)	0.34 (36.72)
70-79	0.93 (3.46)	-30.51* (16.16)	-31.84** (15.82)	4.44 (10.86)	11.29 (33.14)	12.72 (33.02)
ln Index of Deprivation						
Barriers to housing/services	0.14 (0.97)	-1.41 (3.77)	-3.69 (4.04)	-4.34 (3.34)	-12.45 (7.95)	-10.01 (8.26)
Crime	3.12 (1.97)	9.77** (4.27)	9.35 (4.14)	-0.33 (4.97)	-2.87 (8.89)	-2.42 (9.02)
Education	-1.71 (1.76)	0.34 (7.39)	-0.78 (7.61)	6.79 (5.58)	-7.56 (16.40)	-6.36 (16.21)
Employment	12.33** (6.19)	46.56* (26.32)	51.40* (26.77)	-16.75 (14.35)	-92.19** (40.99)	-97.39** (41.23)
Income	-12.26*** (4.41)	-45.07* (24.30)	-46.53* (24.50)	-7.74 (13.29)	114.20 (42.77)	115.77*** (43.65)
Living environment	1.54** (0.77)	-2.28 (2.56)	-4.77 (2.91)	3.57* (2.14)	10.29* (5.54)	12.96** (5.78)
Time-fixed variables & region dummies	Yes	No	No	Yes	No	No
Year 2015	-0.44 (0.73)	2.30* (1.22)	2.09 (1.27)	-2.11** (0.94)	4.64*** (1.77)	4.86*** (1.79)
Year 2016	1.50** (0.63)	3.85*** (1.00)	1.50 (1.47)	-3.01** (1.30)	7.77*** (2.46)	10.29*** (2.97)
Year 2017	0.79 (0.92)	2.54*** (0.66)	2.66*** (0.63)	-5.18*** (1.84)	4.53*** (1.36)	4.40*** (1.34)
Year 2018	1.46 (1.21)	4.24*** (0.55)	4.26*** (0.54)	-6.92*** (2.58)	4.91*** (1.10)	4.88*** (1.08)
Year 2019	-2.07 (1.48)	(omitted)	(omitted)	-13.51*** (3.51)	(omitted)	(omitted)
Constant	-31.31 (41.96)	-2.37** (1.11)	-1.72 (1.25)	104.26 (116.10)	-8.43*** (2.38)	-9.13*** (2.44)
First stage results:						
ln distance from target index			.509***			.509***
			(.045)			(.045)
F-stat excluded instrument			130.68			130.68
Obs	1098	914	914	1098	914	914
R^2	.221	.113	.105	.248	.043	.040

Note: All regressions include a constant term. POLS regressions include time-constant covariates and 8 region dummies not depicted in this table. All regressions include year dummies. The average number of people registered with GP over period 2014-2019 is used as weight. Robust standard errors clustered at CCG level are in the parentheses. Significant at: *10%, **5%, ***1% level.

Table A3 continued. Complete POLS, FD & FDIV regression results for six-week-wait target performance of 13 key diagnostics.

	Colonoscopy (Endoscopy)			Cystoscopy (Endoscopy)		
	POLS	**FD**	**FDIV**	**POLS**	**FD**	**FDIV**
ln hospital spending per capita (£)	9.11 (7.87)	29.14* (17.59)	82.77** (33.55)	18.73** (8.54)	29.63* (17.09)	70.45** (28.17)
ln market forces factor index	-20.88 (26.57)	243.29 (304.08)	184.62 (303.35)	-13.78 (23.86)	539.54* (325.30)	494.88 (314.66)
ln % population aged						
0-9	1.86 (10.80)	-33.16 (32.12)	-35.81 (31.93)	-20.06* (10.34)	-15.61 (19.15)	-17.64 (19.26)
10-19	10.08 (12.14)	-11.02 (34.58)	-8.77 (34.16)	20.78* (12.45)	9.75 (22.60)	11.46 (22.23)
20-29	-4.04 (6.09)	6.07 (34.36)	9.98 (33.96)	-4.74 (6.49)	10.34 (25.10)	13.32 (24.98)
30-39	-3.26 (10.23)	-2.48 (37.63)	-3.52 (37.48)	24.40** (11.01)	-0.21 (29.84)	-1.00 (29.79)
40-49	15.47 (11.46)	19.53 (24.64)	21.08 (24.55)	19.08* (9.94)	31.85 (27.64)	33.03 (27.70)
50-59	-9.79 (11.89)	21.12 (34.64)	28.25 (34.19)	-13.96 (12.92)	56.41** (27.58)	61.83** (27.50)
60-69	-9.00 (12.01	-46.23 (28.93)	-41.55 (28.72)	4.62 (9.32)	-17.33 (22.62)	-13.77 (23.26)
70-79	15.04* (7.93)	-32.87 (22.17)	-34.55 (22.06)	4.90 (7.29)	-7.51 (18.62)	-8.79 (18.65)
ln Index of Deprivation						
Barriers to housing/services	-6.21** (2.47)	1.38 (6.46)	-1.51 (6.33)	-1.36 (2.19)	-7.68 (6.90)	-9.87 (6.72)
Crime	0.71 (3.83)	16.09** (6.63)	15.56** (6.47)	-5.14 (3.47)	-0.63 (4.33)	-1.03 (4.36)
Education	-2.24 (3.54)	36.01** (15.07)	34.59** (14.72)	-2.31 (3.53)	5.18 (13.19)	4.10 (12.88)
Employment	-23.71** (11.70)	-45.87 (38.09)	-39.72 (38.36)	1.59 (10.20)	-12.39 (27.66)	-7.72 (27.47)
Income	14.80 (9.82)	33.76 (29.37)	31.90 (28.97)	-12.63 (9.00)	8.62 (31.61)	7.20 (31.30)
Living environment	0.71 (1.42)	3.95 (4.94)	0.78 (5.14)	0.09 (1.36)	-5.18 (3.19)	-7.58** (3.77)
Time-fixed variables & region dummies	Yes	No	No	Yes	No	No
Year 2015	-2.16*** (0.77)	0.91 (1.32)	0.65 (1.34)	0.25 (0.63)	1.63 (0.99)	1.43 (0.99)
Year 2016	-0.61 (1.05)	3.61* (1.87)	0.63 (2.28)	-0.15 (1.01)	1.09 (1.55)	-1.18 (1.93)
Year 2017	-2.84** (1.36)	1.48 (1.09)	1.63 (1.08)	-1.06 (1.62)	1.28 (0.87)	1.39 (0.86)
Year 2018	-5.96*** (2.00)	0.44 (1.40)	0.47 (1.37)	-2.14 (2.29)	1.66** (0.71)	1.68** (0.70)
Year 2019	-9.31*** (2.54)	(omitted)	(omitted)	-4.81 (2.92)	(omitted)	(omitted)
Constant	96.39 (76.66)	-3.14 (2.06)	-2.32 (2.15)	82.24 (95.40)	-2.02 (1.47)	-1.39 (1.50)
First stage results:						
ln distance from target index			.509***			.509***
			(.045)			(.045)
F-stat excluded instrument			130.68			130.68
Obs	1098	914	914	1098	914	914
R^2	.180	.037	.032	.177	.018	.012
	Flexible Sigmoidoscopy (Endoscopy)			**Gastroscopy (Endoscopy)**		
	POLS	**FD**	**FDIV**	**POLS**	**FD**	**FDIV**
ln hospital spending per capita (£)	3.14 (6.63)	30.10* (18.02)	75.96** (32.92)	2.30 (6.42)	24.77 (17.19)	81.73** (31.77)
ln market forces factor index	-22.85 (20.95)	356.63 (432.95)	306.46 (419.03)	-15.87 (21.33)	382.90 (370.48)	320.58 (364.69)
ln % population aged						
0-9	-3.39 (8.97)	-29.48 (27.02)	-31.75 (26.68)	3.18 (9.33)	-31.32 (28.72)	-34.15 (28.79)
10-19	3.18 (10.31)	-3.48 (34.59)	-1.56 (34.00)	-1.99 (10.06)	-31.40 (31.20)	-29.00 (30.95)
20-29	-2.04 (4.67)	11.20 (30.23)	14.55 (29.78)	-2.09 (5.09)	-9.20 (30.75)	-5.04 (30.64)
30-39	-2.00 (8.73)	-29.55 (29.39)	-30.44 (29.43)	-.72 (8.39)	-42.19 (29.92)	-43.30 (30.23)
40-49	1.94 (10.21)	-11.68 (22.80)	-10.36 (22.89)	5.68 (9.90)	-3.17 (23.08)	-1.53 (23.26)
50-59	-1.02 (10.73)	7.45 (36.90)	13.54 (36.13)	-3.37 (9.89)	-33.30 (31.36)	-25.73 (30.94)
60-69	-4.37 (9.96)	-38.39 (27.63)	-34.39 (27.65)	-5.27 (9.60)	-48.70* (25.06)	-43.74* (25.27)
70-79	5.54 (6.74)	-39.28 (16.71)	-40.72** (16.63)	7.34 (6.25)	-42.40** (19.13)	-44.19** (19.15)
ln Index of Deprivation						
Barriers to housing/services	-4.05* (2.06)	-4.60 (5.77)	-7.06 (5.95)	-2.77 (2.00)	2.69 (5.45)	-0.38 (5.62)
Crime	1.52 (3.08)	12.84** (6.41)	12.39* (6.31)	0.17 (2.81)	12.43** (5.42)	11.86** (5.32)
Education	-4.19 (2.66)	23.69 (14.37)	22.48 (14.20)	-1.23 (3.01)	29.81** (13.14)	28.30** (12.82)
Employment	-11.65 (10.04)	-56.39 (45.37)	-51.14 (45.47)	-9.24 (8.59)	-24.48 (37.77)	-17.96 (37.53)
Income	6.73 (8.32)	39.19 (29.54)	37.60 (29.87)	6.43 (7.47)	18.45 (26.43)	16.47 (26.77)
Living environment	1.69 (1.38)	7.20 (5.45)	4.50 (5.72)	.33 (1.19)	2.35 (4.30)	-1.01 (4.39)
Time-fixed variables & region dummies	Yes	No	No	Yes	No	No
Year 2015	-2.18*** (0.65)	1.14 (1.15)	0.92 (1.17)	-2.10*** (0.65)	1.44 (1.00)	1.17 (1.02)
Year 2016	-0.28 (0.94)	3.90** (1.61)	1.36 (2.07)	0.25 (0.89)	4.87*** (1.59)	1.71 (1.75)
Year 2017	-1.73 (1.25)	2.51** (1.03)	2.64** (1.02)	-1.21 (1.17)	2.62*** (0.85)	2.78*** (0.85)
Year 2018	-4.03** (1.83)	1.44 (1.09)	1.47 (1.07)	-2.81* (1.65)	2.21** (0.92)	2.25** (0.90)
Year 2019	-7.50*** (2.40)	(omitted)	(omitted)	-6.42*** (2.12)	(omitted)	(omitted)
Constant	99.56 (67.21)	-3.56 (2.35)	-2.85 (2.42)	103.34 (67.69)	-2.54 (1.91)	-1.67 (1.92)
First stage results:						
ln distance from target index			.509***			.509***
			(.045)			(.045)
F-stat excluded instrument			130.68			130.68
Obs	1098	914	914	1098	914	914
R^2	.133	.052	.047	.015	.080	.071

Note: All regressions include a constant term. POLS regressions include time-constant covariates and 8 region dummies not depicted in this table. All regressions include year dummies. The average number of people registered with GP over period 2014-2019 is used as weight. Robust standard errors clustered at CCG level are in the parentheses. Significant at: *10%, **5%, ***1% level.

Table A3 continued. Complete POLS, FD & FDIV regression results for six-week-wait target performance of 13 key diagnostics.

	Dual Energy X-ray Absorptiometry - DEXA (Imaging)			Computed Tomography – CT (Imaging)		
	POLS	**FD**	**FDIV**	**POLS**	**FD**	**FDIV**
ln hospital spending per capita (£)	-3.34 (2.76)	-7.63 (6.39)	-2.09 (17.37)	4.22** (1.76)	-1.06 (3.06)	4.11 (5.70)
ln market forces factor index	-3.86 (6.10)	99.25 (81.04)	93.19 (74.79)	7.07 (6.48)	-15.43 (24.01)	-21.08 (23.76)
ln % population aged						
0-9	-0.44 (2.64)	5.952 (8.02)	5.68 (7.87)	-0.19 (2.70)	0.88 (5.66)	0.62 (5.54)
10-19	-1.11 (3.34)	-29.73 (24.93)	-29.50 (24.52)	-0.11 (2.85)	-4.05 (6.08)	-3.84 (5.91)
20-29	0.80 (1.60)	6.91 (11.95)	7.32 (12.06)	-1.68 (1.30)	-15.87 (12.69)	-15.49 (12.47)
30-39	1.22 (3.39)	5.94 (10.48)	5.83 (10.26)	0.91 (2.59)	-22.57 (17.54)	-22.67 (17.32)
40-49	-0.53 (2.38)	18.04 (16.13)	18.20 (15.97)	4.13* (2.19)	-8.92 (12.17)	-8.77 (11.98)
50-59	-3.69 (3.72)	5.33 (10.22)	6.06 (10.54)	-5.70** (2.81)	-9.56 (13.93)	-8.87 (13.59)
60-69	2.30 (2.24)	-3.32 (9.20)	-2.84 (9.31)	-1.72 (2.80)	-11.19* (6.68)	-10.74* (6.52)
70-79	1.91 (1.68)	9.66 (8.26)	9.49 (8.07)	3.99** (1.73)	-5.96 (8.26)	-6.12 (8.16)
ln Index of Deprivation						
Barriers to housing/services	0.59 (0.95)	5.72 (5.09)	5.42 (5.17)	-0.57 (0.55)	1.76 (1.59)	1.49 (1.54)
Crime	-0.95 (1.03)	-0.24 (2.75)	-0.30 (2.71)	0.69 (0.83)	1.79 (2.02)	1.73 (1.98)
Education	-3.28** (1.31)	5.44 (9.17)	5.29 (9.12)	0.73 (0.86)	0.24 (4.41)	0.10 (4.39)
Employment	4.55 (4.30)	-34.39* (20.58)	-33.75* (20.33)	0.86 (2.34)	1.44 (11.18)	2.03 (11.11)
Income	-2.12 (3.90)	44.65*** (15.90)	44.45*** (15.73)	-3.62* (2.18)	-11.52 (8.34)	-11.70 (8.15)
Living environment	0.31 (0.36)	-2.27 (2.98)	-2.60 (3.07)	-0.53* (0.30)	-1.33 (1.03)	-1.63 (1.05)
Time-fixed variables & region dummies	Yes	No	No	Yes	No	No
Year 2015	0.68 (0.48)	0.66 (0.62)	0.63 (0.62)	0.24 (0.19)	0.71*** (0.22)	0.69*** (0.22)
Year 2016	1.13 (0.71)	0.90 (0.67)	0.60 (1.17)	-0.07 (0.24)	0.46 (0.33)	0.18 (0.46)
Year 2017	0.65 (0.93)	-0.61 (0.71)	-0.60 (0.70)	-0.76** (0.34)	-0.04 (0.21)	-0.03 (0.20)
Year 2018	0.88 (1.00)	0.40 (0.50)	0.40 (0.49)	-1.07** (0.50)	0.41 (0.27)	0.42 (0.27)
Year 2019	0.72 (1.19)	(omitted)	(omitted)	-1.75*** (0.62)	(omitted)	(omitted)
Constant	140.68*** (20.88)	-0.30 (0.89)	-0.21 (0.90)	68.24*** (16.42)	-0.74 (0.68)	-0.66 (0.69)
First stage results:						
ln distance from target index			.509***			.509***
			(.045)			(.045)
F-stat excluded instrument			130.68			130.68
Obs	1098	914	914	1098	914	914
R^2	.074	.020	.020	.288	.020	.047
	Non-Obstetric Ultrasound - NOUS (Imaging)			**Echocardiography – ECG (Physiology)**		
	POLS	**FD**	**FDIV**	**POLS**	**FD**	**FDIV**
ln hospital spending per capita (£)	2.30 (1.58)	-2.78 (3.47)	-9.01 (9.07)	8.35** (3.94)	-14.89 (10.16)	-6.88 (16.45)
ln market forces factor index	-3.44 (5.51)	-58.09* (33.79)	-51.28 (33.73)	-20.36 (15.22)	213.99* (128.09)	205.22 (130.29)
ln % population aged						
0-9	1.00 (2.23)	5.05 (5.90)	5.36 (5.86)	-17.81*** (5.96)	11.36 (16.25)	10.96 (15.97)
10-19	-5.67** (2.61)	-7.58 (7.34)	-7.84 (7.30)	8.71 (5.65)	-37.96 (24.84)	-37.62 (24.43)
20-29	-0.78 (1.19)	-5.94 (7.08)	-6.39 (7.10)	-15.01*** (3.20)	-43.87** (21.24)	-43.29** (20.97)
30-39	-1.09 (2.12)	-5.22 (6.86)	-5.10 (6.84)	24.05*** (6.46)	4.05 (22.63)	3.89 (22.31)
40-49	1.11 (2.04)	-6.87 (5.88)	-7.05 (5.80)	-12.66** (5.44)	-19.85 (28.64)	-19.61 (28.20)
50-59	-1.65 (3.02)	2.30 (7.93)	1.47 (7.97)	-2.86 (4.70)	-23.16 (18.86)	-22.10 (18.12)
60-69	-0.79 (2.11)	-2.11 (5.56)	-2.65 (5.69)	1.52 (3.90)	-18.05 (15.75)	-17.36 (15.63)
70-79	0.27 (2.00)	-6.38 (4.91)	-6.19 (4.86)	-4.00 (3.90)	-26.13 (16.20)	-26.38* (15.93)
ln Index of Deprivation						
Barriers to housing/services	-0.61 (0.57)	2.72 (2.17)	3.05 (2.10)	-0.99* (1.17)	-2.30 (6.96)	-2.73 (6.84)
Crime	0.86 (0.89)	1.06 (1.86)	1.13 (1.85)	5.18 (2.72)	12.26* (6.86)	12.18* (6.72)
Education	-1.00 (0.76)	1.38 (5.28)	1.55 (5.11)	-1.15 (2.15)	-9.43 (10.18)	-9.64 (9.97)
Employment	-0.59 (2.39)	-3.57 (8.37)	-4.28 (8.25)	-3.70 (6.10)	26.27 (24.19)	27.19 (23.44)
Income	1.11 (2.15)	6.00 (8.89)	6.21 (8.69)	-5.84 (5.63)	-32.82 (26.21)	-33.10 (25.59)
Living environment	-0.59* (0.30)	0.14 (1.35)	0.50 (1.47)	-0.90 (0.89)	-0.38 (2.90)	-0.85 (3.11)
Time-fixed variables & region dummies	Yes	No	No	Yes	No	No
Year 2015	0.52*** (0.20)	1.15*** (0.27)	1.18*** (0.27)	-0.20 (0.45)	2.01** (0.92)	1.97** (0.91)
Year 2016	0.50* (0.28)	0.80** (0.39)	1.14* (0.66)	-1.59*** (0.58)	2.19** (1.05)	1.75 (1.22)
Year 2017	0.25 (0.38)	0.56*** (0.21)	0.54*** (0.20)	-3.40*** (0.84)	0.89 (1.04)	0.91 (1.03)
Year 2018	0.09 (0.53)	0.69** (0.27)	0.69*** (0.26)	-5.21*** (1.13)	1.32 (1.01)	1.33 (0.99)
Year 2019	-0.75 (0.67)	(omitted)	(omitted)	-8.34*** (1.71)	(omitted)	(omitted)
Constant	68.51*** (16.06)	-0.78* (0.44)	-0.87* (0.47)	-22.73 (41.62)	-1.82 (1.65)	-1.70 (1.64)
First stage results:						
ln distance from target index			.509***			.509***
			(.045)			(.045)
F-stat excluded instrument			130.68			130.68
Obs	1098	914	914	1098	914	914
R^2	.160	.073	.071	.233	.035	.035

Note: All regressions include a constant term. POLS regressions include time-constant covariates and 8 region dummies not depicted in this table. All regressions include year dummies. The average number of people registered with GP over period 2014-2019 is used as weight. Robust standard errors clustered at CCG level are in the parentheses. Significant at: *10%, **5%, ***1% level.

Table A3 continued. Complete POLS, FD & FDIV regression results for six-week-wait target performance of 13 key diagnostics.

		Respiratory Physiology - RP (Physiology)		
		POLS	**FD**	**FDIV**
ln hospital spending per capita (£)		4.89 (7.14)	-21.66 (15.09)	-35.91 (31.70)
ln market forces factor index		-40.19 (25.09)	100.96 (195.01)	116.55 (197.90)
ln % population aged				
	0-9	-17.20 (11.37)	-32.42 (33.91)	-31.71 (33.51)
	10-19	2.88 (11.16)	44.08 (29.56)	43.49 (29.28)
	20-29	3.95 (5.13)	6.03 (31.38)	4.99 (30.81)
	30-39	-1.83 (9.69)	-10.63 (36.80)	-10.35 (36.53)
	40-49	24.19** (10.71)	14.19 (34.81)	13.78 (34.37)
	50-59	-2.00 (10.47)	0.29 (40.38)	-1.60 (39.41)
	60-69	-6.11 (8.40)	-15.54 (21.90)	-16.79 (21.42)
	70-79	-1.74 (6.43)	-1.43 (20.39)	-0.98 (20.24)
ln Index of Deprivation				
	Barriers to housing/services	-0.08 (2.27)	1.25 (4.79)	2.02 (5.19)
	Crime	9.82** (3.86)	-0.72 (6.98)	-0.57 (6.87)
	Education	-6.96 (4.24)	14.15 (12.72)	14.53 (12.63)
	Employment	-14.09 (11.76)	-24.65 (38.37)	-26.28 (37.63)
	Income	9.64 (8.50)	-12.31 (37.50)	-11.82 (36.90)
	Living environment	0.58 (1.32)	2.76 (4.51)	3.60 (4.70)
Time-fixed variables & region dummies		Yes	No	No
Year 2015		-0.18 (0.63)	2.24 (1.40)	2.31* (1.37)
Year 2016		0.34 (0.94)	4.23*** (1.48)	5.03*** (1.93)
Year 2017		-0.14 (1.51)	1.41 (1.02)	1.37 (1.01)
Year 2018		-1.29 (2.17)	0.62 (0.93)	0.61 (0.91)
Year 2019		-2.63 (2.88)	(omitted)	(omitted)
Constant		73.33 (78.33)	-3.55** (1.77)	-3.77** (1.74)
First stage results:				
ln distance from target index				.509***
				(.045)
F-stat excluded instrument				130.68
Obs		1098	914	914
R^2		.286	.031	.030

Note: All regressions include a constant term. POLS regressions include time-constant covariates and 8 region dummies not depicted in this table. All regressions include year dummies. The average number of people registered with GP over period 2014-2019 is used as weight. Robust standard errors clustered at CCG level are in the parentheses. Significant at: *10%, **5%, ***1% level.

Table A4. FDIV regression results of log health spending per capita on waiting time target performance across four geographical regions for outcome measures with a non-significant effect of log health spending in the original model.

	All regions	**North**	**Midlands & East**	**London**	**South**
	N=914	N=295	N=274	N=160	N=182
	Units=187	Units=59	Units=58	Units=32	Units=38
Two-week-wait (TWW)	7.58 (9.25)	-7.27 (31.16)	55.46** (27.50)	-3.14 (11.88)	-18.08 (16.65)
Two-week-wait breast symptoms (TWW)	4.65 (24.34)	-24.12 (53.60)	117.90 (82.10)	-13.47 (18.05)	-25.54 (30.22)
Computed Tomography (CT)	4.11 (5.70)	27.86 (20.54)	-5.21 (9.42)	-3.82 (3.67)	0.90 (8.00)
Non-Obstetric Ultrasound (NOUS)	-9.01 (9.07)	-56.77 (52.89)	6.95 (6.37)	5.70 (6.70)	3.71 (6.74)
Dual Energy X-ray Absorptiometry (DEXA)	-2.09 (17.37)	-61.48 (103.82)	46.89 (30.28)	5.01 (5.88)	15.66 (22.23)
Echocardiography (ECG)	-6.88 (16.45)	-14.23 (69.34)	34.21 (24.04)	-20.81 (18.03)	-31.09 (44.95)
Respiratory Physiology (RP)	-35.91 (31.70)	-25.68 (76.72)	28.82 (79.47)	-68.96 (66.54)	-41.86 (48.75)

Note: FDIV regressions include a constant term and year dummies (for 2015-2018). The average number of people registered with GP over period 2014-2019 is used as weight. Robust standard errors clustered at CCG level are in the parentheses. Significant at: *10%, **5%, ***1% level.

Table A5. FDIV regression results comparing the original model to the model excluding outliers in waiting time target performance scores (scores <70% and of exactly 100% are excluded) for outcome measures with a non-significant effect of log health spending in the original model.

	Original model (N=914)	***p***	**Model excl. outliers**	***p***	**N**
Two-week-wait (TWW)	7.58 (9.25)	.412	7.58 (9.25)	.412	914
Two-week-wait breast symptoms (TWW)	4.65 (24.34)	.849	-25.68* (13.64)	.060	811
Computed Tomography (CT)	4.11 (5.70)	.471	3.84 (5.86)	.512	876
Non-Obstetric Ultrasound (NOUS)	-9.01 (9.07)	.320	-9.39 (9.37)	.317	886
Dual Energy X-ray Absorptiometry (DEXA)	-2.09 (17.37)	.904	-9.06 (20.91)	.665	467
Echocardiography (ECG)	-6.88 (16.45)	.676	-8.60 (17.36)	.620	848
Respiratory Physiology (RP)	-35.91 (31.70)	.257	-34.34 (35.40)	.332	773

Note: **Note:** FDIV regressions include a constant term and year dummies (for 2015-2018). The average number of people registered with GP over period 2014-2019 is used as weight. Robust standard errors clustered at CCG level are in the parentheses. Significant at: *10%, **5%, ***1% level.

Table A6. FDIV regression results comparing the original model to models using different calculation methods of waiting time performance scores for outcome measures with a non-significant effect of log health spending in the original model.

	Original model	***p***	**Arithmetic mean**	***p***	**Year mean**	***p***
Two-week-wait (TWW)	7.58 (9.25)	.412	6.98 (8.91)	.433	6.11 (8.94)	.495
Two-week-wait breast symptoms (TWW)	4.65 (24.34)	.849	-4.59 (18.48)	.804	-5.92 (19.35)	.760
Computed Tomography (CT)	4.11 (5.70)	.471	4.14 (5.62)	.461	4.39 (5.78)	.448
Non-Obstetric Ultrasound (NOUS)	-9.01 (9.07)	.320	-8.77 (8.81)	.320	-9.03 (9.80)	.357
Dual Energy X-ray Absorptiometry (DEXA)	-2.09 (17.37)	.904	-2.90 (15.26)	.849	-6.62 (21.91)	.763
Echocardiography (ECG)	-6.88 (16.45)	.676	-7.47 (15.69)	.634	-5.17 (18.13)	.776
Respiratory Physiology (RP)	-35.91 (31.70)	.257	-33.49 (26.82)	.212	-36.70 (34.85)	.292

Note: **Note:** FDIV regressions include a constant term and year dummies (for 2015-2018). The average number of people registered with GP over period 2014-2019 is used as weight. Robust standard errors clustered at CCG level are in the parentheses. N=914. Significant at: *10%, **5%, ***1% level.

Table A7. FDIV regression results comparing the original model including separate Indices of Deprivation to models using an alternative aggregated measure of deprivation; the Index of Multiple Deprivation (IMD) for outcome measures with a non-significant effect of log health spending in the original model.

	Original model	***p***	**Model with IMD**	***p***
Two-week-wait (TWW)	7.58 (9.25)	.412	7.92 (8.81)	.368
Two-week-wait breast symptoms (TWW)	4.65 (24.34)	.849	21.13 (22.62)	.350
Computed Tomography (CT)	4.11 (5.70)	.471	4.56 (5.36_	.395
Non-Obstetric Ultrasound (NOUS)	-9.01 (9.07)	.320	-8.49 (8.53)	.320
Dual Energy X-ray Absorptiometry (DEXA)	-2.09 (17.37)	.904	-3.12 (16.86)	.853
Echocardiography (ECG)	-6.88 (16.45)	.676	-8.79 (15.73)	.576
Respiratory Physiology (RP)	-35.91 (31.70)	.257	-30.27 (30.79)	.326

Note: **Note:** FDIV regressions include a constant term and year dummies (for 2015-2018). The average number of people registered with GP over period 2014-2019 is used as weight. Robust standard errors clustered at CCG level are in the parentheses. N=914. Significant at: *10%, **5%, ***1% level.

Table A8. FDIV regression results comparing the original model estimated for the unbalanced panel to the same model estimated for a balanced panel for outcome measures with a non-significant effect of log health spending in the original model.

	Unbalanced panel		**Balanced panel**	
	N=914		**N=905**	
	Units=188	***p***	**Units=181**	***p***
Two-week-wait (TWW)	7.58 (9.25)	.412	7.83 (9.20)	.395
Two-week-wait breast symptoms (TWW)	4.65 (24.34)	.849	7.56 (24.10)	.754
Computed Tomography (CT)	4.11 (5.70)	.471	3.29 (5.59)	.555
Non-Obstetric Ultrasound (NOUS)	-9.01 (9.07)	.320	-9.25 (9.11)	.310
Dual Energy X-ray Absorptiometry (DEXA)	-2.09 (17.37)	.904	-1.83 (17.35)	.916
Echocardiography (ECG)	-6.88 (16.45)	.676	-7.99 (16.31)	.624
Respiratory Physiology (RP)	-35.91 (31.70)	.257	-35.29 (31.74)	.266

Note: FDIV regressions include a constant term and year dummies (for 2015-2018). The average number of people registered with GP over period 2014-2019 is used as weight. Robust standard errors clustered at CCG level are in the parentheses. Significant at: *10%, **5%, ***1% level.

Table A9. Average annual patient volume growth rates of the six-week-wait target key diagnostics over the period of analysis (FY2014 to FY2019).

Key diagnostic	**Average annual growth rate (%)**
Computed Tomography (CT)	109.02
Dual Energy X-ray Absorptiometry (DEXA)	109.31
Magnetic Resonance Imaging (MRI)	107.24
Non-Obstetric Ultrasound (NOUS)	107.78
Audiology Assessments (AA)	102.35
Echocardiography (ECG)	111.73
Peripheral Neurophysiology (PN)	107.63
Respiratory Physiology (RP)	112.94
Urodynamics (URO)	105.57
Colonoscopy (COL)	109.03
Cystoscopy (CYS)	104.23
Flexible Sigmoidoscopy (FS)	107.18
Gastroscopy (GAS)	106.60
Imaging aggregate	107.91
Physiology aggregate	108.02
Endoscopy aggregate	107.12
Total aggregate	107.80

Note: Annual growth rates are obtained by calculating the annual growth in absolute patient volumes for FY2014 to FY2019 and taking the average.

Table A10. Average patient prices for six-week-wait key diagnostics in the private health care sector.

Key diagnostic	**Average price of private treatment (£)**
Computed Tomography (CT)	669
Dual Energy X-ray Absorptiometry (DEXA)	103
Magnetic Resonance Imaging (MRI)	1075
Non-Obstetric Ultrasound (NOUS)	358
Audiology Assessments (AA)	135
Echocardiography (ECG)	355
Peripheral Neurophysiology (PN)	325
Respiratory Physiology (RP)	750
Urodynamics (URO)	405
Colonoscopy (COL)	2085
Cystoscopy (CYS)	1760
Flexible Sigmoidoscopy (FS)	1505
Gastroscopy (GAS)	1958

Source: Private Health care UK, 2020.

Table A11. Overview of the procedures included within the six-week-wait key diagnostics. Source: DM01-Guidance (NHS England, 2015).

Magnetic Resonance Imaging (MRI)
Magnetic resonance imaging of whole body
Magnetic resonance imaging of head
Functional magnetic resonance imaging of head
Magnetic resonance imaging of spine
Magnetic resonance imaging of chest
Magnetic resonance imaging of abdomen
Magnetic resonance imaging of pelvis
Cardiac magnetic resonance imaging
Magnetic resonance imaging of bone
Magnetic resonance cholangiopancreatography
Magnetic resonance imaging NEC
Magnetic resonance imaging of kidneys
Magnetic resonance angiography (vascular system)
Computed Tomography (CT)
Computerised tomography of whole body
Computerised tomography of head
Computerised tomography of spine
Computerised tomography of sinuses
Computerised tomography of chest
Computerised tomography of abdomen NEC
Computerised tomography of pelvis
Computerised tomography of calcium scoring
Computerised tomography angiography (vascular system)
Computerised tomography of cerebral vessels
Computerised tomography of bone
Computerised tomography of colon
Computerised tomography NEC
Computerised tomography of pulmonary arteries
Computerised tomography of kidneys
Positron emission tomography NEC
Non-Obstetric Ultrasound (NOUS)
Ultrasound of thyroid gland
Ultrasound of abdomen
Ultrasound of pelvis
Ultrasound of scrotum
Ultrasound of kidneys
Ultrasound of bladder
Ultrasound of bone
Ultrasound of NEC
Ultrasound evaluation of retina
Intravascular ultrasound of coronary artery
Transvaginal ultrasound examination of female genital tract
Diagnostic fibreoptic endoscopic ultrasound examination of oesophagus
Fibreoptic endoscopic ultrasound examination of upper gastrointestinal tract
Laparoscopic ultrasound examination of liver and biopsy of lesion of liver
Laparoscopic ultrasound examination of liver NEC (not elsewhere classified)
Endoscopic ultrasound examination of liver and biopsy of lesion of liver
Other specified endoscopic ultrasound examination of liver
Unspecified endoscopic ultrasound examination of liver
Laparoscopic ultrasound examination of bile duct and biopsy of lesion of bile duct
Other specified laparoscopic ultrasound examination of bile duct
Unspecified laparoscopic ultrasound examination of bile duct
Endoscopic ultrasound examination of bile duct and biopsy of lesion of bile duct
Other specified endoscopic ultrasound examination of bile duct
Unspecified endoscopic ultrasound examination of bile duct
Laparoscopic ultrasound examination of pancreas and biopsy of lesion of pancreas
Other specified laparoscopic ultrasound examination of pancreas
Unspecified laparoscopic ultrasound examination of pancreas
Endoscopic ultrasound examination of pancreas and biopsy of lesion of pancreas
Other specified endoscopic ultrasound examination of pancreas
Diagnostic endoscopic ultrasound examination of peritoneum
Diagnostic endoscopic ultrasound examination of peritoneum and biopsy of intraabdominal organ
Unspecified endoscopic ultrasound examination of pancreas
Barium Enema (BE)
Barium Swallow
Barium Enema
Dual Energy X-ray Absorptiometry (DEXA)
Bone densitometry
Audiology Assessments
Pure tone audiometry
Balance assessment
Hearing assessment
Other specified diagnostic audiology
Unspecified diagnostic audiology
Echocardiography (ECG)
Transthoracic echocardiography (TTE)
Transoesophageal echocardiography (TOE)
Intravascular echocardiography
Epicardial echocardiography
Stress echocardiography
Fetal echocardiography
Other specified diagnostic echocardiography
Unspecified diagnostic echocardiography
Transluminal intracardiac echocardiography
Electrophysiology (EP)
Percutaneous transluminal electrophysiological studies on conducting system of heart
Peripheral Neurophysiology (PN)
Electromyograpy
Nerve conduction studies
Respiratory Studies (RP)
Sleep studies NEC
Polysomnography
Urodynamics (URO)
Urodynamics NEC
Urodynamic studies using catheter
Colonoscopy (COL)
Fibreoptic endoscopic snare resection of lesion of colon
Fibreoptic endoscopic cauterisation of lesion of colon
Fibreoptic endoscopic laser destruction of lesion of colon
Fibreoptic endoscopic laser destruction of lesion of colon
Fibreoptic endoscopic destruction of lesion of colon NEC
Fibreoptic endoscopic submucosal resection of lesion of colon
Fibreoptic endoscopic resection of lesion of colon NEC
Other specified endoscopic extirpation of lesion of colon
Unspecified endoscopic extirpation of lesion of colon
Diagnostic fibreoptic endoscopic examination of colon and biopsy of lesion of colon
Other specified endoscopic examination of colon
Unspecified endoscopic examination of colon
Flexible Sigmoidoscopy (FS)
Endoscopic snare resection of lesion of lower bowel using fibreoptic sigmoidoscope
Endoscopic cauterisation of lesion of lower bowel using fibreoptic sigmoidoscope
Endoscopic laser destruction of lesion of lower bowel using fibreoptic sigmoidoscope
Endoscopic destruction of lesion of lower bowel using fibreoptic sigmoidoscope NEC
Endoscopic submucosal resection of lesion of lower bowel using fibreoptic sigmoidoscope
Endoscopic resection of lesion of lower bowel using fibreoptic sigmoidoscope NEC
Other specified endoscopic extirpation of lesion of lower bowel using fibreoptic sigmoidoscope
Unspecified endoscopic extirpation of lesion of lower bowel using fibreoptic sigmoidoscope
Diagnostic endoscopic examination of lower bowel and biopsy of lesion of lower bowel using fibreoptic sigmoidoscope
Diagnostic endoscopic examination of lower bowel and sampling for bacterial overgrowth using fibreoptic sigmoidoscope
Other specified diagnostic endoscopic examination of lower bowel using fibreoptic sigmoidoscope
Unspecified endoscopic examination of lower bowel using fibreoptic sigmoidoscope

Table A11 continued. Overview of the procedures included within the six-week-wait key diagnostics. Source: DM01-Guidance (NHS England, 2015).

Cystoscopy (CYS)
Endoscopic retrograde pyelography
Endoscopic catheterisation of ureter
Endoscopic ureteric urine sampling
Nephroscopic ureteroscopy
Diagnostic endoscopic examination of ureter and biopsy of lesion of ureter NEC
Diagnostic endoscopic examination of ureter and biopsy of lesion of ureter using rigid ureteroscope
Other specified endoscopic examination of ureter
Unspecified endoscopic examination of ureter
Diagnostic endoscopic examination of bladder and biopsy of lesion of bladder NEC (not elsewhere classified)
Diagnostic endoscopic examination of bladder and biopsy of lesion of prostate NEC (not elsewhere classified)
Diagnostic endoscopic examination of bladder and biopsy of lesion of bladder using rigid cystoscope
Diagnostic endoscopic examination of bladder and biopsy of lesion of prostate using rigid cystoscope
Diagnostic endoscopic examination of bladder using rigid cystoscope
Other specified diagnostic endoscopic examination of bladder
Unspecified diagnostic endoscopic examination of bladder
Diagnostic endoscopic examination of urethra and biopsy of lesion of urethra
Other specified diagnostic endoscopic examination of urethra
Unspecified diagnostic endoscopic examination of urethra
Endoscopic resection of lesion of bladder
Endoscopic cauterisation of lesion of bladder
Endoscopic destruction of lesion of bladder NEC
Other specified endoscopic extirpation of lesion of bladder
Unspecified endoscopic extirpation of lesion of bladder
Endoscopic transection of bladder
Endoscopic hydrostatic distension of bladder
Endoscopic overdistension of bladder NEC
Endoscopic injection of neurolytic substance into nerve of bladder
Other specified endoscopic operations to increase capacity of bladder
Unspecified endoscopic operations to increase capacity of bladder
Endoscopic resection of prostate using electrotome
Endoscopic resection of prostate using punch
Endoscopic resection of prostate NEC
Endoscopic resection of prostate using laser
Other specified endoscopic resection of outlet of male bladder
Unspecified endoscopic resection of outlet of male bladder
Gastroscopy (GAS)
Fibreoptic endoscopic examination of upper gastrointestinal tract and biopsy of lesion of upper gastrointestinal tract
Fibreoptic endoscopic ultrasound examination of upper gastrointestinal tract
Fibreoptic endoscopic insertion of Bravo pH capsule into upper gastrointestinal tract
Fibreoptic endoscopic examination of upper gastrointestinal tract and staining of gastric mucosa
Other specified fibreoptic endoscopic examination of upper gastrointestinal tract
Unspecified fibreoptic endoscopic examination of upper gastrointestinal tract
Diagnostic endoscopic examination of duodenum and biopsy of lesion of duodenum
Other specified diagnostic endoscopic examination of duodenum
Unspecified diagnostic endoscopic examination of duodenum
Diagnostic endoscopic examination of jejunum and biopsy of lesion of jejunum
Other specified diagnostic endoscopic examination of jejunum
Unspecified diagnostic endoscopic examination of jejunum
Diagnostic endoscopic examination of ileum and biopsy of lesion of ileum
Diagnostic endoscopic balloon examination of ileum
Other specified diagnostic endoscopic examination of ileum
Unspecified diagnostic endoscopic examination of ileum
Fibreoptic endoscopic snare resection of lesion of oesophagus
Fibreoptic endoscopic laser destruction of lesion of oesophagus
Fibreoptic endoscopic cauterisation of lesion of oesophagus
Fibreoptic endoscopic injection sclerotherapy to varices of oesophagus
Fibreoptic endoscopic destruction of lesion of oesophagus NEC
Fibreoptic endoscopic submucosal resection of lesion of oesophagus
Fibreoptic endoscopic photodynamic therapy of lesion of oesophagus
Other specified fibreoptic endoscopic extirpation of lesion of oesophagus
Unspecified fibreoptic endoscopic extirpation of lesion of oesophagus
Diagnostic fibreoptic endoscopic examination of oesophagus and biopsy of lesion of oesophagus
Diagnostic fibreoptic endoscopic ultrasound examination of oesophagus
Diagnostic fibreoptic insertion of Bravo pH capsule into oesophagus
Other specified diagnostic fibreoptic endoscopic examination of oesophagus
Unspecified diagnostic fibreoptic endoscopic examination of oesophagus
Endoscopic snare resection of lesion of oesophagus using rigid oesophagoscope
Endoscopic laser destruction of lesion of oesophagus using rigid oesophagoscope
Endoscopic cauterisation of lesion of oesophagus using rigid oesophagoscope
Endoscopic injection sclerotherapy to varices of oesophagus using rigid oesophagoscope
Other specified endoscopic extirpation of lesion of oesophagus using rigid oesophagoscope
Unspecified endoscopic extirpation of lesion of oesophagus using rigid oesophagoscope
Diagnostic endoscopic examination of oesophagus and biopsy of lesion of oesophagus using rigid oesophagoscope
Diagnostic endoscopic insertion of Bravo pH capsule using rigid oesophagoscope
Other specified diagnostic endoscopic examination of oesophagus using rigid oesophagoscope
Unspecified diagnostic endoscopic examination of oesophagus using rigid oesophagoscope
Fibreoptic endoscopic snare resection of lesion of upper gastrointestinal tract
Fibreoptic endoscopic laser destruction of lesion of upper gastrointestinal tract
Fibreoptic endoscopic cauterisation of lesion of upper gastrointestinal tract
Fibreoptic endoscopic sclerotherapy to lesion of upper gastrointestinal tract
Fibreoptic endoscopic destruction of lesion of upper gastrointestinal tract NEC
Fibreoptic endoscopic injection therapy to lesion of upper gastrointestinal tract NEC
Fibreoptic endoscopic rubber band ligation of upper gastrointestinal tract varices
Other specified fibreoptic endoscopic extirpation of lesion of upper gastrointestinal tract
Unspecified fibreoptic endoscopic extirpation of lesion of upper gastrointestinal tract

Table A12. Cumby-Huizinga test for autocorrelation.

		Lags 1-1		Lag 1	
	Obs	**Chi^2**	***p***	**Chi^2**	***p***
Two Week Wait targets					
Two-Week-Wait (TWW)	1098	280.580	.0000	280.580	.0000
Two-Week-Wait Breast Symptoms (TWWBS)	1098	130.832	.0000	130.832	.0000
Six Week Wait targets					
Imaging					
Computerised Tomography (CT)	1098	195.258	.0000	195.258	.0000
Dual Energy X-ray Absorptiometry (DEXA)	1098	7.033	.0000	7.033	.0000
Magnetic Resonance Imaging (MRI)	1098	236.417	.0000	236.417	.0000
Non-obstetric Ultrasound (NOUS)	1098	55.311	.0000	55.311	.0000
Physiology					
Audiology Assessments (AA)	1098	121.153	.0000	121.153	.0000
Echocardiography (ECG)	1098	72.057	.0000	72.057	.0000
Peripheral Neurophysiology (PN)	1098	15.031	.0000	15.031	.0000
Respiratory Physiology (RP)	1098	228.025	.0000	228.025	.0000
Urodynamics (URO)	1098	108.477	.000	108.477	.000
Endoscopy					
Colonoscopy (COL)	1098	136.205	.0000	136.205	.0000
Cystoscopy (CYS)	1098	255.821	.0000	255.821	.0000
Flexible Sigmoidoscopy (FS)	1098	98.254	.0000	98.254	.0000
Gastroscopy (GAS)	1098	94.219	.0000	94.219	.0000

Table A13. Hausman test (using a user-written Stata command [xtoverid] which allows for heteroskedastic-robust standard errors.

	Obs	**Wald chi^2(20)**	***p***
Two Week Wait targets			
Two-Week-Wait (TWW)	1102	161.95	.0000
Two-Week-Wait Breast Symptoms (TWWBS)	1102	141.01	.0000
Six Week Wait targets			
Imaging			
Computerised Tomography (CT)	1102	126.00	.0000
Dual Energy X-ray Absorptiometry (DEXA)	1102	44.04	.0015
Magnetic Resonance Imaging (MRI)	1102	185.15	.0000
Non-obstetric Ultrasound (NOUS)	1102	115.77	.0000
Physiology			
Audiology Assessments (AA)	1102	28.91	.0896
Echocardiography (ECG)	1102	94.50	.0000
Peripheral Neurophysiology (PN)	1102	152.91	.0000
Respiratory Physiology (RP)	1102	73.60	.0000
Urodynamics (URO)	1102	201.60	.0000
Endoscopy			
Colonoscopy (COL)	1102	87.78	.0000
Cystoscopy (CYS)	1102	85.31	.0000
Flexible Sigmoidoscopy (FS)	1102	86.48	.0000
Gastroscopy (GAS)	1102	110.57	.0000

Table A14. Inoue and Solo (2006) LM-test as postestimation (serial correlation of the idiosyncratic error).

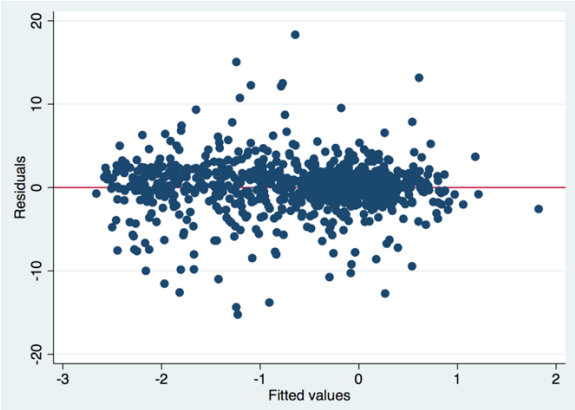
		H0: no autocorrelation of any order Ha: autocorrelation of some order		H0: no autocorrelation of any order Ha: autocorrelation up to order 1	
	Obs	**IS-stat.**	***p***	**IS-stat.**	***p***
Two Week Wait targets					
Two-Week-Wait (TWW)	1102	50.63	.000	48.73	.000
Two-Week-Wait Breast Symptoms (TWWBS)	1102	60.14	.000	51.55	.000
Six Week Wait targets					
Imaging					
Computerised Tomography (CT)	1102	31.68	.000	26.71	.000
Dual Energy X-ray Absorptiometry (DEXA)	1102	22.29	.014	17.78	.003
Magnetic Resonance Imaging (MRI)	1102	27.88	.002	20.21	.001
Non-obstetric Ultrasound (NOUS)	1102	32.07	.000	26.98	.000
Physiology					
Audiology Assessments (AA)	1102	17.65	.061	15.08	.010
Echocardiography (ECG)	1102	30.48	.001	18.79	.002
Peripheral Neurophysiology (PN)	1102	46.42	.000	26.47	.000
Respiratory Physiology (RP)	1102	34.19	.000	29.58	.000
Urodynamics (URO)	1102	45.06	.000	29.97	.000
Endoscopy					
Colonoscopy (COL)	1102	48.45	.000	41.18	.000
Cystoscopy (CYS)	1102	44.52	.000	34.75	.000
Flexible Sigmoidoscopy (FS)	1102	45.43	.000	42.14	.000
Gastroscopy (GAS)	1102	46.50	.000	37.05	.000

Table A15. C-test (also: difference-in-Sargan test) of endogenous regressors. H0: regressor is exogenous.

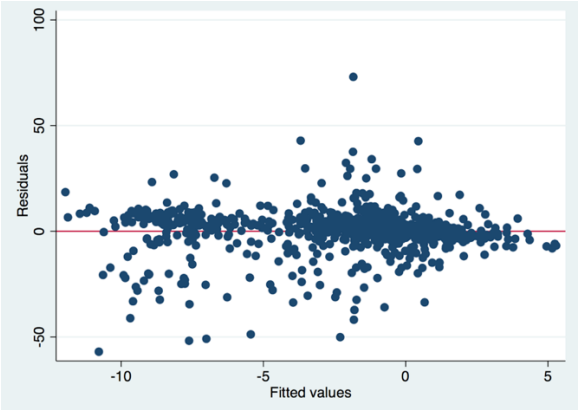
	Obs	**Chi^2(1)**	***p***
Two Week Wait targets			
Two-Week-Wait (TWW)	914	1.895	.1687
Two-Week-Wait Breast Symptoms (TWWBS)	914	0.885	.3468
Six Week Wait targets			
Imaging			
Computerised Tomography (CT)	914	1.427	.2322
Dual Energy X-ray Absorptiometry (DEXA)	914	0.203	.6527
Magnetic Resonance Imaging (MRI)	914	7.142	.0075
Non-obstetric Ultrasound (NOUS)	914	1.012	.3144
Physiology			
Audiology Assessments (AA)	914	7.864	.0050
Echocardiography (ECG)	914	0.336	.5619
Peripheral Neurophysiology (PN)	914	3.886	.0487
Respiratory Physiology (RP)	914	0.295	.5873
Urodynamics (URO)	914	2.078	.1492
Endoscopy			
Colonoscopy (COL)	914	2.817	.0933
Cystoscopy (CYS)	914	4.311	.0379
Flexible Sigmoidoscopy (FS)	914	2.784	.0952
Gastroscopy (GAS)	914	3.517	.0608

Table A16. Residual plots vs fitted values.

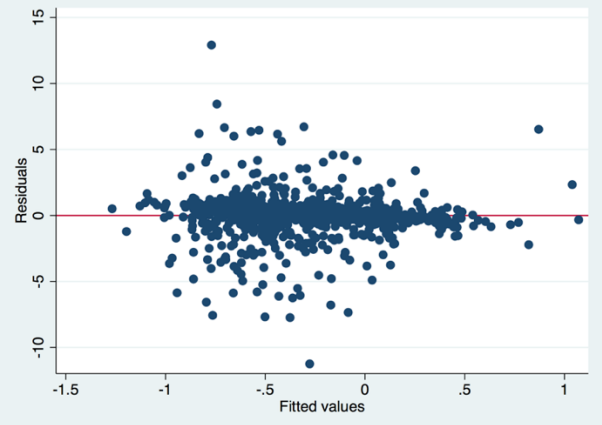
Two-Week-Wait (TWW)



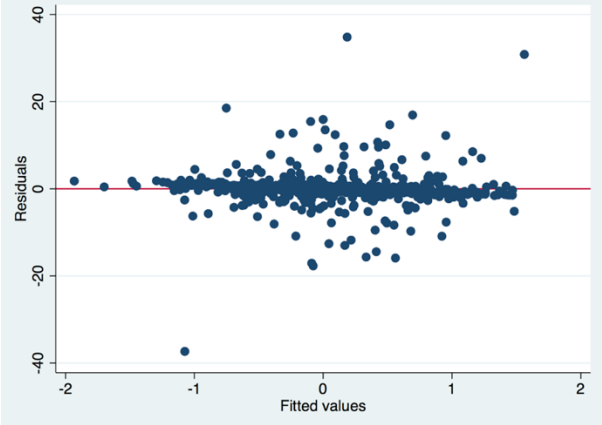
Two-Week-Wait Breast Symptoms (TWWBS)



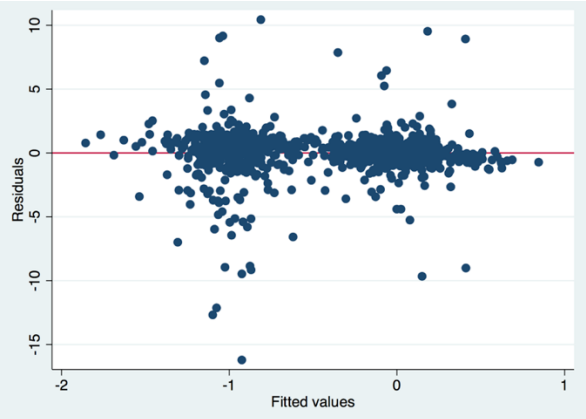
Computed Tomography (CT)



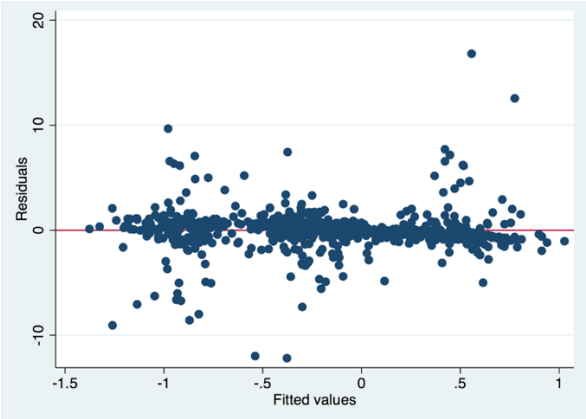
Dual Energy X-ray Absorptiometry (DEXA)



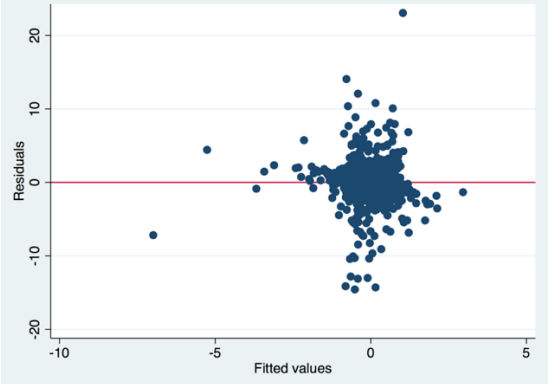
Magnetic Imaging Resonance (MRI)



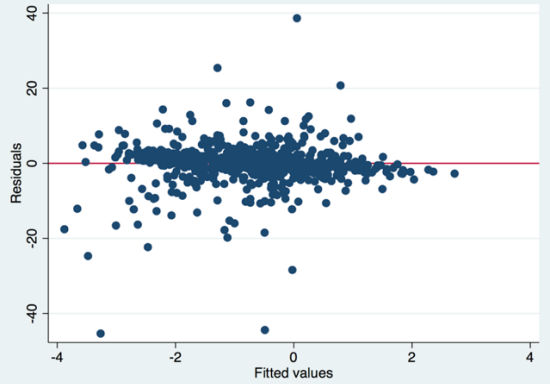
Non-Obstetric Ultrasound (NOUS)



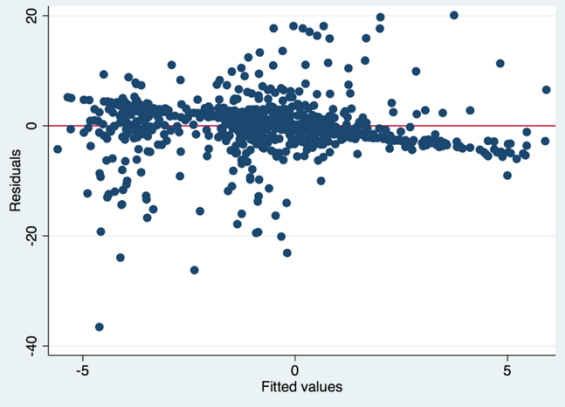
Audiology Assessments (AA)



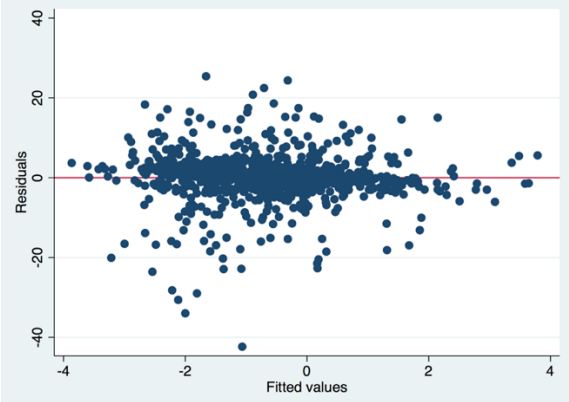
Echocardiography (ECG)



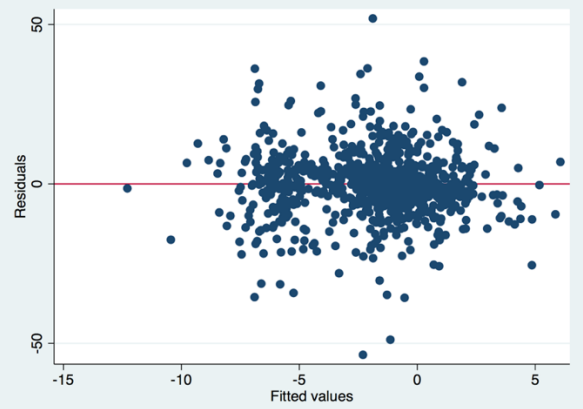
Peripheral Neurophysiology (PN)



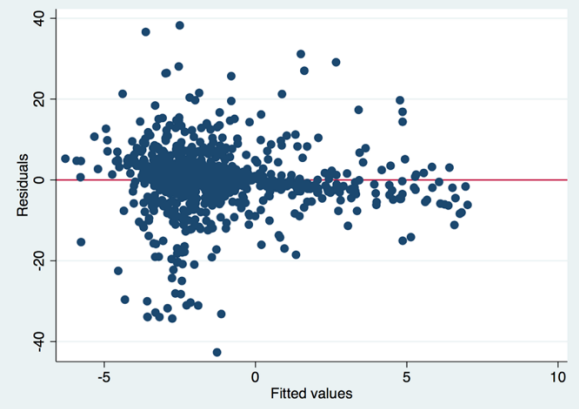
Respiratory Physiology (RP)



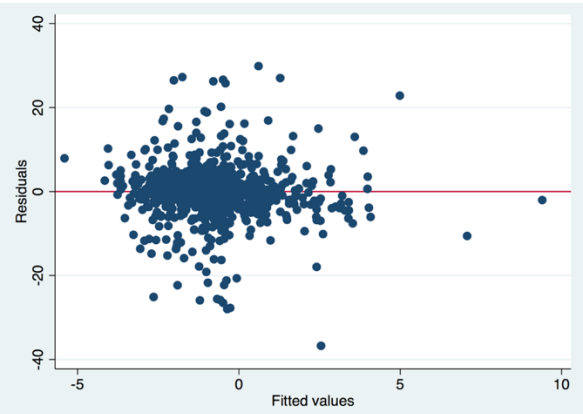
Urodynamics (URO)



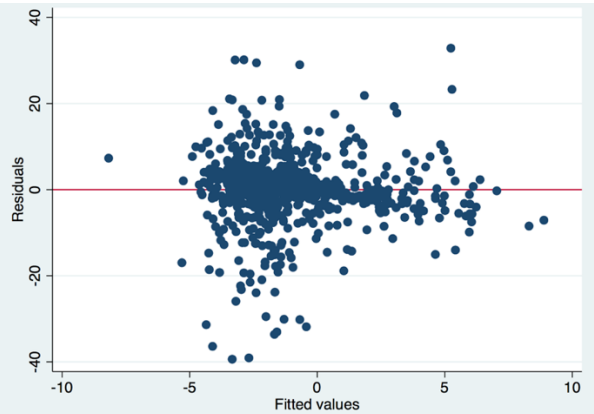
Colonoscopy (COL)



Cystoscopy (CYS)



Flexible Sigmoidoscopy (FS)



Gastroscopy (GAS)

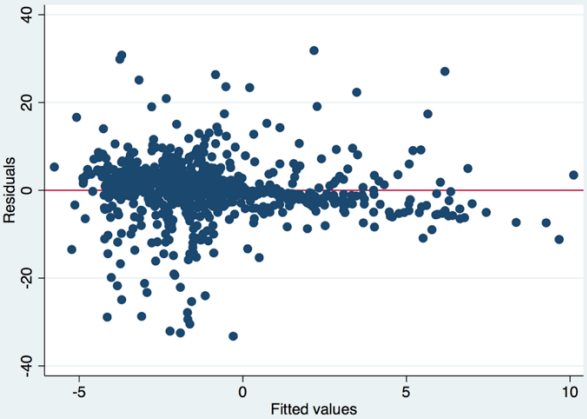


Table A17. Ramsey/Pesaran RESET test. H0: E(y|X) is linear in X.

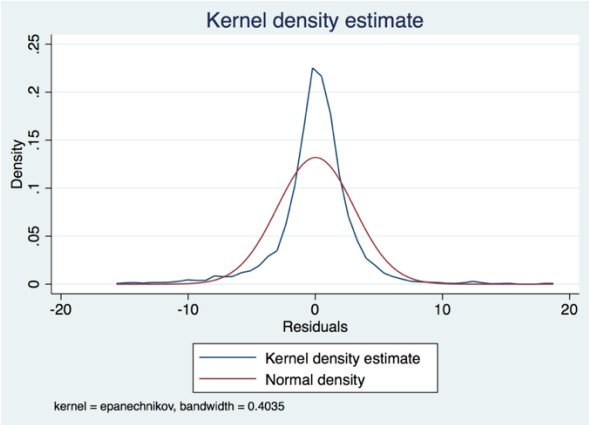
	Obs	$Chi^2(1)$	*p*
Two Week Wait targets			
Two-Week-Wait (TWW)	914	0.26	.6117
Two-Week-Wait Breast Symptoms (TWWBS)	914	6.78	.0092
Six Week Wait targets			
Imaging			
Computerised Tomography (CT)	914	5.23	.0209
Dual Energy X-ray Absorptiometry (DEXA)	914	0.03	.8658
Magnetic Resonance Imaging (MRI)	914	0.09	.7706
Non-obstetric Ultrasound (NOUS)	914	0.00	.9941
Physiology			
Audiology Assessments (AA)	914	24.29	.0000
Echocardiography (ECG)	914	5.31	.0211
Peripheral Neurophysiology (PN)	914	8.34	.0039
Respiratory Physiology (RP)	914	1.20	.2738
Urodynamics (URO)	914	0.05	.8304
Endoscopy			
Colonoscopy (COL)	914	1.66	.1978
Cystoscopy (CYS)	914	0.85	.3562
Flexible Sigmoidoscopy (FS)	914	0.05	.8280
Gastroscopy (GAS)	914	0.27	.6018

Table A18. Shapiro-Wilk test for normal data.

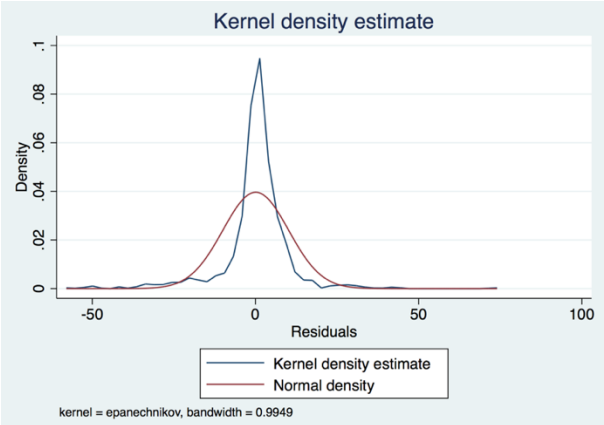
	Obs	W	V	z	p
Two Week Wait targets					
Two-Week-Wait (TWW)	914	0.88923	64.368	10.275	.0000
Two-Week-Wait Breast Symptoms (TWWBS)	914	0.81343	108.417	11.562	.0000
Six Week Wait targets					
Imaging					
Computerised Tomography (CT)	914	0.79389	119.773	11.807	.0000
Dual Energy X-ray Absorptiometry (DEXA)	914	0.56805	251.11	13.633	.0000
Magnetic Resonance Imaging (MRI)	914	0.71943	163.040	12.568	.0000
Non-obstetric Ultrasound (NOUS)	914	0.70519	171.315	12.690	.0000
Physiology					
Audiology Assessments (AA)	914	0.81599	106.930	11.528	.0000
Echocardiography (ECG)	914	0.73070	156.491	12.467	.0000
Peripheral Neurophysiology (PN)	914	0.84327	91.079	11.132	.0000
Respiratory Physiology (RP)	914	0.85646	83.411	10.915	.0000
Urodynamics (URO)	914	0.94314	33.040	8.630	.000
Endoscopy					
Colonoscopy (COL)	914	0.87248	74.103	10.623	.0000
Cystoscopy (CYS)	914	0.87083	75.059	10.654	.0000
Flexible Sigmoidoscopy (FS)	914	0.86679	77.408	10.730	.0000
Gastroscopy (GAS)	914	0.85367	85.033	10.962	.0000

Table A19. Kernel density plots.

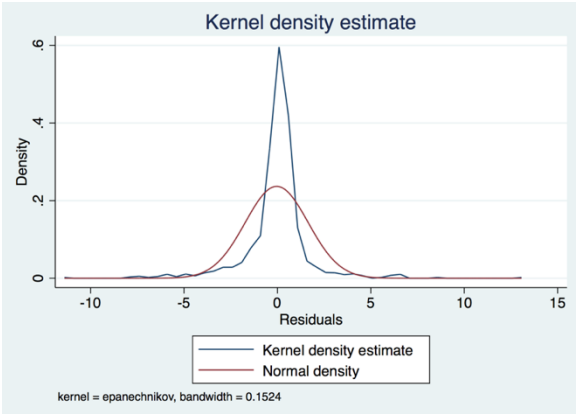
Two-Week-Wait (TWW)



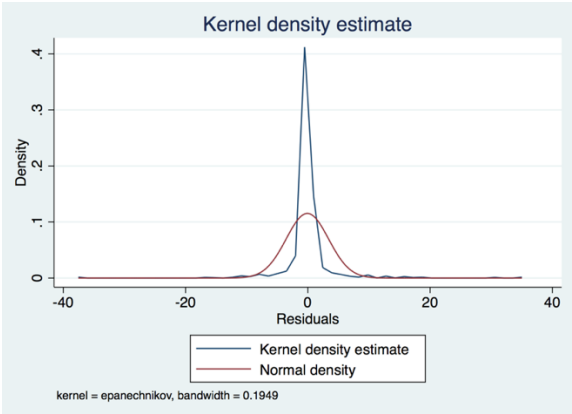
Two-Week-Wait Breast Symptoms (TWWBS)



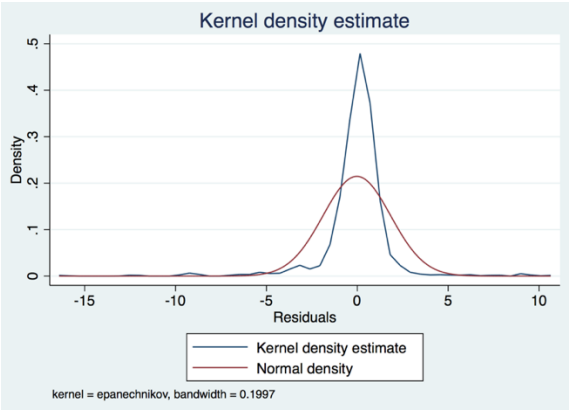
Computed Tomography (CT)



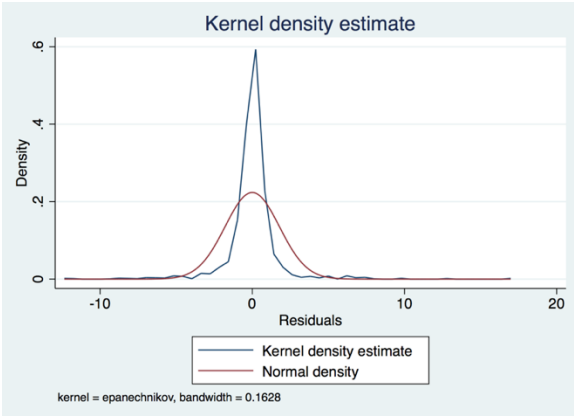
Dual Energy X-ray Absorptiometry (DEXA)



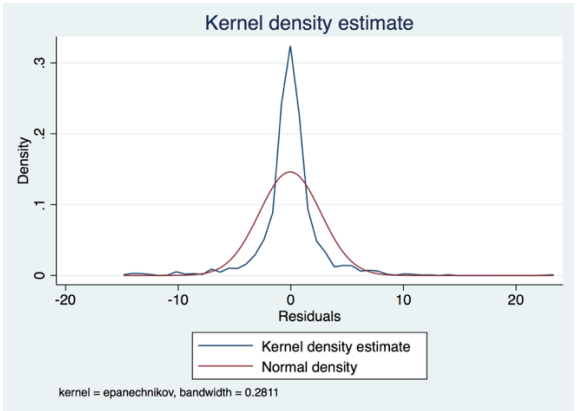
Magnetic Imaging Resonance (MRI)



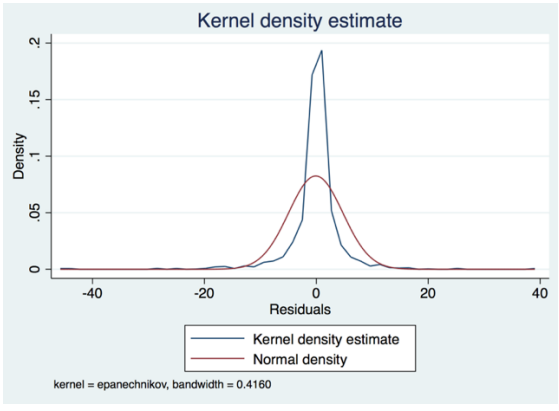
Non-Obstetric Ultrasound (NOUS)



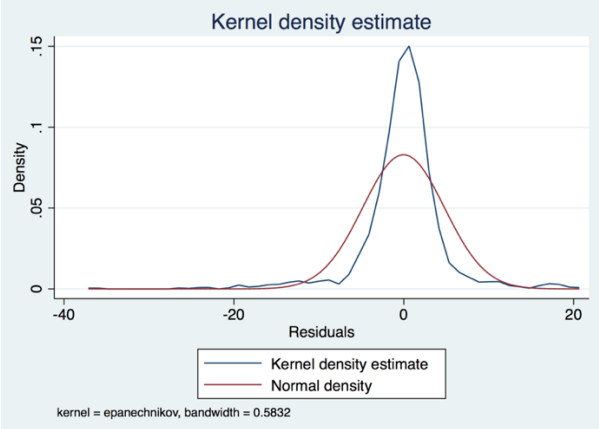
Audiology Assessments (AA)



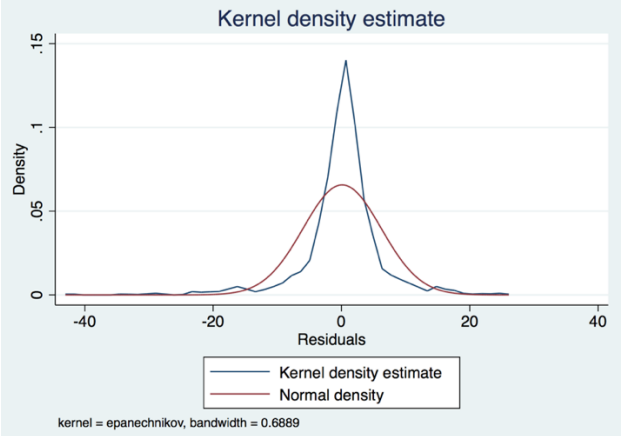
Echocardiography (ECG)



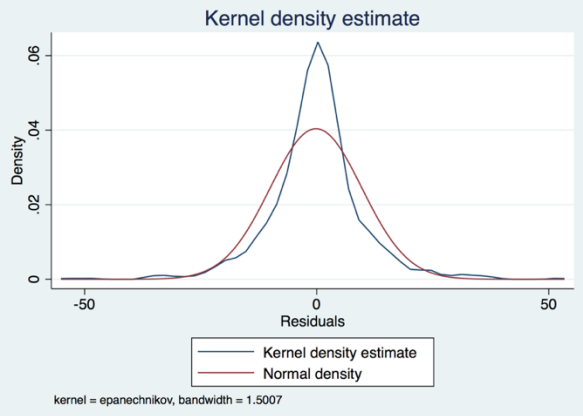
Peripheral Neurophysiology (PN)



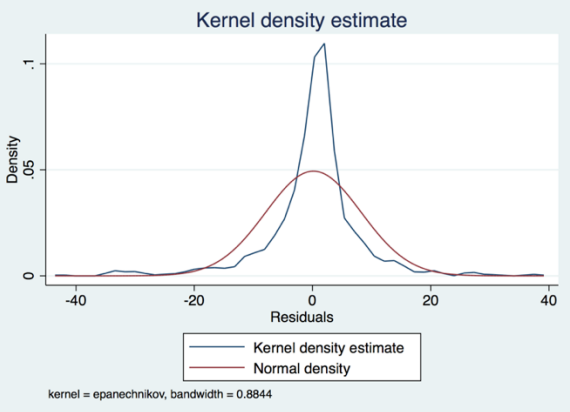
Respiratory Physiology (RP)



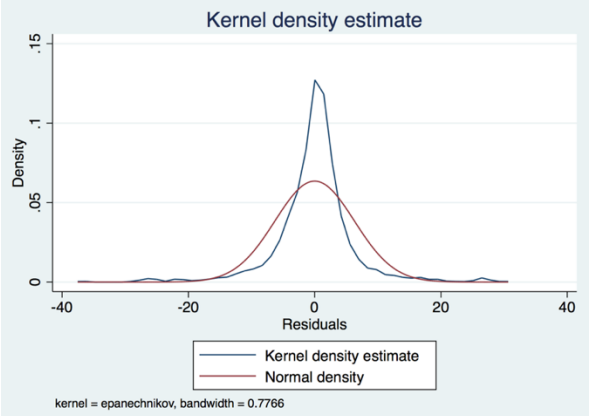
Urodynamics (URO)



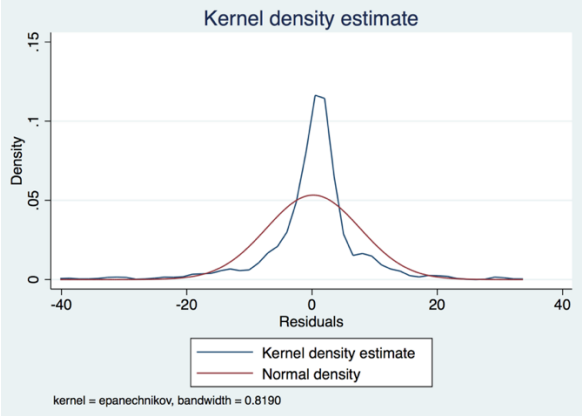
Colonoscopy (COL)



Cystoscopy (CYS)



Flexible Sigmoidoscopy (FS)



Gastroscopy (GAS)