Primary Care Doctors' Shortages and the Role of Financial Incentives *

[for the latest version please click here]

Hanifa Pilvar†

October, 2025

Abstract

Primary care plays a crucial role in ensuring the overall well-being of the population, particularly with the increasing aging demographic. However, the shortage of General Practitioner (GP) doctors remains a pervasive issue across developed countries, prominently in England. Regional disparities in GP access pose significant concerns for policymakers. This paper utilizes a comprehensive dataset of the GP workforce to investigate the impact of financial incentive policies on regional GP shortages in England. The study evaluates three alternative policy proposals using a static demand and supply model for general practices engaged in quality competition. The three policies are financial incentives to new GP registrars, revising practice payment formula to reallocate funds from less deprived to more deprived areas and increasing funding of the practices in the deprived areas. Comparisons reveal that while all scenarios have modest effects on GP per capita in deprived areas, policies directing funds to new entrants demonstrate superior efficacy over those targeting general practices.

Keywords: GP shortage, financial incentives, NHS, primary care, inequality, TERS

JEL classification: I10, I11, I18

1 Introduction

Primary care plays a pivotal role in promoting overall well-being and is widely acknowledged as an effective means to enhance population health (WHO, 2021). This

*I would like to express my sincere gratitude to my supervisors, Francesca Cornaglia and Claire Lim, for their invaluable guidance and insightful feedback throughout this project. I extend my sincere thanks to Pierre Dubois for his valuable advice during my visit to the Toulouse School of Economics. I would like to thank Felipe Gonzalez, Carol Propper, Anna Raute, Mathias Reynaert, Rita Santos, Daniel Wiesen, Yann Videau, Seminar participants at Queen Mary University of London and Toulouse School of Economics for their comments that helped improve this manuscript.

†University of Oxford, hanifa.pilvar@phc.ox.ac.uk

sector provides essential preventive healthcare services (Kirch et al., 2012) and effectively manages chronic diseases (Milani and Lavie, 2015), thereby aiding in the prevention of diseases from reaching critical stages. Within the primary care sector, General Practitioners (GPs) hold a central position. Numerous studies have consistently demonstrated the positive correlation between the availability of general practitioners and improved health outcomes, encompassing lower infant mortality rates, reduced overall mortality, and more effective management of lifestyle-related factors (Starfield and Shi, 2007). Nevertheless, in spite of its significance, GP shortages are widespread. In particular, inadequate GP access is a pressing issue in the UK, which is the subject of this study. England lags behind some other European countries, such as France and Germany, in terms of GPs per population (see Figure A4).

Furthermore, the distribution of GPs is unequal across the country, exacerbating the issue in deprived areas compared to more affluent regions. While numerous studies underscore the potential effectiveness of non-financial incentives, such as flexible working conditions, in encouraging GPs to practice in under-doctored areas (Lee et al., 2019), the predominant focus of policies addressing regional disparities in doctor access revolves around financial incentives. The provision of flexible working conditions requires a level of adaptability that is often lacking in areas where healthcare providers are inundated with patient demands.

Therefore, this paper seeks to provide insights into the effectiveness of financial incentives in addressing GP shortages in England, with a particular focus on regional disparities in the GP-to-population ratio. When comparing financial incentive policies to alleviate the shortage problem in deprived areas, it becomes evident that these policies have a minor impact on regional inequalities in GP access. However, policies that direct funding toward new entrant GPs prove to be more effective than those targeting general practices.

Looking at the count of full-time equivalent GPs per 10,000 patients in the most and least deprived areas, determined by the Index of Multiple Deprivation (IMD).[1] reveals that there exists a 0.5 Full Time Equivalent (FTE) GP gap between these two areas (see Figure A5). Despite an overall increase in the number of doctors, this gap persists over time (the number of FTE fully qualified permanent GPs, defined as GPs excluding those in training and those working under locum contracts, is declining over time; for reference, please consult Figure A6).

In the most deprived areas, GPs bear the responsibility of caring for 400-600 more patients compared to their counterparts in affluent regions, resulting in an overwhelming workload for GPs practicing in under-doctored areas (refer to Figure A11). It is noteworthy that the GP shortage in England, unlike the situation in many other countries, is not confined to rural areas (as illustrated in Figure A7). It is specifically a challenge prevalent in urban deprived areas. For instance, numerous neighborhoods in London grapple with some of the most severe shortages (as demonstrated in Figure A8). This complexity adds to the puzzle as the inadequate supply of GPs in urban deprived areas cannot be solely attributed to geographical remoteness.

The implications of GP shortages are substantial, affecting both the quantity and quality of primary care services. In terms of quantity, a reduced GP-to-patient ratio leads to extended wait times for patients seeking appointments with healthcare practices (refer to Figure A12). This issue is particularly noteworthy in the context of the escalating global burden of non-communicable diseases, primarily driven by the aging population, accounting for 70% of worldwide deaths (88% in the UK) (Reardon, 2011; World Bank, 2021).

Furthermore, the diminished availability of GPs also compromises the quality of care provided. A robust correlation exists between the level of GP workload and GP shortages; each additional GP per 10,000 patients is associated with an increase of

[1]The Index of Multiple Deprivation is a comprehensive relative measure of deprivation that combines seven domains, including income, employment, education, health, crime, housing, and living environment deprivation. In our context, a practice or patient is considered more deprived if it has a higher IMD score.

2.4 daily appointments per doctor (refer to Figure A9). This translates to practices in deprived areas handling a minimum of 1500 additional appointments per doctor annually. The excessive workload significantly influences GPs' decision-making processes concerning diagnostic inputs (Pilvar and Watt, 2023; Shurtz et al., 2018), as well as prescription patterns for antibiotics and opioids (Neprash, 2016). Ultimately, this contributes to lower levels of patient satisfaction with the services they receive, particularly in the most deprived areas (as illustrated in Figure A13).

Moreover, within the context of England and other countries with similar healthcare systems where GPs serve as gatekeepers, GP shortages limit access to secondary and tertiary care services. As noted by Pilvar and Watt (2023), increased GP workloads lead to a reduced likelihood of referral for routine and treatment purposes, while simultaneously raising the probability of referrals for urgent and assessment purposes. Consequently, the strain on primary care has the potential to cascade through other segments of the healthcare system.

In response to the previously mentioned challenges, policymakers have taken steps to alleviate disparities in primary care access. In England, one such initiative is the Targeted Enhanced Recruitment Scheme (TERS), colloquially known as the "Golden Hello," which was introduced in 2017 with the aim of addressing regional GP shortages. This scheme involves a one-time payment of £20,000[2] to GP specialty trainees, referred to as registrars, who commit to working in designated training locations identified by Health Education England. The overarching objective is to entice trainees to these areas, historically plagued by unfilled trainee vacancies, with the expectation that they will subsequently serve as fully qualified GPs upon completing their training. The scale of the scheme amounted to 30% of all registrar vacancies at the outset, and this proportion increased to 50% in the final year of my study (2021).

TERS policy lacks a specific focus on addressing healthcare disparities in deprived

[2] This amount is equivalent to 45% of registrars' average annual income which ranges from £44k to £55k based on the stage of training.

areas, even though the ongoing policy discourse predominantly revolves around neighborhoods characterized by high levels of deprivation and GP shortages. This absence of targeted focus on specific areas may account for the lack of a noticeable change in the trend of the GP-to-patient ratio, as depicted in Figure A5, especially after 2017. It is conceivable that GPs may have chosen to relocate to less deprived regions within the targeted areas.

Furthermore, numerous alternative policies have been proposed to rectify the GP-patient imbalance, yet these proposals have remained unimplemented. For instance, there have been recommendations to adjust the practice funding formula in line with the deprivation levels of the practice's location.[3] However, this policy was met with opposition on two occasions, in 2007 and 2016, by the GPs' union[4] due to concerns that its implementation would lead to reduced funding for many practices (BMA, 2015).

Therefore, I employ a static structural model to assess the impact of various financial incentive policies, such as restricting TERS to deprived areas and increasing funding to practices in deprived areas, on the GP-to-patient ratio. It is important to note that in the public healthcare sector of England, known as the National Health Service (NHS), there is no price. The private sector is quite small, representing only 3% of total consultations (King's Fund, 2014). Consequently, for the majority of patients, healthcare is provided free at the point of use, and general practices receive annual funding from the NHS. These practices are responsible for covering all costs related to the provision of healthcare for their registered patients, including the employment of staff. Therefore, my model take into consideration that patient choose the practice based on the quality of the services rather than the price.

In such a market, it is not immediately evident that practices are competing with

[3]While the current payment adjustments account for factors such as rurality, age mix, and gender mix of registered patients, they do not consider deprivation scores or workloads (Levene et al., 2019).

[4]Approximately 62% voted against reallocating funding based on deprivation scores

one another. However, through an estimation of the demand side of the model, I illustrate that patients place value on the quality of a practice, particularly concerning the number of GPs available. As a result, practices are effectively engaged in competition for quality, and patients may not necessarily choose the practice closest to their place of residency.

Through the utilization of the equilibrium model, I determine that the elasticity of demand is relatively modest. For instance, when reallocating 10% of funds from the least deprived practices to the most deprived practices, the result is a change of less than 2% in the number of GPs per patient. To address the gap depicted in Figure A5, England would need to reallocate approximately 30% of NHS funding from the 10% least deprived practices to the 10% most deprived practices.

It's essential to emphasize that fund reallocation within the NHS maintains a budget-neutral stance. Furthermore, I assess the impact of a policy that boosts funding primarily for primary care, with a specific emphasis on channeling this increase towards the most deprived neighborhoods. The findings reveal that a 10% rise in funding for practices results in a 1% increase in the number of GPs per capita.

When comparing various policy options, I determine that TERS, if exclusively applied to the most deprived areas, represents a more cost-effective means of increasing the GP-to-patient ratio in comparison to augmenting funding to general practices. Policies that allocate funds to deprived areas should prioritize individual doctors over practices. Nevertheless, a fund-neutral policy, while avoiding an increase in the financial burden on the healthcare system, could potentially give rise to tensions between GPs in affluent areas and commissioning bodies. Consequently, policymakers must carefully consider the cost-benefit aspects and political considerations when deciding between these options.

This paper contributes to the existing body of literature on the elasticity of physicians' labor supply. Numerous studies have indicated that the labor supply of physicians does not exhibit a significant response to financial incentives (Dunne et al., 2013; Falcettoni, 2018; Kulka and McWeeny, 2018; Lee et al., 2019; Sloan, 1975; Zhou, 2017). These studies suggest that financial incentives either fail to increase the supply of doctors or result in only marginal increases. However, some research has pointed out that physicians may demonstrate a degree of responsiveness to variations in expected wages, particularly in terms of their specialty choices (Nicholson, 2002; Rizzo and Blumenthal, 1994).

Conversely, other studies have provided evidence of a backward-bending labor supply curve among physicians, in which the income effect dominates the substitution effect (Brown, 1989; Hu and Yang, 1988; Kalb et al., 2015). These studies contend that financial incentives can worsen doctor shortages by further diminishing the labor supply.

This paper, in contrast, does not identify any evidence of a backward-bending labor supply curve. It demonstrates that financial incentive policies can effectively attract new entrant physicians to relocate to specific targeted regions, but they face challenges in mitigating GP shortages in areas with the highest deprivation rates. This aligns with the first strand of the literature mentioned earlier, providing additional support for the limited responsiveness of physicians' labor supply to financial incentives when the policy targets the most deprived regions.

This paper is closely related to the existing literature on the competition among healthcare providers in centrally funded healthcare systems (Gaynor et al., 2016; Hackmann, 2019; Hoxby, 2000; Rothstein, 2007; Santos et al., 2017). It contributes to this specific strand of literature by examining the supply-side response to financial incentives in the absence of a profit function for providers. The study reveals that changes in the funding of practices have a limited impact on the equilibrium level of GPs. This stands in contrast to the findings presented by Hackmann (2019), which focused on the nursing home market where the elasticity of labor supply for nurses is notably higher than that of GPs (Culyer et al., 2000; Hanel et al., 2014).

Furthermore, a study by Choné and Wilner (2022) explores the impact of financial incentives on non-profit hospitals in France. Their findings indicate that incentives aimed at procedure reimbursement may lead to a reduction in quality but can increase market share. This contrast between their study and the results presented in my paper underscores the significance of the target of financial incentive policies. Policies focused on specific aspects of quality, such as the number of doctors, can yield positive effects.

Financial incentives and the issue of regional inequality extend beyond the healthcare sector. Numerous studies highlight the importance of policies offering enhanced rewards to public sector employees who opt to work in underserved areas (Bobba et al., 2021; Dal Bó et al., 2013; Ederer, 2022). However, shortages within the healthcare sector have a significant impact on the overall well-being and life expectancy of individuals, with primary care playing a pivotal role in a well-functioning healthcare system. Therefore, my paper contributes to this broader literature by offering insights into a sector that has an immediate impact on individuals' well-being.

The organization of this paper is as follows: Section 2 provides a brief discussion on the structure of the primary care sector in England. Section 3 introduces the various datasets utilized in this study. Sections 4 and 5 specify and estimate a static model of supply and demand for primary care. In Section 6, counterfactual analyses are presented. Finally, Section 7 concludes the paper and compares different policies.

2 Background

The National Health Service (NHS) represents the healthcare system of the United Kingdom, primarily funded through taxation. Each constituent country within the UK maintains its own distinct NHS entity. For the purpose of this paper, the focus rests upon the scenario within England. As of 2019, NHS England's budget amounted to £148.9 billion, signifying approximately 20% of the government's total budget (King's Fund, 2022). The financial composition comprises 80% from general taxation, 18.5% from National Insurance contributions, and 1.5% from patient charges, such as dental fees (King's Fund, 2021). Noteworthy is the provision that all primary, secondary, and tertiary services are accessible without direct cost at the point of use, with allocation managed through waiting times.

The primary care sector is relatively cost-effective within the system, as only 8% of NHS funding goes to this sector (NHS England, 2017). Simultaneously, the NHS relies extensively on primary care doctors to control the total costs by acting as the gatekeepers of the whole system. The objective is relatively satisfied when comparing healthcare costs across countries. Papanicolas et al. (2019) show that the UK is spending less on healthcare per capita than countries such as the US, Germany, France, and Canada.

Funds are distributed through local health authorities known as Clinical Commissioning Groups (CCGs). These CCGs determine the range of services and establish payment structures through contracts with GP practices. Essential services, which a primary care practice is required to provide for the local population, encompass the routine care of patients with acute, chronic, and terminal conditions during regular working hours (from 8 am to 18:30 pm). In addition, there exist non-essential services which are optional for provision, enabling practices to earn more if they choose to offer these services. Compensation for most essential and non-essential services is provided through capitation payments to the practice, which is a payment per registered patient. The capitation segment of the NHS payment is termed the "global sum" and constitutes 69% of the total net payment to the practice.[5] The remaining 31% of the payment accounts for incentive schemes such as the Quality and Outcome

[5] Payment excluding the reimbursement of premises and information technology equipment costs

Framework,[6] as well as enhanced services. The practice is responsible for covering all costs, including staff expenses. A rough estimate of the practice's income indicates they receive £155 per patient; of this amount, 60% covers the practice's operational expenses (Gravelle et al., 2019), while the remainder constitutes the practice's profit.

Within primary care practices, two categories of permanent fully qualified doctors are partner GPs and salaried GPs. Partner GPs operate as business proprietors, jointly sharing the practice's profits. On the other hand, salaried GPs are employed by the practice partners and function under fixed-term contracts. Alongside these, other doctor roles within the practice encompass locum GPs,[7] who offer their services flexibly, and GP trainees referred to as Registrars. The latter group of trainees engage in patient care under the guidance of a qualified GP as part of their training regimen.

The supply of GPs primarily stems from two sources: medical schools and the migration of foreign doctors to England. Following graduation from medical school and upon receiving a provisional license from the General Medical Council, medical students proceed through a period of 2 years of core training, followed by 3 to 8 years of specialized training. If they opt for general medicine as their specialty, they are required to fulfill a 3-year registrar role under the supervision of a qualified GP within a general practice. International GPs have the option to enter the market either by progressing through the training stages or directly through the International Induction Programme (IIP), which allows them to enter the labor market without undergoing additional training in the UK (refer to Figure A10 for a more detailed illustration).

To bolster the supply of new GPs in regions experiencing shortages, Health Education England (HEE) initiated a policy termed the Golden Hello or Targeted Enhanced Recruitment Scheme (TERS) in 2017. TERS serves as a financial incentive program

[6] The Quality and Outcome Framework involves payment per achieved point, awarded when the practice meets specific targets in chronic disease management

[7] For the purposes of this discussion, locum GPs are treated as a salaried GPs.

designed for trainees who make a commitment to practice in areas within the UK that have historically struggled with recruitment, or are situated in under-doctored or deprived zones. Under TERS, GP Specialty Trainees in selected regions receive a one-time payment of £20,000. The scheme's design anticipates that trainees drawn to these areas by the policy are more inclined to continue practicing there beyond their training period due to the appealing aspects these locations offer.

This payment equates to a £1,100 annual income increase when accounting for a 4% discount factor over a span of 30 years. Considering that the average salaries for GP trainees, often referred to as GP registrars, range from £44,000 to £63,000, contingent on their training stage (BMA, 2023), this one-time payment constitutes a significant boost to their income during their first year. However, it holds a relatively minor impact over the course of their professional lifespan.

3 Data

In this research, I draw upon multiple datasets to inform my analysis. Primarily, I utilize the annual publications of the General Practice Workforce (GPW) dataset. The GPW dataset comprises two distinct data subsets: one centered around individual-level information and another centered around practice-level data. Throughout the paper, I have limited the observations to 2015 onward. Since 2015, GPW annual publications have been based on the NHS Workforce Minimum Data Set, which is a census of NHS employees. Prior to that, GPW was based on a different dataset; therefore, the statistics are not comparable.

In the practice-level workforce data, I am able to access information pertaining to practice characteristics, including the count of GPs categorized by their broad roles. This practice-level data serves as the foundation for my structural analysis, which comprises a static model of the demand for and supply of general practices.

In conjunction with the workforce data, I also leverage the NHS Payments to General Practices dataset. This dataset encompasses comprehensive details about funding allocated to all general practices within England. Additionally, the Quality and Outcome Framework data supplements my analysis by offering annual quality indicators for general practices. Furthermore, the Patients Registered at a GP practice dataset contributes valuable information, specifically the count of patients registered at each GP practice per Lower Super Output Area (LSOA).[8] This latter dataset aids in determining the average distance between patients and their chosen practices which is an important determinant of the choice of health care provider in the literature (see for e.g. Santos et al., 2017).

All the aforementioned datasets operate at the practice-year level and offer valuable insights. I limit the observations analysis to 2015-2019 to exclude the Covid-19 shock to general practice payment structure. A comprehensive overview of the characteristics inherent in these datasets are presented in Table 1.

The average distance between patients and their respective practices is approximately 1.7 kilometers. It's important to acknowledge that my distance measurement might have some degree of uncertainty due to the utilization of Lower Super Output Areas (LSOAs) rather than precise postcodes to gauge patient locations. In my structural analysis, I adopt a choice set of patients within a 10-kilometer radius around the center of the LSOA. This radius is substantially larger than the maximum distance found within my sample, providing ample coverage. Notably, the NHS has reported that around 90% of practices exhibit an average patient-to-practice distance ranging from 800 meters to 3.9 kilometers (NHS, 2018); therefore my choice set is wide enough to include all available options for patients.

The NHS allocates a total of £1.3 million to practices annually. On average,

[8] A Lower Super Output Area (LSOA) is the most finely grained geographic level for which my data holds information about patient registrations at practices. In total, there are 32,844 LSOAs in England as of 2019, with an average population of around 1,600 residents.

practices achieve 96% of the total potential Quality and Outcome Framework (QOF) points[9], which has prompted recent research like the systematic review conducted by Forbes et al. (2017) to assert that QOF no longer effectively incentivizes improved quality of care. Nevertheless, in my analysis, I treat QOF as an exogenous quality indicator in the demand for the practice. This choice is driven by the fact that QOF serves as the most comprehensive quality evaluation metric uniformly assessed by the NHS across all practices.

Further insights highlight that 46% of GPs are female, while 30% are under the age of 40. Additionally, 30% of GPs obtained their primary medical qualification from institutions outside the UK.

The Personal Medical Service (PMS) contract, as opposed to the nationally commissioned General Medical Service (GMS) contract, operates on a local negotiation basis with the practice. This approach allows for customization to meet the specific needs of the local patient population. PMS contracts emerged through pilot studies in 1998 (Campbell et al., 2005), but they are currently in the process of being phased out. Within my sample, 31% of practices operate under a PMS contract. Practices have an average age of 27 years and 16% of them provide drug dispensing services to their patients. I treat these variables as exogeneous quality measures of the practice.

4 Model

We cannot estimate the effect of the TERS policy in reduced form because we do not know which individuals in which practices have received the financial payment. The available information is only at the HEE level (a total of 13 HEE regions in the data), which makes such estimation quite noisy (for a simple analysis at HEE level please

[9]QOF is a reward and incentive programme for all GP practices in England, detailing practice achievement results. The QOF contains five main components: clinical, public health, additional Services, vaccination and immunisation and quality improvement

Table 1: Summary statistics of practice level variables

	Mean	SD	Min	Max
Distance (1km)	1.70	0.86	0.04	8.68
Total NHS payment (£1m)	1.29	0.80	0.01	12.78
QOF(%)	96.34	5.70	2.50	100.00
Number of GP per practice	4.68	3.15	1.00	40.16
Number of patients per practice	8482.65	5037.43	540.00	79647.50
Share female GPs	0.46	0.25	0.00	1.00
Share under 40 GPs	0.30	0.25	0.00	1.00
Share foreign GPs	0.30	0.33	0.00	1.00
PMS contract	0.31	0.46	0.00	1.00
Practice age	27.23	12.48	0.00	71.00
Dispensing practice	0.16	0.37	0.00	1.00
Observations (practice × year)	30,491			
Observations (LSOA × practice × year)	3,129,793			

see the Appendix A1). However, we can estimate a static structural model using the practice level data to compare various active and potential policies.

4.1 Demand

I write a static model of demand for patients registering at a GP practice in time t. Specifically, I assume the following indirect utility function for patient i in LSOA l who chooses practice j at year t:

$$U_{iljt} = \beta_1 D_{lj} + \beta_2 D_{lj}^2 + \beta_{GP} ln(\frac{GP_{jt}}{P_{jt}}) + \beta_x X_{jt} + \beta_{y,x} Y_{lt} \times X_{jt} + \zeta_j + \tau_t + \varepsilon_{iljt} \quad (1)$$

Where D_{lj} represents the distance between the patient's location and the practice. Since I lack the exact postcodes of patients, I measure the distance from the centroid of the LSOA to the practice's postcode. This approach aligns with the method used by Santos et al. (2017). I assume that distance impacts utility in a quadratic form. $ln(\frac{GP_{jt}}{P_{jt}})$ stands for the GP-to-patient ratio at the practice level, while X_{jt} encompasses exogenous practice characteristics. These characteristics include the age of the practice, the proportion of female GPs, the proportion of GPs under 45 years old, the proportion of GPs with qualifications from non-UK universities and the nurse-to-GP ratio. Additionally, I control for two quality indicators, which are considered exogenous in our analysis: whether the practice has a PMS contract (PMS contracts historically aim to enhance healthcare quality via locally commissioned funding) and Quality and Outcome Framework (QOF) points. ζ_j encompasses a set of observable practice characteristics known to patients but unobservable to the econometrician. Finally, ε_{iljt} represents a T1EV error term. I also account for the interaction between LSOA demographic characteristics, Y_{it} (such as the proportion of female individuals, the proportion of individuals under 18, and the proportion of individuals above 65), and exogenous practice characteristics to measure variations in population preferences regarding practice characteristics.

The modeling assumption on ε_{iljt} allows me to specify the practice choice probability as follows:

$$s_{iljt} = \frac{\exp(\beta_1 D_{lj} + \beta_2 D_{lj}^2 + \beta_{GP} ln(\frac{GP_{jt}}{P_{jt}}) + \beta_x X_{jt} + \beta_{y,x} Y_{lt} \times X_{jt} + \zeta_{jt})}{\sum_{k \in CS_l} \exp(\beta_1 D_{lk} + \beta_2 D_{lk}^2 + \beta_{GP} ln(\frac{GP_{kt}}{P_{kt}}) + \beta_x X_{kt} + \beta_{y,x} Y_{lt} \times X_{kt} + \zeta_{kt})} \quad (2)$$

Where CS_l is the choice set of patients in LSOA l which is defined as all practices in a 10Km radius around the centroid of the LSOA.

4.2 Supply

In this section, I present a static model of competition in quality similar to the supply model developed by Choné and Wilner (2022). I assume that the objective function of the practice depends on its revenue and the cost of treating patients and employing GPs:

$$V_{jt}(P_{jt}, r_{jt}, GP_{jt}) = \bar{T}_{jt} + r_{jt} P_{jt} + \alpha_{jt}^{P} P_{jt} + \alpha_{jt}^{GP} GP_{jt} \quad (3)$$

The revenue of practice j at year t is derived from two sources: a lump sum payment, $\bar{T}_{jt}$, and a capitation payment, which is determined by the rate r_{jt} per registered patient, P_{jt}. The cost structure of the practice consists of two components. First, there is $\alpha_{jt}^{P} P_{jt}$, representing the cost associated with delivering a higher quality of care per patient. Second, there is $\alpha_{jt}^{GP} GP_{jt}$, which represents the cost incurred in employing GPs.

Practice chooses the level of GP staffing to maximize its objective function:

$$\max_{GP_{jt}} V_{jt}(P_j(GP_{jt}), r_{jt}, GP_{jt}) \tag{4}$$

F.O.C:

$$\begin{aligned} \frac{\partial P_j}{\partial GP_j} &= \frac{-\alpha_j^{GP}}{r_{jt} + \alpha_j^{P}} \\ \sum_i \frac{\partial s_{ijt}}{\partial GP_{jt}} &= \frac{-\alpha_j^{GP}}{r_{jt} + \alpha_j^{P}} \end{aligned} \tag{5}$$

Define the elasticity of demand w.r.t the number of GPs: $\eta_{jjt} = \frac{\partial P_{jt}}{\partial GP_{jt}} \frac{GP_{jt}}{P_{jt}}$

$$\frac{1}{\eta_{jjt} \frac{P_{jt}}{GP_{jt}}} = -\frac{1}{\alpha_j^{GP}} r_{jt} - \frac{\alpha_j^{P}}{\alpha_j^{GP}} \tag{6}$$

Adding a constant λ_t to account for aggregate shocks and assuming constant α_j^{GP}, yields the estimating supply equation in linear form:

$$\frac{1}{\eta_{jjt} \frac{P_{jt}}{GP_{jt}}} = \gamma r_{jt} + \lambda_j + \lambda_t \tag{7}$$

Where γ is the inverse of the average cost of employing a GP in the practice.

4.3 Identification

To estimate demand parameters, I employ the two-step procedure proposed by Berry et al. (1995). In the first step, I utilize a Maximum-Likelihood approach to estimate taste heterogeneity concerning the distance to the practice, interaction terms, and mean utilities at the practice level. In the second step, I employ an instrumental variable regression to recover mean preferences for observable characteristics at the practice level.

More specifically, in the first step, I recover the common component of utility for each practice and each year ($\delta_{jt} = \beta_{GP} ln(\frac{GP_{jt}}{P_{jt}}) + \beta_x X_{jt}$) and the vector of idiosyncratic preference parameters ($\boldsymbol{\theta} = [\beta_1, \beta_2, \beta_{x,y}]$) via maximum likelihood. Let d_{iljt} take the value 1 if a patient from LSOA l in year t chooses practice j. Then, the log-likelihood function is as follows:

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\delta}) = \frac{1}{N} \sum_i \sum_j \sum_t w_{ljt} d_{iljt} ln(s_{iljt}(\boldsymbol{\theta}, \boldsymbol{\delta})) \tag{8}$$

I weight the log-likelihood function by the number of patients registered from LSOA l at practice j in each year t (w_{ljt}).

I use the technique proposed by Berry et al. (1995) to estimate $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$. They showed that for every value of $\boldsymbol{\theta}$, $\boldsymbol{\delta}$ is the unique fixed point of the contraction mapping:

$$T(\delta_{jt}) = \delta_{jt} + [ln(s_{jt}) - ln(\hat{s}_{jt}(\boldsymbol{\theta}, \boldsymbol{\delta}))] \tag{9}$$

Where s_{jt} is the true share of practice j of all patients in England in year t and $\hat{s}_{jt}(\boldsymbol{\theta}, \boldsymbol{\delta})$ is the estimated share for each iteration of $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$.

In the second step, I use the estimated vector of mean utilities ($\boldsymbol{\delta}$) to perform an IV regression of δ_{jt} on practice characteristics. OLS estimation of β_{GP} could be biased due to unobserved characteristics affecting patient utility. To address the potential endogeneity of the main variable of interest ($\frac{GP_{jt}}{P_{jt}}$), I employ the TERS policy indicator as an instrumental variable for $\frac{GP_{jt}}{P_{jt}}$. As demonstrated in the previous section, I showed that the TERS policy significantly increases the number of FTE GPs per population, making my IV relevant. Furthermore, the policy has been implemented at a higher local level rather than at the practice level, ensuring the exogeneity of the instrument with respect to the utility of patients from the choice of the practice.

In the final step, I recover own elasticities with respect to the number of GPs. I then run a fixed effect regression of $\frac{1}{\eta_{jjt}\frac{P_{jt}}{GP_{jt}}}$ on rate of capitation payment per patient and practice and year fixed effect to recover the main supply parameter (γ).

5 Model estimation

Table 2 presents the the coefficient of parameters that vary by patient and practice from the demand estimation. I have reported certain variables that reveal differential preferences of elderly patients for specific practice characteristics. For a comprehensive list of parameters, please refer to Table A6 in the appendix.

In general, patients tend to favor practices located closer to their residence. Patients in areas with a higher elderly population show a preference for practices that dispense drugs and those with a PMS contract. Some rural practices, due to limited access to pharmacies, dispense drugs, and simultaneously, rural areas often have a higher proportion of elderly residents, making them more inclined towards such practices.

The initiation of PMS contracts in the early 2000s aimed to enhance the quality of healthcare for local populations. Consequently, areas with a significant elderly population tend to have more practices with these contracts. Furthermore, since healthcare services are customized to meet the needs of the local population, they are naturally more appealing to older patients.

Additionally, there is a preference for female, young, and non-UK GPs in areas with a high elderly population, and a corresponding aversion to salaried GPs and a higher nurse-to-patient ratio.

Table 2: Demand taste parameters

Variable		Coefficient	SE
β_1	Distance	-0.444	5.13e-05
β_2	$Distance^2$	0.09	5.54e-04
$\beta_{y3} \times \beta_{x1}$	Share elderly population $\times$ Dispensing practice	3.58	1.76e-03
$\beta_{y3} \times \beta_{x3}$	Share elderly population $\times$ PMS contract	8.92	1.97e-03
$\beta_{y3} \times \beta_{x4}$	Share elderly population $\times$ Practice age	-0.05	5.39e-05
$\beta_{y3} \times \beta_{x2}$	Share elderly population $\times$ Share female GP	7.41	2.49e-03
$\beta_{y3} \times \beta_{x5}$	Share elderly population $\times$ Share young GP	5.06	2.38e-03
$\beta_{y3} \times \beta_{x6}$	Share elderly population $\times$ Share foreign GP	4.89	1.86e-05
$\beta_{y3} \times \beta_{x7}$	Share elderly population $\times$ Share salaried GP	-7.52	2.35e-03
$\beta_{y3} \times \beta_{x8}$	Share elderly population $\times$ ln(nurse to patient)	-1.70	2.86e-04
Likelihood function value			-4.30
Observations			3,260,746

Note: Distance is measured in kilometer. The choice set of patients are considered to be in a 10km radius around the centroid of the LSOA of residence. Dispensing practices are those that dispense drugs to patients. PMS contracts are historically designed to foster higher quality of care in primary care practices. Young GPs are those below 45 years-old. Foreign GPs are those who received their primary medical qualification from outside the UK.

Table 3 displays the results of the OLS and IV regressions for the second-stage demand parameter estimation. Patients exhibit a preference for practices with a higher GP-to-patient ratio, which serves as a key quality indicator in my analysis. This indicates that practices in England engage in quality competition within the primary care sector. Specifically, patients care about the level of staffing of the practice.

By utilizing the indicator of TERS policy as an instrumental variable for the GP-to-patient ratio, the coefficient for this variable increases to 1.99. The instrumental variable exhibits significant strength, with an F-statistic of 32.34.

Finally, I employ demand elasticities to estimate supply-side parameters in Table 4. I utilize NHS payment per patient and per weighted patient as the capitation rate in my supply estimation. Both variables yield similar results for the estimation of γ, which represents the inverse of the cost of GPs for the practice. The estimated cost falls within the range of £132,275 to £133,869. This figure is slightly higher than the national average of GP earnings, reported to be £100,700 in 2019/20 financial year (NHS digital, 2020). My estimates encompass the average costs associated with employing GPs, and their salaries is part of this cost for the practice. Nevertheless,

Table 3: Demand mean utility parameters

	(1) OLS	(2) IV
$ln(\frac{GP}{patient})$	0.403***	1.997**
	(0.0680)	(0.931)
Constant	13.11***	24.34***
	(0.886)	(6.598)
F statistics of the first stage		32.34
Time fixed effect	Yes	Yes
Practice fixed effect	Yes	Yes
Controls	Yes	Yes
Observations	31017	31017

Note: Standard errors are in parentheses and are clustered at practice level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

my estimates are close to national averages which validates my calculations as I do not directly use any measure of cost of the practice or GPs' earnings.

6 Counterfactual analysis

In this section, I present the counterfactual analysis of several policies aimed at increasing the number of GPs per population. First, I demonstrate the effect of the active policy, TERS. Next, I examine hypothetical policies, such as reallocating funds from the least deprived to the most deprived areas, similar to proposals made in 2006 and 2017. Additionally, I consider a policy to augment funding for primary care, with the additional funds directed towards the most deprived practices. For comparative purposes, I focus on the effect on the 10% most deprived practices. However, it is important to note that the analysis can be easily extrapolated to any threshold for deprivation or any desirable regional boundaries.

Table 4: Supply parameters

	(1)	(2)
Payment per patient	7.47e-6***	
	(4.94e-6)	
Payment per weighted patient		7.56e-6***
		(6.11e-6)
Constant	1.15e-3***	1.15e-3***
	(7.44e-5)	(9.21e-5)
Time fixed effect	Yes	Yes
Practice fixed effect	Yes	Yes
Observations	30280	30280

Note: Standard errors are in parentheses and are clustered at practice level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

6.1 TERS policy

In this subsection, my aim is to delve into the relationship between the TERS policy and how it aligns with my model. TERS specifically addresses the challenge of redistributing trainee doctors, known as registrars, to regions facing a shortage of GPs. These GP registrars receive payments from Health Education England during their training while working under the supervision of the training practice. As a consequence, TERS effectively introduces an additional GP doctor to the practice without altering the overall cost of employing GPs. Consequently, adding a free GP to the practice is as if reducing the average cost of GPs in my model by approximately 20%.

Given the low elasticity of demand concerning the GP-to-patient ratio, this cost reduction does not result in an equilibrium increase of 1 FTE of GP per practice. Instead, the model predicts that the compulsory relocation of junior doctors to target areas may prompt current GPs to exit the market or reduce their working hours in the equilibrium. Even if the registrar remains in the practice as a fully qualified GP, the policy alone may not suffice to maintain the desired effect on the number of GPs without a continuous influx of additional registrars into these areas.

Here, I assume that TERS is permanently active for the targeted practices. There-

Table 5: Counterfactual analysis of GP per 10,000 patients (TERS)

	(1) Baseline GP-patient	(2) Counterfactual GP-patient	(3) Change GP-patient	(4) % Change GP-patient	(5) % Change distance
TERS minimum vacancies					
10% most deprived	5.53	5.58	0.050	0.9%	-0.34%
	(0.114)	(0.114)	(0.003)		
TERS- expansion to all practices					
10% most deprived	5.53	5.67	0.14	2%	-0.87%
	(0.114)	(0.114)	(0.001)		

Note: Standard error of means in parentheses. Columns 1 and 2 show the number of GP per 10,000 patients in 2019, the last year of my structural analysis.

fore, even if registrars leave the practice, another registrar will move to this location. Table 5 presents the results of a permanent reduction in the cost of employing GPs in the most deprived areas under the minimum level of vacancies, which was 270 TERS posts in 2017, and under the expansion of TERS to all practices in the most deprived regions (approximately 600 TERS posts). Column 4 shows that TERS increases the number of GPs per 10,000 patients by 1% and 2% under the minimum vacancies and the expansion scenario, respectively.

These estimates are smaller than the estimate from my reduced-form analysis, which shows a 5.6% increase in the GP-to-patient ratio. It's important to note that my reduced-form analysis captures the short-run effect of the policy, while the structural analysis shows the effect at equilibrium. Furthermore, although the treated areas are underserved in terms of the number of primary care doctors, not all of them are necessarily highly deprived. Therefore, the actual take-up might be higher when compared to a scenario where the policy is only offered in the most deprived regions.

Finally, to offer a welfare analysis for patients, Column 5 of Table 5 presents the equivalent change in the distance to the practice that would yield the same improvement in utility as the change in the GP-to-patient ratio for each policy. The TERS policy, with the minimum vacancies and under full expansion, reduces the distance to the practice by 0.34% and 0.87%, respectively

6.2 Fund reallocation

Now, I can proceed to evaluate the effect of some alternative hypothetical policies related to offering financial incentives to doctors to move to under-doctored areas. The first policy under consideration involves reallocating 10% of NHS payments from the least deprived to the most deprived practices. This exercise aims to simulate the fund-neutral policy proposal, which was intended to introduce a correction factor to the payment formula for general practices based on their level of deprivation. Since my model is static, I can compare the results in a steady state. This implies that if the funding change becomes permanent, I anticipate that the GP-to-patient ratio will adjust according to my estimates. However, it's important to acknowledge that my model does not provide estimates regarding the dynamic effect of this policy."

Table 6 illustrates the GP-to-patient ratio under the existing payment scheme and a counterfactual policy for the year 2019. In this policy scenario, where 10% of funding is reallocated from the least deprived to the most deprived practices, I observe a 0.8% decrease in the GP-to-patient ratio for practices experiencing funding reductions. Conversely, practices receiving additional funding witness a 1.45% increase in this ratio. This change effectively narrows the gap by 20%. To provide a welfare analysis similar to the previous subsection, this policy is equivalent to a 0.5% reduction in distance for patients registered at practices that receive extra funding and a 0.22% increase in distance for patients at practices losing funding.

It's important to note that this policy corresponds to reallocating approximately £72 million, which represents around 0.5% of the total funding allocated to primary care in England.

To further narrow the gap, I've introduced a policy that reallocates funds from a substantial number of practices, specifically those in the top 20% of the deprivation distribution to the bottom 10% practices. This particular policy achieves a 30% reduction in the gap. This change is equivalent to 1% reduction in the distance to

Table 6: Counterfactual analysis of GP per 10,000 patients (Fund neutral)

	Baseline GP-patient	Counterfactual GP-patient	Change GP-patient	% Change GP-patient	% Change distance
Transferring 10% of NHS payment from 10% least deprived to 10% most deprived					
10% least deprived	6.14	6.09	-0.05	-0.8%	0.22%
	(0.1054)	(0.1054)	(0.00007)		
10% most deprived	5.53	5.61	0.08	1.45%	-0.50%
	(.1138)	(.1138)	(1.44e-09)		
Transferring 10% of NHS payment from 20% least deprived to 10% most deprived					
20% least deprived	6.22	6.16	-0.05	-0.8%	0.23%
	(0 .0850)	(0.0850)	(0.00005)		
10% most deprived	5.53	5.68	0.16	2.9%	-1.00%
	(0.1138)	(0.1138)	(1.91e-08)		

Note: Standard error of means in parentheses. Columns 1 and 2 show the number of GP per 10,000 patients in 2019, the last year of my structural analysis.

practice for practices gaining additional funding and 0.23% increase in the distance for those losing funds. However, it's important to be aware that implementing this policy would result in around 1300 practices experiencing a reduction in funding. This could potentially create significant turbulence within the primary sector funding landscape.

Based on my analysis, I've found that there is a relatively small elasticity concerning NHS funding. Specifically, a 10% change in funding leads to less than a 2% change in the number of GPs. This elasticity estimate stands in contrast to those for other healthcare workforce, such as nurses (Hackmann, 2019), and aligns with estimates for primary care doctors in the United States (Falcettoni, 2018; Kulka and McWeeny, 2018).

6.3 Increasing Funds

Another alternative policy approach involves augmenting funding for general practices and directing these additional funds exclusively towards the most deprived practices. The TERS policy serves as an exemplar of such an initiative, where an extra budget is allocated to encourage trainee applications in specific areas. However, in this section

Table 7: Counterfactual analysis of GP per 10,000 patients (Fund increasing)

	Baseline GP-patient	Counterfactual GP-patient	Change GP-patient	% Change GP-patient	% Change distance
Increasing payment to 10% most deprived by 10% (Flexible number of GPs)					
10% most deprived	5.53	5.59	0.055	1%	-0.34%
	(0.1138)	(0.1138)	(1.44e-09)		
Increasing payment to 10% most deprived by 10% (Fixed number of GPs)					
10% most deprived	5.53	5.58	0.050	0.9%	-0.32%
	(0.114)	(0.114)	(0.0007)		

Note: Standard error of means in parentheses. Columns 1 and 2 show the number of GP per 10,000 patients in 2019, the last year of my structural analysis.

I examine the effect where funding is allocated to practices in these areas.

In Table 7, I assess the impact of a policy that augments funding to the most deprived areas by 10%, presenting the steady-state results after its implementation in 2019. In the first panel, I allow for an increase in the total number of GPs in the model, while in the second panel, I assume that the total number of GPs remains fixed at the 2019 level. Interestingly, there isn't a substantial difference between the two scenarios, as the labor supply of GPs is relatively inelastic. In both cases, the number of GPs increases by 1%, with negligible changes observed in the number of GPs in the least deprived areas. This results in an 8% reduction in the gap between the most and least deprived areas. Importantly, this policy achieves this reduction in inequality without causing certain practices to lose funding. Furthermore, this policy is equivalent to 0.24% reduction in distance to practices for patients.

7 Concluding remarks

This paper has identified the disparities in GP availability between deprived and affluent areas, showing that there is 10% gap in terms of number of GPs per patient when comparing most with least deprived areas.

The paper has explored alternative policies, including revising the payment formula for practices based on deprivation and increasing funding specifically to primary care in deprived areas. The results suggest that patient demand for the number of GPs is relatively inelastic, and substantial reallocation of funding would be required to bridge the gap in GP availability between deprived and affluent areas.

The estimated elasticity of demand suggests that even a 10% reallocation of funds from the least deprived to the most deprived practices would result in a change of less than 2% in the GP-to-patient ratio. To bridge the existing gap in GP availability between deprived and affluent areas, a substantial reallocation of approximately 30% of NHS funding would be required.

Furthermore, the analysis explores the impact of increasing funding specifically to primary care and directing it towards the most deprived neighborhoods. The findings indicate that a 10% increase in funding to practices would result in a 1% increase in the number of GPs.

Finally, to assess the effectiveness of different policies, I compare the full expansion of TERS with two alternative policies: reallocation of 10% of funding from the 10% most deprived to the 10% least deprived practices, and increased payment to the 10% most deprived practices. TERS policy emerges as the most effective, resulting in a 2% increase in the GP-to-patient ratio and a 0.87% reduction in equivalent distance. It is followed by fund reallocation and then fund increasing policies. Therefore, policies that direct funding towards individual doctors prove more effective than those that increase funding for practices.

Regarding the cost of these policies for the NHS, it is not straightforward to compare fund reallocation with other fund-increasing policies (TERS and fund increasing towards the most deprived practices) because the fund-neutral policy results in some practices losing funds, which is not the case for fund-increasing policies. However, I can compare the full expansion of TERS with the fund-increasing policy towards the most deprived practices. If TERS is expanded to all practices in the most deprived areas with one TERS post per practice, it costs £12 million a year, while increasing funding by 10% costs £72 million a year. Therefore, the fund-increasing policy is not a cost-effective option among the policies assessed in this paper.

It is important to acknowledge some limitations of this study. First, I do not provide any effect on the health outcomes of patients. This requires a patient level data which is absent from this study.

Second, the analysis of the impact of policies is based on a static model of supply and demand for primary care. While this provides valuable insights, it does not capture the dynamic nature of healthcare systems and the potential long-term effects of policy interventions. Future research could consider incorporating a dynamic model of entry and exit to provide a more comprehensive understanding of the impacts of policies on junior and senior doctors.

Lastly, the study primarily focuses on the quantitative aspects of physician shortages and policy effectiveness, with limited exploration of qualitative factors such as job satisfaction, career preferences, and the overall work environment. These qualitative aspects play a crucial role in shaping the decisions and behaviors of healthcare professionals and could provide additional insights into the physician shortage issue (see Lee et al., 2019, for a review).

In conclusion, this research sheds light on the primary care physician shortage issue in England and the challenges associated with addressing it. The findings underscore the need for comprehensive and targeted policies that consider the specific dynamics of primary care provision, workload distribution, and the unique challenges faced by under-doctored regions.

References

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

BMA (2015). Focus on the global sum allocation formula (carr-hill formula). *London: BMA*.

BMA (2023). Pay and conditions circular (M&D).

Bobba, M., T. Ederer, G. Leon-Ciliotta, C. Neilson, and M. G. Nieddu (2021). Teacher compensation and structural inequality: Evidence from centralized teacher school choice in perú. Technical report, National Bureau of Economic Research.

Brown, M. C. (1989). Empirical determinants of physician incomes—evidence from canadian data. *Empirical Economics 14*(4), 273–289.

Campbell, S., A. Steiner, J. Robison, D. Webb, A. Raven, S. Richards, and M. Roland (2005). Do personal medical services contracts improve quality of care? a multi-method evaluation. *Journal of Health Services Research & Policy 10*(1), 31–39.

Choné, P. and L. Wilner (2022). Financial incentives and competitive pressure: The case of the hospital industry. *Journal of the European Economic Association 20*(2), 626–666.

Culyer, A. J., J. P. Newhouse, M. V. Pauly, T. G. McGuire, and P. P. Barros (2000). *Handbook of health economics*. Elsevier.

Dal Bó, E., F. Finan, and M. A. Rossi (2013). Strengthening state capabilities: The role of financial incentives in the call to public service. *The Quarterly Journal of Economics 128*(3), 1169–1218.

De Chaisemartin, C. and X. d'Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review 110*(9), 2964–2996.

Dunne, T., S. D. Klimek, M. J. Roberts, and D. Y. Xu (2013). Entry, exit, and the determinants of market structure. *The RAND Journal of Economics 44*(3), 462–487.

Ederer, T. (2022). Labor market dynamics and teacher spatial sorting. Technical report, Tech. rep.

Falcettoni, E. (2018). The determinants of physicians' location choice: Understanding the rural shortage. *Available at SSRN 3493178*.

Forbes, L. J., C. Marchand, T. Doran, and S. Peckham (2017). The role of the quality and outcomes framework in the care of long-term conditions: a systematic review. *British Journal of General Practice 67*(664), e775–e784.

Gaynor, M., C. Propper, and S. Seiler (2016). Free to choose? reform, choice, and consideration sets in the english national health service. *American Economic Review 106*(11), 3521–3557.

Gravelle, H., D. Liu, C. Propper, and R. Santos (2019). Spatial competition and quality: Evidence from the english family doctor market. *Journal of health economics 68*, 102249.

Hackmann, M. B. (2019). Incentivizing better quality of care: The role of medicaid and competition in the nursing home industry. *American Economic Review 109*(5), 1684–1716.

Hanel, B., G. Kalb, and A. Scott (2014). Nurses' labour supply elasticities: The importance of accounting for extensive margins. *Journal of health economics 33*, 94–112.

Hoxby, C. M. (2000). Does competition among public schools benefit students and taxpayers? *American Economic Review 90*(5), 1209–1238.

Hu, T.-W. and B. M. Yang (1988). The demand for and supply of physician services in the us: a disequilibrium analysis. *Applied Economics 20*(8), 995–1006.

Kalb, G. R., D. Kuehnle, A. Scott, T. C. Cheng, and S.-H. Jeon (2015). What factors affect doctors' hours decisions: comparing structural discrete choice and reduced-form approaches.

King's Fund (2014). The UK private health market.

King's Fund (2021). How the NHS is funded.

King's Fund (2022). The NHS budget and how it has changed.

Kirch, D. G., M. K. Henderson, and M. J. Dill (2012). Physician workforce projections in an era of health care reform. *Annual review of medicine 63*, 435–445.

Kulka, A. and D. McWeeny (2018). Rural physician shortages and policy intervention. In *PAA 2018 Annual Meeting*. PAA.

Lee, T., C. Propper, and G. Stoye (2019). Medical labour supply and the production of healthcare. *Fiscal Studies 40*(4), 621–661.

Levene, L. S., R. Baker, J. Bankart, N. Walker, and A. Wilson (2019). Socioeconomic deprivation scores as predictors of variations in nhs practice payments: a longitudinal study of english general practices 2013–2017. *British Journal of General Practice 69*(685), e546–e554.

Milani, R. V. and C. J. Lavie (2015). Health care 2020: reengineering health care delivery to combat chronic disease. *The American journal of medicine 128*(4), 337–343.

Neprash, H. T. (2016). Better late than never? physician response to schedule disruptions. *Unpubl Work Pap*.

NHS (2018). Digital-first primary care and its implications for general practice payments.

NHS digital (2020). GP earnings and expenses estimates 2019/20. *NHS digital*.

NHS England (2017). Next steps on the NHS Five Year Forward View.

Nicholson, S. (2002). Physician specialty choice under uncertainty. *Journal of labor Economics 20*(4), 816–847.

Papanicolas, I., E. Mossialos, A. Gundersen, L. Woskie, and A. K. Jha (2019). Performance of UK National Health Service compared with other high income countries: observational study. *bmj 367*.

Pilvar, H. and T. Watt (2023). The effect of workload on primary care doctors on referral rates and prescription patterns: Evidence from english NHS. *Working paper*.

Reardon, S. (2011). A world of chronic disease.

Rizzo, J. A. and D. Blumenthal (1994). Physician labor supply: Do income effects matter? *Journal of health economics 13*(4), 433–453.

Rothstein, J. (2007). Does competition among public schools benefit students and taxpayers? comment. *American Economic Review 97*(5), 2026–2037.

Santos, R., H. Gravelle, and C. Propper (2017). Does quality affect patients' choice of doctor? evidence from england. *The Economic Journal 127*(600), 445–494.

Shurtz, I., A. Eizenberg, A. Alkalay, and A. Lahad (2018). Physician workload and treatment choice: the case of primary care.

Sloan, F. A. (1975). Physician supply behavior in the short run. *Industrial and Labor Relations Review 28*(4), 549–569.

Starfield, B. and L. Shi (2007). Commentary: primary care and health outcomes: a health services research challenge. *Health services research 42*(6p1), 2252–2256.

WHO (2021). Primary health care. *Fact sheet*.

World Bank (2021). Cause of death, by non-communicable diseases (% of total). *World Bank Data*.

Zhou, T. J. (2017). The doctor is in/out: Determinants of physician labor supply dynamics. Technical report, Working Paper.

A Appendix

A.1 Reduced form analysis at HEE level

The TERS policy was implemented by Health Education England. It was primarily focused on areas in the north, east of England, and the south west. In contrast, regions such as London, Sussex, and the western midlands were not subject to any TERS posts and consequently serve as control areas in my study.[10] Despite being geographically smaller, these control areas encompass 40% of the total patient population (for visual reference, please see Figure A14).

To assess the efficacy of the TERS policy, I employ the following diff-in-diff model. To facilitate this analysis, I aggregate the observations at the HEE regions level, the level at which treatment takes effect.

$$Y_{ht} = \sum_{k \neq 0} \delta^k TERS_{hk} \times \sigma_k + \phi X_{ht} + \sigma_t + \sigma_h + \epsilon_{ht} \tag{10}$$

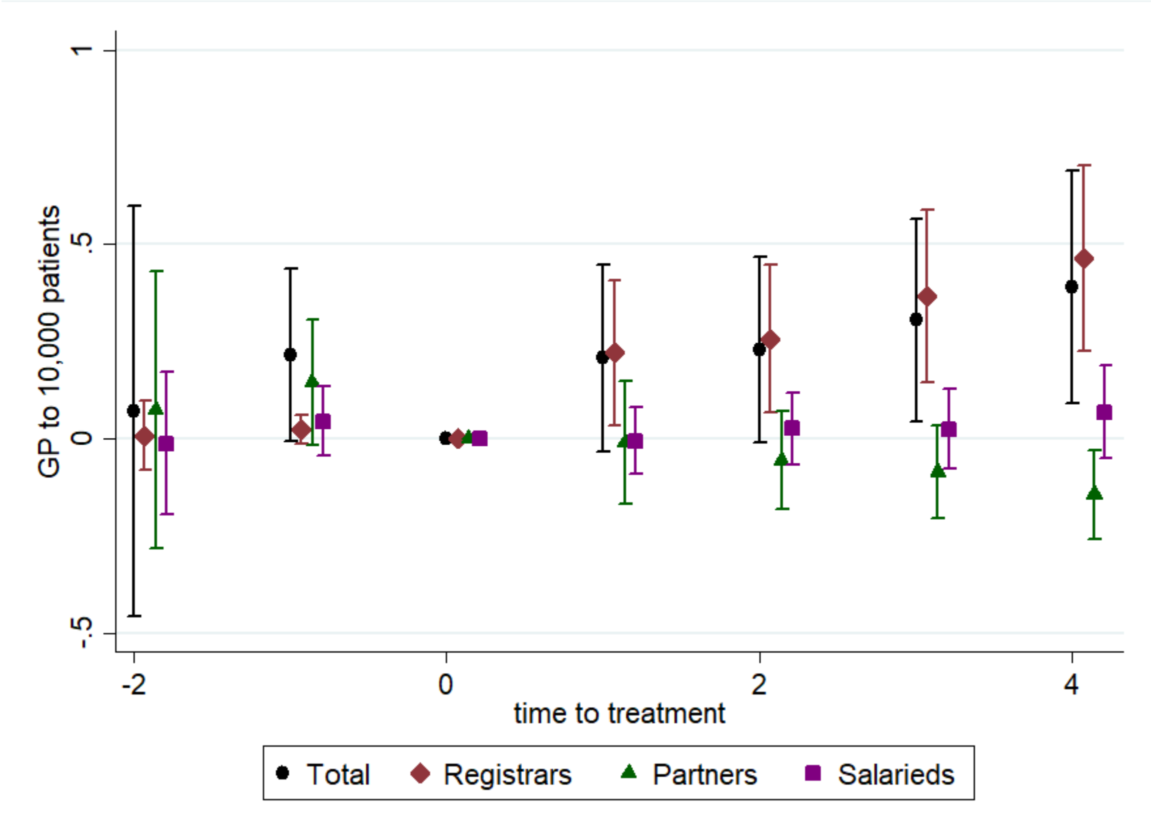
Where Y_{ht} is the number of full-time-equivalent GPs per 10,000 patients at the HEE level h and year t. $TERS_{ht}$ is a binary variable that takes the value of 1 if HEE h in year t is subject to the policy, and 0 otherwise. The control group comprises HEEs that were not exposed to the treatment. X_{ht} encompasses a set of amenity controls. The year-fixed effect is denoted by σ_t, while σ_h represents the fixed effect at the HEE level.

The main coefficient of interest is α_1, which quantifies the causal impact of the policy on the dependent variable. A positive estimate indicates the policy's effectiveness in attaining its objectives. Conversely, a negative estimate suggests the potential occurrence of a crowd-out effect resulting from the policy's enactment.

There are two key assumptions behind my model, one is parallel trend assumption

[10] It's important to note that the policy was not implemented in a staggered manner at HEE level; however, the number of TERS vacancies differ across regions and times.

Figure A1: The effect of TERS on the number of GPs per 10,000 patients



and the other is that the treatment effect should be constant, between regions and over time. Number of TERS posts is different across regions and it expands over time. Therefore, to account for the bias arising from non-constant treatment effect, I use the method proposed by De Chaisemartin and d'Haultfoeuille (2020). To check the validity of the first assumption, I look at the dynamic version of De Chaisemartin and d'Haultfoeuille (2020) method pre-treatment.

Figure A1 displays the results for total FTE of GPs and various GP types, registrars, partners and salaried GPs. First, I look at the pre-treatment trend in the number of GPs and see no significant difference relative to the year before the start of the policy. To emphasis the validity of parallel trend assumption, In Figure A15, I present the results for the Head Count (HC) of doctors which has more observations pre-treatment. I do not see any significant difference between treated and control areas before 2017 in the HC data either.

After the treatment, Figure A1 shows that the policy increases the total FTE of GPs which is mainly driven by the increase in the FTE of registrars. I also look at the qualified GPs to see if the increase in GPs in training had any crowding out effect on them. Following the intervention, there is no significant change in the FTE of salaried GPs although I observe a slight drop in the FTE of partner GPs which is significant 4 years after the start of the policy. To quantify the result , table A2 shows that the Average Treatment Effect (ATE) of the policy. It shows a 0.29 increase in the number of FTE GPs per 10,000 patients which is equivalent to 5.6% increase from the average. Focusing on the main driving force of this increase, the analysis shows that registrars FTE increased by 0.34 following policy implementation, effectively narrowing the gap between registrar entrants in targeted areas and those in other regions by a remarkable increase of 50% from the baseline.

It's crucial to consider the longer-term effects of the policy, particularly regarding registrars transitioning to fully qualified GPs and the potential crowding-out effect on partner GPs. Registrars typically remain in training posts for three years, so the impact on the number of qualified doctors in the target area may become more noticeable after 2020, however, the data I have spans only up to 2021. Although I have limited observations after the policy implementation to determine whether registrars are staying or leaving the region, I can still provide preliminary insights into this aspect. Nevertheless, for a more comprehensive understanding of the policy's effects, especially on qualified GPs, it would be beneficial to look at a longer time trend in the number of GPs per population.

Regardless of inadequate data, it is still worth noting that my results show that the policy led to a decline in the FTE of partner GPs, which could indicate a crowding-out effect. Note that this reduction is likely in the working hours of partner GPs as I do not observe any significant change in the HC of partners (see Figure A15). Over the long term, this effect may be balanced by an increase in the number and working hours of newly qualified GPs if the policy effectively encourages registrars to remain in the regions as either partner or salaried GPs. Ongoing data collection and analysis will be essential in assessing the policy's long-term effectiveness and its implications for the healthcare workforce in the target area.

In order to identify by what degree the TERS opportunities are grabbed by those who will not come to the area in the absence of the policy, I replace TERS dummy by $TERS\_post$ which indicates the number of TERS vacancies in each region. The ATE results are reported in table A3. On average there are 0.02 TERS posts per 10,000 population which increases the number of GPs by 0.01. Therefore, 50% of vacancies are taken up by those who would come to the region even in the absence of the policy, while 50% are taken by new entrants.

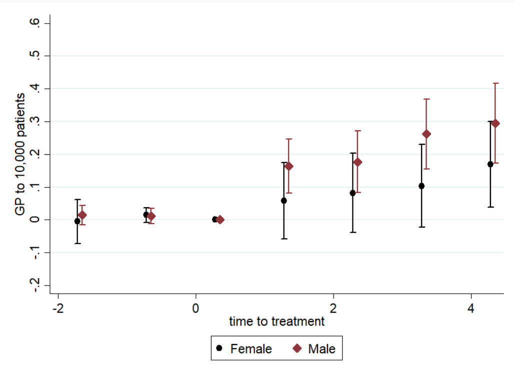
In Figure A2, I examine the effects within Registrars, considering various characteristics. The responses among male, female and non-UK registrars are all positive and statistically significant at the end of the study period. The overall results are primarily influenced by male registrars, with an average response that is more than twice larger than that observed among female registrars (see Table A5). Furthermore, it's noteworthy that the opportunity is primarily seized by registrars with non-UK Primary Medical Qualifications (PMQs). Targeted regions experience a remarkable 100% increase in the number of non-UK registrars (see Table A5).

Finally, reduction in the FTE of partner GPs is an unintended effect of the policy. I am interested to see the heterogeneous response of partners to identify the main drivers of this result. Figure A3 presents the heterogeneous effects among partner GPs. Notably, the results are pronounced among GPs who are over 50 years of age. No difference is observed based on the PMQ of partner GPs (see Figure A16).

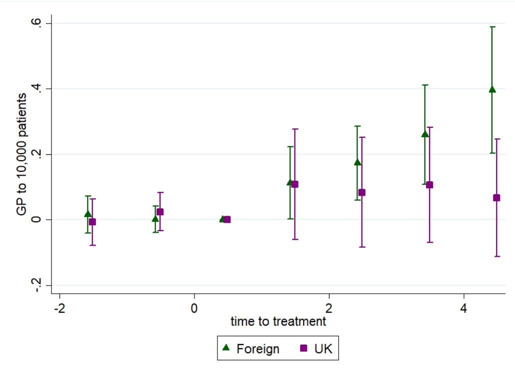
In summary, this section illustrates that financial incentive policies effectively encourage young junior GPs to enter underserved areas. However, this positive outcome

Figure A2: The effect of TERS on the number of Registrars per 10,000 patients

(a) by gender



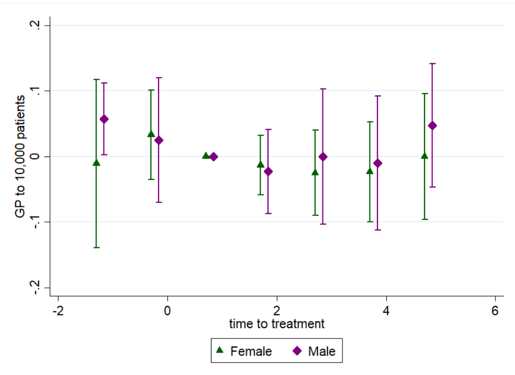
(b) by PMQ



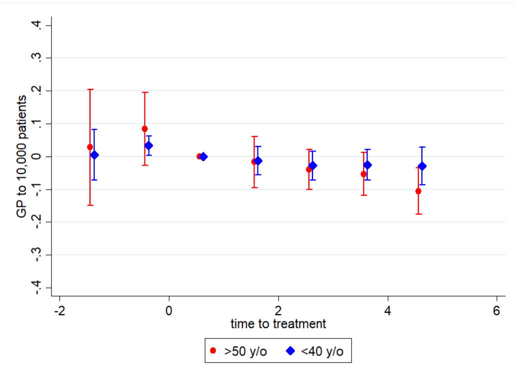
Note: Figure shows the dynamic estimate of equation 10 for registrars; separately by gender and by primary medical qualification.

Figure A3: The effect of TERS on the number of Partners per 10,000 patients

(a) by gender



(b) by age



Note: Figure shows the dynamic estimate of equation 10 for partners; separately by gender and by primary medical qualification.

is counterbalanced by the reduction in the working hours of senior GPs. The policy did not have any significant effect on the share of GPs in the most deprived areas within HEE region (see Figure A17). That is why there is no significant change in the GP per capita gap between deprived and non-deprived regions. Hence the registrars did enter the target areas but located in the least deprived area within the targeted region.

Therefore, to offer a more comprehensive perspective on financial incentives for increasing the GP-to-patient ratio, the following sections employ an equilibrium model to evaluate and compare various policies designed to enhance the number of GPs per population by providing financial incentives.

A.2 Extra tables

Table A1: Summary statistics of individual level GP workforce data

	Mean	SD	Min	Max
Partner GPs				
Share of female GP	0.43	0.49	0.00	1.00
Share of GPs with foreign PMQ	0.25	0.43	0.00	1.00
Age	48.90	9.26	24.00	86.00
Average earnings	115,262.99	11579.71	71583.67	167788.41
Observations	229327			
Salaried GPs				
Share of female GP	0.71	0.45	0.00	1.00
Share of GPs with foreign PMQ	0.28	0.45	0.00	1.00
Age	41.12	9.02	24.00	90.00
Average earnings	57,237.51	6632.60	33835.86	79677.40
Observations	114689			
Registrars				
Share of female GP	0.64	0.48	0.00	1.00
Share of GPs with foreign PMQ	0.33	0.47	0.00	1.00
Age	32.66	5.75	24.00	65.00
Observations	41799			

Table A2: ATE of TERS policy

	(1) Total	(2) Registrar	(3) Salaried	(4) Partner
TERS	0.292**	0.338***	0.030	-0.076
	(0.122)	(0.107)	(0.048)	(0.060)
Observations	91	91	91	91
Mean dep. variable	5.17	0.68	1.47	3.02

Standard errors in parentheses and are clustered at HEE level

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A3: ATE of the policy (FTE)- treatment intensity

	(1) Total	(2) Registrar	(3) Salaried	(4) Partner
TERS posts	0.0104**	0.0121***	0.0010	-0.0027
	(0.0043)	(0.0038)	(0.0017)	(0.0021)
Observations	91	91	91	91
Mean dep. variable	5.17	0.68	1.47	3.02

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A4: ATE of TERS policy (HC)

	(1) Total	(2) Registrar	(3) Salaried	(4) Partner
TERS	0.315	0.342***	-0.011	-0.015
	(0.179)	(0.110)	(0.103)	(0.068)
Observations	91	91	91	91
Mean dep. variable	6.92	0.71	2.63	3.58

Standard errors in parentheses and are clustered at HEE level

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A5: ATE of TERS policy- Registrars

	(1) Female	(2) Male	(3) Foreign PMQ	(4) UK PMQ
TERS	0.106	0.231***	0.243***	0.095
	(0.062)	(0.051)	(0.072)	(0.088)
Observations	91	91	91	91
Mean dep. variable	0.41	0.27	0.23	0.45

Standard errors in parentheses and are clustered at HEE level

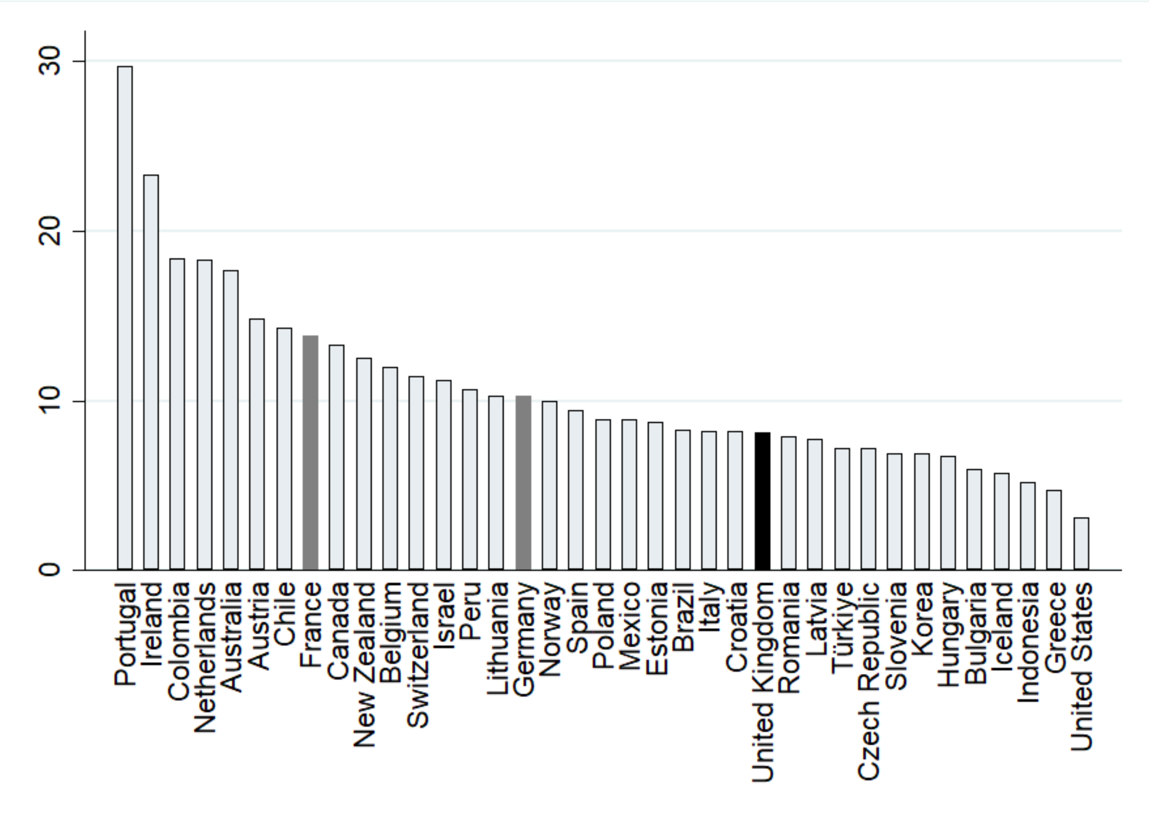
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Demand taste parameters -full list

Variable		Coefficient	SE
β_1	Distance	-0.444	5.13e-07
β_2	$Distance^2$	0.09	5.54e-06
$\beta_{y1} \times \beta_{x1}$	Share female population × Dispensing practice	-3.13	3.36e-05
$\beta_{y2} \times \beta_{x1}$	Share young population × Dispensing practice	5.23	2.14e-05
$\beta_{y3} \times \beta_{x1}$	Share elderly population × Dispensing practice	3.58	1.76e-05
$\beta_{y1} \times \beta_{x2}$	Share female population × Share female GP	7.41	3.41e-05
$\beta_{y2} \times \beta_{x2}$	Share young population × Share female GP	-4.22	2.40e-05
$\beta_{y3} \times \beta_{x2}$	Share elderly population × Share female GP	7.41	2.49e-05
$\beta_{y1} \times \beta_{x3}$	Share female population × PMS contract	1.01	3.81e-05
$\beta_{y2} \times \beta_{x3}$	Share young population × PMS contract	-23.77	2.40e-05
$\beta_{y3} \times \beta_{x3}$	Share elderly population × PMS contract	8.92	1.97e-05
$\beta_{y1} \times \beta_{x4}$	Share female population × Practice age	0.36	6.30e-07
$\beta_{y2} \times \beta_{x4}$	Share young population × Practice age	-0.19	6.15e-07
$\beta_{y3} \times \beta_{x4}$	Share elderly population × Practice age	-0.05	5.39e-07
$\beta_{y1} \times \beta_{x5}$	Share female population × Share young GP	-2.27	3.94e-05
$\beta_{y2} \times \beta_{x5}$	Share young population × Share young GP	-14.35	2.57e-05
$\beta_{y3} \times \beta_{x5}$	Share elderly population × Share young GP	5.06	2.38e-05
$\beta_{y1} \times \beta_{x6}$	Share female population × Share foreign GP	-14.18	3.56e-07
$\beta_{y2} \times \beta_{x6}$	Share young population × Share foreign GP	18.26	2.01e-07
$\beta_{y3} \times \beta_{x6}$	Share elderly population × Share foreign GP	4.89	1.86e-07
$\beta_{y1} \times \beta_{x7}$	Share female population × Share salaried GP	10.52	3.48e-05
$\beta_{y2} \times \beta_{x7}$	Share young population × Share salaried GP	17.09	2.29e-05
$\beta_{y3} \times \beta_{x7}$	Share elderly population × Share salaried GP	-7.52	2.35e-05
$\beta_{y1} \times \beta_{x8}$	Share female population × ln(nurse to patient)	5.17	4.08e-06
$\beta_{y2} \times \beta_{x8}$	Share young population × ln(nurse to patient)	-2.83	3.35e-06
$\beta_{y3} \times \beta_{x8}$	Share elderly population × ln(nurse to patient)	-1.70	2.86e-06
Likelihood function value			-4.30
Observations			3,260,746

Note: Distance is measured in kilometer. The choice set of patients are considered to be in a 10km radius around the centroid of the LSOA of residence. Dispensing practices are those that dispense drugs to patients. PMS contracts are historically designed to foster higher quality of care in primary care practices. Young GPs are those below 45 years-old. Foreign GPs are those who received their primary medical qualification from outside the UK.

Figure A4: Number of head count of GPs per 10,000 patients in 2021



A.3 Extra figures

Figure A5: Patient to GP ratio

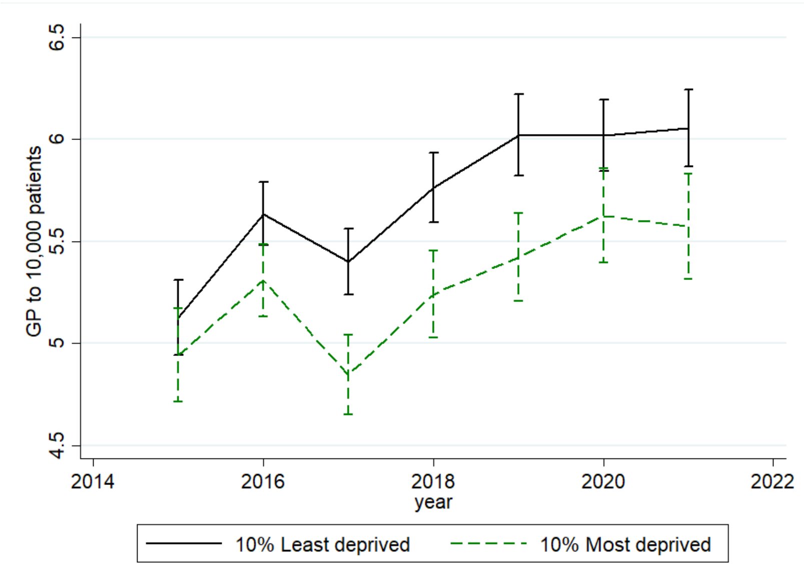


Figure A6: Number of fully qualified permanent GPs per 10,000 patients

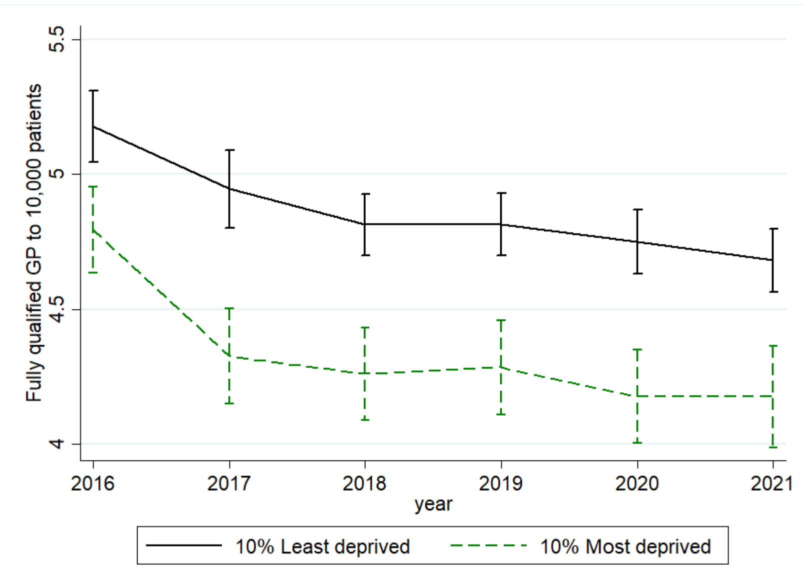


Figure A7: GP per 10,000 patients based on the rurality of the practice

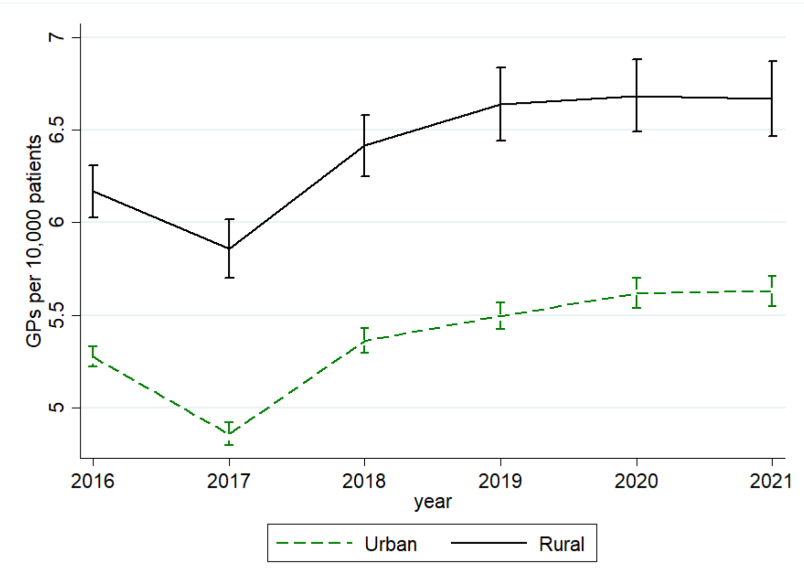


Figure A8: Map of GP per 10,000 patients at CCG level

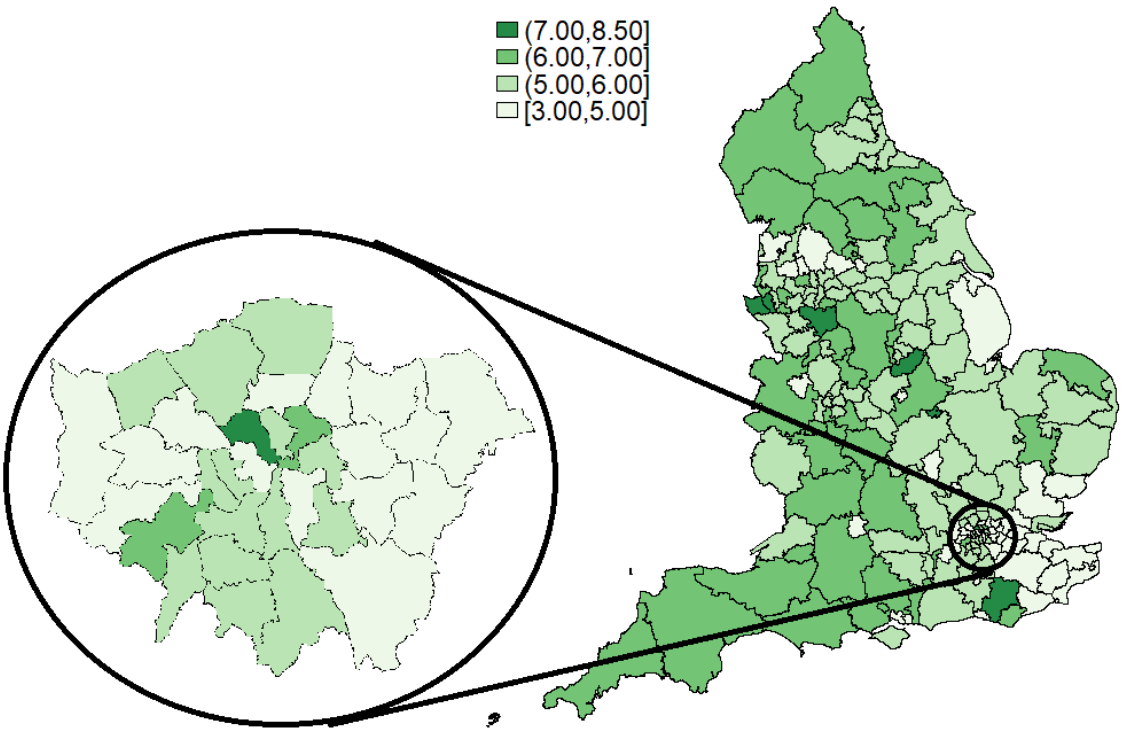


Figure A9: Correlation between GP daily workload and GP shortage

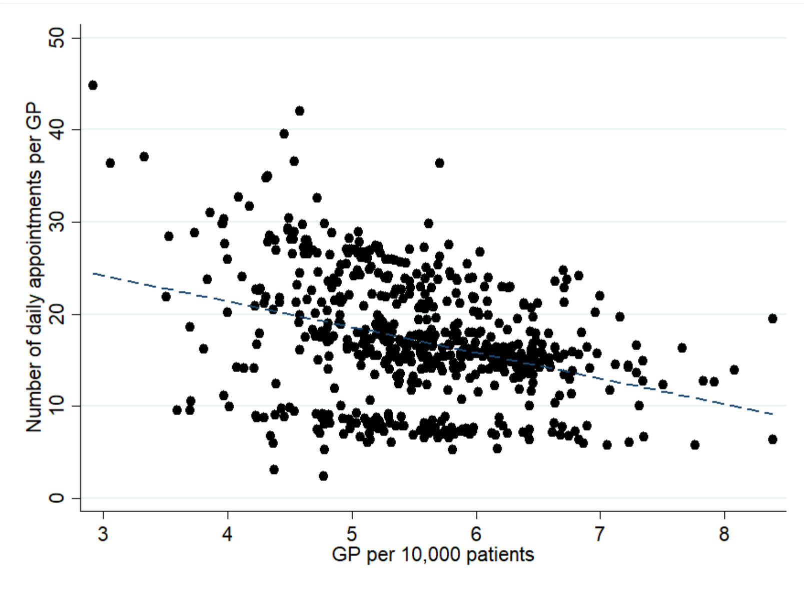


Figure A10: Flowchart of the labor supply of physicians in England

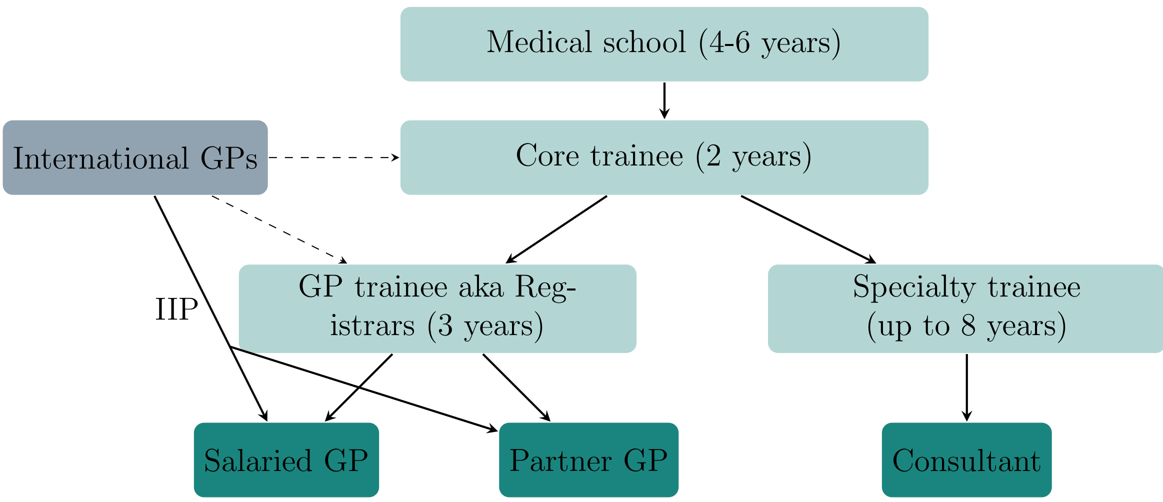


Figure A11: Number of patients per GP

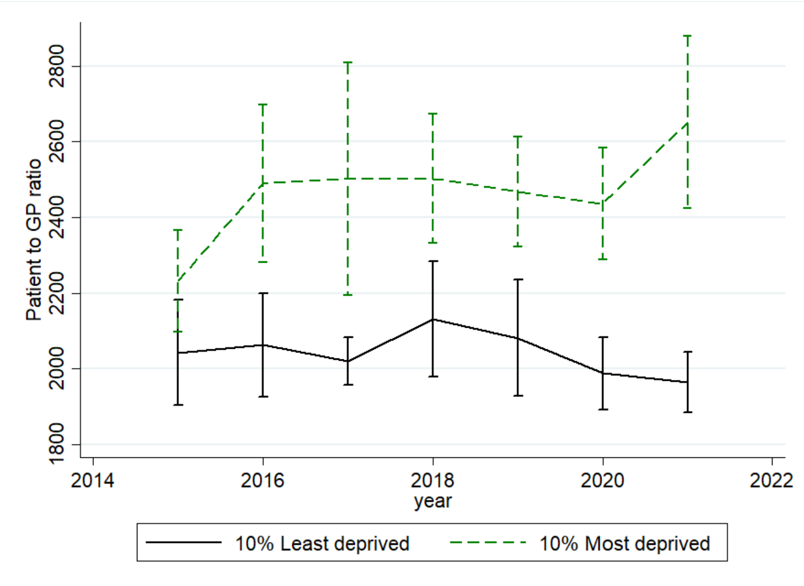


Figure A12: Share of wait times more than two weeks out of total annual appointments

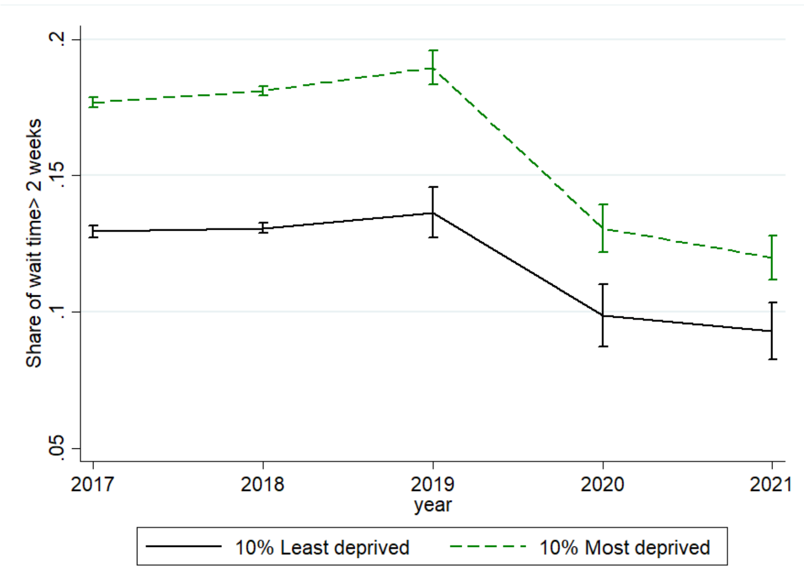


Figure A13: Share of patients with an overall good experience with the GP practice

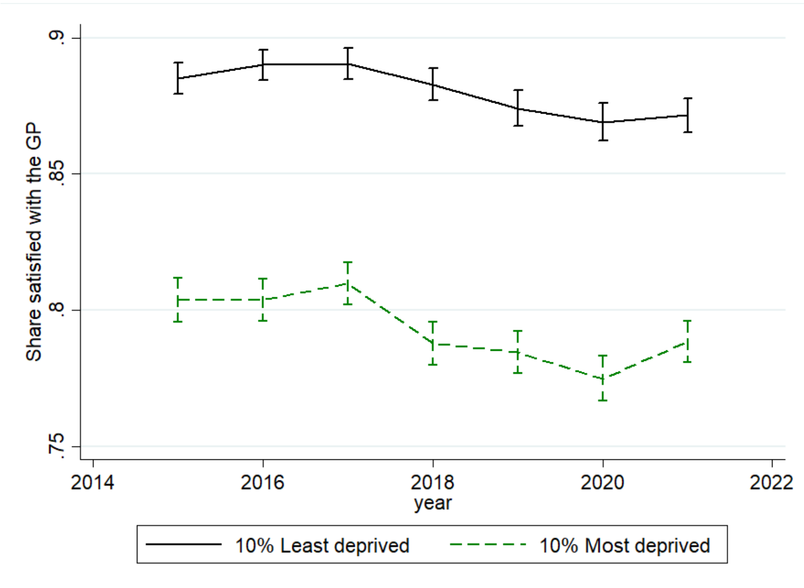
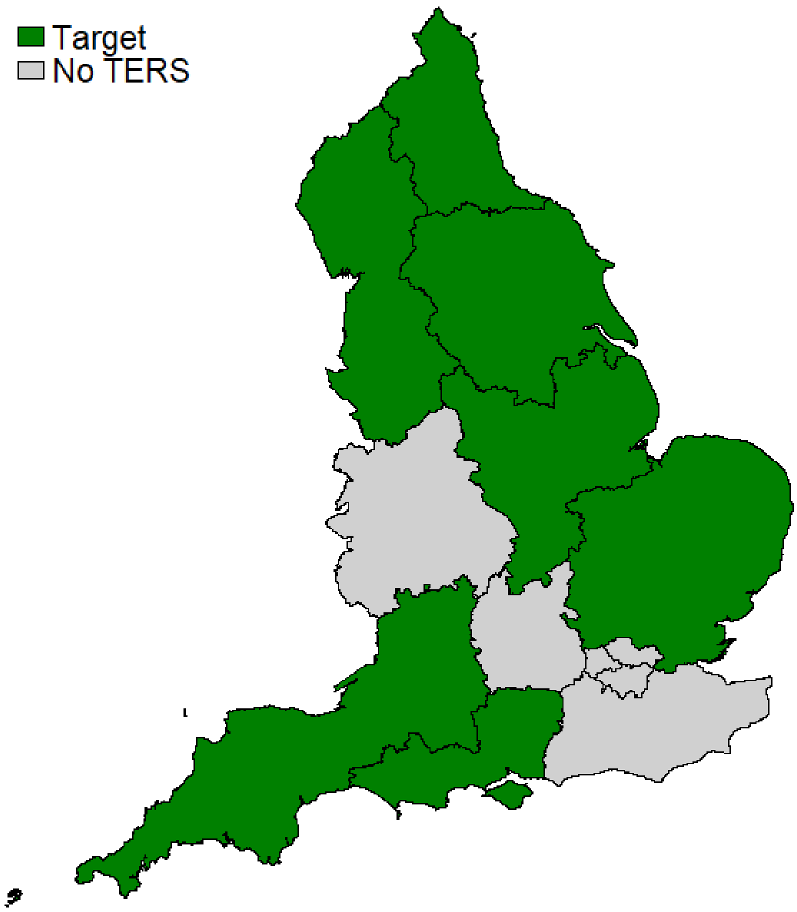


Figure A14: TERS target areas



Note: Target areas are: Health Education East Midlands, Health Education East of England, Health Education North East, Health Education North West, Health Education South West, Health Education Wessex, Health Education Yorkshire and the Humber. Control areas are: Health Education Kent, Surrey and Sussex, Health Education North Central and East London, Health Education North West London, Health Education South London, Health Education Thames Valley, Health Education West Midlands.

Figure A15: Change in the number of HC GPs per 10,000 populations

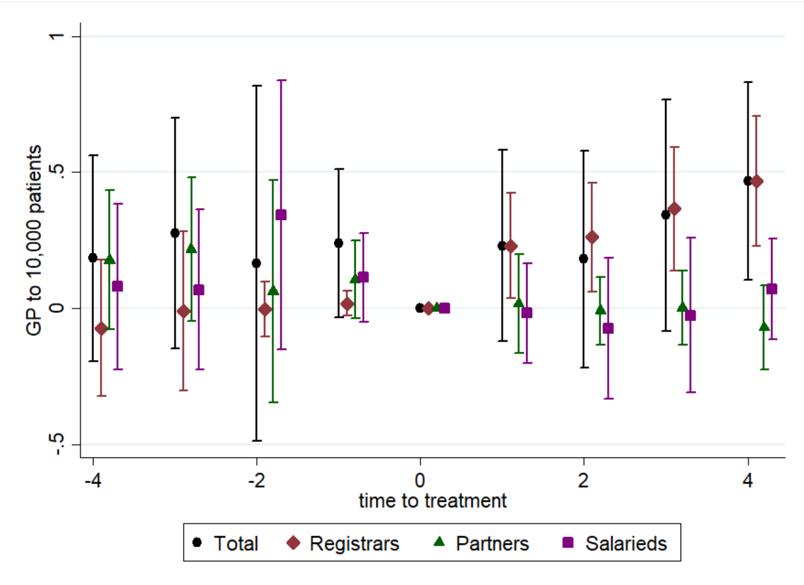


Figure A16: The effect of TERS on the number of Partners per 10,000 patients

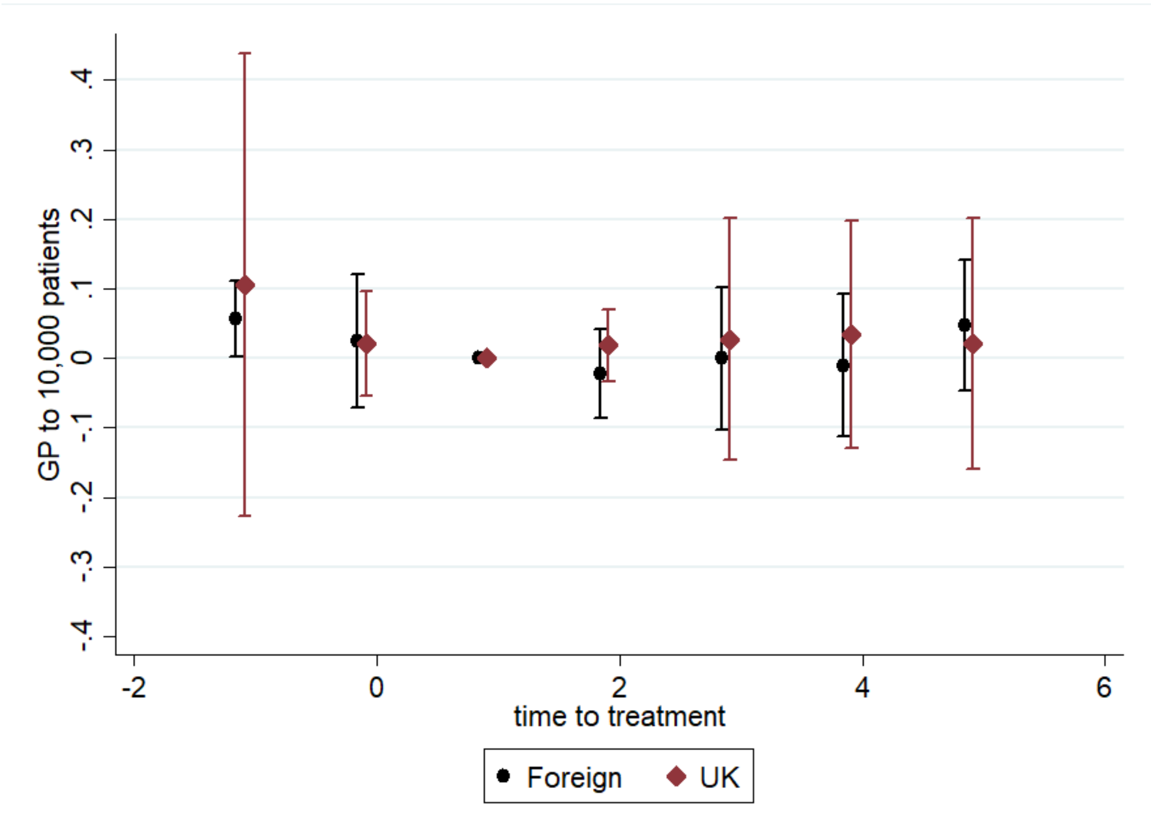


Figure A17: The effect of TERS policy on the share of registrars in the most deprived areas