The Association Between GP Entry with Healthcare Quality in England's NHS*

Click here for the latest version.

Theodore Caputi

March 23, 2020

Contents

*These results are preliminary.

List of Figures

List of Tables

1 Introduction

As patients' first point-of-contact with the healthcare system and the gatekeepers between patients and specialty care, General Practitioners (GPs) form the cornerstone of the England's National Health Service (NHS). As discussed in the Literature Review (Section **??**), improving quality in general practice – interventions to improve patient outcomes, system performance, and professional development – has been a major priority for the NHS for the past several decades (Batalden and Davidoff, 2007). Advocates, academics, and policy makers have designed and tested the effectiveness of several quality improvement interventions, including financial incentives, standard-setting, reminders, inspections, public reporting, educational programs, conferences, and feedback mechanisms (Oxman et al., 1995), and while many of these interventions have garnered evidence of effectiveness, experts agree that meaningfully improving quality in healthcare is still an elusive task (Braithwaite, 2018). It is reasonable, then, to consider new, innovative methods for healthcare quality improvement. One potentially understudied policy lever for quality improvement, based upon the behaviour change literature (Rubenstein et al., 2000; Mittman, Tonesk and Jacobson, 1992), is labor supply moves, i.e., using the movement of physicians around the healthcare system to improve the dissemination and uptake of best practices. A physician entering a new practice could influence the behavior of other physicians at that practice, either through sharing new information on best practices or by exerting peer influence (e.g., physicians may wish to match the practice styles of their proximate peers (Jong et al., 2006)). Past researchers have considered these "GP movers" in the context of improving healthcare capacity in underserved (particularly rural) areas (Dolea, Stormont and Braichet, 2010; Li et al., 2014; Yong et al., 2018), but virtually no studies have considered physician movement in terms of quality improvement. This chapter uses publicly available data on how GPs have moved from practice to practice to provide preliminary evidence that GP movement may be a meaningful determinant of healthcare quality. While this observational study cannot provide causal evidence of an effect, the results may encourage investment in future interventional studies to test the effect of moving GPs on healthcare quality.

1.1 Peer Effects and GP Movers

Over several decades, psychologists, sociologists, economists, and marketers have identified peer effects – the social influence that individuals place on their proximate peers – as one of the strongest forces for behavior change. For example, studies have demonstrated that peer effects influence adolescent's health behaviors (Gaviria and Raphael, 2001; Lundborg, 2006; Ali and Dwyer, 2011), college students' academic achievement (Sacerdote, 2001), the diffusion of technology (Bollinger and Gillingham, 2012), paternity leave participation (Dahl, Løken and Mogstad, 2014), consumer decisions (Moretti, 2011), financial decisions (Bursztyn et al., 2014), and labor productivity (Herbst and Mas, 2015). It is unsurprising, then, that many health services researchers have endorsed peer effects as a possible method for improving healthcare quality (Goodpastor and Montoya, 1996). These scholars theorised that social influence, operationalised in such forms as role models (Kenny, Mann and MacLeod, 2003), opinion leaders (Locock et al., 2001), socially central individuals (Meltzer et al., 2010), and peer feedback (Pronovost and Hudson, 2012), could be a powerful mechanism to encourage physicians to implement evidence-based policies (Phelps and Mooney, 1993).

Indeed, there is some empirical evidence that quality improvement interventions are more successful when they incorporate a component of social influence. For example, Ivers et al. (2012) systematically reviewed the literature on audit and feedback interventions (i.e., quality interventions where a physician is informed of their performance relative to a set standard or the average) and found that feedback and audit interventions are more than three time as as effective when the source of the feedback is a supervisor or colleague (i.e., a peer) (adjusted risk difference [ARD]: 16.50) as opposed to a standards review board or employer (ARD: 2.37) or study investigators (ARD: 5.04). Evidence from several qualitative studies have affirmed that the use of peers in quality improvement interventions increases physicians' receptiveness to change (Ferguson, Wakeling and Bowie, 2014). Aside from audit and feedback interventions, a Cochrane Rewiew conducted by Flodgren et al. (2019) found that, compared with no intervention, local opinion leader based interventions improved adjusted median compliance with evidence-based practice by 10.8%. Further, quality interventions involving opinion leaders improved adjusted median compliance with evidence-based practice by 7.1% to 13.7% relative to other types of quality improvement interventions. Even medical education appears to be more effective when interacted with an element of social influence. An early study found that a quality improvement intervention which paired educational outreach visits with feedback had a greater effect on physician's future performance when the outreach was conducted by a peer physician rather than a trained practice assistant (Van den Hombergh et al., 1999).

Physician moves present another potential quality improvement intervention involving social influence, but to my knowledge, no studies have considered the impact of physician movement on quality improvement. Only a few studies have considered the clinical relevance of entering or exiting physician movers, at all. One study of primary physicians in Norway found that a physician's "practice style" was stable even after he or she moved to a different practice in a different location, suggesting that physicians' preferences play a meaningful role in medical practice variation apart from differences in clientelle (Grytten and Sørensen, 2003). This same method was used by Epstein and Nicholson (2009), who investigated whether individual obstetricians changed their rate of Caesarean-section (C-section) procedures in response to moving to a new peer group. While the authors found some evidence of peer influence, the effect size was exceptionally small; a 2.4 percentage point increase in a physician's peer group's C-section rate driven by moving physicians was associated with a 0.16 percentage point change in the physician's C-section rate. A similar study of cardiologists who moved between two geographic regions in the United States Medicare system found that these physicians adapted to physician behavior in their target region; indeed, the cardiologists adapted their practice, on average, by 0.6 to 0.8 percentage points for every percent point difference between their origin and target region (Molitor, 2018).

1.2 Possible Interventions

There are several potential mover-based interventions that could be used to first test and eventually leverage physician peer effects on healthcare quality. Past studies have implemented interventions incentivising physicians to move from urban to rural settings. These same interventions may be extended to attract physicians from areas with high quality social norms to areas with low quality social norms. Similarly, rotational programs that have traditionally been used to expose training physicians to different areas of medicine (Bell-Dzide, Gokula and Gaspar, 2014) could be modified to expose low-performing practices to physicians from high-performing practices with the intention of improving quality. From the individual practice's perspective, hiring managers may consider the potential peer-based impact that a new physician may have on a practice. Such a hiring practice may wish to find ways to intensify the social influence of a physician moving from practices with higher quality social norms (e.g., ensuring that the new physician interacts with the practice's existing physicians) or mitigate the social influence of a physician moving from a practice with lower quality social norms (e.g., increasing training for the new physician).

2 Aims

The purpose of this study is to assess whether preliminary evidence exists in support of a GP "mover effect", i.e., that GPs who move from practice to practice impact healthcare quality. In addition, this study aims to describe how GPs have moved within England's NHS.

3 Methods

3.1 Setting

The Quality and Outcomes (QOF) scheme, introduced in 2004, is England NHS's primary strategy for improving quality in general practice. Essentially, the QOF scheme functions by setting a series of performance indicators (generally process indicators) based upon the literature for best practices and compensating GP practices based upon their performance on those metrics. For example, the QOF scheme may include a target for the percentage of individuals with heart failure who are prescribed statins, and GP practices would be compensated for how close their performance lies to this target. The over 100 performance indicators in the QOF scheme span several dimensions of healthcare delivery (e.g., patient experience, organisational, clinical, and additional services), with most indicators tied to what is considered best practice for a given diagnosis or circumstance. More information on the implementation of the QOF scheme is available in the Literature Review.

3.2 Data

3.2.1 NHS Quality and Outcomes Framework

The NHS provides a set of publicly available datasets[1] that describe each GP's achievement in terms of the Quality and Outcomes Framework for each year between 2004/2005 and 2012/2013. The data aggregates performance from April 1 of year t to March 31 of year $t+1$, and so while there is overlap in the calendar years of the individual datasets, they refer to strictly different periods. For the datasets between 2006/2007 and 2011/2012, overall achievement scores are reported for five domains: Total, Patient Experience, Clinical, Organisational, and Additional Services.

I combine these datasets so that I have panel data of Total, Patient Experience, Clinical, Organisational, and Additional Services quality for each GP practice for each period. I standardize each quality variable to the maximum for each year and domain and then scale the quality variables to 100. Consequently, each quality variable is comparable across years. These data also provide the listsize for each practice in each year.

3.2.2 GP-by-Practice Tenure Data

The NHS also provides a publicly available dataset[2] that describes each GP's tenure with each practice they've worked at, updated quarterly. Each row of this dataset contains an identifier for the GP, an identifier for the Practice, the date that the GP started working at that practice, and the date the GP stopped working at that practice (if applicable).

	DocID	PRACTICE	parent_org_type	join_date	left_date	amended
1	G8608145	L83647	P	19740401	20160331	1
2	G8435051	C84033	P	20110801	20150211	0
3	G7124307	B81009	P	20140801		0
4	G9043488	A85005	P	20151101		0
5	G9510238	M84037	P	19950904		0
6	G8506841	M85062	P	19740401	19910630	1
7	G9041118	D82012	P	20120601		0
8	G3181182	A86012	P	19740401	19990110	0
9	G8935843	E81065	P	20090401	20120331	0
10	G8509648	A83029	P	19850811	20160222	1

[1] https://digital.nhs.uk/data-and-information/publications/statistical/quality-and-outcomes-framework-achievement-data

[2] https://digital.nhs.uk/services/organisation-data-service/data-downloads/gp-and-gp-practice-related-data

From this dataset, it is possible to construct a dataset of GP moves. First, I filter the dataset by only taking those GPs with more than two rows – these are GPs that have moved practices at least once in their careers. For each GP, I order the observations such that the GP's earliest appointment is first and their latest appointment is last. When a GP has moved more than once, i.e., they have three or more rows in the dataset, I duplicate all rows that are neither their first nor last appointment. This creates pairs of observations in the dataset, where the first row is where the GP started and the second is where the GP went. I assign each pair a unique identifier and then transform the data to describe each individual move, with data related to quality at the first practice ($p = 1$) and the second practice ($p = 2$). This includes variables describing when the GP entered her first practice, when she exited her first practice, when she entered her second practice , and when she exited her second practice (if applicable).

I focus on movers who move within a single "April to March" time period. Because several moves occurred right at the March 31 to April 1 cutoff, I extended each time period by 5 days on either side. I then test whether the GP's date of leaving her first practice and the GP's date of entering her second practice both fall within each time period between April 1, 2006 and March 31, 2012. If so, I treat that period as the period of the move. I exclude moves that occur where the GP left her first practice in one period and entered her second practice in another period.

Because practices with several moves may differ fundamentally from those with zero or one moves (e.g., high turnover) and because discerning an effect could be challenging when the treatment occurs more than once, I focus on practices that had only one GP move within my study period and exclude those practices with more than one entrant from further analysis.

Additionally, I use this dataset to calculate how many GPs work at any given practice on the first date of every month. From this, I calculate the average number of GPs at each practice for each one-year period.

3.2.3 Merging the Datasets

I first merge the quality panel data onto the movers data by the time period of the move and the mover's first practice. From this, I extract the quality metrics for the practice that each mover left during the period that they left. I will use this data to determine whether the mover came from a better (worse) practice in each quality domain.

I then restrict the quality panel dataset to just those practices that had zero or one mover within the time period. I merge the movers dataset back onto the quality panel dataset by the mover's second practice. This panel dataset now contains, for each practice, the quality metrics for the mover's old practice and the date that the GP moved.

I finally merge the panel dataset with the number of GPs for each period. I divide the number of GPs in the practice by the practice's listsize to assess the capacity of the providers in terms of physicians per 1000 patients.

3.2.4 Constructed Variables

Using this merged dataset, I construct several variables pertaining to the second GP practice ($p = 2$) relative to the time period of the move (t_0) that will be useful for the analysis and data visualization.

- "time_to_movement$_{p=2,\, t}$" refers to the difference in time between the current period t and the period where the mover entered t_0. For those practices with no movers, t_0 takes on a random value from 1 to 5.

$$\text{time\_to\_movement}_t = \begin{cases} t - t_0 & \text{Practices with Movers} \\ t - \text{randint}(1,5) & \text{Practice had no movers} \end{cases}$$

- "post$_t$" refers to the time that the mover influenced the practice. Following past work (**?**), it is defined such that it can take on values between 0 and 1 in time periods $t \geq t_0$ for practices that had movers. If the mover moves in midway through the period, post will equal the fraction of days the mover spent in that practice in that period after she arrived. Similarly, if the mover leaves the practice, post will equal the fraction of days the mover spent in that practice in that period before she left. In order to use the groups with no movers as a control, I create random interruptions.

$$\text{post}_t = \begin{cases} 0 & t < t_0 \text{ and Practice had Mover} \\ \frac{\text{Days at Incoming Practice}}{\text{Days in Period} t} & t \geq t_0 \text{ and Practice had Mover} \\ 0 & t < \text{randint(1,5) and Practice had No Movers} \\ 1 & t \geq \text{randint(1,5) and Practice had No Movers} \end{cases}$$

- For each domain d, "move_d" takes on values depending upon whether the mover came from practice with better, worse, or the same quality scores during the time period of the move. The "move" variable is constant within practices (e.g., it doesn't change over time).

$$\text{move}_d = \begin{cases} \text{better} & \text{quality}_{d,\ t_0,\ p=1} > \text{quality}_{d,\ t_0,\ p=2} \\ \text{worse} & \text{quality}_{d,\ t_0,\ p=1} < \text{quality}_{d,\ t_0,\ p=2} \\ \text{same} & \text{quality}_{d,\ t_0,\ p=1} = \text{quality}_{d,\ t_0,\ p=2} \\ \text{no move} & \text{Practice had no movers} \end{cases}$$

- For each domain d, "percent_change_d" is defined using the following formula:

$$\text{percent\_change}_d = \frac{\text{quality}_{d,\ t_0,\ p=1} - \text{quality}_{d,\ t_0,\ p=2}}{\text{quality}_{d,\ t_0,\ p=2}} * 100\%$$

  Note that when $\text{percent\_change}_d > 0$, the mover came from a practice with higher scores in domain d, and when $\text{percent\_change}_d < 0$, the mover came from a practice with lower scores in domain d.

3.3 Statistical Analysis

3.3.1 Descriptive Analysis

First, I will describe practice movement from the perspective of general practitioners in England's NHS. For each year, I describe the number of practicing doctors, as well as how many doctors enter, exit, or move by year. For all moves, I will describe moves by period, including the number of moves; the mean and median difference travelled; and the average length of physicians' stay at the origin practice, the destination practice, and between the two practices. I provide a visualisation of the moves across England and Wales. Then, I focus on the analytical sample (i.e., those GPs who moved between two practices with reported QOF scores and whose destination practice had only one entrant over the study period). For the analytical sample, I report the QOF scores of the movers' origin and destination practices by year of move.

3.3.2 Naiive Difference-in-Difference

Next, I attempt to provide some evidence that GP movers may affect quality. The basis for my approach is a difference-in-difference design. In a simple two-group, two-period difference-in-difference design, the treatment group's outcome value before the treatment and the trend of the control group are used to estimate a counterfactual. Formulaically, a simple difference-in-difference design is estimated with the regression in equation (1).

$$y_{i,t} = \beta_0 + \beta_1 * \text{Period}_t + \beta_2 * 1(\text{Treated})_i + \beta_2 * \text{Period}_t * 1(\text{Treated})_i + X\beta + \epsilon_{i,t} \tag{1}$$

In this equation, $\text{Period}_{i,t}$ takes the value 0 before the intervention and 1 after the intervention, $1(\text{Treated})_i$ takes the value 0 for all participants in the control group and 1 for all participants in the treatment group, and β is the parameter estimate for confounders in the matrix X.

I adjust this model to account for multiple periods before and after the intervention and to include three experimental groups (i.e., GPs who moved from a worse practice to a better practice, GPs who moved from a better practice to a worse practice, and GPs who moved between two practices with the same quality score). The regression model can be estimated with the formula in equation (2). For another example of this method in practice, see (**?**).

$$\text{y}_{it} = \beta_0 + \beta_1 * \text{Period}_t + \beta_{2-4} * \text{move}_i + \beta_{5-7} * \text{Period}_t * \text{move}_i + X\beta + \epsilon_{i,t} \tag{2}$$

3.3.3 Difference-in-Difference with Fixed Effects

A major concern of the Difference-in-Difference design is confounding. While known confounders can be adjusted for in the regression, there is a possibility for unknown confounders, i.e., unobserved differences between practices in the control and the experimental groups that cause the groups to have different trajectories after the intervention. To ameliorate this concern, I extend the naiive difference-in-difference design to include practice-level and period-level intercept fixed effects. The practice-level intercept fixed effect accounts for time invariant confounding and period-level intercept fixed effects account for practice invariant confounding.

Fixed intercept effects can be incorporated in the difference-in-difference model with the formula in equation (3).

$$\text{quality}_{ijt} = \beta_0 + \beta_1 * \text{Period}_t + \beta_{2-4} * \text{move}_{ij} + \beta_{5-7} * \text{Period}_t * \text{move}_{ij} + X\beta + \alpha_j + \lambda_t + \epsilon_{ijt} \quad (3)$$

In this model, α_j takes on a different value for each practice j, and λ_t takes on a different value for each period t. Because these fixed effects account for all time-invariant confounding and all practice-invariant confounding (respectively), the matrix of confounders X should only those confounders that vary within practices across time periods.

One potential confounder is capacity. Practices may (naturally) change the number of physicians working at the practice when hiring a new entrant, and this may vary among the treatment groups and with the outcome. For example, practices with a lower doctor-to-patient ratio may be willing to hire a broader range of physicians (including physicians from worse practices) than practices with a higher doctor-to-patient ratio, and these strained practices may also be less able to invest resources into improving quality. Therefore, I include capacity (in terms of GPs per 1,000 patients) as a confounder in the model.

3.3.4 Difference-in-Difference with Practice-Level Random Slope Effects

The interrupted time series with practice-level fixed intercept effects (equation (3)) assumes that, while the intercept for each practice can be different, the slope (i.e., the relationship between time and the outcome variable) must be parallel. It is possible, however, that individual practices have non-parallel trajectories in the outcomes.

To mitigate this assumption, I incorporate random slope effects into the model. While fixed effects set a different parameter estimate for each group, random effects use data across groups (i.e., partial pooling) to estimate a distribution for a parameter, thereby treating the parameter as a random variable in the model. Consequently, random effects can account for confounding more efficiently than fixed effects but add the assumption that the underlying parameter follows a normal distribution. In this case, random slope effects ensure that the model is adequately efficient to discern meaningful associations.

3.3.5 Difference-in-Difference with Fixed Effects (Movers Only)

While the difference-in-difference design with practice-level and time-level fixed intercept effects accounts for some forms of confounding, it is still possible that practices that accept new GPs are unobservably different from practices that do not accept new GPs. For example, individual practices decide when they need to hire an additional physician to address capacity constraints, and some of the factors for this decision may be unobservable. These differences between practices that hire a new GP and practices that do not hire a new GP may confound the model. Consequently, in the main analysis, I restrict the sample to only those practices that accepted an incoming GP. That is, I compare those practices that had a GP enter from a practice with better QOF scores only to those practices that had a GP enter from a practice with worse QOF scores (or practices that had a GP enter from a practice with the same QOF scores). In this way, only practices that decided they needed a new physician and successfully hired one are compared.

3.3.6 Percentage Change Difference-in-Difference

The main analysis assesses whether having a physician enter from a practice with better QOF scores, a practice with worse QOF scores, or a practice with the same QOF scores is associated with changes in healthcare quality. It is plausible that these relationships are determined not only by the direction of the difference between the physician's origin and destination practice but also by its magnitude. If having a physician enter from a practice with lower QOF scores causally decreases the QOF scores of the practice, then it is likely that having a physician move from a practice with much lower QOF scores will have a more detrimental effect than having a physician move from a practice with only somewhat lower QOF scores.

To assess this possibility, I estimate the relationship between the outcome (the destination practice's QOF score) and the percentage difference between the moving physician's origin and destination practice. This percentage difference is positive when the physician moves from a practice with higher QOF scores and negative when the physician moves from a practice with lower QOF scores.

$$\begin{aligned}\text{quality}_{ijt} = \beta_0 + \beta_1 * \text{Period}_t + \beta_2 * \text{PercentChange}_{ij} + \beta_3 * \text{Post}_t \\ + \beta_4 * \text{Post}_t * \text{PercentChange}_{ij} + X\beta + \\ + \alpha_j + \lambda_t + \epsilon_{ijt}\end{aligned} \tag{4}$$

3.4 Additional Analyses

3.4.1 Sensitivity Analysis: Random Assignment to Treatment Groups

It is possible that a significant finding from the Main Analysis could be due to an overpowered analysis rather than a truly significant association. Consequently, I run a sensitivity analysis in which practices are randomly assigned to treatment groups. If the results of this analysis are significant, it is possible that any result from the main analysis is a statistical artefact.

3.4.2 Naiive Interrupted Time Series Analysis

The difference-in-Difference design, like that in equation (5), can be adapted to decompose the overall effect of the mover into an immediate and gradual effect. This new model, termed an Interrupted Time Series Analysis, can be modelled with the following regression:

$$\begin{aligned}\text{quality}_{ijt} = \beta_0 + \beta_1 * \text{Period}_t + \beta_2 * \text{post}_t + \beta_{3-5} * \text{move}_{ij} \\ + \beta_6 * \text{Period}_t * \text{post}_t + \beta_{7-9} * \text{Period}_t * \text{move}_{ij} \\ + \beta_{10-12} * \text{post} * \text{move} + \beta_{13-15} * \text{Period} * \text{post} * \text{move} + X\beta \\ + \epsilon_{ijt}\end{aligned} \tag{5}$$

I extract the fitted values from this regression[3] and plot mean (expected) quality for each period before and after the GP movement.

[3] For simplicity of the plots, I treat "post" as an indicator variable equal to 0 if it is before the doctor moved in and 1 if it is after the doctor moved.

3.4.3 Interrupted Time Series Analysis with Fixed Effects

The naiive interrupted time series in equation (5) can be extended to include fixed effects, as in equation (6).

$$\begin{aligned} \text{quality}_{ijt} = & \beta_1 * \text{Period}_t + \beta_2 * \text{post}_t + \beta_{3-5} * \text{move}_{ij} \\ & + \beta_6 * \text{Period}_t * \text{post}_t + \beta_{7-9} * \text{Period}_t * \text{move}_{ij} \\ & + \beta_{10-12} * \text{post} * \text{move} + \beta_{13-15} * \text{Period} * \text{post} * \text{move} + X\beta \\ & + \alpha_j + \lambda_t + \epsilon_{ijt} \end{aligned} \tag{6}$$

In this model, β_2 corresponds to the "jump" occurring at the interruption for the control group and β_6 corresponds to the change in slope occuring after the intervention. The parameters of interest are β_{10-12}, which correspond to how the "jump" for the experimental groups relative to the control group (i.e., the immediate effect), and β_{13-14}, which corresponds to how the slope changes after the intervention relative to the control group (i.e., the gradual effect).

3.4.4 Interrupted Time Series Analysis with Fixed Effects (Movers Only)

Like the main analysis in section 3.3.5, it is possible that practices with movers are unobservably different from practices without movers. Consequently, I restrict the sample to only those practices with movers. This method compares only practices that had a mover from a practice with higher QOF scores to practices that had a mover from a practice with lower QOF scores.

4 Results

4.1 Descriptive Results

The number of practicing GPs has increased steadily from 2004 to 2013, increasing from under 35,000 to over 50,000 (Table 1). The number of GPs starting their career has also increased and consistently outpaces the number of GPs who retire. The number of GPs moving from one practice to another has also increased markedly.

There were a total of 8892 physician moves between 1 April 2004 and 30 March 2013 (Figure 1 and 2) (Table 2). Physician moves became increasingly common as time moved on, nearly doubling from 772 in 2004-2005 to 2012-2013. The distance of moves has a positive skew in each period; while the mean distance between a physicians' origin and destination practice is between 18.8 and 26.7 miles, the median distance is between 7.3 and 9.5 miles. Still, in all periods but one, over 20% of moves were over 20 miles and over 10% of moves were over 50 miles. In terms of career trajectory, the median mover had been a physician for between 1000-1500 days and spent less time at their origin practice than they ended up spending at their destination practice (Table 3).

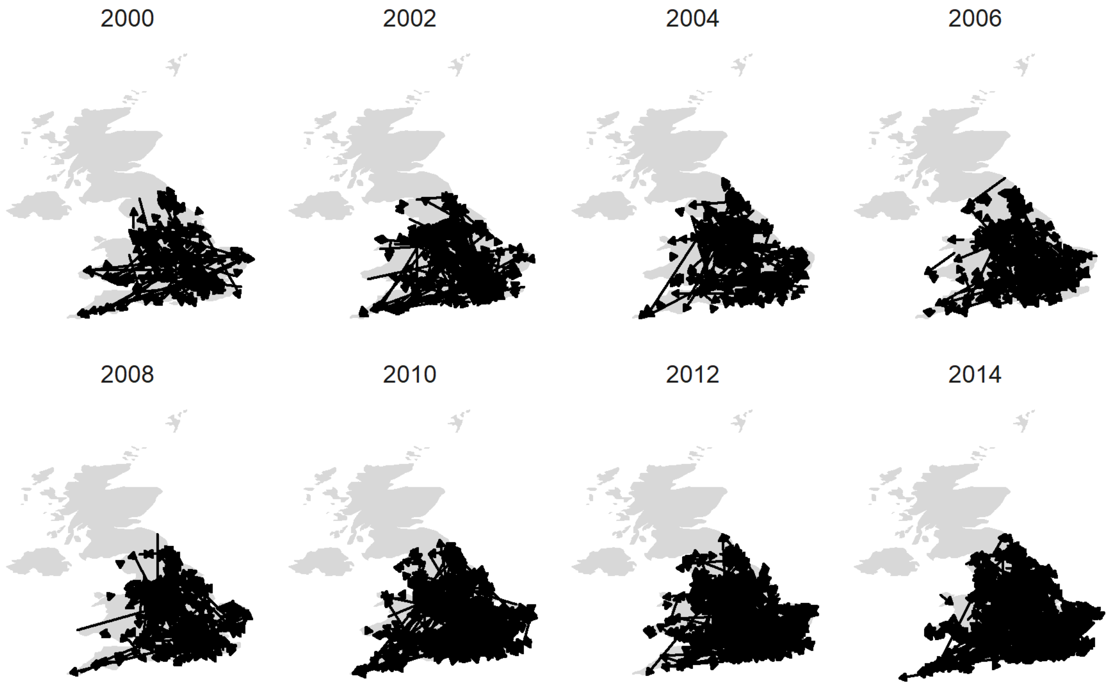
The analytical sample included 2147 physician moves (Table 4). Physicians were slightly more likely to move to practices with higher total, clinical, and patient experience QOF scores, while physicians did not show a clear leaning towards practices with higher or lower organisational or additional services QOF scores. The median percentage difference between the QOF scores at the physicians origin and destination practice is negligible for nearly all periods and QOF domains. Overall, there is little evidence that physicians are systematically moving to practices with higher or lower QOF scores.

Table 1: Overview of GP Starts, Retirements, and Moves

Year	No. Practicing GPs	No. GP Starts	No. GP Retirements	No. GP Movers
2004	34,787	2,374	1,060	733
2005	35,994	2,364	999	721
2006	38,222	3,075	1,282	796
2007	40,013	2,997	1,352	771
2008	41,608	2,878	1,433	830
2009	43,611	3,392	1,657	1,109
2010	45,437	3,409	2,002	1,362
2011	46,681	3,242	2,112	1,424
2012	47,831	3,164	2,417	1,543
2013	50,128	4,528	3,087	1,808

NOTE: This table presents an overview of the starts, retirements, and moves of GPs. "No. GP Starts" refers to the number of GPs who began their careers in a given calendar year. "No. GP Retirements" refers to the number of GPs who stopped practicing until at least 2016 in a given calendar year. (Note that a doctor who stopped practicing in 2012 but continued practicing in 2018 would be considered "retired" in 2012 by this definition). "No. GP Moves" refers to the number of physicians who, within a given calendar year, left at least one practice and joined at least one practice.

Figure 1: GP Moves by Year



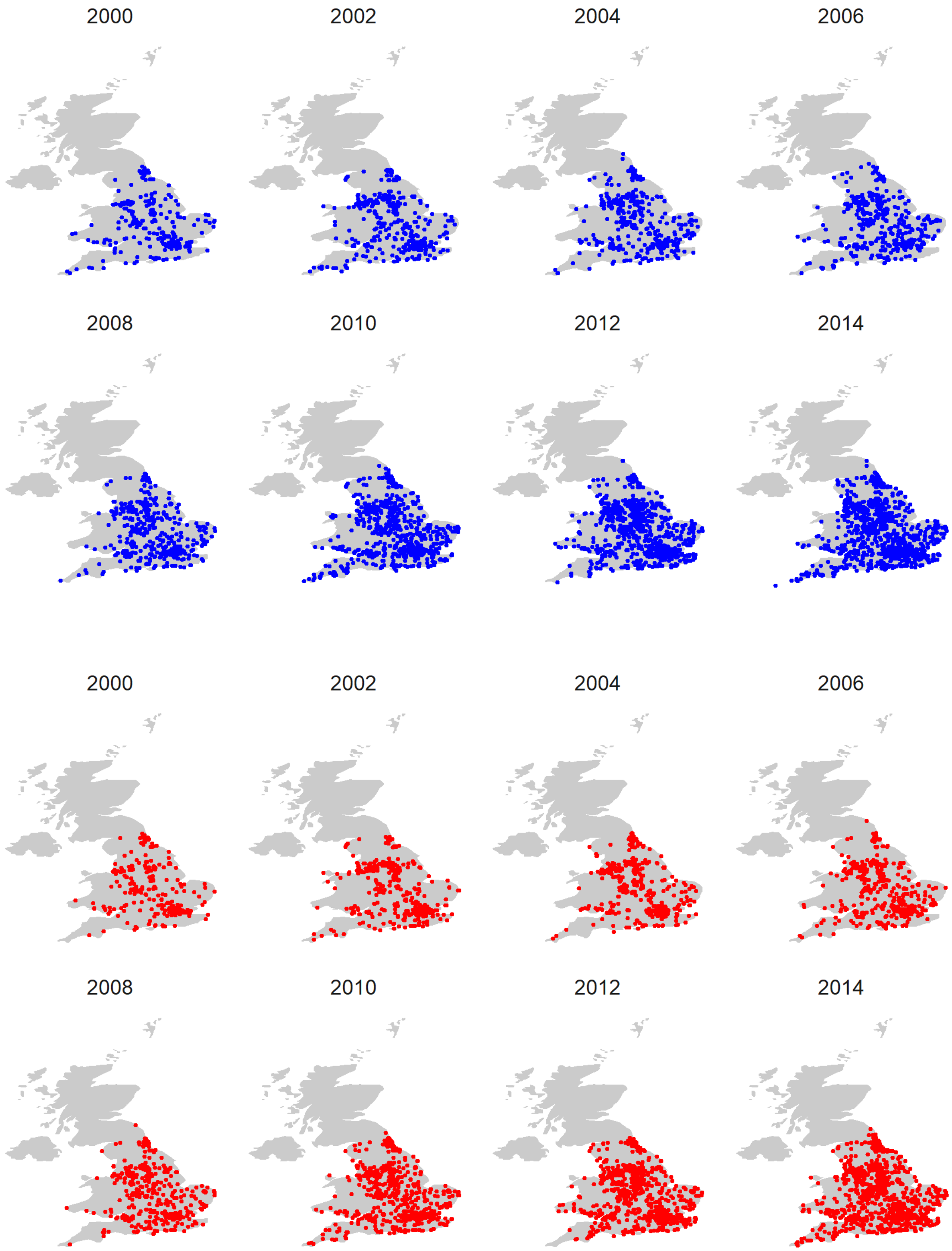
This figure shows GP moves by year from 2000 to 2014. The arrows point in the direction of the mover. GPs who move from or to a postcode that cannot be geocoded are omitted.

Table 2: Distance of GP Moves in Miles

Period	N	Mean Distance	Median Distance (IQR)	Over 20 Miles (%)	Over 50 Miles (%)
2004-2005	772	23.6	7.6 (2.8 to 17.9)	22.9	14.1
2005-2006	665	23.4	8 (2.9 to 21.3)	26.2	14.2
2006-2007	721	24.2	8.6 (3.3 to 20.4)	25.4	15.3
2007-2008	678	26.7	9.3 (3.2 to 21.5)	27.1	14.6
2008-2009	819	25	9.1 (3 to 21.8)	27.3	14.1
2009-2010	1125	23.4	9.5 (3.4 to 20.9)	27	12.4
2010-2011	1254	23.1	8.9 (3.4 to 20.2)	25.2	11.5
2011-2012	1375	21.2	8.5 (3.2 to 18.5)	23.2	10.5
2012-2013	1483	18.8	7.3 (2.7 to 16.1)	19	8.9

NOTE: This table presents data on the distance of moves for each period. Distance was calculated using the latitude and longitude of the GP's origin and destination practice. "N" is the number of moves occurring in that period. "Mean Distance" is the average number of miles for each move within that period. "Median Distance" is the median number of miles for each move within that period, as well as the interquartile range. "Over 20 Miles" and "Over 50 Miles" are the proportion of moves occurring within each period that exceeded 20 and 50 miles, respectively.

Figure 2: GP Moves by Year



This figure shows GP moves by year from 2000 to 2014. Panel A shows where GPs have moved from using red points. Panel B shows where GPs have moved to using blue points. GPs who move from or to a postcode that cannot be geocoded are omitted.

Table 3: Summary of GP Moves

Period	N	Entry to Move (Median Days)	Time at Origin (Median Days)	Time at Destination (Median Days)
2004-2005	772	1806	1036	3220.5
2005-2006	665	1186	717	2918
2006-2007	721	1099	641	2454
2007-2008	678	992.5	668	2493
2008-2009	819	1005	716	2067
2009-2010	1125	1086	720	1816
2010-2011	1254	1279	777.5	1841
2011-2012	1375	1443	851	1552
2012-2013	1483	1513	762	2831

NOTE: This table presents data on the career trajectories of GP movers. "Entry to Move" is the median number of days between the date the physician became a registered NHS physician and the date they joined a new practice from an old practice. "Time at Origin" is the median number of days the mover spent at their original practice, and "Time at Destination" is the median number of days the mover spent at their destination practice.

Table 4: Summary of Moves in Terms of Domain-Specific QOF Scores

Period	To Better	To Same	To Worse	Mean Origin Score	Mean Destination Score	Median Pct. Difference (IQR)
Total						
2006-2007	176 (66.4%)		89 (33.6%)	95.6	95.9	0% (-2.2 to 2.11)
2007-2008	182 (75.8%)		58 (24.2%)	96.8	97.4	-0.1% (-1.4 to 1.39)
2008-2009	109 (39.8%)		165 (60.2%)	95	95.6	0.8% (-1.3 to 3.28)
2009-2010	128 (32.5%)		266 (67.5%)	93.8	93.3	0.3% (-3.5 to 3.38)
2010-2011	148 (31.8%)		317 (68.2%)	94.3	95.3	1% (-1.8 to 3.98)
2011-2012	346 (68%)		163 (32%)	97.1	97.3	0% (-1.4 to 1.66)
Clinical						
2006-2007	127 (47.9%)	11 (4.2%)	127 (47.9%)	96.5	96.7	0% (-1.6 to 1.35)
2007-2008	124 (51.7%)	13 (5.4%)	103 (42.9%)	97.3	97.8	0% (-1.7 to 0.66)
2008-2009	116 (42.3%)	20 (7.3%)	138 (50.4%)	98	98.3	0% (-0.7 to 0.83)
2009-2010	191 (48.5%)	8 (2%)	195 (49.5%)	96.6	95.8	0% (-2.4 to 2.25)
2010-2011	202 (43.4%)	7 (1.5%)	256 (55.1%)	96.8	97.5	0.2% (-1 to 2.54)
2011-2012	263 (51.7%)	10 (2%)	236 (46.4%)	97.1	97.3	-0.1% (-1.6 to 1.75)
Organisational						
2006-2007	130 (49.1%)	12 (4.5%)	123 (46.4%)	92.4	92.6	0% (-4 to 2.95)
2007-2008	110 (45.8%)	7 (2.9%)	123 (51.2%)	94.5	95.4	0.2% (-2.2 to 3.22)
2008-2009	129 (47.1%)	11 (4%)	134 (48.9%)	96.4	95	0% (-2.7 to 1.6)
2009-2010	196 (49.7%)	27 (6.9%)	171 (43.4%)	97.3	96.3	0% (-2.1 to 1.41)
2010-2011	202 (43.4%)	34 (7.3%)	229 (49.2%)	97.5	98	0% (-1.4 to 1.54)
2011-2012	224 (44%)	15 (2.9%)	270 (53%)	96.6	97	0.1% (-1.3 to 1.61)
Patient Experience						
2006-2007	225 (84.9%)	22 (8.3%)	18 (6.8%)	95.8	96.7	0% (0 to 0)
2007-2008	221 (92.1%)	10 (4.2%)	9 (3.8%)	97.5	98.6	0% (0 to 0)
2008-2009	108 (39.4%)	5 (1.8%)	161 (58.8%)	79.5	84.3	6.6% (-4.2 to 22.53)
2009-2010	108 (27.4%)	4 (1%)	282 (71.6%)	65.4	67.4	5.5% (-23.5 to 40.61)
2010-2011	131 (28.2%)	1 (0.2%)	333 (71.6%)	67.2	71.9	3.3% (-19 to 44.2)
2011-2012	463 (91%)	43 (8.4%)	3 (0.6%)	99.4	99.4	0% (0 to 0)
Additional Services						
2006-2007	60 (22.6%)	145 (54.7%)	60 (22.6%)	97	97.2	0% (0 to 0)
2007-2008	49 (20.4%)	131 (54.6%)	60 (25%)	97.2	98.1	0% (0 to 0.01)
2008-2009	47 (17.2%)	173 (63.1%)	54 (19.7%)	98.4	97.4	0% (0 to 0)
2009-2010	179 (45.4%)	66 (16.8%)	149 (37.8%)	96.5	95.6	0% (-3.8 to 2.92)
2010-2011	150 (32.3%)	151 (32.5%)	164 (35.3%)	97.8	97.8	0% (-1 to 1.42)
2011-2012	175 (34.4%)	162 (31.8%)	172 (33.8%)	97.7	97.6	0% (-1.3 to 1.22)

NOTE: This table summarises the moves that are included in the analytical sample, i.e., moves into practices that had only one entrant between April 2006 and March 2012. “Period” is the April 1 to March 30 year. “To Better” are those GPs who moved from a practice with lower scores to a practice with higher scores. “To Worse” are those GPs who moved from a practice with a higher score to a practice with a lower score. “To Same” are those GPs who moved from a practice with the same score as the practice they moved from. “Mean Origin Score” is the average domain-specific QOF score for the GP mover’s original practice. “Mean Destination Score” is the average domain-specific QOF score for the GP mover’s destination practice. “Mean Difference” is the difference between the mover’s original practice QOF score and the mover’s destination practice QOF score.

4.2 Naiive Difference-in-Difference

The naiive difference-in-difference analysis does not show a clear association between whether a practice had a GP move from a better or worse practice (Table 5). However, the assumption that the control and treatment groups are relatively similar at baseline is violated, and the naiive difference-in-difference fails to account for several potential forms of confounding. Therefore, the results of the naiive difference-in-difference are inconclusive.

Table 5: Naiive Difference-in-Difference (No Fixed Effects)

	Dependent variable:				
	Total	Patient Experience	Clinical	Organizational	Additional Services
	(1)	(2)	(3)	(4)	(5)
Common Intercept	95.216***	88.663***	96.561***	94.361***	96.140***
	(95.127, 95.306)	(88.378, 88.949)	(96.474, 96.647)	(94.228, 94.494)	(96.007, 96.274)
Δ Intercept for Practices with Incoming GP from Same		11.363***	2.618***	3.066***	2.906***
		(9.093, 13.632)	(1.815, 3.420)	(2.150, 3.981)	(2.532, 3.281)
Δ Intercept for Practices with Incoming GP from Worse	2.198***	1.597***	1.907***	2.665***	2.552***
	(1.986, 2.410)	(0.833, 2.360)	(1.700, 2.115)	(2.347, 2.983)	(2.168, 2.936)
Δ Intercept for Practices with Incoming GP from Better	−0.808***	−3.102***	−0.572***	−0.704***	−1.340***
	(−1.025, −0.592)	(−3.748, −2.456)	(−0.785, −0.359)	(−1.038, −0.371)	(−1.723, −0.958)
Any Move	−0.042	−4.026***	−0.059	1.753***	−0.018
	(−0.181, 0.097)	(−4.471, −3.581)	(−0.194, 0.076)	(1.546, 1.960)	(−0.226, 0.190)
Effect of Move of GP from Same		4.032*	−1.156*	−0.176	−0.053
		(−0.221, 8.286)	(−2.419, 0.107)	(−1.898, 1.546)	(−0.650, 0.543)
Effect of Move of GP from Worse	−0.193	−0.133	0.212	−0.493*	−0.096
	(−0.567, 0.180)	(−1.480, 1.214)	(−0.146, 0.571)	(−1.047, 0.062)	(−0.797, 0.604)
Effect of Move of GP from Better	1.098***	1.635***	0.741***	0.416	−0.284
	(0.731, 1.465)	(0.538, 2.733)	(0.371, 1.111)	(−0.151, 0.984)	(−1.006, 0.438)
Practice FE	NO	NO	NO	NO	NO
Year FE	NO	NO	NO	NO	NO
Practice RE	NO	NO	NO	NO	NO
Observations	43,193	43,193	43,193	43,193	43,193

NOTE: All regressions are adjusted for confounding by the doctor-to-patient ratio.

4.3 Difference-in-Difference with Fixed Effects

Adding practice-level and time-level fixed intercept effects to the analysis accounts for several potential sources of confounding. According to this model, having a GP move from a worse practice significantly decreases patient experience and organisational QOF scores but significantly increases clinical QOF scores relative to practices with no moves (Table 6). Further, having a GP move from a better practice significantly increases total, clinical, organisational, and additional services QOF scores but significantly decreases patient experience QOF scores relative to practices with no moves.

Table 6: Difference-in-Difference with Practice and Time Fixed Intercept Effects

	Dependent variable:				
	Total	Patient Experience	Clinical	Organizational	Additional Services
	(1)	(2)	(3)	(4)	(5)
Any Move	0.136^{**}	0.558^{**}	0.031	0.199	0.139
	(0.0002, 0.271)	(0.054, 1.063)	(−0.095, 0.157)	(−0.038, 0.436)	(−0.095, 0.374)
Effect of Move of GP from Same		1.976	0.055	−0.114	−0.121
		(−2.741, 6.693)	(−0.902, 1.012)	(−1.623, 1.394)	(−0.674, 0.432)
Effect of Move of GP from Worse	−0.060	-1.103^{**}	0.407^{***}	-0.558^{**}	0.029
	(−0.331, 0.212)	(−2.170, −0.035)	(0.144, 0.669)	(−1.064, −0.052)	(−0.583, 0.642)
Effect of Move of GP from Better	1.096^{***}	-1.853^{***}	1.078^{***}	1.612^{***}	0.741^{**}
	(0.791, 1.401)	(−2.943, −0.763)	(0.800, 1.356)	(1.093, 2.131)	(0.117, 1.364)
Practice FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES
Practice RE	NO	NO	NO	NO	NO
Observations	43,193	43,193	43,193	43,193	43,193

NOTE: All regressions are adjusted for confounding by the doctor-to-patient ratio.

4.4 Difference-in-Difference with Practice-Level Random Slope Effects

The model with practice-level random slope effects shows having an GP enter from a better practice is correlated with a significant increase in total, clinical, organisational, and additional services QOF scores (Table 7). Having a GP enter from a worse practice is correlated with a significant decrease in organisational QOF scores. Having a GP enter from a worse or better practice is associated with a significant decline in patient experience QOF scores.

Table 7: Difference-in-Difference with Practice Random Slope Effects

	Dependent variable:				
	Total	Patient Experience	Clinical	Organizational	Additional Services
	(1)	(2)	(3)	(4)	(5)
Δ Intercept of Practices with Incoming GP from Same		11.931^{***}	2.566^{***}	3.096^{***}	3.242^{***}
		(9.398, 14.465)	(1.137, 3.995)	(1.679, 4.513)	(2.632, 3.851)
Δ Intercept of Practices with Incoming GP from Worse	2.376^{***}	2.616^{***}	2.310^{***}	2.501^{***}	2.777^{***}
	(1.992, 2.761)	(1.808, 3.424)	(1.919, 2.701)	(2.006, 2.997)	(2.127, 3.427)
Δ Intercept of Practices with Incoming GP from Better	-0.556^{***}	-1.363^{***}	-0.651^{***}	-0.782^{***}	-1.755^{***}
	(−0.937, −0.175)	(−2.067, −0.659)	(−1.049, −0.254)	(−1.298, −0.266)	(−2.400, −1.109)
Any Move	0.131^{**}	0.375	0.057	0.209^{*}	0.075
	(0.002, 0.260)	(−0.082, 0.832)	(−0.063, 0.178)	(−0.002, 0.420)	(−0.143, 0.294)
Effect of Move of GP from Same		0.732	−0.208	−0.281	−0.296
		(−3.261, 4.724)	(−1.249, 0.833)	(−1.839, 1.278)	(−0.859, 0.268)
Effect of Move of GP from Worse	−0.100	-0.970^{*}	0.197	-0.525^{**}	−0.167
	(−0.403, 0.203)	(−2.034, 0.094)	(−0.095, 0.488)	(−1.041, −0.009)	(−0.805, 0.471)
Effect of Move of GP from Better	0.856^{***}	-1.116^{**}	1.116^{***}	1.171^{***}	0.666^{**}
	(0.536, 1.176)	(−2.100, −0.133)	(0.811, 1.421)	(0.639, 1.703)	(0.020, 1.313)
Practice FE	NO	NO	NO	NO	NO
Year FE	YES	YES	YES	YES	YES
Practice Intercept RE	YES	YES	YES	YES	YES
Practice Slope RE	YES	YES	YES	YES	YES
Observations	43,193	43,193	43,193	43,193	43,193
Akaike Inf. Crit.	253,368.600	359,030.200	247,656.500	295,798.800	298,463.500
Bayesian Inf. Crit.	253,507.400	359,186.400	247,812.600	295,954.900	298,619.600

NOTE: All regressions are adjusted for confounding by the doctor-to-patient ratio.

4.5 Difference-in-Difference with Fixed Effects (Movers Only)

When restricting the sample to only those practices with movers, the association becomes more clear. Relative to those practices that had a GP move from a worse practice, practices that had a GP enter from a better practice saw increases in total, clinical, organisational, and additional services QOF scores (Table 8).

Table 8: Difference-in-Difference with Fixed Intercept Effects, Only Movers

	Dependent variable:				
	Total	Patient Experience	Clinical	Organizational	Additional Services
	(1)	(2)	(3)	(4)	(5)
Effect for Move of GP from Worse	0.032	0.221	0.365^{***}	-0.410^{*}	0.155
	$(-0.235, 0.299)$	$(-0.776, 1.218)$	$(0.119, 0.611)$	$(-0.880, 0.060)$	$(-0.234, 0.545)$
Δ Effect for Move of GP from Better	1.195^{***}	-1.009	0.676^{***}	2.131^{***}	0.771^{***}
	$(0.855, 1.535)$	$(-2.240, 0.221)$	$(0.366, 0.986)$	$(1.543, 2.720)$	$(0.189, 1.353)$
Practice FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES
Practice RE	NO	NO	NO	NO	NO
Observations	12,521	12,272	12,521	12,521	12,521

NOTE: All regressions are adjusted for confounding by the doctor-to-patient ratio.

4.6 Sensitivity Analysis: Treatment Groups Randomly Assigned

In contrast, when the treatment groups are randomly assigned, there are few observed significant effects (Table 9). The results of this sensitivity analysis suggests that a meaningful association exists and that observed correlations are not statistical artefacts.

Table 9: Sensitivity Analysis: Treatment Groups Randomly Assigned

	Dependent variable:				
	Total	Patient Experience	Clinical	Organizational	Additional Services
	(1)	(2)	(3)	(4)	(5)
Any Move	0.249**	0.489	0.010	0.361**	0.235
	(0.047, 0.451)	(−0.273, 1.252)	(−0.178, 0.198)	(0.001, 0.721)	(−0.122, 0.592)
Effect of Move of GP from Same	0.038	0.181	0.225*	0.137	−0.053
	(−0.224, 0.300)	(−0.792, 1.153)	(−0.015, 0.466)	(−0.316, 0.591)	(−0.505, 0.398)
Effect of Move of GP from Worse	−0.035	−0.416	0.252**	−0.276	−0.157
	(−0.294, 0.223)	(−1.381, 0.549)	(0.011, 0.493)	(−0.736, 0.183)	(−0.608, 0.293)
Effect of Move of GP from Better	−0.032	−0.768	0.247**	−0.096	0.002
	(−0.289, 0.226)	(−1.742, 0.206)	(0.006, 0.488)	(−0.556, 0.363)	(−0.453, 0.458)
Practice FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES
Practice RE	NO	NO	NO	NO	NO
Observations	43,193	43,193	43,193	43,193	43,193

NOTE: All regressions are adjusted for confounding by the doctor-to-patient ratio.

4.7 Percentage Change Difference-in-Difference (Movers Only)

When accounting for the percentage change between the entrant's old and new practice, there is a significant positive correlation between the percentage change and the destination practices' total, clinical, organisational, and additional services QOF scores. These results suggest a sort of "dose-dependent" response – the greater the difference between the entering GP's origin practice and the destination practice, the greater the effect.

Table 10: Percentage Change Difference-in-Difference

	Dependent variable:				
	Total	Patient Experience	Clinical	Organizational	Additional Services
	(1)	(2)	(3)	(4)	(5)
Any Move * % Δ with GP's Last Practice	0.171^{***}	-0.017^{*}	0.055^{***}	0.178^{***}	0.001^{***}
	(0.147, 0.195)	(-0.038, 0.003)	(0.034, 0.076)	(0.159, 0.197)	(0.001, 0.002)
Practice FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES
Practice RE	NO	NO	NO	NO	NO
Observations	12,521	12,272	12,521	12,521	12,521

NOTE: All regressions are adjusted for confounding by the doctor-to-patient ratio.

4.8 Naiive Interrupted Time Series Analysis

The naiive interrupted time series analysis fail to show any clear pattern between having a physician enter from a practice with higher or lower QOF scores and practices' subsequent QOF scores (Table 11). Importantly, the parallel trends assumption (i.e., that the trajectory across control and experimental groups are parallel) appears violated for several QOF domains (Figures 3 and 4).

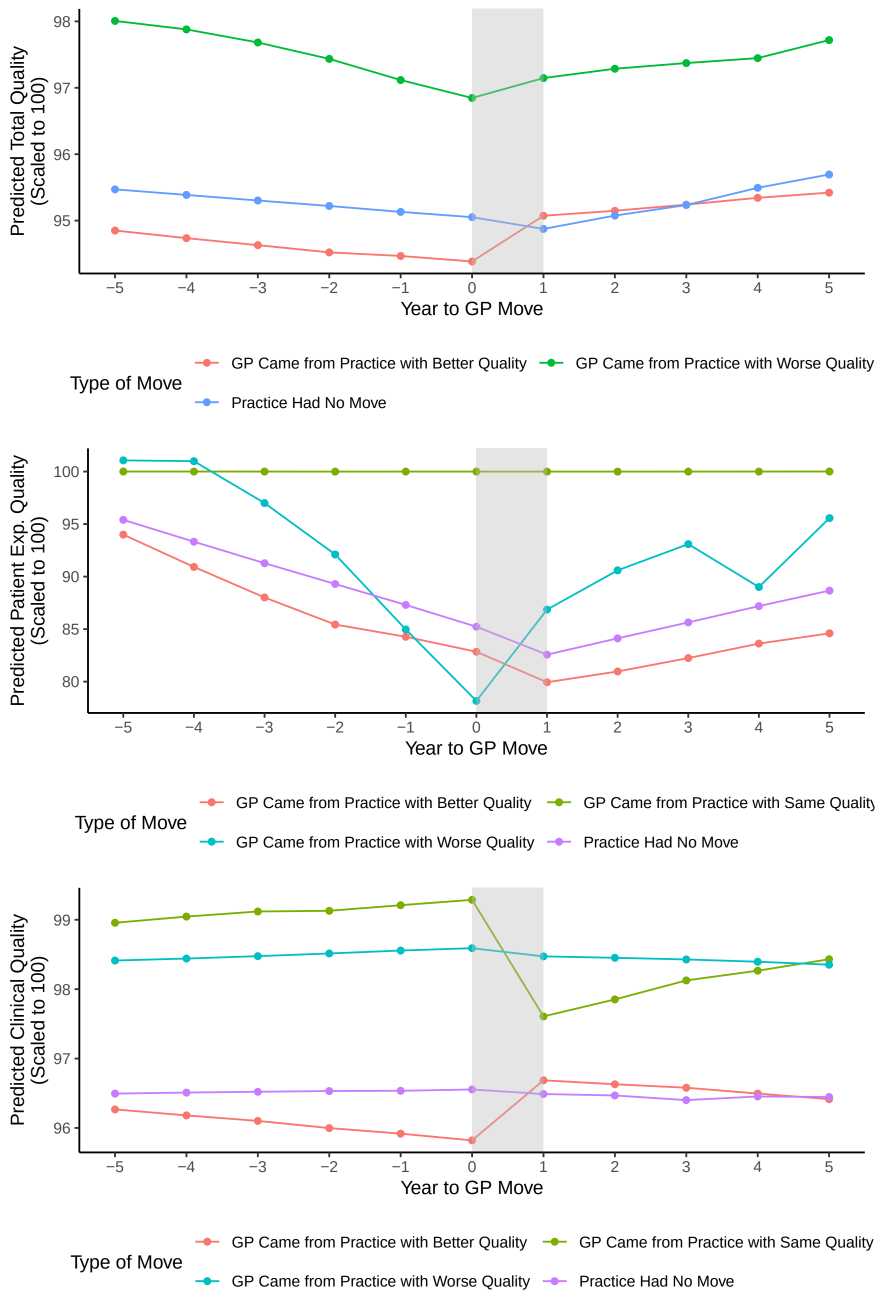
Table 11: Naiive Interrupted Time Series Analysis (No Fixed Effects)

	Dependent variable:				
	Total	Patient Experience	Clinical	Organizational	Additional Services
	(1)	(2)	(3)	(4)	(5)
Common Intercept	95.998***	107.762***	96.446***	89.542***	95.811***
	(95.706, 96.290)	(106.865, 108.658)	(96.162, 96.730)	(89.111, 89.974)	(95.374, 96.248)
Δ Intercept for Practices with Incoming GP from Same		−7.745**	3.454***	3.382**	3.534***
		(−14.830, −0.659)	(0.973, 5.935)	(0.351, 6.412)	(2.332, 4.737)
Δ Intercept for Practices with Incoming GP from Worse	3.251***	8.264***	1.944***	3.884***	3.374***
	(2.538, 3.964)	(5.711, 10.817)	(1.255, 2.633)	(2.842, 4.926)	(2.082, 4.666)
Δ Intercept for Practices with Incoming GP from Better	−0.611*	0.316	0.434	−1.079*	0.783
	(−1.314, 0.092)	(−1.702, 2.334)	(−0.267, 1.135)	(−2.169, 0.011)	(−0.497, 2.063)
Common Initial Slope	−3.504***	−43.075***	0.197	3.839***	−1.132**
	(−4.152, −2.855)	(−45.066, −41.084)	(−0.433, 0.827)	(2.881, 4.797)	(−2.102, −0.162)
Δ Initial Slope for Practices with Incoming GP from Same		43.086***	−5.343**	0.324	1.299
		(27.050, 59.121)	(−10.607, −0.080)	(−7.496, 8.144)	(−1.236, 3.834)
Δ Initial Slope for Practices with Incoming GP from Worse	−1.849**	−45.287***	0.651	1.177	0.670
	(−3.666, −0.032)	(−53.103, −37.471)	(−0.968, 2.270)	(−1.295, 3.648)	(−2.615, 3.955)
Δ Initial Slope for Practices with Incoming GP from Better	3.097***	11.477***	−0.178	−3.501***	−8.597***
	(1.531, 4.663)	(7.024, 15.930)	(−1.835, 1.478)	(−6.012, −0.991)	(−12.048, −5.145)
Common Change in Level	−0.168***	−4.112***	0.025	1.038***	0.071
	(−0.228, −0.108)	(−4.296, −3.929)	(−0.033, 0.083)	(0.949, 1.126)	(−0.019, 0.160)
Δ Change in Level for Practices with Incoming GP from Same	0.571***	7.113***	−0.046	−0.627***	0.146*
	(0.467, 0.676)	(6.792, 7.433)	(−0.147, 0.055)	(−0.781, −0.472)	(−0.010, 0.303)
Δ Change in Level for Practices with Incoming GP from Worse	−0.029	−0.387*	−0.207***	0.044	−0.432***
	(−0.166, 0.108)	(−0.777, 0.003)	(−0.345, −0.070)	(−0.172, 0.260)	(−0.682, −0.181)
Δ Change in Level for Practices with Incoming GP from Better	−0.214***	−1.412***	−0.009	−0.297***	−0.173
	(−0.358, −0.071)	(−1.943, −0.880)	(−0.146, 0.128)	(−0.503, −0.091)	(−0.428, 0.083)
Common Change in Slope		4.114***	−0.168	−0.118	−0.134
		(2.801, 5.427)	(−0.655, 0.319)	(−0.707, 0.470)	(−0.373, 0.105)
Δ Change in Slope for Practices with Incoming GP from Same	−0.295**	−1.485***	0.198	0.587***	1.344***
	(−0.550, −0.040)	(−2.213, −0.757)	(−0.066, 0.462)	(0.185, 0.988)	(0.813, 1.875)
Δ Change in Slope for Practices with Incoming GP from Worse	0.298**	6.711***	−0.062	−0.153	−0.063
	(0.016, 0.580)	(5.562, 7.860)	(−0.320, 0.195)	(−0.546, 0.241)	(−0.575, 0.449)
Δ Change in Slope for Practices with Incoming GP from Better		−7.114***	0.695	−0.027	−0.166
		(−9.795, −4.432)	(−0.172, 1.562)	(−1.235, 1.181)	(−0.582, 0.251)
Practice FE	NO	NO	NO	NO	NO
Year FE	NO	NO	NO	NO	NO
Practice RE	NO	NO	NO	NO	NO
Observations	43,193	43,193	43,193	43,193	43,193

NOTE: All regressions are adjusted for confounding by the doctor-to-patient ratio.

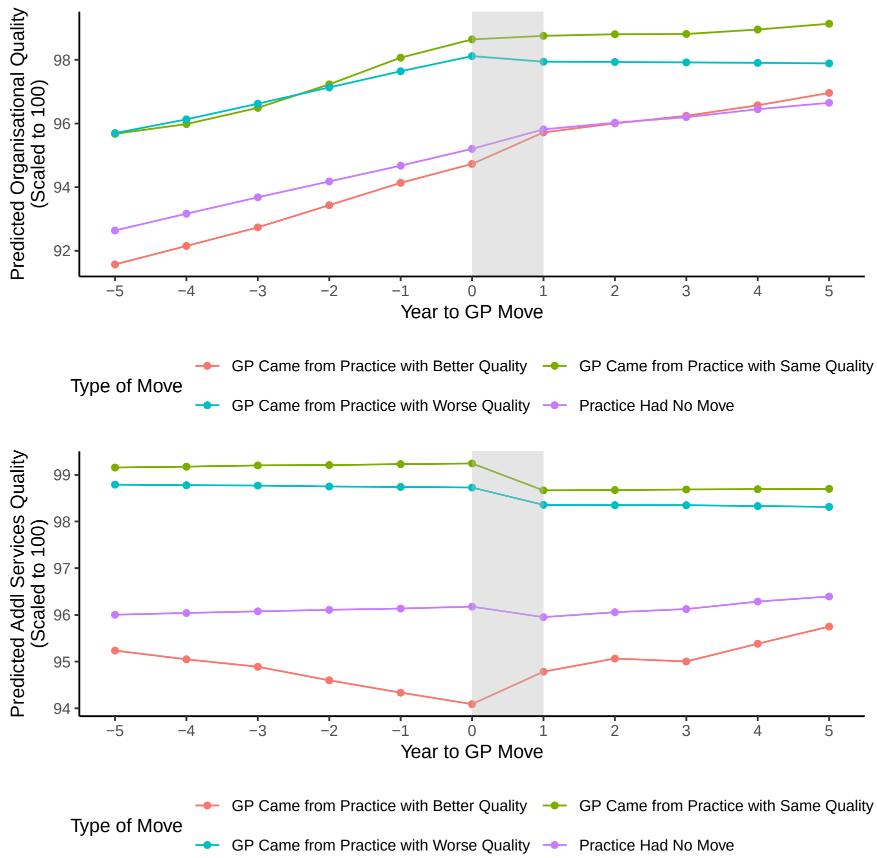
4.9 Interrupted Time Series Analysis with Fixed Effects

Figure 3: Difference-in-Difference Estimates



These figures represent the mean fitted values from the extended Interrupted Time Seriesregression (Equation (5)). The shaded region is the time period when the GP moved to the new practice. Note that the x-axis standardizes the interruption to 0. Consequently, values that are presented far from the move (i.e., near the edges) are based on a limited number of observations. For example, only practices where the move occurred in the last period would have an observation for -5 years, and only practices where the move occurred in the first period would have an observation for +5 years.

Figure 4: Difference-in-Difference Estimates



These figures represent the mean fitted values from the Interrupted Time Series regression (Equation (5)). The shaded region is the time period when the GP moved to the new practice. Note that the x-axis standardizes the interruption to 0. Consequently, values that are presented far from the move (i.e., near the edges) are based on a limited number of observations. For example, only practices where the move occurred in the last period would have an observation for -5 years, and only practices where the move occurred in the first period would have an observation for +5 years.

Table 12: Interrupted Time Series Analysis with Practice and Time Fixed Intercept Effects

	Dependent variable:				
	Total	Patient Experience	Clinical	Organizational	Additional Services
	(1)	(2)	(3)	(4)	(5)
Δ Initial Slope for Practices with Incoming GP from Same		3.342***	−0.033	−0.016	0.112
		(2.268, 4.416)	(−0.352, 0.285)	(−0.477, 0.444)	(−0.076, 0.301)
Δ Initial Slope for Practices with Incoming GP from Worse	−0.126**	−0.220	0.112**	−0.091	0.0001
	(−0.222, −0.030)	(−0.647, 0.207)	(0.024, 0.200)	(−0.255, 0.073)	(−0.197, 0.198)
Δ Initial Slope for Practices with Incoming GP from Better	0.042	−1.118***	−0.007	0.092	−0.335***
	(−0.051, 0.134)	(−1.437, −0.799)	(−0.096, 0.082)	(−0.081, 0.266)	(−0.528, −0.142)
Common Change in Level	0.513*	3.845***	−0.026	0.629	0.336
	(−0.012, 1.037)	(1.891, 5.800)	(−0.513, 0.461)	(−0.288, 1.546)	(−0.572, 1.243)
Δ Change in Level for Practices with Incoming GP from Same		22.340***	−0.299	−1.653	1.515
		(8.048, 36.632)	(−3.786, 3.189)	(−8.787, 5.482)	(−0.532, 3.563)
Δ Change in Level for Practices with Incoming GP from Worse	−0.787	1.733	1.121*	1.833*	1.422
	(−2.119, 0.544)	(−4.946, 8.413)	(−0.004, 2.245)	(−0.274, 3.940)	(−1.392, 4.236)
Δ Change in Level for Practices with Incoming GP from Better	0.268	−9.903***	0.578	−2.794**	−6.227***
	(−0.837, 1.374)	(−13.701, −6.105)	(−0.564, 1.720)	(−4.941, −0.646)	(−9.226, −3.228)
Common Change in Slope	−0.062	−0.608***	0.016	−0.066	−0.039
	(−0.146, 0.022)	(−0.921, −0.295)	(−0.062, 0.094)	(−0.213, 0.081)	(−0.185, 0.106)
Δ Change in Slope for Practices with Incoming GP from Same		−4.424***	0.066	0.221	−0.283*
		(−6.662, −2.187)	(−0.504, 0.635)	(−0.824, 1.266)	(−0.607, 0.041)
Δ Change in Slope for Practices with Incoming GP from Worse	0.157	−0.276	−0.148*	−0.294*	−0.190
	(−0.039, 0.352)	(−1.221, 0.669)	(−0.317, 0.021)	(−0.612, 0.025)	(−0.603, 0.222)
Δ Change in Slope for Practices with Incoming GP from Better	0.101	1.637***	0.073	0.577***	1.107***
	(−0.071, 0.273)	(1.045, 2.229)	(−0.100, 0.246)	(0.250, 0.903)	(0.668, 1.546)
Practice Intercept FE	YES	YES	YES	YES	YES
Practice Slope FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES
Practice RE	NO	NO	NO	NO	NO
Observations	43,193	43,193	43,193	43,193	43,193

NOTE: All regressions are adjusted for confounding by the doctor-to-patient ratio.

4.10 Interrupted Time Series Analysis with Fixed Effects (Movers Only)

Table 13: Interrupted Time Series Analysis with Fixed Intercept Effects, Only Movers

	Dependent variable:				
	Total	Patient Experience	Clinical	Organizational	Additional Services
	(1)	(2)	(3)	(4)	(5)
Δ Initial Slope for Practices with Incoming GP from Better	0.144^{***}	-1.105^{***}	-0.131^{**}	0.186^{*}	-0.334^{***}
	(0.037, 0.251)	(−1.565, −0.645)	(−0.231, −0.030)	(−0.007, 0.379)	(−0.571, −0.098)
Change in Level for Practices with Incoming GP from Worse	0.619	21.371^{***}	1.361^{***}	2.754^{***}	1.374
	(−0.616, 1.854)	(14.597, 28.145)	(0.338, 2.384)	(0.826, 4.682)	(−1.341, 4.088)
Δ Change in Level for Practices with Incoming GP from Better	0.849	-21.961^{***}	−0.493	-4.597^{***}	-7.637^{***}
	(−0.602, 2.300)	(−29.067, −14.854)	(−1.828, 0.843)	(−7.114, −2.080)	(−11.294, −3.980)
Change in Slope for Practices with Incoming GP from Worse	−0.053	-3.299^{***}	-0.174^{**}	-0.408^{***}	−0.167
	(−0.238, 0.132)	(−4.277, −2.321)	(−0.330, −0.017)	(−0.703, −0.112)	(−0.574, 0.240)
Δ Change in Slope for Practices with Incoming GP from Better	−0.016	3.368^{***}	0.217^{**}	0.865^{***}	1.296^{***}
	(−0.233, 0.201)	(2.346, 4.390)	(0.017, 0.418)	(0.485, 1.244)	(0.763, 1.828)
Practice Intercept FE	YES	YES	YES	YES	YES
Practice Slope FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES
Practice RE	NO	NO	NO	NO	NO
Observations	12,521	12,070	12,115	11,916	7,657

NOTE: All regressions are adjusted for confounding by the doctor-to-patient ratio.

5 Discussion

5.1 Validity of Quality Measures

In addition to the potential for this analysis to inform future quality improvement interventions, investigation of a "mover effect" could also serve to inform theoretical approaches to quality improvement and/or validate the QOF performance indicators.

The dominant framework for improving quality in general practice in England's NHS is the "Hierarchy and Targets" theoretical model (Bevan and Fasolo, 2013). This model, in which physicians are rewarded based upon the practice's achievement relative to preset performance indicators, is predicated on the notion that physicians have some agency over the practice's achievement on those performance indicators. That is, if a practice's performance on a set of indicators is entirely (or heavily) dependent upon factors outside of a physician's control, then offering incentives to the physician will not improve the practice's performance. Indeed, in this case, incentives may even reduce healthcare equity, as the most fortunate practices – those practices with the best external conditions – will disproportionatley reap the incentives.

Physicians point to this possibility when criticising the current "Hierarchy and Targets" theoretical model (Oliver, 2009), and although targets are currently well-ingrained in the NHS's scheme for quality improvement, it is not the only model that has been proposed or implemented throughout the history of the NHS. Other models of quality improvement account for this possibility. For example, the "Altruism" theoretical model, which assumes that performance is entirely dictated by outside forces, implies that quality can be improved by allocating resources to the worst performing practices in hopes that the additional resources will improve the circumstances and performance of those practices (Bevan and Fasolo, 2013).

Consequently, in order to determine whether the capacity for the current targets-based model to achieve its goal of improving healthcare quality, it is necessary to evaluate whether (and the extent that) GPs have meaningful agency over their practice's performance on the QOF indicators. Given the complexity of the health production function, however, it can be challenging to disentangle the effects of physicians from those of external factors, such as the health of the population, community health priorities, and practice-level resources. Focusing on movers into practices, then, presents an opportunity to plausibly observe the effect of physicians rather than other context variables. That is, if a physician moves into a new practice, it is unlikely that the context variables for that practice would meaningfully or systematically change – for example, the patient population, community priorities, and practice-level resources should be constant. If this assumption is met, a change in the practice's performance after an entering GP can reasonably be ascribed to the physician, including her medical practice choices and social influence.

The direction of a "mover effect" can then be used to help resolve which model is most appropriate for quality improvement. Under the "Hierarchy and Targets" model, physicians are generally capable of improving their performance given adequate motivation. However, under the "Altruism" theoretical model, physicians generally act to the best of their ability given the information and resources available to them. While a mover is unlikely to meaningfully change the practice's tangible resources, she may bring new information to the practice. In this case, given that physicians are well aware of the quality metrics, it is possible that if a new physician with better information enters the practice, that information may be shared, and the other physicians in the practice may begin incorporating the new information to improve the quality of their care. On the other hand, under the "Altruism" model, when a physician who enters the practice from a worse performing practice, only the entering GP will adapt her practices to the new information available at the new practice; there should, at least theoretically, be no effect of the entering GP, who now has access to the target practice's information and resources, on the quality of the practice. Indeed, the addition of any new physician, even one from a worse performing practice, should relieve the patient burden of the other physicians in the practice and, in turn, improve the practice's quality. Consequently, if GPs entering from better practices positively impact the quality of their incoming practices and GPs entering from worse practices negatively impact the quality of their incoming practices,[4], this would contradict the tenets of the "Altruism" theoretical model. This finding would, instead, constitute evidence then the physician's actions – beyond the resources or information available to the practice and the increased capacity from the new physician – affect quality, thereby providing support to the "Hierarchy and Targets" model.

[4]It is at least hypothetically possible that physicians moving from better or worse practices move the target practice's quality in an unexpected direction. For example, it is possible that physicians may resent a high-performing newcomer and resist adopting her practices. This finding would still suggest that physicians adjust their quality of care in reaction to social influences, but the implications for interventions would be different. In that instance, it might be worthwhile to consider exposing high-quality practices to low-quality GPs through incentivised moves or rotational programs.

GP movement is a potentially important and understudied feature of the GP labor supply. GP movement has become more common in the past several years (Figures **??** and **??**), though it is unknown whether and how GP movement may affect quality of care. GP moves may decrease quality of care by disrupting relationships built between patients and GPs. On the other hand, GP movement could fuel an exchange of ideas and practices. I will assess the effect of GP movement on GP practice quality outcomes.

References

Ali, Mir M., and Debra S. Dwyer. 2011. "Estimating peer effects in sexual behavior among adolescents." *Journal of Adolescence*, 34(1): 183–190.

Batalden, P. B, and F. Davidoff. 2007. "What is "quality improvement" and how can it transform healthcare?" *Quality and Safety in Health Care*, 16(1): 2–3.

Bell-Dzide, Dodzi, Murthy Gokula, and Phyllis Gaspar. 2014. "Effect of a long-term care geriatrics rotation on physician assistant students' knowledge and attitudes towards the elderly." *The Journal of Physician Assistant Education: The Official Journal of the Physician Assistant Education Association*, 25(1): 38–40.

Bevan, Gwyn, and Barbara Fasolo. 2013. "Models of governance of public services: empirical and behavioural analysis of 'econs' and 'humans'." In *Behavioural Public Policy..* , ed. Oliver Angus, 38–62. Cambridge, UK:Cambridge University Press.

Bollinger, Bryan, and Kenneth Gillingham. 2012. "Peer Effects in the Diffusion of Solar Photovoltaic Panels." *Marketing Science*, 31(6): 900–912.

Braithwaite, Jeffrey. 2018. "Changing how we think about healthcare improvement." *BMJ*, k2014.

Bursztyn, Leonardo, Florian Ederer, Bruno Ferman, and Noam Yuchtman. 2014. "Understanding Mechanisms Underlying Peer Effects: Evidence From a Field Experiment on Financial Decisions." *Econometrica*, 82(4): 1273–1301.

Dahl, Gordon B., Katrine V. Løken, and Magne Mogstad. 2014. "Peer Effects in Program Participation." *American Economic Review*, 104(7): 2049–2074.

Dolea, Carmen, Laura Stormont, and Jean-Marc Braichet. 2010. "Evaluated strategies to increase attraction and retention of health workers in remote and rural areas." *Bulletin of the World Health Organization*, 88: 379–385.

Epstein, Andrew J., and Sean Nicholson. 2009. "The formation and evolution of physician treatment styles: An application to cesarean sections." *Journal of Health Economics*, 28(6): 1126–1140.

Ferguson, Julie, Judy Wakeling, and Paul Bowie. 2014. "Factors influencing the effectiveness of multisource feedback in improving the professional practice of medical doctors: a systematic review." *BMC Medical Education*, 14(1): 76.

Flodgren, Gerd, Mary Ann O'Brien, Elena Parmelli, and Jeremy M. Grimshaw. 2019. "Local opinion leaders: effects on professional practice and healthcare outcomes." *Cochrane Database of Systematic Reviews*, , (6).

Gaviria, Alejandro, and Steven Raphael. 2001. "School-Based Peer Effects and Juvenile Behavior." *Review of Economics and Statistics*, 83(2): 257–268.

Goodpastor, Walter A., and Isaac D. Montoya. 1996. "Motivating physician behaviour change: social influence versus financial contingencies." *International Journal of Health Care Quality Assurance*, 9(6): 4–9.

Grytten, Jostein, and Rune Sørensen. 2003. "Practice variation and physician-specific effects." *Journal of Health Economics*, 22(3): 403–418.

Herbst, D., and A. Mas. 2015. "Peer effects on worker output in the laboratory generalize to the field." *Science*, 350(6260): 545–549.

Ivers, Noah, Gro Jamtvedt, Signe Flottorp, Jane M Young, Jan Odgaard-Jensen, Simon D French, Mary Ann O'Brien, Marit Johansen, Jeremy Grimshaw, and Andrew D Oxman. 2012. "Audit and feedback: effects on professional practice and healthcare outcomes." *Cochrane Database of Systematic Reviews.*

Jong, Judith D. de, Gert P. Westert, Ronald Lagoe, and Peter P. Groenewegen. 2006. "Variation in Hospital Length of Stay: Do Physicians Adapt Their Length of Stay Decisions to What Is Usual in the Hospital Where They Work?" *Health Services Research*, 41(2): 374–394.

Kenny, Nuala P., Karen V. Mann, and Heather MacLeod. 2003. "Role Modeling in Physicians??? Professional Formation: Reconsidering an Essential but Untapped Educational Strategy:." *Academic Medicine*, 78(12): 1203–1210.

Li, Jinhu, Anthony Scott, Matthew McGrail, John Humphreys, and Julia Witt. 2014. "Retaining rural doctors: Doctors' preferences for rural medical workforce incentives." *Social Science & Medicine*, 121: 56–64.

Locock, L., S. Dopson, D. Chambers, and J. Gabbay. 2001. "Understanding the role of opinion leaders in improving clinical effectiveness." *Social Science & Medicine (1982)*, 53(6): 745–757.

Lundborg, Petter. 2006. "Having the wrong friends? Peer effects in adolescent substance use." *Journal of Health Economics*, 25(2): 214–233.

Meltzer, David, Jeanette Chung, Parham Khalili, Elizabeth Marlow, Vineet Arora, Glen Schumock, and Ron Burt. 2010. "Exploring the use of social network methods in designing healthcare quality improvement teams." *Social Science & Medicine*, 71(6): 1119–1130.

Mittman, Brian S., Xenia Tonesk, and Peter D. Jacobson. 1992. "Implementing Clinical Practice Guidelines: Social Influence Strategies and Practitioner Behavior Change." *QRB - Quality Review Bulletin*, 18(12): 413–422.

Molitor, David. 2018. "The Evolution of Physician Practice Styles: Evidence from Cardiologist Migration." *American Economic Journal: Economic Policy*, 10(1): 326–356.

Moretti, Enrico. 2011. "Social Learning and Peer Effects in Consumption: Evidence from Movie Sales." *The Review of Economic Studies*, 78(1): 356–393.

Oliver, A. 2009. "Can financial incentives improve health equity?" *BMJ*, 339(sep24 2): b3847–b3847.

Oxman, A. D., M. A. Thomson, D. A. Davis, and R. B. Haynes. 1995. "No magic bullets: a systematic review of 102 trials of interventions to improve professional practice." *CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne*, 153(10): 1423–1431.

Phelps, Charles, and Cathleen Mooney. 1993. "Variations in Medical Practice Use: Causes and Consequences."

Pronovost, Peter J, and Daniel W Hudson. 2012. "Improving healthcare quality through organisational peer-to-peer assessment: lessons from the nuclear power industry." *BMJ Quality & Safety*, 21(10): 872–875.

Rubenstein, Lisa V., Brian S. Mittman, Elizabeth M. Yano, and Cynthia D. Mulrow. 2000. "From Understanding Health Care Provider Behavior to Improving Health Care: The QUERI Framework for Quality Improvement." *Medical Care*, 38(6): I129–I141.

Sacerdote, B. 2001. "Peer Effects with Random Assignment: Results for Dartmouth Roommates." *The Quarterly Journal of Economics*, 116(2): 681–704.

Van den Hombergh, P., R. Grol, H. van den Hoogen, and W. van den Bosch. 1999. "Practice visits as a tool in quality improvement: mutual visits and feedback by peers compared with visits and feedback by non-physician observers." *Quality and Safety in Health Care*, 8(3): 161–166.

Yong, Jongsay, Anthony Scott, Hugh Gravelle, Peter Sivey, and Matthew McGrail. 2018. "Do rural incentives payments affect entries and exits of general practitioners?" *Social Science & Medicine*, 214: 197–205.