HEDG

Health, Econometrics and Data Group

The University of York

WP 23/13

Non-monetary interventions, workforce retention and hospital quality: evidence from the English NHS

Giuseppe Moscelli; Melisa Sayli; Jo Blanden; Marco Mello;

Henrique Castro-Pires and Chris Bojke

August 2023

http://www.york.ac.uk/economics/postgrad/herc/hedg/wps/

Non-monetary interventions, workforce retention and hospital quality: evidence from the English NHS

Giuseppe Moscelli[*†‡], Melisa Sayli[*†], Jo Blanden[†‡], Marco Mello[¶], Henrique Castro-Pires[†], and Chris Bojke[§]

[†]School of Economics, University of Surrey

[§]School of Medicine, University of Leeds

[‡]IZA

[¶]Economics Department, Business School, University of Aberdeen

Abstract

Excessive turnover can significantly impair an organization's performance. Using high-quality administrative data and staggered difference-in-differences strategies, we evaluate the impact of a programme that encouraged public hospitals to increase staff retention by providing data and guidelines on how to improve the non-pecuniary aspects of nursing jobs. We find that the programme has decreased the nurse turnover rate by 4.49%, decreased exits from the public hospital sector by 5.38%, and reduced mortality within 30 days from hospital admission by 3.45%, preventing 11,400 deaths. Our results are consistent with a theoretical model in which information is provided to managers of multi-unit organizations, who trade off coordinating decisions across units and adapting them to local conditions.

Keywords: labor supply, workforce retention, non-monetary incentives, hospital care, staggered difference-in-differences.
JEL Codes: J32, J38, J45, J63, I11, C22.

---

[*]Corresponding author: g.moscelli@surrey.ac.uk.
This research has been funded by The Health Foundation under the "Efficiency Research Programme – Round 3" grant scheme (Award ID: 1327076). The findings and opinions expressed in this study do not represent any views from The Health Foundation. We thank the Department of Health and Social Care and NHS England respectively for access to NHS Electronic Staff Record (ESR) and NHS Staff Survey (NSS) data. Hospital Episode Statistics are Copyright 2009–2020, re-used with the permission of NHS Digital; all rights reserved. The authors are thankful for comments and useful suggestions to Adrian Boyle, Holger Breinlich, Hyein Cho, Igor Francetic, Guido Friebel, Willian Fuchs, Hugh Gravelle, Alessandro Iaria, Esteban Jaimovich, Elaine Kelly, Ambar Laforgia, Sandra McNally, Matthias Parey, Raffaella Sadun, and seminars participants to the Universities of Exeter, Southampton, Glasgow, Bristol, UK HESG Winter Meeting 2022, and Royal Economic Society 2022, International Association of Applied Econometrics 2022, American-European Health Economics Study Group 2022, European Health Economics Association 2022, EALE 2022, NBER Organizational Economics 2023 and SOLE 2023 conferences. The usual disclaimer applies.

1 Introduction

Work is the engine of society, and understanding how individuals are incentivized to work is a central concern among social scientists (Gibbons 1998; Bandiera et al. 2010, 2013; Duflo et al. 2012, among many others). In the last three decades, economics research has widened its scope from the estimation of wage elasticities to investigating how a broader range of factors influence whether, how long and how hard employees work (Mas and Pallais, 2017; Cassar and Meier, 2018). Employee responsiveness to non-financial aspects of job quality is pertinent to public sector occupations (Dixit, 2002), and especially so when the potential for substantial wage increases is limited by the combination of labor-intensiveness of production, a large workforce and tight government budget constraints. In this study, we evaluate the effectiveness of changing the non-pecuniary aspects of public-sector jobs on a large scale, by exploiting a labor market policy aimed at decreasing the turnover of nurses working in all public hospitals in England.

Nurses are a vital part of the healthcare sector, constituting about one-third of healthcare employees in countries like the United States and the United Kingdom. Despite an increasing number of nurses trained and employed, nurse vacancy rates are high across the OECD countries and the supply of nurses fails to keep up with rising demand. In the UK, even before the COVID-19 pandemic, there was a 50,000 nurse staffing gap (Buchan et al., 2020), and vacancy rates for registered nurses increased from 6% to 11% between 2013 and 2016 (Helm and Bungeroth, 2017). In the US, about 1.1 million new registered nurses are needed by 2030 (Bureau of Labor Statistics, U.S. Department of Labor, 2021); analogously, serious nursing shortages are faced by Canada, Germany and Japan (Marć et al., 2019). Given the connection between nurse shortages and poor patient care and outcomes (Ball et al., 2014, 2018; Duffield et al., 2011; Griffiths et al., 2019, 2018; Rafferty et al., 2007), policy makers are seeking sustainable, cost-effective ways to reduce nurse shortages and high turnover rates.

Compared to training new nurses, which takes three to four years in the UK, improving retention is a time and cost-efficient solution to staff shortages (Shields, 2004; Duffield et al., 2014), and retains specific human capital within the employing organization. A possible strategy to improve retention is to increase wages. However, the empirical literature supports a limited role for wages to increase labor supply among nurses (see Lee et al. 2019 for a comprehensive review). Even if nurse retention was highly responsive to wages, a conspicuous pay rise across a sector as large as UK public health care is expensive.[1] An alternative approach to reducing turnover rates is to improve the non-financial elements of work that are valued by workers. Cassar and Meier (2018) emphasize the importance of mission, autonomy, competence and relatedness, which together make work "meaningful". Understanding the role played by non-monetary factors in shaping workers' labor supply is important, because workers with mission-driven preferences, such as front-line healthcare workers, may value the various facets of jobs differently compared to those in the private sector (Ellingsen and Johannesson, 2008; Brekke and Nyborg, 2010; Lee et al., 2019).

In this work, we evaluate the effects of the Retention Direct Support Programme (henceforth, *RDSP* or 'the *Programme*'), which was launched in July 2017. The Programme's objective was to improve retention among nurses working in NHS hospital organizations (HO), called Trusts in the English NHS. HOs were not given a numeric target to meet, and the Programme had no other direct goals. While data and guidelines were issued centrally, actions were chosen locally. NHS HOs' supervisory body, NHS Improvement (NHSI), provided tailored retention data to identify areas for development, as well as liaison officers to help form and execute action plans. The provision and analysis of retention data is a crucial feature of this intervention, because before the RDSP policy NHS HOs did "not collect data on retention in a consistent and robust way and so any national drive to improve nurse retention would have to address this" (Marangozov et al., 2016). The localized approach used by NHS policy-makers to enact this Programme is in line with the substantial unexplained variation in workforce retention found across NHS HOs (Kelly, Stoye and Warner, 2022). Prompted by the data they were given, the HOs were tasked to build their own retention strategies, and implement them under the guidance of the NHS supervisory body.

To evaluate the effect of RDSP we construct a unique dataset, consisting of a monthly panel of English NHS hospitals, by combining four different administrative datasets, respectively holding records on: NHS hospital nurses' employment spells

[1] NHS nurses have been involved in recent industrial action during which pay rises were resisted by Government due to affordability concerns (The Guardian, 2023).

at individual-level; nurses' working conditions from NHS staff surveys at individual-level; timing and organizational content of the intervention; and health outcomes of patients admitted to NHS hospitals. The RDSP was implemented in a staggered fashion, with the HOs split into five cohorts and each cohort starting the Programme at different times. In order to achieve causal identification of the impact of the Programme, we exploit new methodological advances in the estimation of difference-in-difference (DiD) with staggered treatment adoption. We exploit the Callaway and Sant'Anna (2021) estimator to investigate the direct effect of RDSP on nurse retention outcomes, whereas we rely on the Arkhangelsky et al. (2021) synthetic DiD estimator to gauge the effect of RDSP on health outcomes, for which the likely presence of other contemporaneous interventions might invalidate the DiD parallel trend assumption; we find very similar RDSP effects on nurse retention outcomes obtained through the two aforementioned methods, as one would reasonably expect if the RDSP was the predominant intervention in place targeting nurse retention in NHS HOs.

Our work brings the following contributions to the labor, health and organizational economics literatures.

First, we provide a simple theory model describing the responsiveness of employee turnover to an intervention that provides information to managers. Retention is influenced by managers choosing actions to reflect the local conditions in the organization they manage. Information improves this match, so our model predicts the effectiveness of the RDSP. The model also predicts that managers in charge of multiple hospital sites will respond less to information, so retention will increase less in these HOs. Moreover, the model predicts heterogeneous RDSP effects across treatment cohorts, driven by the differences in managers' actions with and without detailed information about their employee retention.

Second, we test the model's hypothesis that the intervention improves employee retention within the (healthcare) organization where they work, and also within the whole (public hospital care) sector. Overall, we find that the RDSP has improved nursing retention by 0.78 percentage points (ppt), equivalent to retaining, on average, 1,697 nurses and midwives who would have left their HO otherwise. This is around a quarter to a half of the standard deviation of the retention rate across HOs in the pre-period. Our results hold when we use alternative estimators, such as the interaction-weighted estimator of Sun and Abraham (2021), to capture the dynamic treatment effects of interest.

Third, we investigate the mechanisms that underlie the Programme's success by exploring how the treatment effect varies across different dimensions. The results relating to hospital characteristics are in line with the predictions of the model, indicating that the provision of information to local HOs is an important aspect of the Programme's effectiveness. We also attempt to unpack the "black box" of the RDSP by exploiting information about the Programme areas of intervention used by each HO. Although establishing a precise causal link between HOs action plans and outcomes is more difficult in this case, our results provide suggestive evidence that managers are making effective choices for their own HO, as there is little evidence of strong differences in the estimated treatment effects with respect to the areas of interventions made. Furthermore, we consider how the effectiveness of the RDSP depends on staff characteristics such as seniority, which has been shown to be important for patient outcomes (Kelly, Propper and Zaranko, 2022).

Finally, we explore the broader consequences of the programme by investigating its effects on patient outcomes and hospital productivity. The intervention was designed to improve workforce retention at the hospital organization level, but it is possible that patient outcomes, such as mortality, would also improve as vacancies and staff turnover reduce. Moreover, if the intervention led to a general improvement in management and productivity, we might expect that participation improved patient outcomes directly. An alternative hypothesis is that focusing on employee retention could distract hospital managers and nursing staff from activities with more direct benefit for patients, in line with the classical multitasking trade-off (Holmstrom and Milgrom, 1991). We document a reduction in 30-day risk-adjusted mortality for patients admitted to hospital and patients' waiting times for planned treatments at the RDSP-treated HOs.

The paper proceeds as follows. Section 2 presents the related literature and the institutional settings of the English NHS, its nursing workforce and the RDSP policy, whereas Section 3 illustrates the theoretical framework and its predictions. Section 4 describes the data sources and the empirical strategy, while Section 5 and Section 6 respectively report the main results and the robustness checks. Section 7 concludes.

2 Background

2.1 Related Literature

The structure of the RDSP means that its evaluation makes important contributions to the recent literature on how to improve human resource management and productivity. First, it is a light-touch intervention that encourages managers to act based on data, information on best practice and their own judgment. Friebel et al. (2022)'s work is particularly relevant in this context, as it finds that simply asking managers to 'do what they can' to reduce turnover is effective among retail workers in Eastern Europe. Gosnell et al. (2020) show that providing workers with data on their outputs has a substantial impact on their performance. Our findings demonstrate that goal-setting alone can also be sufficient to encourage managers to act to improve retention in the public sector. A wide literature shows that targets and incentives can influence performance in the NHS (Propper et al., 2010; Cooper et al., 2011; Gaynor et al., 2013; Bloom et al., 2015) and other public services (Burgess et al., 2017). However, strong incentives can be counter-productive when agents are pro-socially motivated (Ellingsen and Johannesson, 2008), raising the question of the optimal strength of incentives in the public sector. We show that low-powered incentives are sufficient to motivate NHS managers to make changes to working conditions that influence workers' retention.

Our study also speaks to the literature on effective management (Bloom and Van Reenen, 2007; Hoffman and Tadelis, 2021; Alan et al., 2023). Data, monitoring and people management are emphasized as strongly associated with good management and performance in public sector organizations (Bloom et al., 2014; McNally et al., 2022). The intervention assessed can be characterized as a provision of information and a "nudge" towards adopting best practice in human resource management, therefore providing further support to the idea that better management can be achieved in the public sector without great cost .

Organizational culture is also functional to reduce employee turnover. Alan et al. (2023) shows that an extremely specific intervention to improve company culture reduced voluntary quits among white-collar workers in Turkey. Linos et al. (2022) demonstrates that an intervention to encourage workers to share best practice and the formation of wider work communities can reduce resignations among police call-handlers by half. Positive human resource management practices, such as having a positive attitude, setting clear expectations and internal consultations, or providing career support and coaching, may have a substantial effect on staff retention (Hoffman and Tadelis, 2021). Our results demonstrate that workplace improvements matter in the public sector and affect the decision to quit, an important margin in understaffed public services.

Lastly, our contribution is closely related to the literature about the effects of healthcare workers' labor supply on healthcare providers' performance and patient outcomes (e.g. Gruber and Kleiner 2012; Chan 2018; Chan Jr and Chen 2022; Kelly, Propper and Zaranko 2022, among others), and to two previous studies in particular. Propper and Van Reenen (2010) show that labor market (pay) regulation can affect hospital performance, as they find that in-hospital mortality for patients admitted due to a heart-attack is positively associated with higher outside wages, a known determinant of hospital nurses' attrition. Friedrich and Hackmann (2021) investigate the effects of a policy which caused unintended nurse shortages by extending the allowed statutory maternity leave period; they find that the unplanned emergency readmission risk increased for heart-attack patients admitted to hospitals facing more severe shortages. Complementarily to the aforementioned works, we report a mortality decrease in hospitals 'treated' by the RDSP policy, and particularly in the treatment cohorts that experienced larger gains in nurses' retention; such a result is not obvious if the effects of healthcare staff shortages on hospital performance are asymmetric and larger in magnitude than the effects of staffing increases, which may occur due economies of scale and indivisibility of labor tasks.

2.2 The English NHS and its nursing workforce

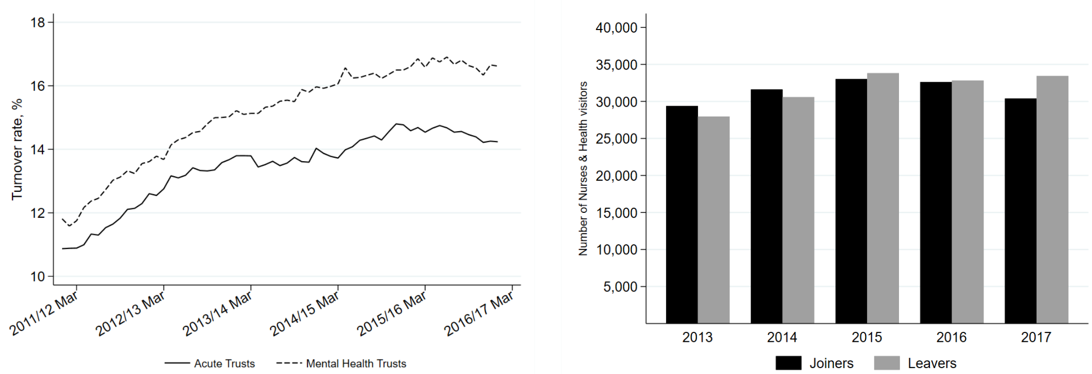
The NHS is publicly funded through general taxation and provides free comprehensive primary and secondary healthcare services to over 56 million people in England and a further 11 million in the other devolved nations of the UK (Scotland, Wales, and Northern Ireland).[2] The NHS budget for England in 2017 was approxi-

[2]Recent estimates suggest that only around 10.5% of the UK population hold voluntary private health insurance (Tikkanen et al., 2020)

mately £110bn. Public hospitals providing secondary care are run by organizations called NHS hospital Trusts, or simply NHS Trusts, that we henceforth call *hospital organizations* (HOs). In March 2020 about 564,000 nurses, midwives and nursing associates living in England were registered with the Nursing and Midwifery Council (NMC).[3] The English NHS employs around 330,000 of these registered nurses and midwives, who make up almost half of the professionally qualified clinical staff.[4]

Figure 1. NHS nursing workforce occupational time trends

(a) Turnover (b) Joiners to and leavers from the NHS



Notes. Panel (a): Authors' calculation from Electronic Staff Records 2009-2020. Turnover is measured for each month from one year to the following. Panel (b): Headcounts of Nurses and Health Visitors in NHS Hospital and Community Health Services (HCHS), from NHS Digital, NHS Hospital & Community Health Service workforce statistics (NHS Digital, 2018; National Audit Office, 2020) data.

The nursing workforce has been under significant pressure from growing demand for healthcare combined with high turnover rates. Levels of work-related stress have increased alongside staff turnover rates (Perreira et al., 2018). The National Audit Office (2020) notes that the increase in the full-time equivalent nursing numbers between 2010/11 and 2018/17 was not enough to meet NHS needs. Figure 1(a) and Figure 1(b) show that turnover rates have increased in recent years and that in 2017 more nurses left the NHS than joined. In the same period, vacancy rates were also high, with about 38,000 full-time equivalent open posts in the first quarter of

[3]The NMC is the professional body for nurses and midwives in the UK. To practice their profession, nurses and midwives need to register with the NMC and qualify to the NMC's standards.

[4]For brevity, in this work the terms "nursing staff" and "nurses" are referred to both nurses and midwives. Clinical staff includes HCHS doctors, qualified nurses and health visitors, midwives, qualified scientific, therapeutic and technical staff, and qualified ambulance staff.

2017/18 (June 2017) (NHS Digital, 2021) or 10.9% of the nursing workforce employed in NHS hospitals. Because of the high number of nursing vacancies, the NHS relies significantly also on temporary and agency staff, which cost NHS HOs approximately £1.46 billion per year (The Open University, 2018); improving nurses' retention would also reduce these labor costs, on top of preventing losses in human capital.

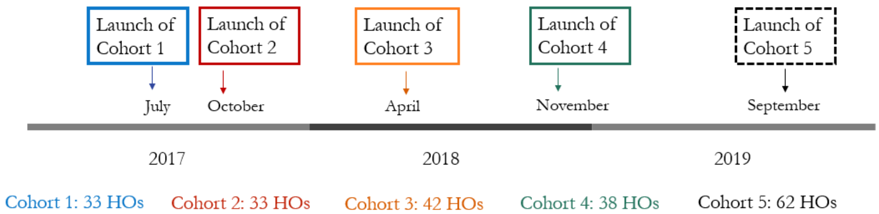
2.3 The Retention Direct Support Programme

The RDSP was designed by NHS Improvement with the aim of improving nursing retention in English NHS acute and mental health care HOs (NHS England, 2019) to tackle the nursing supply challenge. The Programme was clinically-led, involving at least one member of the nursing team from the HO, and focused on factors that were under HOs' control (NHS Improvement, 2017). The RDSP implementation consisted of a common organizational structure of the Programme and performance monitoring process for all HOs, alongside the development of retention improvement plans tailored to address each HO's retention needs (see details below). The RDSP was rolled out in a staggered fashion over 5 cohorts from 2017 until 2020, as shown in Figure 2. The Programme ended *de facto* in Spring 2020 due to COVID-19 pandemic onset. NHSI, the hospital monitoring body, allocated HOs into cohorts to be treated, starting with HOs that had above-average leaver rates.[5] In the weeks following the first contact from NHSI, clinical and workforce leads from the HOs were invited to participate to a "Retention Masterclass workshop", scheduled with approximately six weeks' notice. The HO leads received data packs from the NHSI a few days before the workshop, which contained HO-specific retention measures with regional benchmarks (NHS Improvement, 2017) to help HOs understanding their retention profile and potential for improvement. The Retention Masterclass workshops were scheduled close to the official launch of the RDSP cohorts to introduce the Programme to the HOs leads and function as an interactive platform to review and discuss the barriers to retention in their organizations. During the workshop, NHSI also presented potential areas the HOs might focus on to reduce turnover rates, showcased best practices,

[5]The selection was based on several factors, but more weight was given to HO's turnover rates and trends in the five years preceding the RDSP. HOs did not know which Cohort they were allocated to, until only a few weeks before they were contacted to join the Programme.

Figure 2. Organization of the Retention Direct Support Programme

(a) Timeline



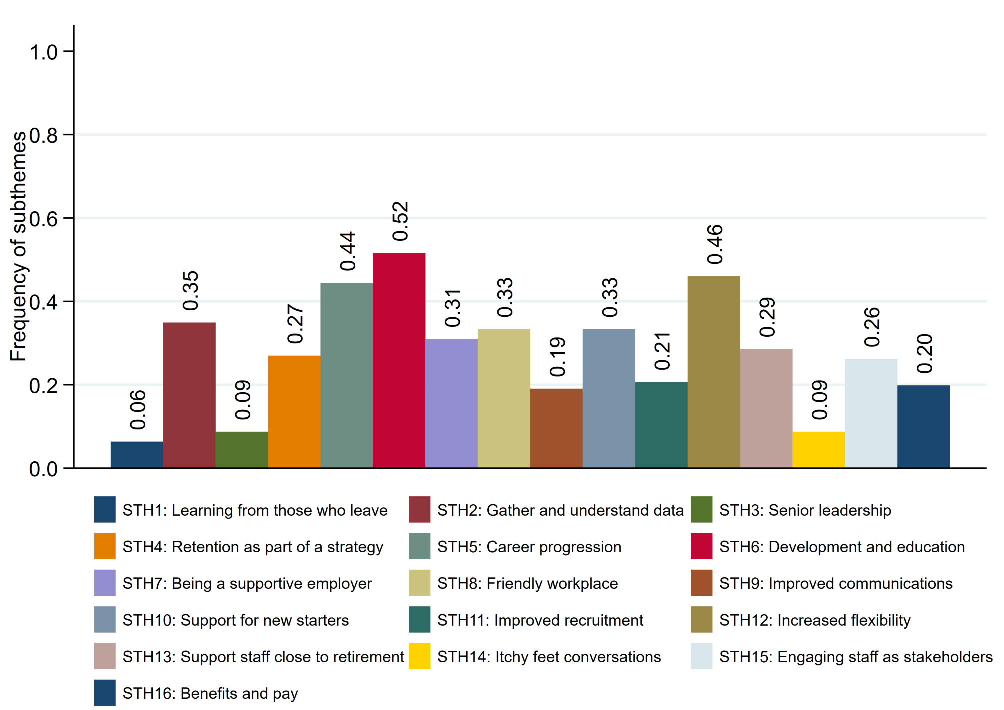
(b) Structure

and demonstrated how to use data to inform decision-making. HOs were also given guidelines on how to develop their action plans.

Each HO was matched with an existing NHSI officer acting as lead and collaborating with the HO for the whole duration of the RDSP. HOs were given 90 days to develop and submit a retention improvement Action Plan, and expected to use this period to review their data, identify areas of improvement, and set clear and measurable actions to reduce turnover rates. In the 12 months following the launch of the RDSP, NHSI officers monitored the progress of the HOs, provided quarterly data packs, and supported HOs that lagged behind their agreed targets.

Differently from other nationwide NHS policies, such as the 18-week waiting time target for planned surgery or the 4-hour waiting time target for urgent emergency hospital care, the RDSP did not set any specific turnover rate targets for HOs to achieve. NHSI's expectation was to see an improvement in turnover rates in the 12 months following the start of the Programme. Moreover, rather than a one-fits-all approach, within RDSP HOs were encouraged to focus on retention challenges endemic to their workforce and to set their turnover goals accordingly. Thus, the Programme also enabled HOs to incorporate existing and planned workforce governance initiatives into their action plans.

Figure 3. Frequency of subthemes chosen, from RDSP Action Plans



Notes. Authors' calculations from NHSI thematic coding matrix from Cohorts 1 to 4. The subthemes are categorized by NHSI using Action Plans submitted by the Hospital Organizations.

Ex-post, i.e. after the RDSP implementation, NHSI identified 16 recurring subthemes from the Action Plans submitted by HOs. Figure 3 shows the frequency of recurring subthemes in all Action Plans. No subtheme (STH) was included in all Action Plans, but some subtheme choices were more popular than others. About 35% of HOs focused on gathering and understanding workforce turnover data (STH2), consistent with the fact that before RDSP HOs lacked "a consistent and robust way" to collect data on employee retention (Marangozov et al., 2016). Respectively 52% and 44% of HOs included initiatives aimed at improving professional development (STH6) and career progression (STH5), which meant including strategies to develop clear and attainable career paths, re-design appraisal processes, and provide career coaching, consistent with the fact that NHS nurses perceive a lack of continuous learning and development opportunities offered in their roles (House of Commons Health Committee, 2018; NHS England and NHS Improvement, 2019).[6]

Several HOs included actions to make the organization a friendly workplace (STH8; 33%) and supportive for new starters (STH10; 33%), but even more HOs focused on improving work flexibility (STH12; 46%), another aspect emphasized by the House of Commons Health Committee (2018) and the 2019 NHS Long-term Plan (NHS England and NHS Improvement, 2019) as necessary for improved retention, through initiatives aimed at offering flexibility in rotations, improving online shift-scheduling, and facilitating transfer schemes. While pay is a contentious topic among nurses and midwives (Nelson and McLaughlin, 2020), it was not a recurring theme in the retention improvement plans. Only 13 out of 122 Action Plans explicitly mentioned pay, and the related initiatives were classified under the "Benefits and pay" subtheme (STH16), chosen by 20% of the HOs. This finding perhaps is not surprising, as NHS nurses' and midwives' wages are negotiated and determined at the national level, with no scope for individual bargaining. To summarise, a wide range of interventions was adopted across HOs, in line with what we would expect if HO managers tailored their actions to best fit their HO's local conditions.

3 Conceptual Framework

In this section we develop a simple conceptual framework that illustrates how a non-monetary intervention might affect workforce retention. Our model closely follows the canonical literature on organizational economics that examines the trade-off between adapting actions to local conditions and coordinating decisions across multiple units of the organization (see Dessein and Santos (2006) and Alonso et al. (2008)). In particular, we study how the effect of an informational intervention varies with an organization's local conditions, number of units, and managerial skill. We

[6]The lack of development opportunities was further exacerbated by cuts to the Continuing Professional Development (CPD) from the Health Education England's budget. The workforce development budget is mostly used for nurses' training and it suffered a 60% cut from 2015/16 to 2017/18 (Bungeroth et al., 2018).

tailor the terminology used in our model to our application of interest — nurses' retention in the NHS, but the insights of our model can be easily and broadly applied to employee retention in other settings with multi-unit organizations.

Setup. Consider a HO that controls m hospitals, each indexed by $i \in \mathcal{M} := \{1, .., m\}$. Each hospital has n nurses on its staff,[7] and all hospitals in the same HO share a unique management board, which for brevity's sake we refer to as *the HO manager* in charge. Each nurse might stay or leave their current job, and whether each of them stays is independently drawn from a Bernoulli distribution with success probability p, where p depends on an action $a_i \in [0, 1]$ taken by the HO manager and on the hospital-specific local condition $\theta_i \in [1/2, 1]$. In particular, for a given pair (a_i, θ_i), each nurse stays on their job with probability $p(a_i, \theta_i) = \theta_i - (a_i - \theta_i)^2$. That is, when the action taken a_i matches the local condition θ_i, the retention probability p is maximized and equal to θ_i.

For each nurse leaving, the HO manager must hire a replacement at a cost $r > 0$. Therefore, for a given vector of actions $a = (a_1, ..., a_m)$ and a vector of local conditions $\theta = (\theta_1, ..., \theta_m)$ the expected replacement cost is given by

$$R(a, \theta) = \sum_{i \in \mathcal{M}} \sum_{k=0}^{n} r \cdot \binom{n}{k} [1 - p(a_i, \theta_i)]^k [p(a_i, \theta_i)]^{n-k} = r \cdot n \cdot \sum_{i \in \mathcal{M}} (1 - p(a_i, \theta_i)).$$

In addition, we assume it is costly for the manager to take distinct actions in different locations (i.e., hospitals). It is cheaper for the manager to adopt standardized procedures across all hospitals than to tailor her actions to each location. Formally, if the manager chooses different actions across different hospitals, she incurs a miscoordination cost

$$C(a) = \frac{1}{m} \sum_{i \in \mathcal{M}} \sum_{j \in \mathcal{M}} \frac{(a_i - a_j)^2}{2}.$$

The manager's payoff is then given by $U(a, \theta) = -R(a, \theta) - (1/\gamma) \cdot C(a)$, where $\gamma > 0$ denotes the relative importance of miscoordination costs and can be interpreted as managerial skill. The larger the γ, the less costly for the manager to take distinct actions in different locations.[8]

[7]For simplicity, we assume all hospitals have the same staff size. The analysis carries through if they have distinct staff sizes.

[8]For simplicity, we assumed that each index i and pair (a_i, θ_i) refers to one entire hospital. An

The manager then faces a trade-off between adapting her choices to each hospital's local conditions and reducing miscoordination. An additional challenge is that the manager might not perfectly observe local conditions. We assume that each θ_i is independently and identically distributed with c.d.f. denoted by $F \in \Delta([1/2, 1])$, where θ_F and σ_F^2 denote, respectively, the expected value and the variance of the distribution F. Given her information, the manager's problem is choosing a vector of actions to maximize her expected utility.

$$\max_{a \in [0,1]^m} \mathbb{E}_\mu[U(a, \theta)] = \max_{a \in [0,1]^m} \sum_{i \in \mathcal{M}} \left\{ - r \cdot n \cdot \mathbb{E}_\mu \Big[(1 - p(a_i, \theta_i)) \Big] - \frac{1}{m\gamma} \sum_{j \in \mathcal{M}} \frac{(a_i - a_j)^2}{2} \right\}.$$

The optimal manager's action depends on her beliefs about the vector of local conditions θ. Denote by μ the manager's belief about θ, where μ_0 describes the case in which she has no information beyond the prior.[9] We denote by μ_I the full-information case (when the manager perfectly observes each θ_i).

Proposition 1. *The optimal manager's action as a function of her belief μ is*

$$a_i^*(\mu) = \frac{2\gamma rn}{1 + 2\gamma rn} \mathbb{E}_\mu[\theta_i] + \frac{1}{1 + 2\gamma rn} \sum_{j \in \mathcal{M}} \frac{\mathbb{E}_\mu[\theta_j]}{m} \quad \textit{for all } i \in \{1, ..., m\}, \tag{1}$$

and the staff retention rate as a function of the vector of local conditions θ and the manager's belief μ is

$$S(\theta, \mu) := \frac{1}{m} \sum_{i \in \mathcal{M}} p(a_i^*(\mu), \theta_i) = \frac{1}{m} \sum_{i \in \mathcal{M}} \Big[\theta_i - \big(a_i^*(\mu) - \theta_i \big)^2 \Big].$$

The optimal manager's action is a convex combination between the expected local condition of hospital i and the average expected condition across all other hospitals within the HO. The manager optimally trades off adaptation (matching each local action to the expected local condition) and coordination (matching actions across different units); the higher the manager's skill γ, the more she adapts to local conditions.

alternative interpretation is that each hospital has multiple staff groups, and that each pair (a_i, θ_i) represents one staff group within that hospital. In this alternative interpretation, coordination matters even for HOs that control a single hospital.

[9] μ_0 denotes the distribution of the vector θ when each θ_i is independently and identically distributed according to F.

better**The effect of an informational intervention.** We study the effect of a policy that provides managers with information about the vector of realized local conditions θ. For simplicity, we focus on the case in which full information about θ is provided to the HO manager.[10] Before the policy, managers would choose their actions based on the prior, while after the intervention, they are fully informed about θ. That is, the manager's pre-policy choices are $a_i^*(\mu_0) = \theta_F$, while the post-policy are

$$a_i^*(\mu_I) = \frac{2\gamma rn}{1+2\gamma rn}\theta_i + \frac{1}{1+2\gamma rn}\theta_{AVG} \quad \text{for all } i \in \{1, ..., m\},$$

where $\theta_{AVG} := \sum_{j\in\mathcal{M}} \theta_j/m$ denotes the average realized θ.

The realized treatment effect at a given HO is the difference between staff retention rates when the manager is informed versus uninformed. That is, the *realized treatment effect* at a HO with local conditions θ is $RTE(\theta) := S(\theta, \mu_I) - S(\theta, \mu_0)$, while the *average treatment effect* is given by the expected RTE across all possible θ's. That is, $ATE := \mathbb{E}_F[RTE(\theta)]$.

Proposition 2. *For any $\theta \in [1/2, 1]^m$, the realized treatment effect is positive and equal to*

$$RTE(\theta) = \Big(\theta_F - \theta_{AVG}\Big)^2 + \Big[1 - \frac{1}{(1+2\gamma rn)^2}\Big]\theta_{VAR} \geqslant 0,$$

where $\theta_{VAR} := \sum_{i\in\mathcal{M}}(\theta_{AVG} - \theta_i)^2/m$ denotes the dispersion of local conditions in a given HO. Moreover, the average treatment effect is

$$ATE(\gamma, m) = \sigma_F^2\Big[1 - \frac{1}{(1+2\gamma rn)^2}\Big(\frac{m-1}{m}\Big)\Big].$$

Proposition 2 has three immediate implications. First, the realized treatment effect is strictly positive for almost any vector of local conditions.[11] The intuition is that an informed manager can better adapt her action to the HO's local conditions and, hence, achieve a higher staff retention rate. Second, the realized and the average treatment effects are strictly decreasing in managerial skill. A higher-skill manager can more easily adjust each local action and, hence, respond more to information.

[10]The results carry through if only partial information is provided.
[11]It is equal to zero only if all hospitals have local conditions equal to θ_F.

Third, the average treatment effect is strictly decreasing in the number of hospitals of the HO. The larger the HO, the harder it is for the manager to adapt her choices to each hospital's local conditions.

Average treatment effect on the treated across cohorts. In an empirical analysis, often we can only estimate the average treatment effect on the treated (ATT). Suppose that the informational intervention described above is implemented in a group (we also refer to it as a cohort) of HOs. The ATT for a given group of hospital organizations is then the average realized treatment effect given the distribution of types in such a cohort. Each HO has its own local conditions θ, and each cohort might have a different distribution of local conditions. Suppose that a given cohort has distribution G of local conditions. Then, the ATT in such a cohort is given by $ATT(G) := \mathbb{E}_G\big[RTE(\theta)\big]$, where θ is distributed according to G.

Proposition 3. *For a given distribution of local conditions G, the average treatment effect on the treated is*

$$ATT(G) = \Big(\theta_F - \mathbb{E}_G[\theta_{AVG}]\Big)^2 + \mathbb{V}_G\Big[\theta_{AVG}\Big] + \Big[1 - \frac{1}{(1+2\gamma rn)^2}\Big]\mathbb{E}_G\Big[\theta_{VAR}\Big], \quad (2)$$

where $\mathbb{V}_G[\theta_{AVG}]$ denotes the variance of the local average conditions, and $\mathbb{E}_G[\theta_{VAR}]$ denotes the expected within HO local conditions dispersion under distribution G.

Proposition 3 implies that the average treatment effect on the treated is larger for cohorts where the expected average local condition is further away from the prior mean, where there is a larger variance of local average conditions, or where there is a higher expected within-HO variance. That is, distinct distributions of local conditions generate heterogeneous ATTs. In the Corollary below, we highlight two important comparisons: one regarding a lower initial staff retention rate but with constant within cohort average local conditions dispersion, the second with constant initial retention but distinct within cohort average local conditions dispersion.

Corollary 1. *Let G and $\hat{G}$ be two distributions of local conditions. Suppose that both distributions have the expected average type below the prior mean, and the within-HO dispersion of local conditions is the same across them.*
That is, $max\big\{\mathbb{E}_G\big[\theta_{AVG}\big], \mathbb{E}_{\hat{G}}\big[\theta_{AVG}\big]\big\} \leqslant \theta_F$ and $\mathbb{E}_G\big[\theta_{VAR}\big] = \mathbb{E}_{\hat{G}}\big[\theta_{VAR}\big]$. Then, $ATT(G) > ATT(\hat{G})$ if any of the two following conditions hold:

1. $\mathbb{V}_G[\theta_{AVG}] = \mathbb{V}_{\hat{G}}[\theta_{AVG}]$ *and* $\mathbb{E}_G[S(\theta, \mu_0)] < \mathbb{E}_{\hat{G}}[S(\theta, \mu_0)]$;
2. $\mathbb{E}_G[S(\theta, \mu_0)] = \mathbb{E}_{\hat{G}}[S(\theta, \mu_0)]$ *and* $\mathbb{V}_{\hat{G}}[\theta_{AVG}] < \mathbb{V}_G[\theta_{AVG}]$.

Part 1 of Corollary 1 implies that among cohorts with initial retention below the mean, the policy will have a larger effect on cohorts with the lowest average initial retention. Part 2 implies that for a given average initial retention level, the ATT is larger for cohorts with larger dispersion of local conditions.

Empirical Predictions (EP). Our model yields five empirical predictions describing the effect of an informational intervention on staff retention, which we refer to in the results sections:

EP1. The intervention positively affects retention for all local conditions.

EP2. Fixing the within cohort average local conditions' dispersion, the ATT is larger for cohorts with lower initial staff retention rates.

EP3. Fixing the average initial retention rates, the ATT is larger for cohorts with higher within cohort average local conditions' dispersion.

EP4. The effect is larger for HOs with more skilled managers.

EP5. The effect is smaller for HOs with more hospitals.

4 Methods

4.1 Data sources

We construct our monthly panel of NHS HOs by linking several micro-level datasets. The measures of retention are built from the individual-level monthly Electronic Staff Records (ESR) 2009-2020 dataset. The ESR is an administrative dataset that contains monthly payroll information, along with basic demographic characteristics (e.g. age, gender and ethnicity) of all employees working in the English NHS. The information on the RDSP implementation comes from NHS Improvement (NHSI), which was the NHS monitoring body responsible for the development and execution of the intervention, and it contains details about the timing of the Programme's roll-out and the subthemes chosen by HOs.[12]

[12] Before the July 2022 merger with NHS England (NHSE), NHSI worked alongside NHSE and the Department of Health and Social Care to monitor, oversee and support NHS HOs to guarantee the provision of healthcare services to patients.

We complement our panel with the information on nurses' attitudes toward work and perceptions of their workplace using individual-level data from the NHS Staff Survey (NSS) 2014-2018, which we re-aggregate at HO level. The NSS are annual staff surveys commissioned by the NHS to collect information on NHS employees' experiences and well-being at work (NHS England, 2022). The NSS data is a valuable resource to understand the differences in nursing staff's beliefs and perceptions about their workplace, which can be correlated with the ability of a given HO to retain its nursing workforce. In the regression analysis, we exploit HO-level variables from the NSS data before the RDSP was launched as baseline covariates to enforce one of our main robustness checks, i.e. the difference-in-differences under *conditional* parallel trend assumption. Figure A1 illustrates the relative timings of the measurement of NSS variables and retention outcomes in our data setup.

Finally, to investigate the impact of the RDSP on hospital quality and productivity, we construct a monthly HO-level panel using the Hospital Episode Statistics (HES) data from 2009/10 to 2019/20, which provide detailed information on patient admissions to acute care English NHS hospitals and has been used in many economics studies (e.g. Propper and Van Reenen 2010; Propper et al. 2010; Cooper et al. 2011; Gaynor et al. 2013; Bloom et al. 2015). We use HES Admitted Patient Care data linked to Office of National Statistics (ONS) mortality data at the patient level to calculate: 30-day monthly standardized hospital mortality indicators (SHMI) and emergency re-admission rates for planned admissions, as patient outcomes; and the number of admissions to acute care hospitals and waiting times for planned treatment, as measures of hospital-level productivity.[13]

4.2 Measures of retention

We measure nursing retention in two ways: with the stability rate and with the NHS leaver rate, both are computed for each month on a year-on-year basis by HO.

[13]The 30-day risk-adjusted mortality and readmission HO quality measures have been computed through the standardization method used by NHS Digital, the official data provider and statistics office of the NHS. The risk adjustment is achieved by estimating a logit on binary mortality or emergency readmission patient level outcomes, separately for each year, and using patient gender, age categories, and Charlson co-morbidities as covariates in the specification. The ratio between observed and predicted outcomes is then computed for each HOs-month pair, and rescaled by the unconditional mean outcome across all HOs during the said month.

More specifically, we define the stability rate for nurses and midwives' in HO h at calendar time measured in month t, S_{ht}, as

$$S_{ht} = \left(\frac{\sum_h \mathbb{I}_i\big(\text{ employed in HO } h \text{ at } t \big| \text{ employed at } t-12\big)}{\sum_h \mathbb{I}_i\big(\text{employed in HO } h \text{ at } t-12\big)} \right) \times 100$$

The stability rate indicates the percentage of the nurses and midwives who were actively employed in HO h at $t-12$ and were still employed in the same HO at t. This outcome measures how many nurses and midwives are retained in a HO on a year-on-year basis for each month and it accounts for seasonality by comparing the same month a year apart.[14] The stability rate S_{ht} reflects the leaving decisions that occurred between $t-12$ and t, which may have implications for the impact evaluation of the Programme, as we further discuss in Section 6.

The complement to the stability rate is the turnover rate, $100 - S_{ht}$, which we split into churn, i.e. the rate of nurses and midwives' movements between NHS HOs, and the NHS leaver rate (or leaver rate in short), i.e. the rate of nurses and midwives who leave the NHS. Some organizational changes instigated by the RDSP may also discourage nurses from leaving the NHS, therefore we also evaluate the impact of the RDSP on the NHS leaver rates. We calculate the NHS leaver rate, L_{ht}, as the percentage of nurses and midwives who left their organizations at t and have not reappeared in the NHS payroll within the following six months, $t+6$, i.e.

$$L_{ht} = \left(\frac{\sum_h \mathbb{I}_i\big(\text{ left HO } h \text{ between } t-12 \text{ and } t \big| \text{not in ESR until } t+6\big)}{\sum_h \mathbb{I}_i\big(\text{employed in HO } h \text{ at } t-12\big)} \right) \times 100$$

Both retention measures that we consider are policy-relevant. A lower nurse stability rate implies an increase in recruiting costs at the HO-level to hire staff employed elsewhere in the NHS, the private sector or abroad. Instead, high NHS leaving rates imply systemic costs for the public hospital sector, which has to fill the resulting workforce shortage by training new workers or recruiting healthcare professionals from abroad.

[14]For example, if one HO had 100 nurses and midwives in April 2017 and of those nurses and midwives 85 of them remained in the same HO in April 2018, the stability index in April 2018, $S_{\text{April 2018}}$, is 85%.

4.3 Empirical Strategy

We employ difference-in-differences (DiD) designs to evaluate the effects of RDSP on NHS HOs nursing staff retention. Specifically, we estimate two-way fixed effects (TWFE) regressions

$$Y_{ht} = \mu_h + \lambda_t + \beta^{TWFE} D_{ht} + \varepsilon_{ht}, \tag{3}$$

and also event-study TWFE specifications

$$Y_{ht} = \mu_h + \lambda_t + \sum_{k=-T}^{-2} \delta_k D_h^k + \sum_{k=0}^{T} \delta_k D_h^k + \varepsilon_{ht}; \tag{4}$$

where: $Y_{ht} = \{S_{ht}, L_{ht}\}$ are the retention outcomes of interest, i.e. stability rates S_{ht} and NHS leaver rates L_{ht}, in HO h at calendar time t; μ_h and λ_t are respectively HO and calendar time fixed effects. In Eq. (3), D_{ht} is the treatment indicator taking value 1 in all months after the RDSP start in a given cohort, and β^{TWFE} identifies the overall average treatment effect on the treated (ATT) under the canonical DiD assumptions. In Eq. (4), D_{ht}^k is the event-time indicator for the relative time (in months) to/from RDSP, k; the parameters $\delta_{k\geqslant 0}$ are the estimates of the ATT at times k, and $\delta_{k<-1}$ are pre-treatment estimates conventionally used for testing the parallel trends assumption.[15]

Our empirical setting is suitable to employ a DiD design because there is a substantial number of HOs exposed to the policy, a 'parallel trends' assumption is plausible because RDSP was main policy at national level devoted to decrease NHS nurse turnover, and we can we can rely on a sufficient number of pre- and post-policy periods to control for selection effects through the inclusion of additive HO-specific and time-specific fixed effects.

However, recent methodological advances in the DiD literature have shown that the estimates of β^{TWFE} in Eq. (3) and δ_s in Eq. (4) may be biased when there is a staggered treatment adoption with multiple periods and heterogeneous treatment effects (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021; Sun and Abraham,

[15] As frequently done in the literature, we exclude the month before the RDSP launch, $k = -1$, as the reference period, and we bin together the periods before and after 12 months to/from the RDSP when estimating (Eq. (4)).

2021; de Chaisemartin and D'Haultfœuille, 2020; Borusyak et al., 2021). The bias in $\hat{\beta}^{TWFE}$ stems from several potential sources: the variance weighting of the OLS; using the early-treated units as controls for later-treated units, i.e. making "forbidden" comparisons (Goodman-Bacon, 2021); or treatment effect heterogeneity across units treated at different times (Sun and Abraham, 2021). To avoid the bad comparisons biasing the TWFE, different heterogeneity-robust DiD estimators have been proposed.[16] In our analysis, we evaluate the RDSP's impact on nurse retention outcomes using the methodology proposed by Callaway and Sant'Anna (2021) (CSA) and exploit the variation in the timing of the RDSP across different groups of NHS HOs for identification.[17] We also present estimation results from the traditional TWFE regressions, and check the robustness of our results using Sun and Abraham (2021)'s (SA) interaction weighted approach.

Staggered DiD and heterogeneity-robust estimation

The staggered-DiD approach by Callaway and Sant'Anna (2021) is based on a series of average treatment effects at time t for the cohort first treated at time c, $ATT(c,t)$s. In the context of our study, the cohort-time ATTs are the average treatment effects at calendar time t for HOs that started RDSP at time c. Borrowing notation from Roth et al. (2023), under parallel trends and no anticipation, the average treatment effects in post-treatment period when $c \leqslant t$:

$$ATT(c,t) = E[y_{ht} - y_{h,c-1}|C_h = c] - E[y_{ht} - y_{h,c-1}|C_h = NT]$$

which is the difference in the retention between time t and $c-1$, i.e. the period the RDSP started in cohort c for treated HOs in cohort c and the control group, NT. The control group consists of HOs that are either *never-treated*, or *not-yet-treated* up to time t. Likewise, the pre-treatment effects for the period when $c > t$ are

$$ATT(c,t) = E[y_{ht} - y_{h,t-1}|C_h = c] - E[y_{ht} - y_{h,t-1}|C_h = NT]$$

[16] Roth et al. (2023) and de Chaisemartin and D'Haultfœuille (2022) provide comprehensive reviews of recent DiD estimation methods

[17] We use the **did** package version 2.0.0 (Callaway and Sant'Anna, 2020) and the **csdid** package (Rios-Avila et al., 2021) to estimate CSA models respectively in R and Stata 17.

The reference period for comparison during pre-treatment periods is the preceding calendar month, $t-1$. These short-differences are in contrast to the universal reference period of dynamic TWFE or Sun and Abraham (2021), that typically use the last period before the treatment as the reference period for all differences (Callaway, 2021).

CSA also allows for parallel trends to hold after conditioning on covariates through a doubly robust estimation method.[18] Conditioning on trends of observed covariates can be considered as a more credible approach compared to an unconditional parallel trends assumption if the retention trajectories depend on factors that would determine HOs' allocation into cohorts. As the HOs allocation into RDSP cohorts was not randomized, but based on past employee retention outcomes, we provide additional results by assuming parallel trends only conditional on past retention values and other covariates. We use the difference between the average retention rates in 2016 and 2012 of each HO to control for differential trends in past retention. To capture baseline differences across workplace environments, we control for the average level of hospital work stress as of reported by nurses and midwives in the 2015 and 2016 NHS Staff Surveys. To account for hospital size and organizational complexity, we control for the average number of hospital nurses in 2016 and the number of hospital sites.

An additional advantage of the CSA approach is the possibility to aggregate dynamic $ATT(c,t)$'s estimates into a single policy effect or by cohort, i.e. treatment timing:

$$\overline{ATT} = \sum_{c \in C} \sum_{t} \omega(c,t) \cdot ATT(c,t), \tag{5}$$

where the relevant weights, $\omega(c,t)$, are respectively chosen by different lengths of treatment exposure (event-study $\overline{ATT}$s), or by the time each cohort spends under treatment (cohort-specific $\overline{ATT}$s).

HOs in Cohort 5 started the RDSP only in September 2019, which was shortly before the *de facto* end of the RDSP in January 2020 and also very close to the last period of our observation window. Thus, we consider HOs allocated to the first four cohorts as treated, and HOs in Cohort 5 as controls, i.e. *never-treated* HOs.[19]

[18]This method combines inverse probability weighting (matching) and outcome regression method to minimize mis-specification bias. The doubly-robust approach of Sant'Anna and Zhao (2020) is adapted to work under multi-period and multi-group settings in Callaway and Sant'Anna (2021).

[19]Using Cohort 5 as the never-treated comparison group has the advantage to allow us the estimation of the RSDP impact on each cohort for at least 12 months. To do so, we restrict the estimation

The setup of the RDSP as a staggered treatment fits well with the identifying assumptions of CSA's difference-in-differences approach: the RDSP was irreversible, and once the RDSP action plans were implemented by HOs, the policies remained in place for the remainder of the sample period. There should be no anticipation effects of the treatment on treated organizations, as HOs were informed about their involvement in the RDSP only 6 to 8 weeks in advance, with limited information about the scope and the extent of the Programme as this was clarified by the initial Retention Masterclass workshop. The short notice period minimizes the risk of potential anticipation for individual HOs.

Measuring the effects on quality

NHS HOs are frequently subject to localized policy interventions aimed at improving patient care, these may occur at HO level or amongst selected groups of HOs. These policies often prioritize HOs falling short in terms of minimum care quality standards, which is frequently also negatively associated with high staff turnover. In this case, and specifically when the pre-intervention trends in hospital performance indicators are affected by the presence of interventions different from RDSP and aimed at raising care quality standards or hospital productivity, the assumptions required by the Callaway and Sant'Anna (2021) estimator may not be fully satisfied.[20]

For this reason, to study the effects of the RDSP on hospital care quality and productivity we rely on the synthetic difference-in-difference (SDiD) approach (Arkhangelsky et al., 2021; Porreca, 2022). Intuitively, the SDiD estimator allows us to compare the change in quality performance of treated HOs over time with that of a re-weighted set of never-treated HOs, where weights are chosen with the explicit objective of satisfying the parallel trend assumption prior to the RDSP start. SDiD is appealing not only because it allows a staggered treatment, which is required by the structure of the

period to end in November 2019, which coincides with the end of action plan submissions for Cohort 5. We also check the robustness of our results by defining *not-yet-treated* HOs, i.e. all the HOs belonging to cohorts that have not yet started the intervention, as the control group.

[20]While it is realistic to assume that the parallel trend assumption (PTA) holds for labor supply outcomes such as stability and NHS leaving rates, given that RDSP was the main policy rolled-out in that domain during that time, it is less likely that PTA holds with respect to HO quality and productivity outcomes, due to other local interventions frequently occurring in the NHS and not systematically tracked at national level.

RDSP, but also because of its desirable properties to recover valid counterfactuals and relatively unbiased treatment effects with respect to canonical DiD estimators including the Callaway and Sant'Anna (2021) one, according to recent simulation (Porreca, 2022) and analytical studies (Chen, 2023). Our SDiD minimization problem can be expressed as

$$(\widehat{ATT}, \hat{\beta}, \hat{\mu}, \hat{\lambda}) = \underset{(\widehat{ATT}, \hat{\beta}, \hat{\mu}, \hat{\lambda})}{\arg\min} \left\{ \sum_{h=1}^{H} \sum_{t=-T}^{T} \left(Q_{h,t} - \beta \cdot RDSP_{h,t} - \mu_h - \lambda_t\right)^2 \hat{\omega}_h \hat{\tau}_t \right\},$$

where $Q_{h,t}$ denotes a monthly hospital quality indicator. Specifically, we investigate whether the RDSP affected quality metrics such as risk-adjusted mortality within 30 days from admission, 30-day risk-adjusted emergency readmission rates for patients previously treated for a planned procedure by hospital h, hospital waiting times for planned procedures and number of hospital admissions.

$RDSP_{h,t}$ is a cohort-specific policy indicator, taking value one after the start of the RDSP (e.g. June and October 2017 for cohorts 1 and 2, respectively) and zero otherwise. To identify the average treatment effect of the RDSP on quality, hospital organizations in the control group are weighted by the cross-sectional weights ω_h, which are estimated to ensure parallel trends between treated and control units over the entire pre-RDSP period (i.e. for $t \in \{-T, ..., 0\}$). τ_t are instead time weights, whose estimation is meant to make constant the difference in quality of control hospitals between the post- and pre-policy period, which is the other key requirement of a standard DiD analysis.

Sample restrictions

In the retention outcomes analysis, the sample includes nursing and midwifery staff working in both acute (including Community and Specialist care) and mental health care HOs. We exclude: a small number of HOs that have undergone mergers and acquisitions processes from July 2016 onward, because in this case it is impossible to unequivocally assign a HO to a certain cohort; one HO with very few nursing staff; one HO with missing workforce information records in ESR; and one HO for which ESR records were collected only towards the end of the sample period.

The retention outcomes sample includes 193 NHS HOs observed from June 2016 to November 2019. This time window allows us to observe all HOs in the treated group for at least 12 months before and after their enrollment into the RDSP assigned cohorts (Cohort 1 in July 2017 and Cohort 4 in November 2018). The HO performance analysis sample includes only acute care HOs, as risk-adjusted health outcomes (i.e., 30-day mortality and unplanned readmissions rates) can be reliable computed only for these organizations, due to data missingness and small number of admissions in both mental health and community HOs.

4.4 Descriptive statistics

The RDSP included a similar number of HOs in each cohort. Table 1 presents an overview of the NHS HOs in each RDSP cohort. The estimation sample is a balanced panel in terms of calendar time, but unbalanced with respect to the time into treatment, due to the staggered and irreversible nature of the RDSP (e.g. earlier cohorts have shorter pre-RDSP periods and longer post-RDSP periods compared to later cohorts).

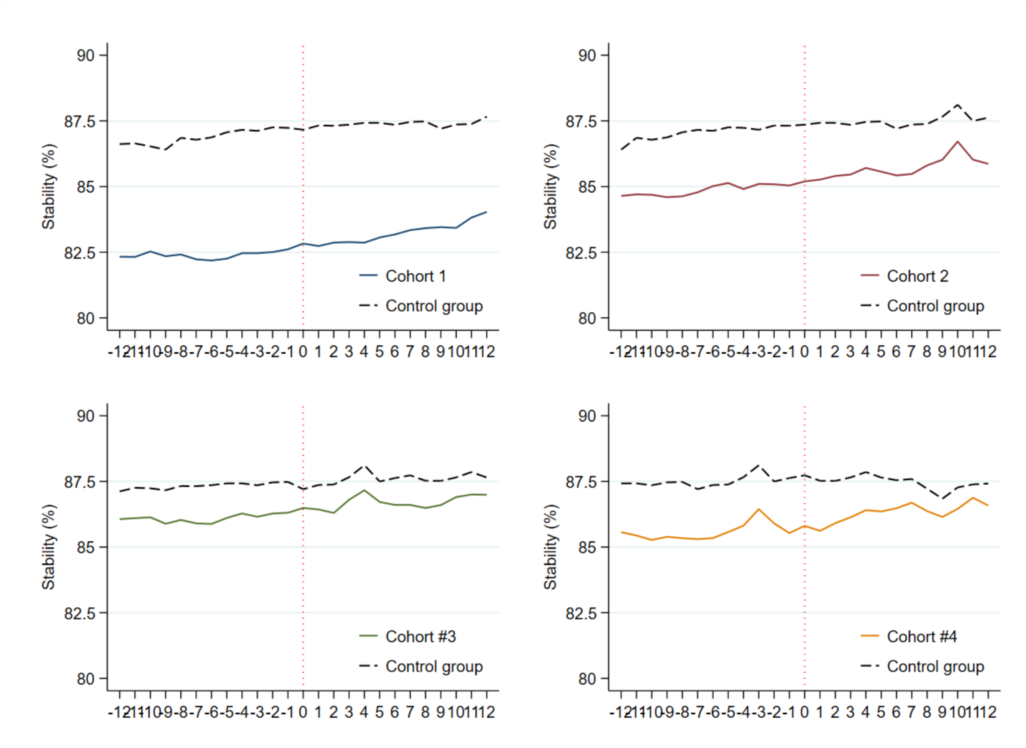
Table 1. Overview of Cohorts in the RSDP

	Cohort 1	**Cohort 2**	**Cohort 3**	**Cohort 4**	**Control group**
RDSP launch (treatment start)	July 2017	October 2017	April 2018	November 2018	-
Number of pre-RDSP periods	13	16	22	29	41
Number of post-RDSP periods	28	25	19	12	
Number of Trusts	31	29	35	37	61
Trust-month observations	1302	1218	1470	1554	2562
Average monthly NHS-leaver rates over past 5 years, 2011/12-2015/16	8.78% (1.95)	8.07% (2.21)	7.79% (1.89)	6.94% (1.83)	6.54% (1.40)
Average monthly stability rates over past 5 years, 2011/12-2015/16	83.57% (2.69)	85.79% (2.91)	86.93% (2.43)	86.53% (3.27)	87.92% (2.35)
	Distribution of past average monthly stability rates				
Bottom quartile	67.74%	27.59%	20.00%	18.92%	9.84%
Second quartile	19.35%	42.38%	22.86%	29.73%	18.03%
Third quartile	9.68%	10.34%	37.14%	24.32%	32.79%
Top quartile	3.23%	20.69%	20.00%	27.03%	39.34%

Notes. Control group consists of HOs in Cohort 5 and two additional HOs that were not included in the RDSP list. Past retention is the average monthly stability rates between 2011/12 - 2015/16, and the table shows the share of HOs in each quartile within cohorts.

In the five years before the start of the RDSP (2011/2 to 2015/16), the average monthly stability rate of nurses and midwives stood at 86.46% and the NHS leaving rate for nurses was 7.15%. The first cohort of the RDSP had the lowest average stability rate over this period, with more than two-thirds of its HOs in the bottom quartile of the pre-RDSP stability distribution. Compared to the first Cohort, only 27.59% of HOs in Cohort 2 were in the bottom quartile in terms of stability (Table 1), and more than half of the HOs in Cohorts 3 and 4 were from the top two quartiles. Nevertheless, Figure A2 shows that the distributions of pre-RDSP retention measures substantially overlap among the treated and the control groups.

Figure 4. Common trends between treated and control cohorts



Notes. Figures are centered at the time RDSP was launched in Cohorts, HOs, and are balanced for relative time periods. The vertical dashed line indicates the timing of the RDSP, and the figures show 12 months before and 12 months after the RDSP.

A close visual inspection of the pre-trends in Figure 4 (Appendix Figure A3), reveals two important facts. All treated cohorts exhibit stability rate (NHS leaver rate) trends similar to the control cohort in the months leading up to the RDSP, which is reassuring for the plausibility of the parallel trend assumption. Moreover, the average pre-trends of HOs belonging to different cohorts present remarkably similar trends, which is reassuring for the fact that the treatment allocation of HOs into cohorts was based on HOs' pre-RDSP retention *levels* but not on retention *trends*.[21]

We also test the mean differences in retention outcomes between *treated* and *control* cohorts for the month before the RDSP launched and at the end of the sample period (Appendix Table A1): during the pre-RDSP period all cohorts had significantly lower stability rates and higher NHS leaver rates than the control group (except for Cohort 4). Finally, Appendix Table A2 presents summary statistics for selected characteristics from the month before the RDSP was first launched (June 2017).

5 Main Results

5.1 Effects on nurse retention outcomes

Figure 5 is the event-study plot presenting the estimated monthly $\overline{ATT}$s aggregated over different cohorts, during the 12 months leading and following the launch of the RDSP; we report estimates from alternative difference-in-differences estimators: Callaway and Sant'Anna (2021), Sun and Abraham (2021) and the naive TWFE. The parallel trends assumption is supported by the pre-trend test proposed by CSA[22], and, according to the formal sensitivity test proposed by Rambachan and Roth (2023), the pre-trends are not sensitive to potential linear extrapolation of parallel trends.[23]

On average, the impact of the RDSP took off after the first three months, with increasing RDSP impacts over time. This is consistent with the 90 day period when HOs designed their retention improvement Action Plans. We find similar results using both Sun and Abraham's (2021) interaction-weighted estimator and dynamic TWFE.

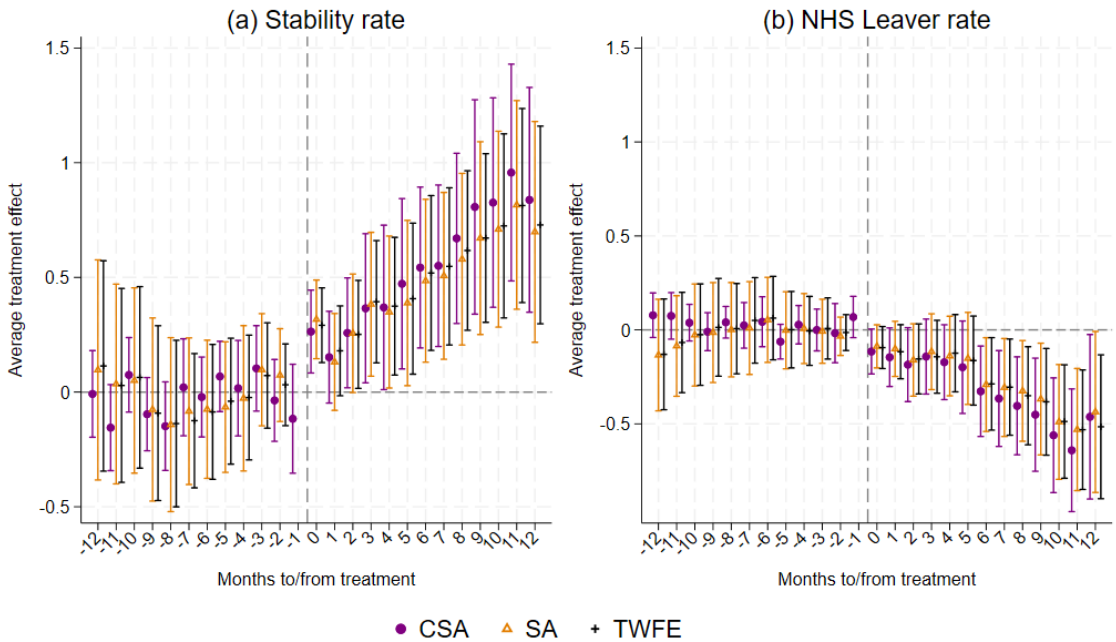
The overall and cohort-specific $\overline{ATT}$s of RDSP on retention outcomes, under unconditional (columns 1 and 3) and conditional (columns 2 and 4) parallel trends assumption, are presented in Table 2. The Programme improved nurses' and midwives'

[21] This fact is further supported by the monthly trends of stability rates (NHS leaver rates) from early 2011/12 to 2016/17 reported in Appendix Figures A4, panels (a) and (b)).

[22] We test the hypothesis whether the pre-treatment estimates for the lead 6 (12) months are jointly equal to zero using an augmented Wald test. We fail to reject with a p-value of 0.273 (0.080).

[23] We test the sensitivity of the slope (smoothness) restriction that the post-treatment violations in parallel trends cannot deviate much from a linear extrapolation of pre-treatment trends by a factor M. $M = 0$ indicates that counterfactual difference in trends is linear and $M > 0$ means non-linearity. Figure A5 presents the results from the sensitivity analysis for different exposure periods at different values of M.

Figure 5. Event-study estimates of the RDSP effects on nurse retention outcomes



Notes. Event study estimates based on aggregated cohort-time ATTs under unconditional parallel trends assumption. Both the Sun and Abraham (2021) and TWFE specifications use universal reference period, i.e omits the month before the RDSP, $k = -1$ and bin the periods for $k < -12$ and $k > 12$, rather than trimming the sample.

Table 2. Effects on retention outcomes (TWFE and Callaway and Sant'Anna (2021))

	Stability rate		**NHS leaver rate**	
β^{TWFE}	0.472*** (0.168)		-0.230* (0.126)	
Overall ATT	0.775*** (0.188)	0.621*** (0.234)	-0.439*** (0.131)	-0.302* (0.155)
Cohort 1 ATT	0.950*** (0.353)	0.772* (0.419)	-0.488** (0.236)	-0.434 (0.269)
Cohort 2 ATT	0.677** (0.295)	0.658* (0.349)	-0.425* (0.223)	-0.268 (0.218)
Cohort 3 ATT	0.557 (0.356)	0.230 (0.490)	-0.455* (0.267)	-0.006 (0.333)
Cohort 4 ATT	0.912*** (0.268)	0.834*** (0.231)	-0.393** (0.184)	-0.443** (0.207)
Conditional parallel trends (PTA)	✗	✓	✗	✓
PTA p-value (12 months)	0.080	0.689	0.135	0.214
PTA p-value (6 months)	0.273	0.656	0.623	0.646

Notes. Standard errors are bootstrapped using 1,000 replications and clustered at HO level. Significance levels: *p<0.1; **p<0.05; ***p<0.01. The estimation period ends in November 2019 for stability rate and October 2019 for leaving the NHS rate.

stability by 0.78ppt (column 1), which is equivalent to a 0.92% increase in the average pre-treatment stability rate or to a 4.49% decrease in the nurse turnover rate.[24]

[24] Estimating the model with stability rate as the outcome is equivalent using turnover rate as the outcome, since the turnover rate is the complement to 1 of the stability rate.

The doubly-robust $\overline{ATT}$ estimate is only slightly smaller and equal to 0.621ppt. The overall effect under unconditional parallel trends is almost twice the magnitude of the TWFE DiD estimate (0.47ppt), which does not account for the differential treatment timing.[25] The RDSP led also to a decrease in the NHS leaver rates of nurses and midwives in the treated cohorts. The NHS leaver rate dropped on average by 0.32ppt (0.60ppt) at 6 (11) month into the Programme. By aggregating the cohort-specific $\overline{ATT}$s across all cohorts, we find that the RDSP decreased the NHS leaver rates on average by 0.439ppt, which is a 5.38% reduction in the pre-RDSP NHS leaver rates.[26] Moreover, the results under conditional parallel trends are still negatively signed, but one-third smaller in magnitude.

Effects by treatment cohort and reversion to mean

Consistently with EP1, we find that the RDSP increased the stability of nursing staff in all cohorts, particularly in the first and last cohorts. On average, Cohort 1 increased its nursing stability rate by 1.15% (=0.95/82.6%) in 28 months, which is equivalent to a one-third of a standard deviation based on their pre-RDSP stability rate, and decreased the NHS leavers rate by 5.6% (=-0.49/8.73%). Despite spending less than half of Cohort 1's time under RDSP, Cohort 4 HOs improved their nurses' stability (NHS leaving rate) by 1.06% (5.65%) in 12 (11) months. Appendix Figure A7 reports the monthly event-study $ATT(c,t)$ estimates of the Programme by each cohort, for both retention outcomes of interest.

The results by cohort provide support to Proposition 3 and empirical predictions EP2 and EP3 (see Section 3) of our conceptual framework, in that we estimate larger $\overline{ATT}$ for the group of HOs with pre-policy staff retention farther away from the average (e.g. Cohort 1) and with larger initial dispersion (e.g. Cohort 4). Moreover, Cohort 4's large significant $\overline{ATT}$s on stability and NHS leaving rates, which are of the same magnitude of Cohort 1's $\overline{ATT}$s, suggest that the RDSP estimated positive

[25]Figure A6 provides the Bacon decomposition of the TWFE DiD estimate. Both comparing only treated cohorts that are treated at different timings (*Timing groups*) and comparing each treated cohort with the never-treated cohort (*Never treated vs timing*) returns very similar average treatment effects.

[26]Similar results are provided in Table A3, which replicates the CSA analysis on the retention rates of nurses and midwives that are younger than 65.

retention gains are not simply mechanical and due to reversion to the mean.[27]

Effects by pre-policy managerial quality and organizational complexity

We also investigate how the effects of RDSP varied depending on the HO managerial quality (proxied by nursing staff satisfaction with line managers care for nurse wellbeing and provision of feedback on nurse work) and HO organization complexity (proxied by the number of hospital sites within each HO), by splitting the sample based on whether the pre-policy management quality variables or organization complexity were above or below the median in the two years before the RDSP launch. Table 3 results, consistently with Proposition 2, EP4 and EP5 of our conceptual framework, show that the RDSP effects on nurses' stability rates were larger in HOs with a higher perceived line manager quality, and also in HOs with a smaller number of hospital sites. These findings support the theoretical models insights about the importance of information distribution and managerial action in the RDSP.

Table 3. Effects on stability rates by managerial quality and organizational complexity

HO characteristics	HO value vs distribution	ATT	S.E.	6 months pre-trend test p-value
Line manager caring for staff well-being	$< 50^{\text{th}}$ percentile	0.476*	(0.261)	0.181
Line manager caring for staff well-being	$> 50^{\text{th}}$ percentile	1.156***	(0.266)	0.147
Line manager providing feedback	$< 50^{\text{th}}$ percentile	0.621**	(0.268)	0.400
Line manager providing feedback	$> 50^{\text{th}}$ percentile	0.944***	(0.266)	0.165
Hospital number of sites	$< 50^{\text{th}}$ percentile	0.918***	(0.235)	0.604
Hospital number of sites	$> 50^{\text{th}}$ percentile	0.546*	(0.301)	0.000

Notes. Managerial quality variables computed at HO level from NSS 2015 & 2016 data on nurses and midwives only. Standard errors bootstrapped (1,000 replications), clustered at HO-level. $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Effects on retention outcomes by RDSP subthemes

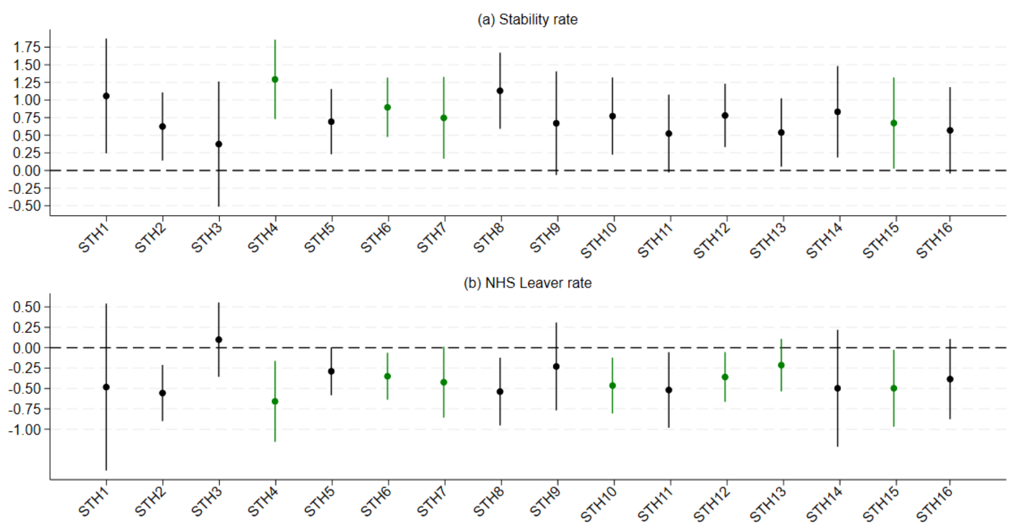
We would like to understand if some HO intervention strategies were more successful than others in improving retention outcomes. To do so, we exploit detailed

[27] If the RDSP effect was due to mean-reversion, Cohort 4's estimated $\overline{ATT}$ should be close to 0. Similarly, Figure 4 reassures that there were no apparent structural breaks (i.e., either drops in stability or jumps in NHS leaving rates) in the retention outcomes of the control group Cohort 5, after the launch of RDSP in the treated cohorts.

information about the HO strategies agreed with NHSI in the HO action plans. The data on action plans is available only for a large subset of HOs from Cohort 1 to 4 (i.e. 112 out of 132 treated HOs).[28] As discussed in Section 2, the subthemes groups were created ex post by NHSI, based on the approved RDSP plans.[29] Table A5 presents the frequency of the subthemes adopted by HOs within cohorts.

We evaluate the impact of each subtheme on nursing staff retention outcomes by estimating separate regressions for each subtheme using the Callaway and Sant'Anna (2021) method. The treatment $d_{hca} = 1$ is a binary indicator taking value 1 if a HO h in cohort c adopted an action plan including strategies falling under subtheme a; $d_{hca} = 0$ for the Cohort 5 control group.

Figure 6. Effects of RDSP subthemes on retention outcomes



Notes. Effects of RDSP action plan subthemes on nurses and midwives' stability (panel a) and NHS leaver rates (panel b), under unconditional parallel trends assumption. 95% confidence bands from boostrapped standard errors (1,000 replications) clustered at HO level. ATT coefficients are coloured in green if the corresponding sub-analysis by RDSP subtheme satisfies the parallel trend assumption in the 6 months preceding the policy at the 5% level. Subthemes classification as in Figure 3.

Figure 6 presents the overall associations between subtheme adoption and nursing

[28]We further check the robustness of our main results by estimating the RDSP effects using only the 112 treated HOs with available action plans data. Under unconditional parallel trends, the overall $\overline{ATT}$s of the Programme are a 0.81ppt increase in stability and a 0.44ppt reduction in NHS leaver rates, therefore very similar to those in Table 2.

[29]Table A4 reports the list of action points defining each subtheme.

staff's stability (panel a) or NHS leaver rates (panel b).[30] The majority of subthemes appears positively associated with higher retention of nurses and midwives in treated HOs.[31] Furthermore, we investigate whether HOs experiencing large gains in retention outcomes, after the RDSP launch, adopted different action plan strategies from HOs with smaller gains. For each HO in the treated cohorts and with an available action plan data, we compute changes in retention outcomes defined as the 12 months change in HO h retention after RDSP minus the average 12 months change in Cohort 5 HOs h retention during the same period – effectively a naive HO-by-HO DiD. We then rank the 12 months changes in ascending order of retention gains, and plot the frequencies of subthemes adoption in the top and bottom quintiles of the retention gains distribution. According to Appendix Figure A8, which reports the results for gains in stability rates, HOs with larger gains over 12 months in stability rates focused more on retention as part of a wider HO strategy (STH4) and improving recruitment (STH11), and less on gathering and understanding data (STH2), being a supporting employer (STH7) and improving communications (STH9).[32]

The HO action plans with larger gains in the stability (NHS leaver) rates included action points related to: learning from those who leave; retention as part of a wider HO strategy; friendly workplace; itchy feet conversations; support for new starters; increased flexibility (gathering and understanding data; retention as part of a wider HO strategy; friendly workplace; improved recruitment; itchy feet conversations). The above results can be only interpreted as suggestive associations, because the subthemes adopted are most likely endogenous and possibly due to baseline unobservable differences between each treated HO, other treated HOs and the control group HOs. Nevertheless, these results provide some guidance on which HO interventions might be most effective. It is notable that none of the subthemes are having a negative effect on retention outcomes. Moreover, apart a handful of cases, the 'treatment effects by subthemes' are rather similar, having overlapping 95% confidence intervals. Overall, the evidence above suggests that managers chose action plan strategies that ultimately proved effective in their local context.

[30] Table A6 reports the estimated $\overline{ATT}$s for each subtheme.

[31] Based on a lead 6 months pre-trend test, the unconditional parallel trends assumption holds for a subset of the estimates (those in green).

[32] Similar results, shown in Figure A9, are found with respect to the distribution of NHS leaving rate gains.

Heterogeneous effects by workers' and job characteristics

Our main analysis is based on the whole population of nurses and midwives actively employed by NHS HOs. We further assess whether the impact of the policy differed for particular groups of nurses and midwives. We recompute the stability rates for nursing staff based on their job seniority, ethnic minority (BAME) status and nationality (British versus non-British). We define nurses' and midwives' seniority either by pay bands levels or by their age. Pay bands broadly define two skill groups, with senior nurses in pay bands 6 and higher.[33] As experience is highly correlated with age, we also divide the sample by age, assuming that nursing staff who are 41 and older are more experienced than their younger colleagues. As shown in the event-study Figure A10 plots, the intervention had quicker and stronger effects on retention for less experienced nursing staff, although the RDSP improved both junior and senior nursing staff's stability rates, respectively by 1 ppt and 0.57ppts. We find similar results when we use the age split, with increases in the stability rates of younger (older) nurses by 0.98ppt (0.62ppt) on average.

Figure A11 presents the event-study plots by ethnicity and nationality status. The Programme improved the retention of nurses and midwives from a white background by 0.66ppt on average. Only one-fifth of the nursing staff in treated cohorts are from an ethnic minority background (see Table A2); the Programme had some positive, albeit not statistically significant, effects on the stability rates of Black, Asian and other minority ethnic groups. RDSP also increased the stability rates of British nursing staff and had no significant effect on non-British nurses and midwives.

5.2 Effects on hospital quality and productivity

There is limited empirical evidence on the impact of non-monetary policies aimed at improving staff retention on patients' health outcomes and hospital productivity. Moreover, the direction of the effect is a priori ambiguous. The RDSP improved nursing retention in treated hospitals; this is likely to affect hospital performance positively by increasing the level of hospital workforce resources available to deliver timely patient care. On the other hand, RDSP may have also resulted in a 'distraction

[33] In all treated cohorts, there were more senior nurses and midwives and the stability rates for junior nursing staff was significantly lower than senior nurses' in the pre-treatment period.

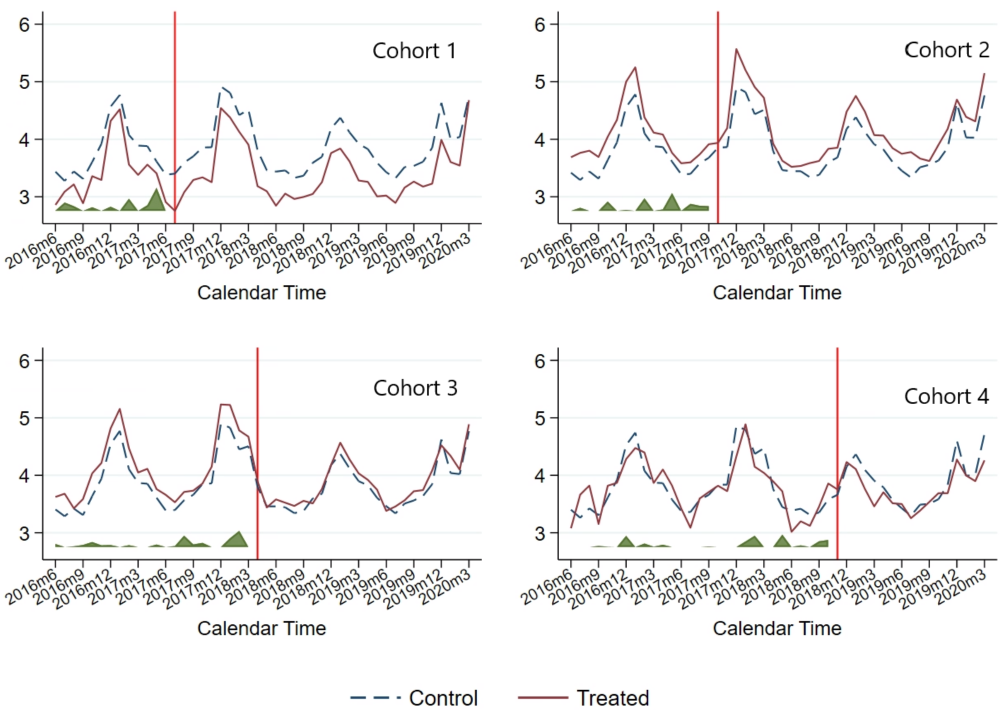
of efforts', if the set of actions delivered through RDSP redirected some effort that might have been dedicated to patients towards improving nurses' working conditions.

Table 4. Effects on health outcomes and hospital productivity measures

	Quality		Productivity		Retention	
	30-day in/outside hospital mortality (all patients)	30-day emergency readmission (planned patients)	Waiting times (in days; planned patients)	Total Admissions	Stability Index	NHS Leaving rate
	(1)	(2)	(3)	(4)	(5)	(6)
ATT (Std. err.)	-0.132** (0.065)	0.100 (0.132)	-4.147* (2.272)	7.530 (63.754)	0.622*** (0.204)	-0.321** (0.140)
Outcome mean (std. dev.)	3.820 (1.302)	6.800 (1.840)	69.249 (45.303)	4982.007 (2667.210)	85.964 (3.236)	7.211 (1.953)

Notes. Average treatment effects estimated using the Synthetic difference-in-difference approach outlined in section 4.3. Bootstrapped standard errors (1,000 replications) in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$ *** $p < 0.01$.

Figure 7. 30-day risk-adjusted mortality rates trends (treated vs synthetic cohorts)



Notes. Synthetic difference-in-difference trends in all patients 30-day mortality from hospital admission. Sample: acute care NHS HOs.

To understand the impact of the RDSP on patient outcomes and hospital productivity, we focus on acute care HOs only. To proxy HO quality we use 30-day standardized HO mortality rates (SHMI) for any HO admission (NHS Digital, 2023) and 30-day emergency readmission rates of patients previously discharged following a planned treatment (NHS Digital, 2022), two widely used measures in the economics literature (Gaynor et al., 2013; Doyle Jr et al., 2015; Moscelli et al., 2018; Friedrich and Hackmann, 2021). To proxy HO productivity, we use HO waiting times (in days) for planned patients and total number of monthly HO admissions.

The estimated ATTs, based on the Arkhangelsky et al. (2021) synthetic DiD estimator, are reported in Table 4. The RDSP effects on mortality within 30 days from admission is negative and significant at 5% level; the -0.132 estimate implies a 3.45% decrease in the all-cause mortality rate in treated HOs with respect to control group HOs, based on a baseline mean mortality rate of 3.82%.[34] Figure 7 plots the 30-day SHMI trends for each HOs cohort and their synthetic counterfactual; the overall HO mortality decrease after RDSP treatment appears to be resulting from an increase in the negative mortality gap in Cohorts 1 and 4, and a tiny decrease in relative mortality in Cohorts 2 and 3, with respect to the respective synthetic controls. The magnitude of the overall mortality effect might appear tiny, but based on the 8,681,000 (i.e. 5,222,548 planned and 3,458,452 emergency) patient admissions occurred in RDSP-treated NHS HOs during financial year 2019/20, i.e. the period when all four cohorts had received the RDSP intervention, our results implies that about 11,441 fewer patients died.[35] Assuming that the value of one year in full health is of £60,000 (Glover and Henderson, 2010)[36], and that each patient has only one year of residual life after discharge from hospital, the effect of RDSP on 30-day all-cause mortality is worth £686.460 ($851.210) millions.[37]

In contrast, we do not find a significant effect on emergency readmissions for planned patients.[38] We also find a 4.1 days decrease in waiting times for planned care, although significant only at 10% level, and we do not find a significant effect on the total number of patient admissions per HO. Finally, we re-estimate the RDSP

[34] -0.132/3.82 = 3.45%. Gaynor et al. (2013) find an effect of comparable magnitude, 0.099 (Table 3, pp. 251), on all-cause mortality rates within a month from HO admission when evaluating the effect of competition in NHS HOs; this is equivalent to a 1% all-cause death rate increase for a 10% increase in the Herfindahl-Hirschman index

[35] 8,681,000 × 3.82% × 3.45%.

[36] Similar values are found by Cutler and McClellan (2001) and Ryen and Svensson (2015).

[37] Even under a much more conservative approach, assuming that patients are not discharged in full health and so with a reduced value of £30,000 per year, i.e. half the full health stock, the estimated RDSP effect on all-cause mortality is large and worth £343.230 ($425.605) millions.

[38] The estimated $\overline{ATT}$ is positive, consistently with the known negative association between mortality and unplanned readmissions rates due to survival bias of frail patients (Laudicella et al., 2013).

effects on retention outcomes using the Arkhangelsky et al. (2021) SDiD estimator to make sure that these are consistent with those estimated with the Callaway and Sant'Anna (2021) approach. The results, in the last two columns of Table 4, show statistically significant estimates with the same sign and point estimates within the 95% uniform confidence bands of the Callaway and Sant'Anna (2021) estimates.

Heterogeneous effects on 30-day mortality by ICD-10 chapters

Table A7 presents the effects of RDSP on risk-adjusted 30-day mortality, estimated separately by homogeneous groups of diagnosis at HO admission through the International Classification of Diseases (ICD-10) chapters. While most of the estimated $\overline{ATT}$s are imprecisely estimated, all but two are negatively signed, as the all-cause 30-day mortality. Moreover, $\overline{ATT}$s on mortality due to diseases related to the respiratory system or pregnancy, childbirth and puerperium are significant at 5% level; therefore, RDSP decreases mortality related to diseases where the UK NHS performance is known to be poor in international comparisons (Salciccioli et al., 2018; Diguisto et al., 2022). Finally, the $\overline{ATT}$s on mortality due to conditions originating in the perinatal period is significant at 10% level.

5.3 Cost-benefit and welfare anaysis

A policy achieving its targets is not necessarily also cost effective. NHSI estimates that it costs £11,400 to replace a nurse (NHS Improvement, 2018), which implies the Programme saved £19,345,800 (i.e. $11,400 \times 1,697$ nurses who did not change hospital organization or leave the NHS) from the NHS budget. However, little information is available about the cost of the Programme. By liaising with NHSI officers, we were made aware that: no additional funding was made available to the treated NHS HOs; the Programme was implemented by existing NHSI liaison officers at Deputy Director level and NHS Trusts officers at Chief Nurse and Chief Nurse Assistant levels; the time spent by NHSI officers was about one full working day per HO per year. Based on the above inputs, publicly available information on NHS workers' salaries and a minimal set of assumptions on the time spent by NHS HOs officers, we compute the back-of-the-envelope opportunity cost estimates for the labor inputs of NHSI and NHS HOs' officers reported in Table A8. In the first instance, we could assume that only NHSI officers' opportunity-cost of time ($\pounds 324,511.3$) should be accounted for, as HO officers should have dedicated some of their time local employee retention policies at HO-level even without the RDSP implementation. If we relax this assumption, and include also the HO officers' opportunity-cost of time ($\pounds 1,118,120.6$ and $\pounds 1,762,015.5$ respectively for HOs' Chief Nurses and Chief Nurse Assistants), an estimate of the total opportunity-cost of time amounts to $\pounds 3,204,647.4$, or £24,277.6 per treated HO. As the value of NHS staff time spent per HO was less than £141,500 (i.e. £19,345,800 divided by the 132 treated NHS HOs), we find that the RDSP was indeed cost-effective. If we consider also the significant reduction of all-cause 30-day mortality, resulting in 11,441 fewer patients died in RDSP-treated HOs and equivalent to an estimated monetary value of £686.5 millions of life gained, the RDSP cost-effectiveness is further strengthened.

Finally, according to the Marginal Value of Public Funds (MVPF) (Finkelstein and Hendren, 2020; Hendren and Sprung-Keyser, 2020) framework, which is estimated as the beneficiaries' willingness to pay for the policy divided by the net cost to government, policies with non-negative effects on beneficiaries welfare and 'pay for themselves' have an infinite MVPF. Given that patient outcomes improve (and very likely nurses welfare too) and the net cost to the UK Government is negative, the RDSP appears to fall into this category and can therefore be considered Pareto improving.

6 Robustness checks

Falsification tests

To check the validity of our results, we perform a series of falsification tests. Table A10 presents estimates of the RDSP effects on nursing staff retention outcomes in which the treatment is placebo obtained by preponing the RDSP start. We use three different time periods before July 2017 and set the RDSP roll-out in the same months but during earlier years.[39] In none of the placebo periods we find a positive effect of

[39]In some of the placebo periods, a few HOs were not open yet. Therefore, there are fewer HOs in some placebo periods. However, as shown in Table A9, the relative size of the cohorts is very

RDSP on stability rates. In a similar fashion, using the same placebo periods, we do not find any significant policy effect on 30-day risk-adjusted mortality (Table A11).

In a different falsification test, we use randomization inference to evaluate the likelihood that our estimates on retention outcomes are false positives. We use the original observation period, from June 2016 to November 2019, randomly allocate HOs into cohorts while keeping the cohort sizes equal to the original allocation, and estimate the effects of interest on retention outcomes. This estimation exercise is repeated for 500 times. Figure A12 presents the overall $\overline{ATT}$ estimates from the randomized cohort allocation: *panel a.* shows $\overline{ATT}$s from all replications, while *panel b.* shows $\overline{ATT}$s from the replications whose random allocation does not violate the parallel pre-trends test at 12 months. Compared to the baseline $\overline{ATT}$, none of the point estimates from the replicated samples is significantly different from zero. We repeat the same randomization inference exercise with 30-day all-cause mortality rates as the outcome of interest (Figure A13): the $\overline{ATT}$ is negative and significant at 5% level only in 3% of the 500 random allocations.

Workers' mobility and spillover effects

DiD designs identify valid causal effects under the Stable Unit Treatment Value Assumption (SUTVA), therefore a potential identification concern in our analysis is whether retention spillover effects occur across HOs from different cohorts, e.g. in case the policy affects HOs who are untreated but geographically close to treated HOs. This could happen either if retention in control group HOs is positively affected by being close to treated areas because they benefit from aspects of the Programme, or alternatively if the control group HOs are negatively affected because their neighbors become a more attractive place to work.

There are several reasons to discount this issue. First, the customized nature of the retention strategies adopted by each HO implies that the strategy adopted by one may not be suitable elsewhere. Additionally, the not-yet-treated HOs would not have access to the bespoke data and support from the supervising NHSI lead officer. Third, the RDSP is a sequential hierarchical intervention, in which a HO action plan is defined by HO managers in the first 90 days, but the labor supply effect that we

similar to the original treatment distribution in all placebo periods.

measure depends on the aggregate responsiveness of the turnover behaviour of individual nursing staff, once the action plan is implemented in a given HO. Nevertheless, we provide some empirical tests to check if joining or leaving patterns of control HOs are affected by the introduction of the policy.

Initially, we focus on workers who changed their HO at least once between June 2016 and November 2019, and look at the cross-cohort transition frequencies of nursing staff, before and after the RDSP was launched in their origin HO. If employees' churn across HOs was motivated by the RDSP introduction, we should find higher transition rates from untreated to treated HOs. Each row in Figure A14 illustrates the transition rates of nursing staff switching HO before (panel a) and after (panel b) the RDSP was launched in their HO (in y-axis), taking into account RDSP's staggered adoption. The green bars indicate transitions rates to not-yet treated HOs, and the diagonal bars show the within cohort transitions, i.e. the share of nursing staff who switched hospitals which were in the same cohort. We find similar transition patterns before and after RDSP was launched.

Next, we look at the impact of the RDSP on the share of new HO joiners, which we further split into new NHS joiners and joiners from other NHS HOs (churn joiners). As shown in Table A12, the impact of RDSP on new joiners (and its further breakdown in NHS and churn joiners) is small in magnitude, negatively signed and never significant for the overall $\overline{ATT}$ as well as in each cohort. These results are an indirect confirmation of our main findings, as fewer nurses and midwives would be hired if the RDSP effectively increased retention rates in treated HOs, resulting in fewer new joiners. We also find no significant effect of RDSP on staff levels and the estimated $\overline{ATT}$s are also negative (Table A12 last column), proving that the improved retention in treated HOs is not mechanically due to increases in staffing levels.

Choice of the reference control group

In our main analyses, we use HOs in Cohort 5 as the *never-treated* control group as they were the last treated cohort, but the treatment lasted only six months due to the onset of the COVID-19 pandemic. To check whether our findings are sensitive to the definition of the comparison group, we re-estimate the effects on retention outcomes with the Callaway and Sant'Anna (2021) estimator, but setting *not-yet-treated* HOs as the control group. Table A13 presents estimation results for stability and leaving the NHS rates. Columns (II) shows the estimates using Cohort 5 HOs as the *never-treated* control group and restricting the sample to end in August 2019 to match the sample period with the *not-yet treated* control group in columns (III). Columns 3 and 6 report the $\overline{ATT}$ estimates using *not-yet treated* HOs as the DiD control group.[40] We find that on average the RDSP increased (decreased) nursing stability (NHS leavers) rates by 0.68ppt (0.37ppt), which is only slightly smaller in absolute value than our baseline estimates. The difference is likely due to the truncation of the post-treatment period up to September 2019: this may particularly affect Cohort 4's $\overline{ATT}$, whose effects are large, as we know from Table 2, but whose the post-treatment outcomes are observed only for nine months when using *not-yet treated* HOs as control group.

Measurement of retention outcomes

As described in Section 4.2, the retention measures are calculated by computing (in each month and HO) the share of nurses who work at time t and were employed in the same HO 12 months ago, $t-12$. This means that until the end of the first 12 months of the Programme, some months fall under pre-treatment period and some to post-treatment, which we can regard as a partial treatment. From 12 months onward, the retention is calculated only using post-RDSP periods, allowing the full effect of the treatment to be observed. This measurement of the retention outcome variables may lead to an underestimation of the $\overline{ATT}$. To check whether this is the case, we can compare the baseline $\overline{ATT}$s in Table 2 with the censored $\overline{ATT}$s in Table A14, which we obtain by averaging the cohort-specific $\overline{ATT}$s from 12^{th} to 19^{th} months into the Programme or from 12^{th} months to the final observations in our sample. Although larger in abdsolute values, the truncated RDSP $\overline{ATT}$ estimates on stability (NHS leaver) rate are not statistically different from the main $\overline{ATT}$s reported in Table 2. From this analysis we conclude that the $\overline{ATT}$s in Table 2 are likely to be conservative point estimates of the Programme's impact, which might have been even more effective in increasing nursing workforce retention.

[40] For comparison purposes, in columns 2 and 5 we report $\overline{ATT}$ estimates using Cohort 5 HOs as *never-treated* control group but restricting the sample period to August 2019, while Columns 1 and 4 report the baseline $\overline{ATT}$s.

7 Conclusions

Staffing pressures are intense in the public sector as demand continues to grow while turnover rates increase. Public sector workers may be responsive to non-financial aspects of their jobs (Ashraf et al., 2014), although relatively little is known about how working conditions can be improved to increase employee retention. This paper examines the impact of the Retention Direct Support Programme (RDSP), a large scale, national-level intervention, which aimed to increase nurse retention in NHS hospital providers. The nature of the policy enables us to investigate the effect of non-financial working conditions on retention, in addition to exploring the role of information sharing for effective management in complex organizations.

We find that the RDSP achieved its objective in terms of reducing nurses' turnover rates. Our most conservative estimates show that the Programme improved the stability rates of nurses and midwives by 0.78ppt on average, or almost a quarter of the between-HO standard deviation in nursing retention. The RDSP led to the retention of 1,697 nurses and midwives who would have otherwise left their hospital organizations. There is a positive, but limited, impact of the Programme in reducing quits from the NHS. These estimates are likely to be conservative due to the nature of our retention outcomes, computed over 12 months. When we focus our attention on the post-treatment period beginning 12 months after the RDSP enrollment, we find even larger proportionate retention gains in terms of HO-specific stability rates and exits from the NHS.

The RDSP succeeded in improving retention, although it was insufficient to resolve the retention problem in NHS hospitals. Nevertheless, it should be emphasized that we might be surprised that the policy had *any* measurable effect: this is a very light-touch intervention, which appears to have worked primarily by filling information gaps on the scale of the problem at the single hospital organization level, and by providing some examples of best practice about how it could be solved. Such an approach has the additional advantage of being relatively cheap, cost-effective and potentially complementary to other policies designed to alleviate workforce pressures. The success of the intervention seems to be driven, in part, by its effect on managerial behavior, as the largest effects on retention are concentrated in hospital organizations that were either better managed pre-policy or with fewer sites. We speculate, although we cannot prove, that the RDSP success is also due middle managers being the main actors – as these have been proved effective in curbing employee turnover in other context (Friebel et al., 2022) – rather than top hospital managers (Janke et al., 2019).

The impact of RDSP on hospital quality through reductions in the mortality risk is also positive, and is concentrated in disease-specific areas (respiratory conditions and maternity services) where the UK NHS performs poorly. Our empirical setting allows only for the estimation of the total effect of RDSP on hospital quality. We cannot evaluate the extent to which the total effect on quality stems from an indirect effect through improvements in staff retention or directly through the increased productivity of existing nurses reacting to improved working conditions brought about by the RDSP. Our results on quality of care are likely driven by both factors, although interestingly we find evidence of no significant increase in the number of patients admitted to hospital and of an imprecise decrease in waiting times for planned patients, implying that direct productivity effects are concentrated in certain dimensions.

Finally, our work contributes to the debate on the trade-off between centralization and decentralization in the management of public organizations (Marschak, 1959; Sah and Stiglitz, 1991; Alonso et al., 2008). From our results, it appears that preserving a certain level of centralization, in terms of disseminating information and providing guidance on best practices, may help decentralized units to overcome information asymmetries. Hence, our findings suggest that an effective configuration of service providers in the public sector may be achieved through an organizational structure in which centralized and decentralized decision-making units cooperate while retaining distinctive functions. Such cooperation, based on a constant exchange of information flows, can be useful to monitor and evaluate the organizational performance of the decentralized branches, and, as demonstrated by the RDSP, may lead to widespread improvements in the targeted outcomes.

References

Alan, S., Corekcioglu, G. and Sutter, M. (2023), 'Improving workplace climate in large corporations: A clustered randomized intervention', *The Quarterly Journal of Economics* **138**(1), 151–203.

Alonso, R., Dessein, W. and Matouschek, N. (2008), 'When does coordination require centralization?', *American Economic Review* **98**(1), 145–79.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W. and Wager, S. (2021), 'Synthetic difference-in-differences', *American Economic Review* **111**(12), 4088–4118.

Ashraf, N., Bandiera, O. and Jack, B. K. (2014), 'No margin, no mission? a field experiment on incentives for public service delivery', *Journal of Public Economics* **120**, 1–17.

Ball, J. E., Bruyneel, L., Aiken, L. H., Sermeus, W., Sloane, D. M., Rafferty, A. M., Lindqvist, R., Tishelman, C. and Griffiths, P. (2018), 'Post-operative mortality, missed care and nurse staffing in nine countries: A cross-sectional study', *International Journal of Nursing Studies* **78**, 10–15.

Ball, J. E., Murrells, T., Rafferty, A. M., Morrow, E. and Griffiths, P. (2014), '"Care left undone" during nursing shifts: associations with workload and perceived quality of care', *BMJ Quality & Safety* **23**(2), 116–125.

Bandiera, O., Barankay, I. and Rasul, I. (2010), 'Social incentives in the workplace', *The review of economic studies* **77**(2), 417–458.

Bandiera, O., Barankay, I. and Rasul, I. (2013), 'Team incentives: Evidence from a firm level experiment', *Journal of the European Economic Association* **11**(5), 1079–1114.

Bloom, N., Liang, J., Roberts, J. and Ying, Z. J. (2014), 'Does Working from Home Work? Evidence from a Chinese Experiment', *The Quarterly Journal of Economics* **130**(1), 165–218.

Bloom, N., Propper, C., Seiler, S. and Van Reenen, J. (2015), 'The impact of competition on management quality: evidence from public hospitals', *The Review of Economic Studies* **82**(2), 457–489.

Bloom, N. and Van Reenen, J. (2007), 'Measuring and explaining management practices across firms and countries', *The Quarterly Journal of Economics* **122**(4), 1351–1408.

Borusyak, K., Jaravel, X. and Spiess, J. (2021), 'Revisiting event study designs: Robust and efficient estimation'. *mimeo.*

Brekke, K. A. and Nyborg, K. (2010), 'Selfish bakers, caring nurses? A model of work motivation', *Journal of Economic Behavior & Organization* **75**(3), 377–394.

Buchan, J., Ball, J., Shembavnekar, N. and Charlesworth, A. (2020), 'Building the NHS nursing workforce in England', *London: The Health Foundation* .

Bungeroth, L., Fennell, E. and S., A. (2018), 'Investing in a Safe and Effective Workforce: Continuing professional development for nurses in the UK'.

Bureau of Labor Statistics, U.S. Department of Labor (2021), 'Registered Nurses'. *Occupational Outlook Handbook*. `https://www.bls.gov/ooh/healthcare/registered-nurses.htm`.

Burgess, S., Propper, C., Ratto, M. and Tominey, E. (2017), 'Incentives in the public sector: Evidence from a government agency', *Economic Journal* **127**(605), F117–F141.

Callaway, B. (2021), 'Universal vs. varying base period in event studies'. *mimeo*.

Callaway, B. and Sant'Anna, P. H. (2020), 'did: Difference in differences'. R package version 2.0.0.

Callaway, B. and Sant'Anna, P. H. (2021), 'Difference-in-differences with multiple time periods', *Journal of Econometrics* **225**(2), 200–230.

Cassar, L. and Meier, S. (2018), 'Nonmonetary incentives and the implications of work as a source of meaning', *Journal of Economic Perspectives* **32**(3), 215–38.

Chan, D. C. (2018), 'The efficiency of slacking off: Evidence from the emergency department', *Econometrica* **86**(3), 997–1030.

Chan Jr, D. C. and Chen, Y. (2022), The productivity of professions: Evidence from the emergency department, Technical report, National Bureau of Economic Research.

Chen, J. (2023), 'Synthetic control as online linear regression', *Econometrica* **91**(2), 465–491.

Cooper, Z., Gibbons, S., Jones, S. and McGuire, A. (2011), 'Does hospital competition save lives? Evidence from the English NHS patient choice reforms', *The Economic Journal* **121**(554), F228–F260.

Cutler, D. M. and McClellan, M. (2001), 'Is technological change in medicine worth it?', *Health affairs* **20**(5), 11–29.

de Chaisemartin, C. and D'Haultfœuille, X. (2020), 'Two-way fixed effects estimators with heterogeneous treatment effects', *American Economic Review* **110**(9), 2964–96.

de Chaisemartin, C. and D'Haultfœuille, X. (2022), 'Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey'. NBER Working Paper Series: 29691.

Dessein, W. and Santos, T. (2006), 'Adaptive organizations', *Journal of Political Economy* **114**(5), 956–995.

Diguisto, C., Saucedo, M., Kallianidis, A., Bloemenkamp, K., Bødker, B., Buoncristiano, M., Donati, S., Gissler, M., Johansen, M., Knight, M. et al. (2022), 'Maternal mortality in eight European countries with enhanced surveillance systems: descriptive population based study', *BMJ* **379**.

Dixit, A. (2002), 'Incentives and organizations in the public sector: An interpretative review', *Journal of human resources* pp. 696–727.

Doyle Jr, J. J., Graves, J. A., Gruber, J. and Kleiner, S. A. (2015), 'Measuring returns to hospital care: Evidence from ambulance referral patterns', *Journal of Political Economy* **123**(1), 170–214.

Duffield, C., Diers, D., O'Brien-Pallas, L., Aisbett, C., Roche, M., King, M. and Aisbett, K. (2011), 'Nursing staffing, nursing workload, the work environment and patient outcomes', *Applied Nursing Research* **24**(4), 244–255.

Duffield, C. M., Roche, M. A., Homer, C., Buchan, J. and Dimitrelis, S. (2014), 'A comparative review of nurse turnover rates and costs across countries', *Journal of advanced nursing* **70**(12), 2703–2712.

Duflo, E., Hanna, R. and Ryan, S. P. (2012), 'Incentives work: Getting teachers to come to school', *American economic review* **102**(4), 1241–1278.

Ellingsen, T. and Johannesson, M. (2008), 'Pride and prejudice: The human side of incentive theory', *American economic review* **98**(3), 990–1008.

Finkelstein, A. and Hendren, N. (2020), 'Welfare analysis meets causal inference', *Journal of Economic Perspectives* **34**(4), 146–167.

Friebel, G., Heinz, M. and Zubanov, N. (2022), 'Middle managers, personnel turnover, and performance: A long-term field experiment in a retail chain', *Management Science* **68**(1), 211–229.

Friedrich, B. U. and Hackmann, M. B. (2021), 'The returns to nursing: Evidence from a parental-leave program', *The Review of Economic Studies* **88**(5), 2308–2343.

Gaynor, M., Moreno-Serra, R. and Propper, C. (2013), 'Death by market power: reform, competition, and patient outcomes in the National Health Service', *American Economic Journal: Economic Policy* **5**(4), 134–66.

Gibbons, R. (1998), 'Incentives in organizations', *Journal of Economic Perspectives* **12**(4), 115–132.

Glover, D. and Henderson, J. (2010), 'Quantifying health impacts of government policies', *London: Department of Health* .

Goodman-Bacon, A. (2021), 'Difference-in-differences with variation in treatment timing', *Journal of Econometrics* **225**(2), 254–277.

Gosnell, G. K., List, J. A. and Metcalfe, R. D. (2020), 'The impact of management practices on employee productivity: A field experiment with airline captains', *Journal of Political Economy* **128**(4), 1195–1233.

Griffiths, P., Maruotti, A., Saucedo, A. R., Redfern, O. C., Ball, J. E., Briggs, J., Dall'Ora, C., Schmidt, P. E. and Smith, G. B. (2019), 'Nurse staffing, nursing assistants and hospital mortality: retrospective longitudinal cohort study', *BMJ Quality & Safety* **28**(8), 609–617.

Griffiths, P., Recio-Saucedo, A., Dall'Ora, C., Briggs, J., Maruotti, A., Meredith, P., Smith, G. B., Ball, J. and the Missed Care Study Group (2018), 'The association between nurse staffing and omissions in nursing care: A systematic review', *Journal of Advanced Nursing* **74**(7), 1474–1487.

Gruber, J. and Kleiner, S. A. (2012), 'Do strikes kill? evidence from new york state', *American Economic Journal: Economic Policy* **4**(1), 127–157.

Helm, C. and Bungeroth, L. (2017), 'Safe and effective staffing: the real picture'.

Hendren, N. and Sprung-Keyser, B. (2020), 'A unified welfare analysis of government policies', *The Quarterly Journal of Economics* **135**(3), 1209–1318.

Hoffman, M. and Tadelis, S. (2021), 'People management skills, employee attrition, and manager rewards: An empirical analysis', *Journal of Political Economy* **129**(1), 243–285.

Holmstrom, B. and Milgrom, P. (1991), 'Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design', *Journal of Law, Economics & Organization* **7**, 24.

House of Commons Health Committee (2018), The nursing workforce: Second Report of Session 2017-19, Technical report, House of Commons. https://publications.parliament.uk/pa/cm201719/cmselect/cmhealth/353/353.pdf.

Janke, K., Propper, C. and Sadun, R. (2019), The impact of CEOs in the public sector: Evidence from the English NHS, Technical report, National Bureau of Economic Research.

Kelly, E., Propper, C. and Zaranko, B. (2022), 'Team composition and the returns to human capital: Evidence from nursing teams'.

Kelly, E., Stoye, G. and Warner, M. (2022), 'Factors associated with staff retention in the NHS acute sector', *The Institute for Fiscal Studies* .

Laudicella, M., Donni, P. L. and Smith, P. C. (2013), 'Hospital readmission rates: signal of failure or success?', *Journal of Health Economics* **32**(5), 909–921.

Lee, T., Propper, C. and Stoye, G. (2019), 'Medical labour supply and the production of healthcare', *Fiscal Studies* **40**(4), 621–661.

Linos, E., Ruffini, K. and Wilcoxen, S. (2022), 'Reducing burnout and resignations among frontline workers: a field experiment', *Journal of Public Administration Research and Theory* **32**(3), 473–488.

Marangozov, R., Williams, M. and Buchan, J. (2016), 'The labour market for nurses in the UK and its relationship to the demand for, and supply of, international nurses in the NHS', *Brighton: Institute for Employment Studies* .

Marć, M., Bartosiewicz, A., Burzyńska, J., Chmiel, Z. and Januszewicz, P. (2019), 'A nursing shortage–a prospect of global and local policies', *International nursing review* **66**(1), 9–16.

Marschak, T. (1959), 'Centralization and decentralization in economic organizations', *Econometrica: Journal of the Econometric Society* pp. 399–430.

Mas, A. and Pallais, A. (2017), 'Valuing alternative work arrangements', *American Economic Review* **107**(12), 3722–59.

McNally, S., Schmidt, L. and Valero, A. (2022), 'Do management practices matter in further education?'. *mimeo*.

Moscelli, G., Gravelle, H., Siciliani, L. and Gutacker, N. (2018), 'The effect of hospital ownership on quality of care: evidence from england', *Journal of Economic Behavior & Organization* **153**, 322–344.

National Audit Office (2020), 'The NHS nursing workforce', *HC 109 Session 2019-*

2021. Department of Health & Social Care . https://www.nao.org.uk/report/nhs-nursing-workforce/.

Nelson, N. and McLaughlin, C. (2020), 'NHS faces fresh crisis as 250,000 nurses consider quitting because of low pay', *Mirror*. Accessed 19/05/2022, available at https://www.mirror.co.uk/news/uk-news/nhs-faces-fresh-crisis-250000-22416470.

NHS Digital (2018), 'Nurse and Health Visitor joiners and leavers from the NHS'. https://digital.nhs.uk/data-and-information/find-data-and-publications/supplementary-information/2018-supplementary-information-files/leavers-and-joiners/nurse-and-health-visitor-joiners-and-leavers-from-the-nhs. Accessed on 15/09/2021.

NHS Digital (2021), 'NHS Vacancy Statistics England April 2015 – June 2021 Experimental Statistics'. https://digital.nhs.uk/data-and-information/publications/statistical/nhs-vacancies-survey/april-2015---june-2021-experimental-statistics. Accessed on 14/09/2021.

NHS Digital (2022), Emergency readmissions within 30 days of discharge from hospital - Specification v1.4, Technical report. Technical report. https://digital.nhs.uk/data-and-information/publications/statistical/ccg-outcomes-indicator-set/specifications/3.2-emergency-readmissions-within-30-days-of-discharge-from-hospital_1_4. Accessed on 10-06-2023.

NHS Digital (2023), Summary Hospital-level Mortality Indicator (SHMI), Technical report. Technical report. https://digital.nhs.uk/data-and-information/publications/ci-hub/summary-hospital-level-mortality-indicator-shmi. Accessed on 10-06-2023.

NHS England (2019), 'Interim NHS People Plan'. https://www.longtermplan.nhs.uk/wp-content/uploads/2019/05/Interim-NHS-People-Plan_June2019.pdf. Accessed on 14/06/2021.

NHS England (2022), 'NHS Staff Survey in England'. https://www.england.nhs.uk/statistics/statistical-work-areas/nhs-staff-survey-in-england/. Accessed on 14/02/2022.

NHS England and NHS Improvement (2019), The NHS Long Term Plan, Technical report. https://www.longtermplan.nhs.uk/publication/nhs-long-term-plan/.

NHS Improvement (2017), 'Retention Support Programme'.

NHS Improvement (2018), 'Treasuring the NHS's greatest asset:retaining our workforce'.

Perreira, T. A., Berta, W. and Herbert, M. (2018), 'The employee retention triad in health care: Exploring relationships amongst organisational justice, affective commitment and turnover intention', *Journal of Clinical Nursing* **27**(7-8), e1451–e1461.

Porreca, Z. (2022), 'Synthetic difference-in-differences estimation with staggered treatment timing', *Economics Letters* **220**, 110874.

Propper, C., Sutton, M., Whitnall, C. and Windmeijer, F. (2010), 'Incentives and targets in hospital care: evidence from a natural experiment', *Journal of Public Economics* **94**(3-4), 318–335.

Propper, C. and Van Reenen, J. (2010), 'Can pay regulation kill? panel data evidence on the effect of labor markets on hospital performance', *Journal of Political Economy* **118**(2), 222–273.

Rafferty, A. M., Clarke, S. P., Coles, J., Ball, J., James, P., McKee, M. and Aiken, L. H. (2007), 'Outcomes of variation in hospital nurse staffing in English hospitals: cross-sectional analysis of survey data and discharge records', *International Journal of Nursing Studies* **44**(2), 175–182.

Rambachan, A. and Roth, J. (2023), 'A more credible approach to parallel trends', *Review of Economic Studies* .

Rios-Avila, F., Sant'Anna, P. and Callaway, B. (2021), 'Csdid: Stata module for the estimation of difference-in-difference models with multiple time periods'. Statistical Software Components S458976, Boston College Dept. of Economics, revised 25-02-2023.

Roth, J., Sant'Anna, P. H., Bilinski, A. and Poe, J. (2023), 'What's trending in difference-in-differences? a synthesis of the recent econometrics literature', *Journal of Econometrics* .

Ryen, L. and Svensson, M. (2015), 'The willingness to pay for a quality adjusted life year: a review of the empirical literature', *Health Economics* **24**(10), 1289–1301.

Sah, R. K. and Stiglitz, J. E. (1991), 'The quality of managers in centralized versus decentralized organizations', *The Quarterly Journal of Economics* **106**(1), 289–295.

Salciccioli, J. D., Marshall, D. C., Shalhoub, J., Maruthappu, M., De Carlo, G. and

Chung, K. F. (2018), 'Respiratory disease mortality in the United Kingdom compared with EU15+ countries in 1985-2015: observational study', *BMJ* **363**.

Sant'Anna, P. H. and Zhao, J. (2020), 'Doubly robust difference-in-differences estimators', *Journal of Econometrics* **219**(1), 101–122.

Shields, M. A. (2004), 'Addressing nurse shortages: What can policy makers learn from the econometric evidence on nurse labour supply?', *The Economic Journal* **114**(499), F464–F498.

Sun, L. and Abraham, S. (2021), 'Estimating dynamic treatment effects in event studies with heterogeneous treatment effects', *Journal of Econometrics* **225**(2), 175–199.

The Guardian (2023), 'Health unions urge Sunak to resolve pay dispute before unprecedented strike', *The Guardian* .

The Open University (2018), Tackling the nursing shortage, Technical report. `https://cdn.ps.emap.com/wp-content/uploads/sites/3/2018/05/The_Open_University_Tackling_the_nursing_shortage_Report_2018.pdf`.

Tikkanen, R., Osborn, R., Mossialos, E., A, D. and GA, W. (2020), 'International health care system profiles'.

Online Appendix A. Proofs

Proof of Proposition 1. Note that

$$\mathbb{E}_\mu[U(a,\theta)] = \sum_{i\in\mathcal{M}} \Bigg\{ -r\cdot n\cdot \mathbb{E}_\mu\Big[1-\theta_i-(a_i-\mathbb{E}_\mu[\theta_i]+\mathbb{E}_\mu[\theta_i]-\theta_i)^2\Big] - \frac{1}{m\gamma}\sum_{j\in\mathcal{M}}\frac{\big(a_i-a_j\big)^2}{2}\Bigg\}$$

$$= \sum_{i\in\mathcal{M}} \Bigg\{ -r\cdot n\cdot \Big[1-\mathbb{E}_\mu[\theta_i]+(a_i-\mathbb{E}_\mu[\theta_i])^2+\mathbb{E}_\mu\big[(\mathbb{E}_\mu[\theta_i]-\theta_i)^2\big]\Big] - \frac{1}{m\gamma}\sum_{j\in\mathcal{M}}\frac{\big(a_i-a_j\big)^2}{2}\Bigg\}.$$

Hence, maximizing $\mathbb{E}_\mu[U(a,\theta)]$ is equivalent to

$$\underset{a\in[0,1]^m}{min} \sum_{i\in\mathcal{M}}\Bigg\{rn(a_i-\mathbb{E}_\mu[\theta_i])^2+\frac{1}{m\gamma}\sum_{j\in\mathcal{M}}\frac{\big(a_i-a_j\big)^2}{2}\Bigg\}.$$

The first-order condition with respect to a_i is

$$2rn(a_i^*-\mathbb{E}_\mu[\theta_i]) - \frac{1}{\gamma}\sum_{j\in\mathbb{M}}\frac{(a_j^*-a_i^*)}{m} = 0 \text{ for all } i\in\{1,...,m\}. \tag{6}$$

Summing over all i, we get $\sum_{i\in\mathcal{M}} a_i^* = \sum_{i\in\mathcal{M}}\mathbb{E}_\mu[\theta_i]$. Replacing it into (6) we get equation (1). To calculate the stability rate substitute each $a_i^*(\mu)$ into p and compute the average across hospitals. □

Proof of Proposition 2.

$$RTE(\theta) = R(\theta,\mu_I)-R(\theta,\mu_0) = \frac{1}{m}\sum_{i\in\mathcal{M}}(a_i^*(\mu_0)-\theta_i)^2 - \frac{1}{m}\sum_{i\in\mathcal{M}}(a_i^*(\mu_I)-\theta_i)^2$$

$$= \frac{1}{m}\sum_{i\in\mathcal{M}}\Big(\theta_F-\theta_i\Big)^2 - \sum_{i\in\mathcal{M}}\Big(\frac{2\gamma rn}{1+2\gamma rn}\theta_i+\frac{1}{1+2\gamma rn}\theta_{AVG}-\theta_i\Big)^2$$

$$= \frac{1}{m}\sum_{i\in\mathcal{M}}\Big[\Big(\theta_F-\theta_{AVG}\Big)^2+\Big(\theta_{AVG}-\theta_i\Big)^2-\frac{1}{(1+2\gamma rn)^2}\Big(\theta_{AVG}-\theta_i\Big)^2\Big]$$

$$= \Big(\theta_F-\theta_{AVG}\Big)^2+\Big[1-\frac{1}{(1+2\gamma rn)^2}\Big]\theta_{VAR} \geqslant 0.$$

While for the average treatment effect

$$ATE(\gamma,m) = \mathbb{E}_F\Big[\Big(\theta_F-\theta_{AVG}\Big)^2+\Big(1-\frac{1}{(1+2\gamma rn)^2}\Big)\theta_{VAR}\Big] = \sigma_F^2\Big[1-\frac{1}{(1+2\gamma rn)^2}\Big(\frac{m-1}{m}\Big)\Big].$$

□

Proof of Proposition 3.

$$
\begin{aligned}
ATT(G) &= \mathbb{E}_G\Big[\Big(\theta_F - \theta_{AVG}\Big)^2\Big] + \Big[1 - \frac{1}{(1+2\gamma rn)^2}\Big]\mathbb{E}_G\Big[\theta_{VAR}\Big] \\
&= \mathbb{E}_G\Big[\Big(\theta_F - \mathbb{E}_G[\theta_{AVG}]\Big)^2\Big] + \mathbb{E}_G\Big[\Big(\mathbb{E}_G[\theta_{AVG}] - \theta_{AVG}\Big)^2\Big] + \Big[1 - \frac{1}{(1+2\gamma rn)^2}\Big]\mathbb{E}_G\Big[\theta_{VAR}\Big] \\
&= \Big(\theta_F - \mathbb{E}_G[\theta_{AVG}]\Big)^2 + \mathbb{V}_G\Big[\theta_{AVG}\Big] + \Big[1 - \frac{1}{(1+2\gamma rn)^2}\Big]\mathbb{E}_G\Big[\theta_{VAR}\Big].
\end{aligned}
$$

□

Proof of Corollary 1. Note that for $\tilde{G} \in \{G, \hat{G}\}$

$$
\mathbb{E}_{\tilde{G}}[S(\theta,\mu_0)] = \mathbb{E}_{\tilde{G}}\Big[\theta_{AVG} - \frac{1}{m}\sum_{i\in\mathcal{M}}\big(\theta_F - \theta_i\big)^2\Big] = \mathbb{E}_{\tilde{G}}\big[\theta_{AVG}\big] - \big(\theta_F - \mathbb{E}_{\tilde{G}}\big[\theta_{AVG}\big]\big)^2 - \mathbb{V}_{\tilde{G}}\big[\theta_{AVG}\big].
$$

1. Then, $\mathbb{E}_G[S(\theta,\mu_0)] < \mathbb{E}_{\hat{G}}[S(\theta,\mu_0)]$ and $\mathbb{V}_G[\theta_{AVG}] = \mathbb{V}_{\hat{G}}[\theta_{AVG}]$ imply $\mathbb{E}_G[\theta_{AVG}] < \mathbb{E}_{\hat{G}}[\theta_{AVG}]$. Therefore, by equation (2), we get the result.
2. Then, $\mathbb{V}_{\hat{G}}[\theta_{AVG}] < \mathbb{V}_G[\theta_{AVG}]$ and $\mathbb{E}_{\hat{G}}[S(\theta,\mu_0)] = \mathbb{E}_G[S(\theta,\mu_0)]$ imply $\mathbb{E}_{\hat{G}}[\theta_{AVG}] < \mathbb{E}_G[\theta_{AVG}]$. Moreover, this implies that

$$
\mathbb{V}_{\hat{G}}[\theta_{AVG}] + \big(\theta_F - \mathbb{E}_{\hat{G}}[\theta_{AVG}]\big)^2 < \mathbb{V}_G[\theta_{AVG}] + \big(\theta_F - \mathbb{E}_G[\theta_{AVG}]\big)^2.
$$

   Therefore, $ATT(\hat{G}) < ATT(G)$.

□

Online Appendix B. Additional Tables & Figures

Table A1. Average retention before and after the RDSP by cohort

	Stability rate		NHS-leaver rate	
	Pre-RDSP	End of RDSP	Pre-RDSP	End of RDSP
Cohort 1	82.613	84.580	8.725	7.363
	(2.838)	(2.758)	(1.805)	(1.876)
Control	87.238	87.419	6.576	6.415
	(2.770)	(3.105)	(1.569)	(1.754)
$\Delta(C1 - Control)$	-4.625***	-2.839***	2.149***	0.948**
	(0.616)	(0.660)	(0.364)	(0.396)
Cohort 2	85.042	86.569	7.702	6.72
	(3.204)	(2.662)	(1.905)	(1.554)
Control	87.319	87.419	6.564	6.415
	(2.811)	(3.105)	(1.776)	(1.754)
$\Delta(C2 - Control)$	-2.277***	-0.851	1.139***	0.306
	(0.664)	(0.670)	(0.410)	(0.382)
Cohort 3	86.307	87.562	7.836	6.791
	(3.210)	(2.286)	(2.303)	(1.823)
Control	87.481	87.419	6.611	6.415
	(2.711)	(3.105)	(1.904)	(1.754)
$\Delta(C3 - Control)$	-1.175*	0.142	1.226***	0.376
	(0.615)	(0.601)	(0.436)	(0.377)
Cohort 4	85.534	86.581	6.957	6.168
	(3.239)	(3.124)	(1.557)	(1.772)
Control	87.629	87.419	6.379	6.415
	(2.619)	(3.105)	(1.796)	(1.754)
$\Delta(C4 - Control)$	-2.095***	-0.838	0.579	-0.247
	(0.597)	(0.648)	(0.356)	(0.367)

Notes. Pre-RDSP averages are calculated for the month before the RDSP was launched in HOs, i.e. the timings for each cohort are June 2017, September 2017, March 2018, and October 2018, respectively. The end of the RDSP statistics are based on the stability rates in November 2019, and on NHS leaver rates in October 2019. For cohorts, standard deviations are in parentheses, and for $\Delta(C - Control)$ standard errors are in parentheses with p-values *p<0.1; **p<0.05; ***p<0.01.

Table A2. Sample summary statistics by cohort, pre-RDSP (measured in June 2017)

	Cohort 1		Cohort 2		Cohort 3		Cohort 4		Control cohort	
	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev
Nursing workforce composition in Trust										
Share of female nurses and midwives, %	87.339	6.296	84.909	9.629	85.913	8.244	91.691	4.663	92.538	2.906
Average age	42.705	2.372	43.648	2.280	43.410	2.166	42.217	2.504	42.676	2.114
Share from the EU, %	9.648	7.002	6.339	4.814	6.138	6.215	6.940	5.032	5.864	5.450
Share from Overseas, %	12.123	7.983	11.295	8.853	8.179	8.072	8.477	6.063	6.467	4.752
Share from ethnic minority background, %	27.635	18.570	25.035	20.766	14.914	14.280	21.275	17.283	12.624	10.050
Other Trust characteristics and outcomes										
All staff headcount (size of Trust)	4,801	2,993	4,914	2,710	5,063	2,598	5,331	3,530	5,280	2,864
Number of nurses and midwives	1,632	1,165	1,557	860	1,671	929	1,707	1,057	1,659	850
Trust age (years from foundation)	18.194	6.660	17.483	6.027	18.200	5.925	20.351	5.554	20.016	6.201
Sickness absence rate, %	4.037	1.137	4.241	0.805	4.357	0.935	4.269	1.019	4.204	0.787
Average hours worked (full-time, $\geqslant 0$)	166.631	4.524	166.341	4.831	167.430	3.475	166.374	3.928	166.824	2.814
Share of Bank hours in average hours worked, %	1.716	2.098	1.564	1.913	1.791	1.868	1.568	1.821	1.489	1.430
Monthly SHMI, emergency patients	2.722	1.320	2.876	0.506	2.963	0.695	2.579	1.201	2.753	0.782
Monthly emergency readmission rate, electives	1.167	0.656	0.882	0.330	1.046	0.298	1.249	1.202	0.989	0.533
Number of emergency admissions	2,780	2,046	3,305	1,803	3,272	1,757	2,691	2,038	2,814	1,648
Number of elective admissions	4,804	4,493	5,162	3,318	4,765	2,992	4,991	4,239	4,866	3,393
NSS 2015 items										
Overall engagement score	7.204	0.275	7.096	0.321	6.999	0.345	7.224	0.315	7.207	0.333
Share of nursing staff (%) who										
Worked at least 11 hours additional unpaid hours per week	6.984	3.001	6.411	2.708	5.200	2.452	5.115	2.325	4.433	1.985
Recognised for good work	53.307	4.284	54.042	5.230	52.962	6.802	51.788	7.268	52.647	7.168
Felt unwell due work stress in the last 12 months	41.076	5.685	42.248	6.357	43.216	7.021	40.234	5.077	39.668	6.414
Satisfied with the support from immediate manager	69.014	4.327	70.055	5.309	70.148	5.807	67.745	6.237	68.494	5.224
Satisfied with the support from colleagues	83.694	4.728	85.245	4.272	85.416	3.242	83.587	4.104	85.800	3.695
NHS regions										
East of England	0.194	0.402	0.172	0.384	0.114	0.323	0.000	0.000	0.098	0.300
London	0.290	0.461	0.276	0.455	0.057	0.236	0.216	0.417	0.082	0.277
Midlands	0.097	0.301	0.172	0.384	0.200	0.406	0.189	0.397	0.180	0.388
North East and Yorkshire	0.032	0.180	0.103	0.310	0.143	0.355	0.189	0.397	0.180	0.388
North West	0.129	0.341	0.103	0.310	0.114	0.323	0.216	0.417	0.180	0.388
South East	0.194	0.402	0.103	0.310	0.257	0.443	0.135	0.347	0.082	0.277
South West	0.065	0.250	0.069	0.258	0.114	0.323	0.054	0.229	0.197	0.401

Notes. Nursing workforce compositions are averages from previous financial year and calculated using the ESR. Staff headcounts come from NHS Workforce Statistics. NSS 2015 items are calculated from individual level data for nurses and midwives. † SHMI and admission numbers are calculated for acute care NHS HOs only, thus the sample sizes for each cohort is smaller than for other summary statistics.

Table A3. RDSP effects on retention outcomes, nurses below legal retirement age

	Stability rate		NHS Leaver rate	
	ATT	Std.Error	ATT	Std.Error
Overall	0.724***	(0.185)	-0.408***	(0.131)
Cohort 1	0.931***	(0.348)	-0.498**	(0.243)
Cohort 2	0.592**	(0.276)	-0.416*	(0.213)
Cohort 3	0.481	(0.355)	-0.394	(0.263)
Cohort 4	0.882***	(0.264)	-0.341**	(0.172)

Notes. CSA estimates under unconditional parallel trends assumption. Retention measures are computed only among nurses that are younger than 65. Bootstrapped standard errors (1,000 replications) clustered at HO level are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A4. Subthemes classification

Subtheme 1 (Learning from those who leave): Improve leaver experience including exit interviews; Review leaver process.
Subtheme 2 (Gathering and understanding data): Improving exit data; Manage processes for collecting leaver data; Identify high turnover 'hot spots'; Look at risk profiles for future turnover; Understanding what improvement is achievable; Understand why people leave within 12 months; Manage the natural churn of band 5 nurses; Identifying areas where it is hard to recruit; Discuss all leavers at nursing, midwifery and AHP taskforce; Metrics and reporting; Review employer of choice status against competition; Review the effectiveness of current systems and processes for capturing leaver data and to implement improvements to the capture and use of such data.
Subtheme 3 (Senior leadership): Effective visible leadership and management; Compassionate management; Corporate social responsibility; Manager engagement; band 5 nurse support; Relationship with manager; Support from management.
Subtheme 4 (Retention as part of wider strategy): Effective appraisal and talent management; Equity of experience and opportunity; Focused approach to retention initiatives; Clear clinical pathways; Ensuring inclusive leadership at all levels; Make substantive employment more attractive than bank work: flexibility, self-rostering, family-friendly policies, study leave, mandatory training; Support for existing staff: improve flexibility, work-life balance, reduce flow of leavers ; identify key difficulties around recruitment Equality and diversity Exploring good practice from within (services that are good) and externally other and retention respectively; Embed retention as BAU activity; Trusts / NHSI; Flexible working; To reduce work pressures on staff which may lead to poor morale and staff turnover; Implement additional methods to effectively share best practices across the Trust; Management capacity developed and increased.
Subtheme 5 (Career progression): Career progression/development/pathways/coaching; Increased opportunities for staff development, CPD and career progression; establishment of new roles/posts/career pathways; Career clinics; Career management; Pathways for those working less than full time; Implement new roles that provide career development opportunities to attract and retain key staff; Enable staff to access career pathways that also support succession planning; Retire and return; Refresh NHS careers campaign; Career clinics with a view to potentially include 'itchy feet' conversations; Redeployment and careers centre; Development of a personal career pathways post-preceptorship; Promote an improved understanding of career development opportunities; Career coach and stay conversations; Careers marketplace.
Subtheme 6 (Development and Education): Career development; Continuous professional development; Design and implement a development programme for Band 5 Clinicians; Improve training / career development opportunities; Improving Career Development and CPD rotations scheme, coaching and buddying scheme, career clinics; Nursing leadership & line management: improve ownership for staff retention engagement, quality of appraisals; training and development; support the development of talented individuals at all levels; support the ageing workforce; market CPD and leadership opportunities Development and career planning: explicit vertical and horizontal career pathways, improved internal movement, improved unqualified to qualified pathway; Improved clinical management and development; Maximising potential talent; Support managers to get the best out of and develop individual staff; Targeted support and development for key groups of staff; support our staff to explore and pursue career progression within the trust; Flexible working: Improved retire and return, flexible working offer; Retaining and progressing BME nursing and midwifery staff; Supporting line managers to develop leadership skills and create a supportive environment for staff; deliver a leadership programme that supports development towards compassionate leadership in practice (CLIP); Celebrate difference; increase diversity and ensure greater inclusion; Support the movement and development of colleagues.
Subtheme 7 (Being a supportive employer): Improved support for staff health and wellbeing; reduce workplace stress; to promote the health and wellbeing offer to all; improved health and wellbeing/resilience; improve physical health; to provide pastoral support; improved supportive culture within the trust; Health and wellbeing of band 5 registered nurses,

Subtheme 8 (Friendly workplace): Culture and leadership; staff feeling valued; Staff feel valued and can make a difference; Values based organisational culture; Reducing experiences of bullying, harrassment and discrimination; Making the work feel manageable: workforce, patient flow, managing stress; Develop a culture of team identity; Developing organisational values and cultures; positive and flexible working environment; revise the appraisal process to enable talent management and succession planning to be part of the process and then act on this information; Support around violence and aggression; Reduce violence, bullying and harassment; develop and empower our line managers to retain their staff; Keeping in contact: keeping in touch after staff leave for targeted recruitment campaigns; Staff can identify and discuss their career aspirations.
Subtheme 9 (Improved communication): Attraction and recruitment; Making the trust a place where people want to work - workplace attractiveness, empowered; Employer brand; Marketing of trust as a place to work; Reputation of employer; Implement an innovative programme of marketing and recruitment events to increase UK/EU and overseas applicants for nursing vacancies; Widening participation publicity; Recognition of people: communicate successes.
Subtheme 10 (Support for new starters): Focus on making first year of employment supportive, nurturing and fulfiling experience; Graduate nurse programme; Greater understanding of the experience of colleagues joining the team; Improve retention for staff with less than 2 years' service; Support of nurse preceptorship programme; Smoother onboarding; to provide further support for newly qualified nurses; Improved induction; To develop pre-start strategies to keep staff engaged.
Subtheme 11 (Improved recruitment: proposed actions): Appropriate staff are appointed into appropriate roles; Better selection decisions to recruit people who will stay and deliver on what is promised to staff; Improved recruitment practices at all levels for a better and greater candidate experience; Attractive employer, apps and engaging more widely with the community to widen participation as well as being socially inclusive; Providing career opportunities; Customer focussed recruitment; Employment at first point of registration; Recruit and retain staff who share the trust;'s values; Effective management of recruitment.
Subtheme 12 (Increased flexibility): Approach to flexible working which is consistently applied; Improve staff work-life balance; Improve access and understanding of flexible working options; Flexible and happy workforce; expansion of clinical and leadership rotation opportunities.
Subtheme 13 (Support staff close to retirement): Improve retire and return offer in over 55 years old group; Target staff due to retire in 2 years; To support our older workforce to remain in work ensuring they feel valued by the trust; To review the issues associated with retention of staff over 50 and make recommendations in relation to retire and return; Develop the role of the legacy nurse; Flexible retirement; Improve access, support and understanding of flexible retirement options; Support our ageing workforce; individual health and wellbeing plans; Enhanced flexbile working options; Retirement plans.
Subtheme 14 (Itchy feet conversations): Promotion of stay discussions; within six months of starting role at the trust, line manager to meet with staff member to undertake a 'stay' survey; review and redesign career pathways within the organisation; redeployment; Professional development - access to information of what is available, how you access it and how it can be funded.
Subtheme 15 (Engaging staff as stakeholders): Increase nursing engagement to improve retention of new and established nursing staff; Survey staff with 5-10 years service; staff engagement programme; Engage, support and retain student nurses; Staff voice and engagement; Improving our understanding of staff; Develop and maintain nurse resilience; focus groups, learning organisation; promote the positive; staff satisfaction – listening to our nurses; monitor engagement activity; recognition and engagement.
Subtheme 16 (Benefits and pay): Improving staff offer – promote Reward and pay, health and wellbeing opportunities, career support; Recognition, reward and appreciation; Improved visibility of total reward package and incentives; Improve and increase the promotion and uptake of staff benefits that are on offer and to develop new ones; Accessing charitable funds.

Table A5. Frequency of RDSP subthemes adoption by cohort

	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
STH1: Learning from those who leave	0.057	0.118	0.125	0.061	0.092
STH2: Gathering and understanding data	0.343	0.353	0.425	0.273	0.352
STH3: Senior leadership	0.143	0.176	0.050	0.030	0.099
STH4: Retention as part of wider HO strategy	0.286	0.235	0.250	0.212	0.246
STH5: Career progression	0.371	0.500	0.475	0.455	0.451
STH6: Development and education	0.514	0.618	0.500	0.455	0.521
STH7: Being a supportive employer	0.343	0.353	0.300	0.242	0.310
STH8: Friendly workplace	0.343	0.382	0.275	0.273	0.317
STH9: Improved comunications	0.143	0.265	0.225	0.121	0.190
STH10: Supporting new starters and newly qualified staff	0.343	0.324	0.325	0.364	0.338
STH11: Improved recruitment	0.286	0.176	0.225	0.091	0.197
STH12: Increased flexiblity	0.400	0.529	0.400	0.545	0.465
STH13: Supporting staff approaching retirement	0.229	0.294	0.275	0.364	0.289
STH14: Stay conversations and itchy feet conversations	0.000	0.059	0.175	0.152	0.099
STH15: Engaging staff as stakeholders	0.171	0.382	0.325	0.212	0.275
STH16: Staff benefits, rewards and pay	0.286	0.412	0.125	0.061	0.218

Table A6. Effects on retention outcomes by of RDSP subthemes

	Overall ATT			
	Stability		NHS-leaver rate	
STH1: Learning from those who leave	1.057***	(0.397)	-0.483	(0.522)
STH2: Gathering and understanding data	0.624**	(0.251)	-0.556***	(0.176)
STH3: Senior leadership	0.374	(0.438)	0.099	(0.233)
STH4: Retention as part of wider HO strategy	1.293***	(0.297)	-0.658***	(0.253)
STH5: Career progression	0.692***	(0.235)	-0.289*	(0.150)
STH6: Development and education	0.896***	(0.221)	-0.350**	(0.147)
STH7: Being a supportive employer	0.747***	(0.288)	-0.423*	(0.222)
STH8: Friendly workplace	1.131***	(0.282)	-0.538**	(0.212)
STH9: Improved comunications	0.670*	(0.370)	-0.230	(0.274)
STH10: Supporting new starters and newly qualified staff	0.772***	(0.269)	-0.464***	(0.174)
STH11: Improved recruitment	0.524*	(0.291)	-0.518**	(0.236)
STH12: Increased flexiblity	0.780***	(0.225)	-0.359**	(0.156)
STH13: Supporting staff approaching retirement	0.539**	(0.246)	-0.214	(0.164)
STH14: Stay conversations and itchy feet conversations	0.833**	(0.337)	-0.497	(0.366)
STH15: Engaging staff as stakeholders	0.673**	(0.304)	-0.497**	(0.240)
STH16: Staff benefits, rewards and pay	0.570*	(0.321)	-0.385	(0.251)

Notes. Estimates from separate regressions, under unconditional parallel trends assumption. Bootstrapped standard errors (1,000 replications) clustered at HO level are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A7. RDSP effects on 30-day mortality, by ICD-10 chapter

ICD chapter	Disease	ATT	Std. Error	N	Average SHMI	Std. dev.
Chapter 1	Certain infectious and parasitic diseases	0.267	(0.426)	5704	12.539	6.322
Chapter 2	Neoplasms	-0.389	(0.376)	6164	14.467	8.519
Chapter 3	Diseases of the blood, blood-forming organs and the immune mechanism	-0.054	(0.077)	5060	1.046	1.531
Chapter 4	Endocrine, nutritional and metabolic diseases	0.113	(0.174)	5796	2.764	4.476
Chapter 5	Mental and behavioural disorders	-0.148	(0.132)	5198	2.112	2.130
Chapter 6	Diseases of the nervous system	-0.102	(0.170)	5888	1.712	2.772
Chapter 7	Diseases of the eye and adnexa	-0.006	(0.011)	2300	0.026	0.109
Chapter 8	Diseases of the ear and mastoid process	-0.002	(0.004)	2530	0.012	0.069
Chapter 9	Diseases of the circulatory system	-0.700	(0.658)	5796	8.108	10.623
Chapter 10	Diseases of the respiratory system	-0.484**	(0.235)	5842	12.090	4.903
Chapter 11	Diseases of the digestive system	-0.038	(0.083)	5796	2.564	1.872
Chapter 12	Diseases of the skin and subcutaneous tissue	-0.034	(0.044)	5612	0.635	0.610
Chapter 13	Diseases of the musculoskeletal system and connective tissue	-0.017	(0.016)	5842	0.133	0.274
Chapter 14	Diseases of the genitourinary system	-0.032	(0.068)	5796	1.715	1.585
Chapter 15	Pregnancy, childbirth and the puerperium	-0.0000999**	(0.0000465)	5520	0.00012	0.00093
Chapter 16	Certain conditions originating in the perinatal period	-0.282*	(0.170)	5428	1.730	2.403
Chapter 17	Congenital malformations, deformations and chromosomal abnormalities	-0.332	(0.300)	1886	1.206	3.159
Chapter 18	Symptoms, signs and abnormal findings, not elsewhere classified	-0.038	(0.035)	5888	0.510	0.486
Chapter 19	Injury, poisoning and consequences of external causes	-0.039	(0.084)	6026	1.089	1.963

Notes. SDID estimates on standardized hospital mortality by ICD-10 chapter. Bootstrapped standard errors (1,000 replications) clustered at HO level are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A8. Cost-Benefit Analysis

		Cohort1	Cohort2	Cohort3	Cohort4	Total opportunity costs of labor time
(A)	Number of Trusts	33	33	42	38	
(B)	RDSP Start	01/07/2017	01/10/2017	01/04/2018	01/11/2018	
(C)	RDSP End	30/11/2019	30/11/2019	30/11/2019	30/11/2019	
(D = C-B)	Time into treatment (in years)	2.42	2.16	1.67	1.08	
(E)	Number of working days in a year	264	264	264	264	
		NHSI liaison officers (Deputy Director level) labor time costs				
(F = A*D*4)	Time worked on RSDP (in days)	318.97	285.70	279.85	164.08	
(G)	Yearly salary (in £)	£ 81,701	£ 81,701	£ 81,701	£ 81,701	
(H=G/E)	Yearly salary per day (in £)	£ 309	£ 309	£ 309	£ 309	
(I=F*H)	Labor time opportunity-cost (in £)	£ 98,712.72	£ 88,416.15	£ 86,605.09	£ 50,777.39	£ 324,511.35
		NHS Trusts officers (Chief Nurse and Deputy Chief Nurse) labor time costs				
(J)	Yearly Chief Nurse Salary (in £)	£ 93,835.00	£ 93,835.00	£ 93,835.00	£ 93,835.00	
(K=J/264)	Daily Chief Nurse Salary (in £)	£ 355.44	£ 355.44	£ 355.44	£ 355.44	
(L=1*12)	Yearly FTE worked on RSDP (in days)	12	12	12	12	
(M=A*D*K*L)	FTE Labor time opportunity-cost (in £)	£ 340,119.74	£ 304,642.40	£ 298,402.31	£ 174,956.12	£ 1,118,120.57
(N)	Yearly Deputy Chief Nurse Salary (in £)	£ 73,936.00	£ 73,936.00	£ 73,936.00	£ 73,936.00	
(O=N/264)	Daily Deputy Chief Nurse Salary (in £)	£ 280.06	£ 280.06	£ 280.06	£ 280.06	
(P=2*12)	Yearly FTE worked on RSDP	24.00	24.00	24.00	24.00	
(Q=A*D*O*P)	FTE Labor time opportunity-cost (in £)	£ 535,985.36	£ 480,077.59	£ 470,244.01	£ 275,708.54	£ 1,762,015.50
	Grand Total Opportunity Cost of Time					£ 3,204,647.42

Notes. NHSI Deputy Director salary from `https://www.glassdoor.co.uk/` updated on 13-07-2023. HO Chief Nurse and Chief Nurse Assistant salaries from `https://www.nhsemployers.org/system/files/media/NHS-TCS-2019-pay-poster_0.pdf`, respectively at Band 9 (spline 52) and Band 8D (spline 47). Row F assumes one FTE day of work for HO each quarter. Row L assumes one FTE day of work each month. Row P assumes two FTE days of work each month.

Table A9. HO allocation into cohorts during pre-RDSP periods (Falsification tests 1 and 2)

	Size	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Cohort 5
Original	193	16.06%	15.03%	18.13%	19.17%	31.61%
Placebo 1	175	15.43%	14.86%	18.86%	19.43%	31.43%
Placebo 2	192	15.63%	15.10%	18.23%	19.27%	21.77%
Placebo 3	176	15.34%	14.77%	18.75%	19.32%	31.82%

Notes. Placebo timings are June 2010 - December 2013 (Placebo 1), June 2013 - December 2016 (Placebo 2), and June 2011 - December 2014 (Placebo 3). Cohort 5 is considered as the never-treated control group.

Table A10. Effects on stability rate preponing RDSP implementation (Falsification test 1)

	Placebo 1		Placebo 2		Placebo 3	
	Placebo	Original	Placebo	Original	Placebo	Original
Overall ATT	-0.676§ (0.234)	0.683§ (0.185)	0.219 (0.288)	0.749§ (0.184)	0.050 (0.259)	0.693§ (0.198)
Cohort 1	-1.385§ (0.524)	0.871 (0.352)	0.389 (0.419)	0.843 (0.382)	-0.493 (0.607)	0.868§ (0.361)
Cohort 2	-0.822 (0.405)	0.614 (0.338)	-0.245 (0.465)	0.677 (0.308)	-0.042 (0.580)	0.575 (0.354)
Cohort 3	-0.232 (0.484)	0.367 (0.327)	-0.148 (0.377)	0.557 (0.335)	-0.288 (0.269)	0.372 (0.319)
Cohort 4	-0.434 (0.323)	0.894§ (0.270)	0.792 (0.403)	0.912§ (0.261)	0.878 (0.487)	0.955§ (0.287)
PTA p-value (12 months)	0.013	0.136	0.001	0.067	0.630	0.168
PTA p-value (6 months)	0.010	0.430	0.003	0.217	0.515	0.449

Notes. Estimated under unconditional parallel trends for stability of nurses and midwives. Bootstrapped standard errors are in parentheses. § indicates that the 95% simultaneous confidence band does not cover 0. Placebo timings are June 2010 - December 2013 (Placebo 1), June 2013 - December 2016 (Placebo 2), and June 2011 - December 2014 (Placebo 3).

Table A11. Effects on 30-day mortality preponing RDSP implementation (Falsification test 2)

	All-cause 30-day risk-adjusted mortality rate		
	Placebo 1	Placebo 2	Placebo 3
ATT	-0.043	0.037	-0.025
	(0.047)	(0.057)	(0.051)

Notes. Synthetic difference-in-difference treatment effect estimates for total mortality, under 3 different placebo policy timings. Placebo timings are June 2010 - December 2013 (Placebo 1), June 2013 - December 2016 (Placebo 2), and June 2011 - December 2014 (Placebo 3). Bootstrapped standard errors obtained using 1,000 replications are reported in parentheses. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

Table A12. Effects of RDSP on new joiners and staff levels

	Joiners	NHS joiners	Churn joiners	Staff levels
Overall ATT	-0.067 (0.394)	-0.031 (0.33)	-0.039 (0.151)	-4.563 (10.16)
Cohort 1	-0.778 (1.366)	-0.48 (1.283)	-0.303 (0.388)	-12.617 (23.948)
Cohort 2	0.427 (0.465)	0.604 (0.394)	-0.186 (0.235)	21.281 (18.916)
Cohort 3	0.062 (0.506)	-0.157 (0.393)	0.216 (0.23)	-14.734 (13.813)
Cohort 4	0.019 (0.456)	-0.032 (0.328)	0.057 (0.227)	-8.45 (12.325)
lead 12 months p	0.0011	0.0000	0.0669	0.268
lead 6 months p	0.1011	0.0107	0.4322	0.337

Notes. Estimates under unconditional parallel trends. Bootstrapped standard errors (1000 replications) in parentheses.

Table A13. Effects on stability rates using not-yet-treated cohorts as control group

	Stability rate			NHS-leaver rate		
	(1)	(2)	(3)	(4)	(5)	(6)
Control group	Never treated	Never treated	Not yet treated	Never treated	Never treated	Not yet treated
Post treatment until	November 2019	August 2019	August 2019	October 2019	August 2019	August 2019
Overall ATT	0.775***	0.656***	0.677***	-0.439***	-0.387***	-0.402***
	(0.188)	(0.183)	(0.177)	(0.131)	(0.125)	(0.123)
partially aggregated						
Cohort 1	0.950***	0.851***	0.971***	-0.488**	-0.444*	-0.483**
	(0.353)	(0.328)	(0.349)	(0.236)	(0.229)	(0.223)
Cohort 2	0.677**	0.579*	0.599**	-0.425*	-0.391*	-0.403*
	(0.295)	(0.296)	(0.269)	(0.223)	(0.209)	(0.215)
Cohort 3	0.557	0.424	0.378	-0.455*	-0.402	-0.412*
	(0.356)	(0.314)	(0.297)	(0.267)	(0.250)	(0.239)
Cohort 4	0.912***	0.773***	0.773***	-0.393**	-0.323*	-0.323*
	(0.268)	(0.270)	(0.251)	(0.184)	(0.182)	(0.172)

Notes. Aggregated treatment effect parameters with different estimation windows and control group definitions. Bootstrapped standard errors (1,000 replications) clustered at the HO level are reported in parentheses. Significance levels: p-values *p<0.1; **p<0.05; ***p<0.01.

Table A14. Breakdown of $\overline{ATT}$ estimates on retention outcomes over time

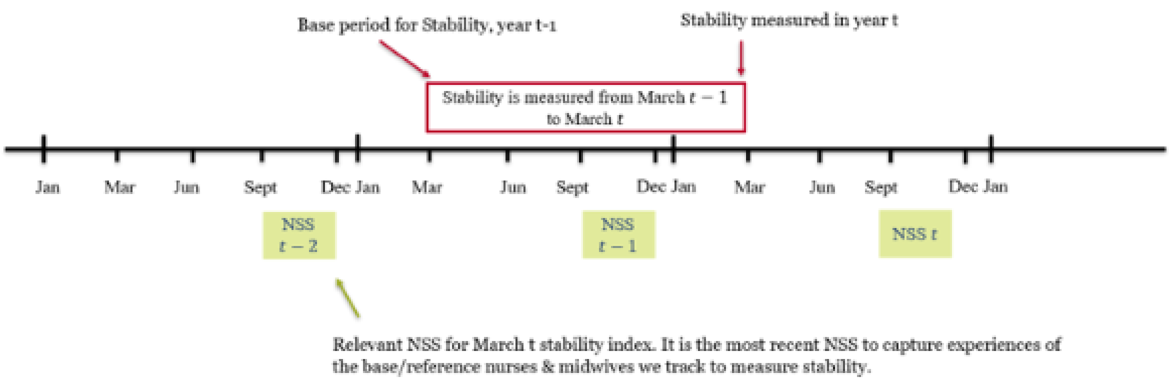
(a) Stability rates						
	ATT [0,11]		ATT [12,19]		ATT [12,τ]	
Overall	0.520	(0.157)§	0.915	(0.275)§	1.096	(0.289)§
Cohort 1	0.427	(0.280)	1.104	(0.455)§	1.319	(0.425)§
Cohort 2	0.473	(0.277)	0.544	(0.407)	0.852	(0.414)§
Cohort 3	0.256	(0.327)	1.009	(0.454)§	1.009	(0.454)§
Cohort 4	0.884	(0.261)§	1.257	(0.412)§	1.257	(0.412)§

(b) NHS leaver rate						
	ATT [0,11]		ATT [12,18]		ATT [12,$\tau-1$]	
Overall	-0.309	(0.105)§	-0.575	(0.216)§	-0.644	(0.229)§
Cohort 1	-0.249	(0.244)	-0.528	(0.310)	-0.667	(0.317)§
Cohort 2	-0.244	(0.177)	-0.511	(0.311)	-0.592	(0.349)
Cohort 3	-0.327	(0.228)	-0.675	(0.377)	-0.675	(0.377)
Cohort 4	-0.393	(0.193)§				

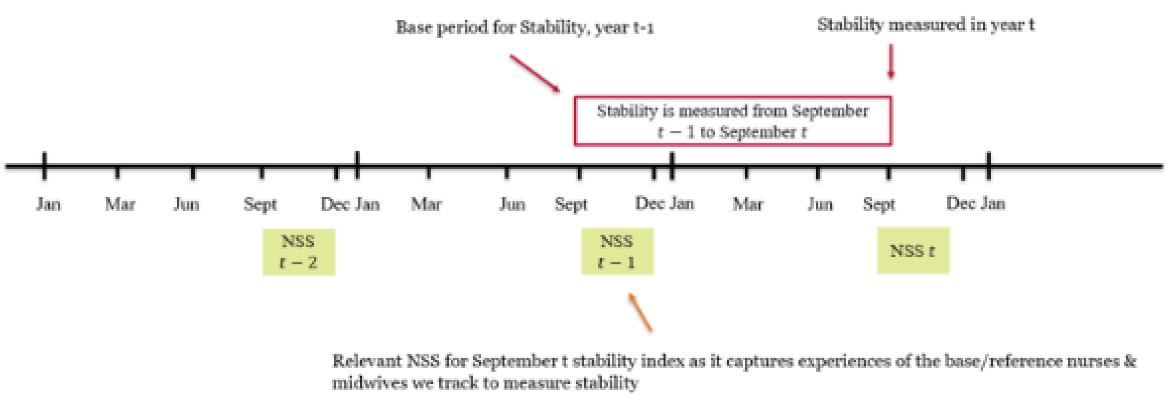
Notes. The censored ATTs are obtained using the same unconditional PTA model as in Table 2, but instead of aggregating all post-treatment periods, the aggregation is based on a subset of post-treatment periods. Bootstrapped standard errors are clustered at HO level. § indicates that the 95% simultaneous confidence band does not cover 0. τ indicates the relative time from the RDSP, which corresponds to November 2019 for stability rate and October 2019 for the NHS leaver rate. τ = 28 for Cohort 1, τ = 25 for Cohort 2, τ = 19 for Cohort 3, and τ = 12 for Cohort 4.

Figure A1. Data setup

(a) Before September

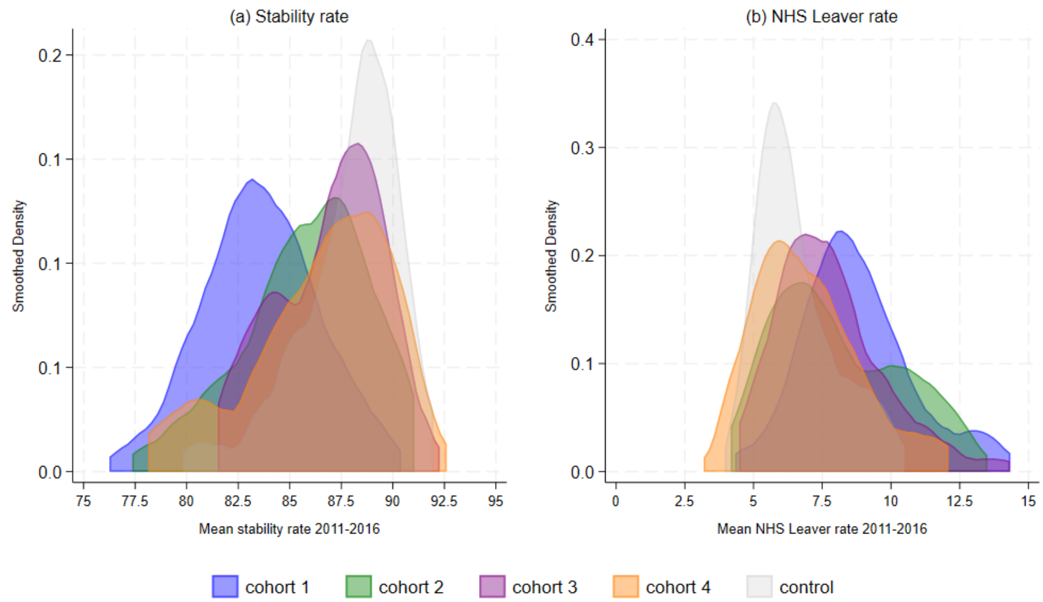


(b) After September



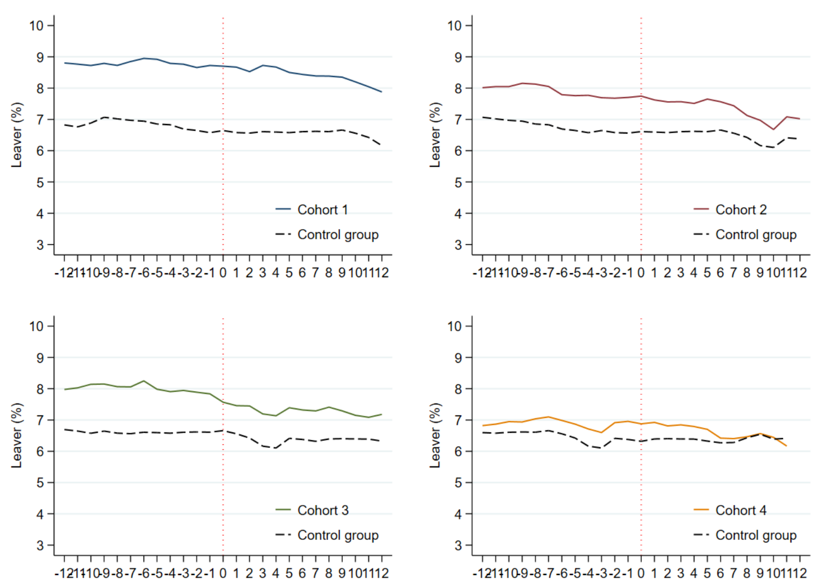
Notes. The same data structure holds for NHS leaver rates. t refers to the analysis year, $t-1$ is the base year. NSS refers to the NHS Staff Survey which is conducted every year in autumn since 2003. Staff working in HOs in 1st September are eligible to respond to the NSS. The NSS runs from the mid-September and remains open on average 8 weeks.

Figure A2. Distribution of average monthly retention measures (2011/12 - 2015/6)



Notes. Smooth histograms calculated using a kernel density smoother.

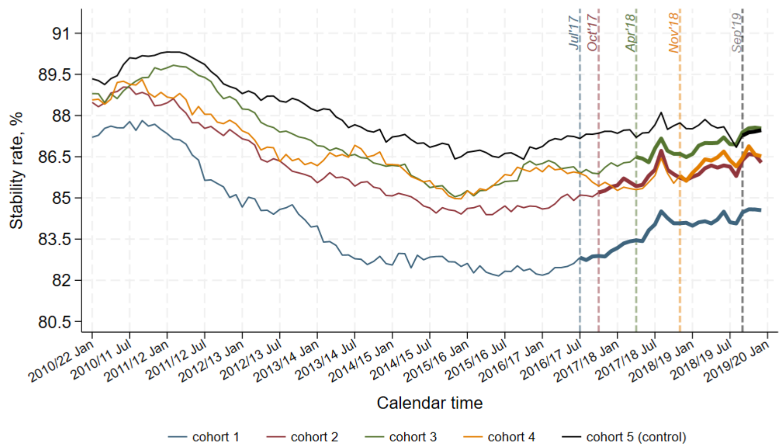
Figure A3. Common trends between treated and control cohorts (NHS Leaver rates)



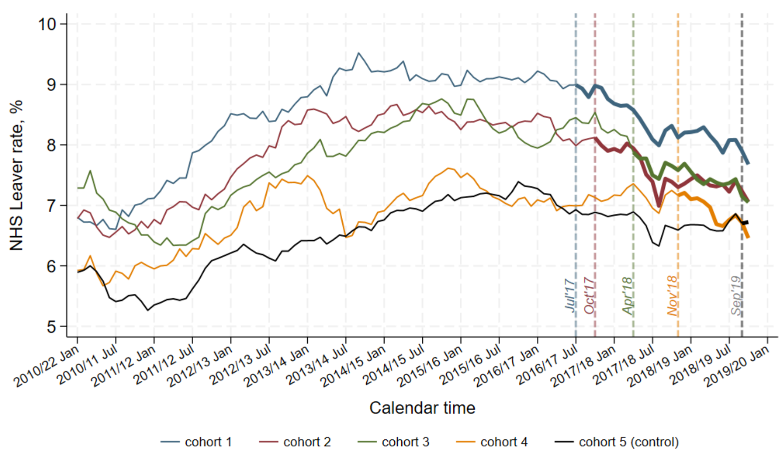
Notes. Figures are centered at the time RDSP was launched in Cohorts, HOs, and are balanced for relative time periods. The vertical dashed line indicates the timing of the RDSP, and the figures show 12 months before and 12 months after the RDSP.

Figure A4. Trends in nurses' retention and RDSP launch dates, by cohort

(a) Stability rates from 2010/11 to 2019/20

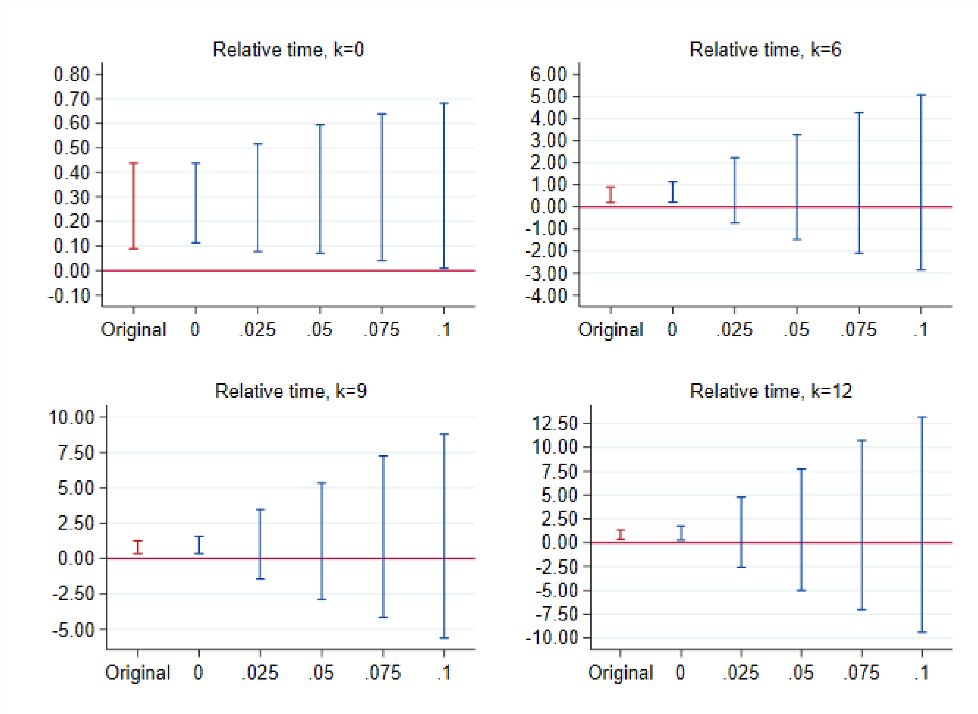


(b) NHS Leaver rates from 2010/11 to 2019/20



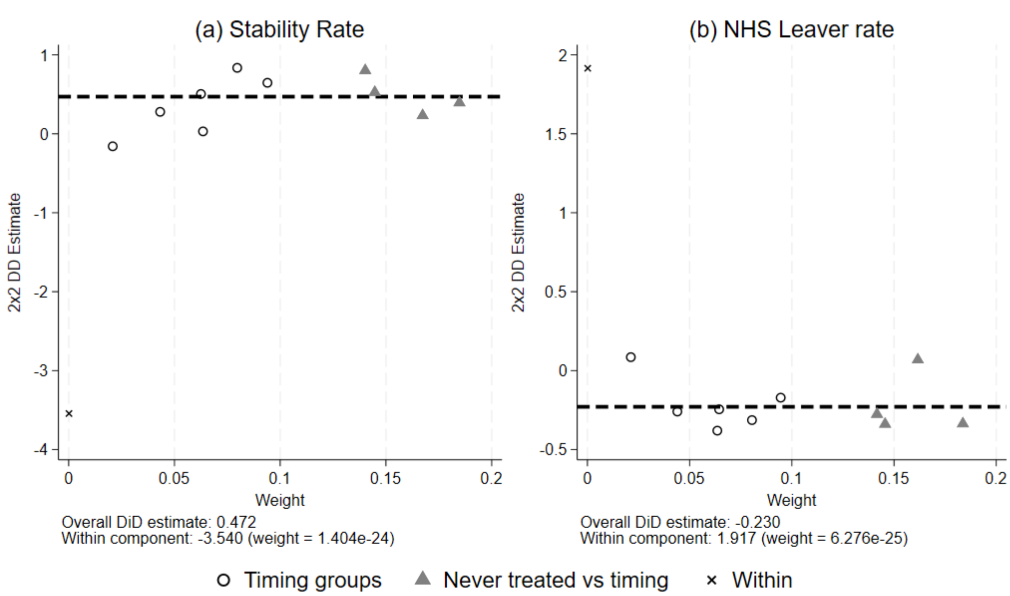
Notes. Cohort 5 includes 2 additional HOs that were not in the NHSI allocation. Vertical lines show the RDSP start dates for each cohort, and the thicker horizontal lines indicate post-RDSP period in each cohort.

Figure A5. Sensitivity of parallel trends test (Rambachan and Roth, 2023)



Notes. The x-axis shows different values for M, the smoothness restriction, ranging from 0 to 0.1, and the y-axis shows the 95% robust confidence interval. The original indicates the confidence interval for ATT at exposure month k using asymptotic standard errors (as in panel (b) of Figure 5).

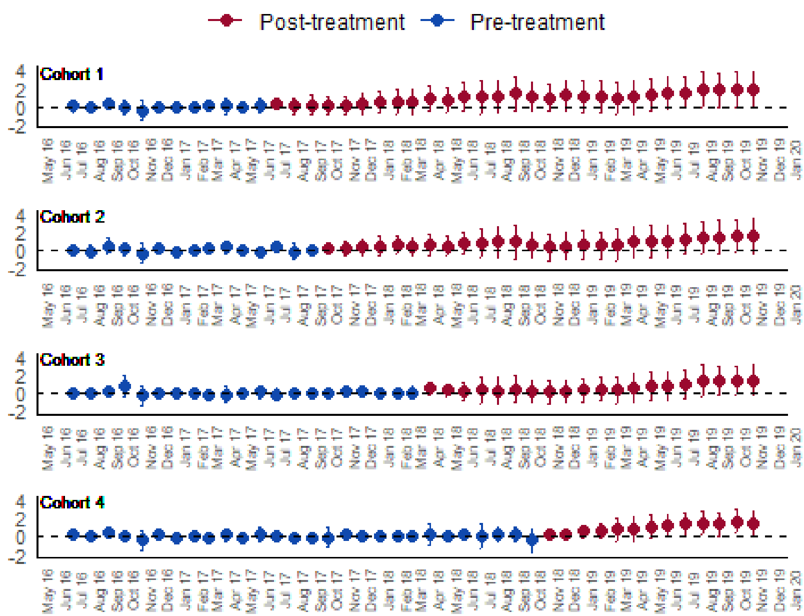
Figure A6. Goodman-Bacon TWFE decomposition



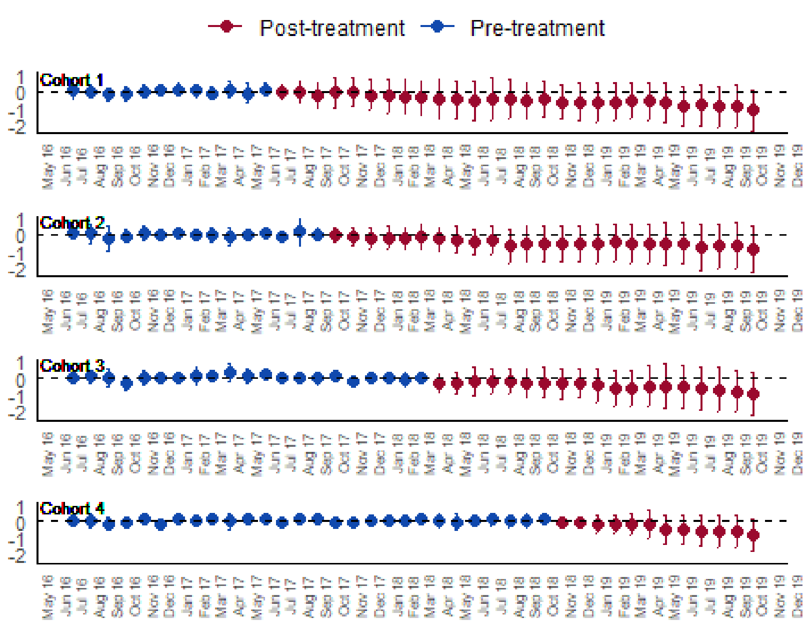
Notes. Goodman-Bacon decomposition of the TWFE DiD estimates reported in the top panel of Table 2.

Figure A7. $ATT(c,t)$s estimates by cohort and RDSP launch month

(a) Stability rates

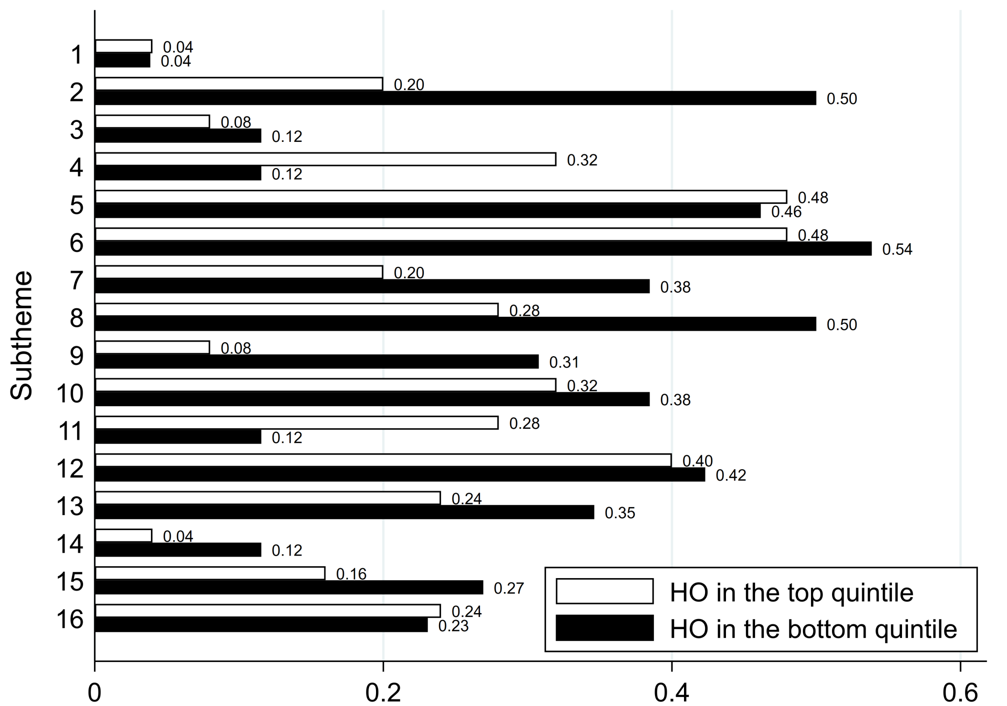


(b) NHS Leaver rates



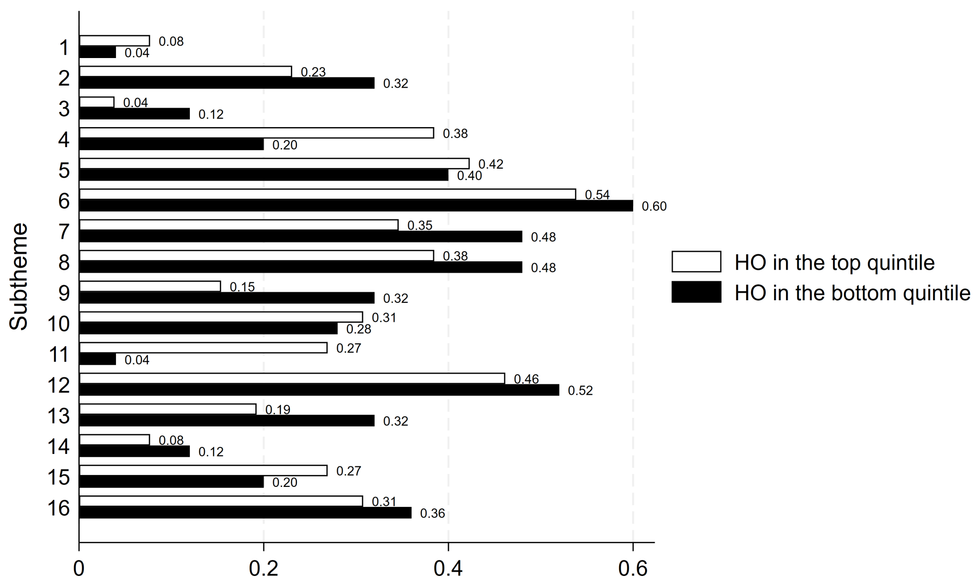
Notes. The cohort-time estimates, $\widehat{ATT}(c,t)$s, are estimated under unconditional parallel trends assumption and shown with simultaneous 95% confidence bands from bootstrapped standard errors clustered at HO level.

Figure A8. Frequency of subthemes in top and bottom quintiles of stability rate gains



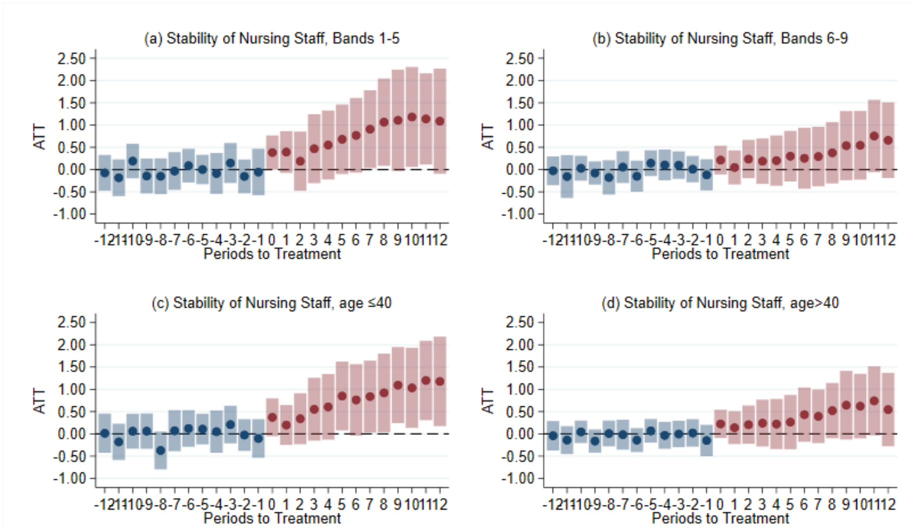
Notes. Frequency of subthemes adoption among NHS HOs belonging to the top or bottom 20% of the distribution of changes in stability rates in the first 12 months from RDSP launch. STH1: Learning from those who leave; STH2: Gathering and understanding data; STH3: Senior leadership; STH4: Retention as part of a wider HO strategy; STH5: Career progression; STH6: Development and education; STH7: Being a supportive employer; STH8: Friendly workplace; STH9: Improved communications; STH10: Support for new starters; STH11: Improved recruitment; STH12: Increased flexibility; STH13: Support staff close to retirement; STH14: Itchy feet conversations; STH15: Engaging staff as stakeholders; STH16: Benefits and pay.

Figure A9. Frequency of subthemes in top and bottom quintiles of NHS leaving rate gains



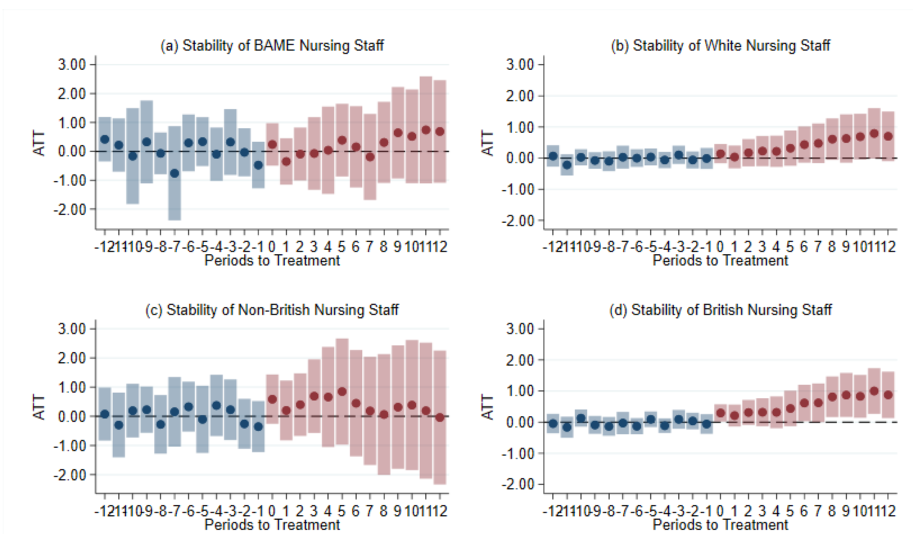
Notes. Frequency of subthemes adoption among NHS HOs belonging to the top or bottom 20% of the distribution of changes in NHS leaving rates in the first 12 months from RDSP launch. STH1: Learning from those who leave; STH2: Gathering and understanding data; STH3: Senior leadership; STH4: Retention as part of a wider strategy; STH5: Career progression; STH6: Development and education; STH7: Being a supportive employer; STH8: Friendly workplace; STH9: Improved communications; STH10: Support for new starters; STH11: Improved recruitment; STH12: Increased flexibility; STH13: Support staff close to retirement; STH14: Itchy feet conversations; STH15: Engaging staff as stakeholders; STH16: Benefits and pay.

Figure A10. Seniority of nurses and midwives - an event study approach



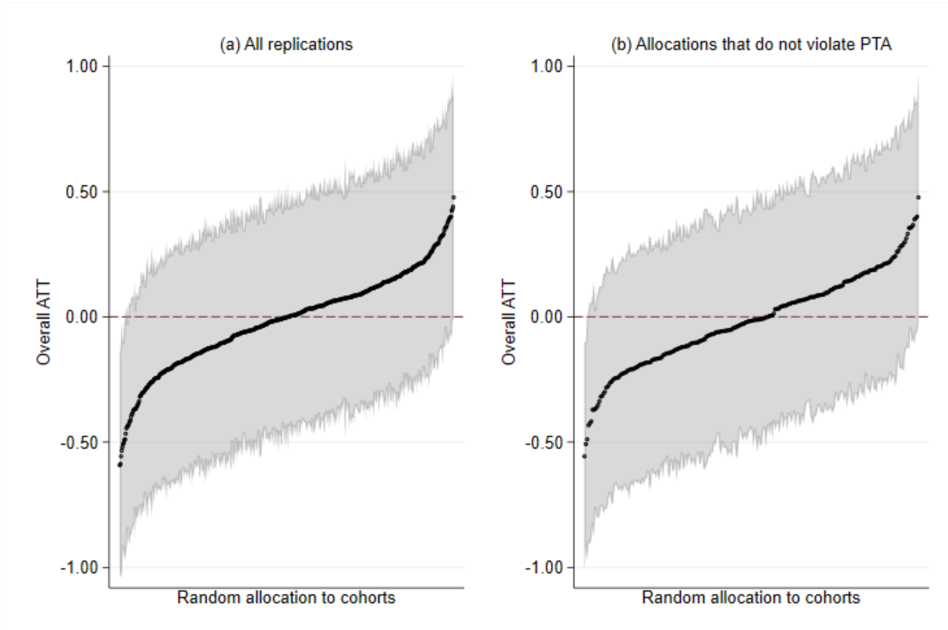
Notes. Event-study estimates under unconditional parallel trend assumption. Uniform 95% confidence intervals using bootstrapped standard errors clustered at the HO level. All models satisfy unconditional parallel trends for 6 months pre-trends with p-values 0.1481 (a), 0.4140 (b), 0.8711 (c), 0.0726 (d).

Figure A11. Ethnicity and nationality of nursing staff - an event study approach



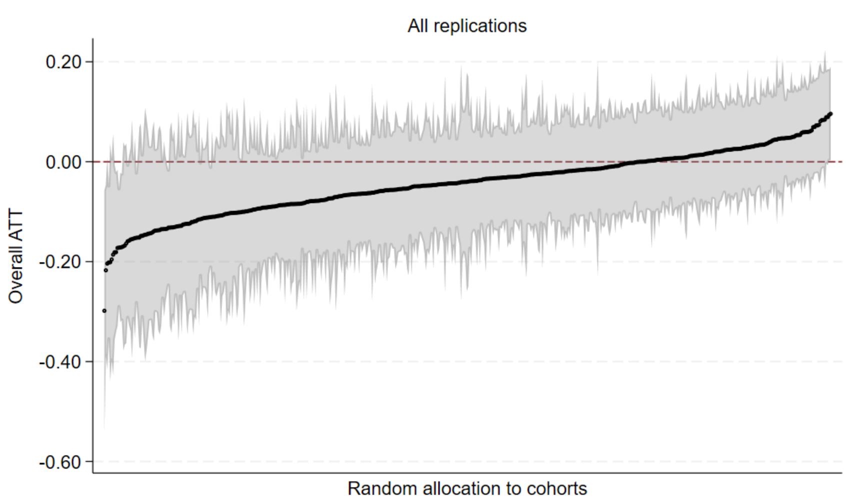
Notes. Event-study estimates under unconditional parallel trend assumption. Uniform 95% confidence intervals using bootstrapped standard errors clustered at the HO level. All models satisfy unconditional parallel trends for 6 months pre-trends with p-values 0.2140 (a), 0.9082 (b), 0.3269 (c), 0..1629 (d)

Figure A12. Effect on nurse stability rate with random allocation of HOs into Cohorts (Falsification test 3)



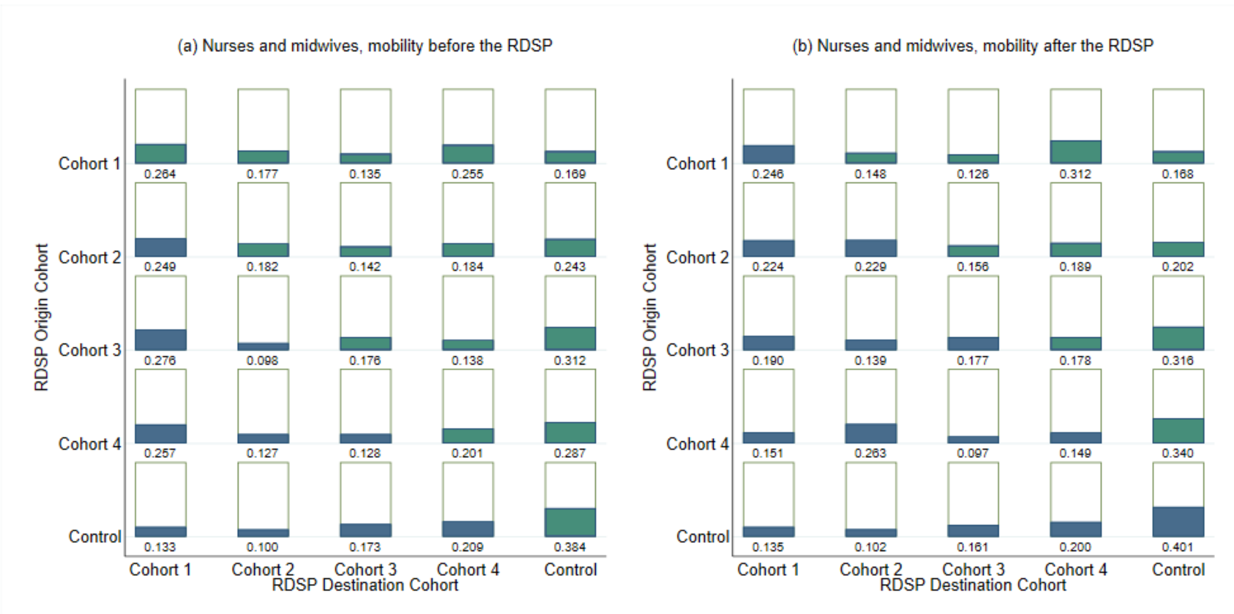
Notes. Aggregate $\overline{ATT}$s from 500 randomized allocations of HOs to cohorts, under unconditional parallel trends. Uniform 95% confidence intervals using bootstrapped standard errors clustered at HO level. Panel (a): all replication results; Panel (b) results subset (248 of 500) from replications with unconditional parallel trends 12 months pre-test $p > 0.05$.

Figure A13. Effect on 30-day risk-adjusted mortality with random allocation of HOs into Cohorts (Falsification test 4)



Notes. Aggregate $\overline{ATT}$s for the SDiD model on total hospital mortality, from 500 randomized allocations of HOs to cohorts. 95% confidence intervals using bootstrapped standard errors (1,000 replications) clustered at HO level.

Figure A14. Nurses and midwives' mobility before and after the RDSP

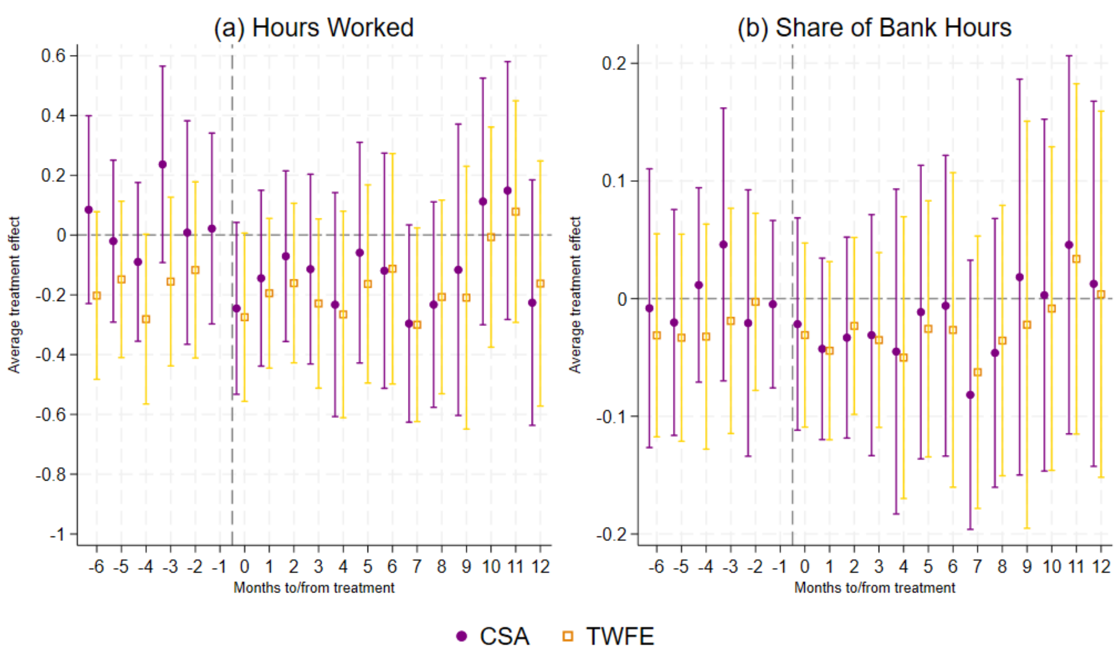


Notes. 28,558 nurses and midwives changed HO at least once (6.7%) from June 2016 to November 2019 (34,472 switches in total). The period before/after the RDPS in panels (a) and (b) varies depending on each cohort RDSP launch. Green bars indicate the transitions to hospitals in the not-yet treated cohorts. For instance, Cohort 3 HOs introduced the RDSP in April 2018, when RDSP had been already launched in Cohort 1 and 2, but not in Cohort 4.

Online Appendix C. Effects on intensive margins

The explicit aim of the RDSP was to reduce turnover rates, within and across HOs. Nevertheless, some strategies outlined in action plans, such as e-Rostering, might have led to a re-allocation of working hours and encouraged nurses to work more.[41] Therefore, we focus on the average monthly hours worked by nurses and midwives in full-time contracts. We exclude workers with zero hours from the sample, and define a job as full-time when the total monthly work-time equivalent (WTE) is at least 95%.[42]

Figure A15. RDSP effects on labor intensive margins



Notes. CSA event-study estimates under unconditional parallel trend assumption. Panel (a): average monthly hours worked at the HO level. Panel (b): Share of Bank hours worked at the HO level.

In 2016, the average monthly working hours for a full time nursing staff was 166.8 hours, which is 4 hours more than the full-time contractual hours of 37.5 per week.[43]

[41] E-Rostering is an electronic shift system that provides information on staffing levels to meet healthcare demands and also facilitate workforce flexibility.

[42] For instance, if a nurse has 2 part-time jobs in a HO with 0.55 WTE and 0.40 WTE jobs, their total monthly WTE is 0.95, and they qualify as a full-time nurse even though they hold part-time jobs.

[43] The ESR is a payroll data, thus it does not have information on unpaid hours. Nurses and midwives

We do not find any effect of the RDSP on the average hours worked by full-time nurses and midwives in treated HOs, as shown by panel A of Figure A15, consistently with the fact that the Programme primary aim was to improve working conditions affecting staff retention, but it did not provide any strong incentive to work longer hours. Furthermore, as shown in panel B of Figure A15, we do not find any evidence that the RDSP had an impact on the share of additional paid working hours, usually referred to as "Bank" hours in the English NHS.[44]

are likely to work additional unpaid hours to cover shifts and provide quality patient care.

[44] "Bank" work is carried out by employees who are registered to provide shifts on a temporary basis, mostly on a zero-hours contract, with no further obligation for regular work at hospital Trusts. It is different from other temporary nursing staff which are on fixed-term, non zero-hours contracts and provide regular work shifts. Bank work is very common among NHS nurses and midwives, with an average of 16% of nurses and midwives registered as bank workers in each month in 2016. Bank staff may come either from the existing nursing staff of a HO (in-house bank) or from employees of an outside organization, who are only contracted as Bank workers within the HO (Bank-only). The difference is that Bank-only staff may leave the HO once their period of temporary employment terminates, whereas in-house Bank staff are nurses already employed within the HO and providing additional labor time.