Do non-monetary interventions improve staff retention? Evidence from English NHS hospitals

Giuseppe Moscelli*†‡, Melisa Sayli*†, Jo Blanden†‡, Chris Bojke§, and Marco Mello†

†School of Economics, University of Surrey

§School of Medicine, University of Leeds

‡IZA

Abstract

Excessive turnover reduces the stock of an organization's human capital. In the public sector, where salary increases are often constrained, managers need to leverage on non-monetary working conditions to retain their employees. We investigate whether workers are responsive to improvements in non-wage aspects of their job by evaluating the impact on nurse retention of a programme that encouraged public hospitals to increase staff retention through data monitoring and improving the non-pecuniary aspects of nursing jobs. Employing rich employee-level administrative data from the universe of English NHS hospitals, and a staggered difference-in-difference design, we find that the programme has improved nursing retention within hospitals, decreased exits from the public hospital sector, and decreased mortality by preventing 11.400 patient deaths within 30 days from hospital admission. Our results indicate that a light-touch intervention can shift management behavior and improve hospital workforce turnover. These findings are important in sectors affected by labor supply shortages, and they are especially policy-relevant in the health care context, where such shortages have potentially negative effects on patient outcomes.

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

---

*Corresponding authors: g.moscelli@surrey.ac.uk; m.sayli@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, Igor Francetic, Hugh Gravelle, Esteban Jaimovich, Elaine Kelly, Sandra McNally, Matthias Parey, Joao Santos Silva and participants to the iHEA World Congress (July 2021), Australian Health Economics Society Online Conference (September 2021), Portuguese National Conference on Health Economics (October 2021), University of Exeter (November 2021), UK HESG Winter Meeting 2022 (January 2022), REAL Centre at the Health Foundation, University of Southampton (February 2022), and Royal Economic Society (April 2022), International Association of Applied Econometrics (June 2022), American-European Health Economics Study Group and European Health Economics Association (July 2022), European Labour Economist Association (September 2022) 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. 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. The responsiveness of labor supply to the non-financial aspects of job quality is especially pertinent in public sector occupations, where large workforces combined with budget constraints limit the potential for wage increases. With 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. We provide causal evidence about the nexus between management practices and employee retention, by showing that a low-powered incentive to monitor retention and improve job quality led to significant reductions in turnover among public sector hospital nurses, and deliver insights about which non-monetary incentives can be exploited to stimulate labor supply.

Limiting excessive workforce turnover is important for the efficient functioning of any large and complex organizations, but it is paramount in a labor-intensive sector like health care. Nurses are a vital part of the healthcare sector workforce, constituting about one-third of its employees, both in the United States and the United Kingdom. Nurse vacancy rates are high across the OECD countries: although the numbers of nurses have been increasing, the supply of nurses is failing 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). Moreover, nurse shortages have been linked to lower quality of health care provided, for example, higher patient mortality (Rafferty et al., 2007; Ball et al., 2018; Griffiths et al., 2019), higher likelihood of missed care (Ball et al., 2014; Griffiths et al., 2018), and increased length of hospital stay (Duffield et al., 2011). Therefore, policy makers are seeking sustainable, cost-effective ways to reduce nurse shortages and high turnover rates. Compared to training new nurses, which takes 3 to 4 years, improving retention is a time and cost-efficient solution to staff shortages (Shields, 2004), with the additional benefit of retaining specific human capital within the employing organization. A possible strategy to improve retention is to increase wages. However, as reviewed in Lee et al. (2019), the empirical literature supports a limited role for wages to increase labor supply among nurses. Even if nurse retention was highly responsive to wages, a conspicuous pay rise across a sector as large as UK public health care[1] is expensive.[2] 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 result in "meaningful" work. Similar factors are also emphasized in the human resources management literature (Bakker and Demerouti, 2007). Understanding the role played by non-monetary factors in shaping workers' labour supply is also 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 (Besley and Ghatak, 2005; 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 by NHS Improvement (NHSI), the monitoring body of English National Health Service (NHS) hospital organizations. The explicit purpose of this intervention was to reduce turnover rates among nurses working in NHS hospitals. Our evaluation makes use of a unique dataset, consisting of a

[1]The National Health Service, or NHS, studied here, was the fifth largest employer in the world in 2015 (Forbes, 2015)

[2]In February 2023, NHS nurses are undertaking industrial action to reverse real-terms pay declines. Government is resisting due to affordability concerns (Campbell, 2023).

monthly panel of English NHS hospitals, which we constructed by combining different sources of micro-level administrative data (Electronic Staff Records 2015-2019 and NHS Staff Surveys 2013-2018) with organization-level data from NHSI (timing, cohorts and themes of the RSDP intervention). The RDSP was implemented in a staggered fashion: hospital organizations, called Trusts in the English NHS, were split into five cohorts, with each cohort starting the Programme at different times. In order to achieve causal identification of the impact of the Programme, we exploit the new methodological advances in the estimation of difference-in-difference (DiD) with staggered treatment adoption. For the evaluation of impact on labour supply we exploit the DiD estimator proposed by Callaway and Sant'Anna (2021), and for the evaluation of the impact on health outcomes we exploit the synthetic DiD estimator proposed by Arkhangelsky et al. (2021). Overall, we find that the RDSP has improved nursing retention by 0.78 percentage points (ppt) leading to the retention, on average, of 1,697 nurses and midwives who would have left their Trust otherwise. This is around a quarter to a half of the standard deviation of the retention rate across Trusts 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. Moreover, using the Arkhangelsky et al. (2021) synthetic difference in differences estimator, we find evidence of reductions in both 30-day risk-adjusted mortality for any patient admitted and waiting times for planned patients admitted at the RDSP-treated hospital organizations.

Our work brings the following contributions to the labour, health and organizational economics literatures. First, we test the hypothesis of whether a low-cost non-wage intervention improves employee retention within the (healthcare) organization where they work, and also within the whole (public hospital care) sector. There was no specific direction issued about how the NHS hospital organizations should achieve these goals, although managers were prompted by NHSI if no improvement was observed. Hospital providers were tasked to build their own retention strategies: NHSI provided tailored retention data to identify areas for improvement as well as liaison officers to help develop and execute action plans. The provision and analysis of retention data is an especially important feature of this intervention because NHS hospital providers "do 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 localised approach used by NHS policy-makers to enact this Programme is in line with the recent finding that there is substantial unexplained variation in NHS workforce retention across providers (Kelly et al., 2022).

Second, we investigate possible 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 participation to improve patient outcomes directly. An alternative hypothesis is that the specific focus on retention could distract hospital managers and senior clinicians from activities with more direct benefit for patients, in line with the classical principle of equal compensation in incentives (Holmstrom and Milgrom, 1991), such that agents reallocate their efforts to those activities boosting performance close to standards. This chimes with results from Friebel et al. (2022), who find a positive impact on employee retention, but no effect on store-level profits. We find no statistically significant evidence of either positive or negative effects on patient outcomes (standardized mortality and unplanned readmission rates) or productivity (proxied by the number of admissions) in the treated organizations.

Third, we try to unpack the "black box" of the RDSP by exploiting information about the Programme areas of intervention used by each hospital organization. Although it is not possible to establish precise causal links between individual activities and outcomes, our results provide suggestive evidence that embedding retention in the organization's strategy leads to stronger outcomes.

Retated 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 recent work is especially relevant in this context, as it finds effects of simply asking managers to 'do what they can' to reduce turnover among retail workers in Eastern Europe. Gosnell et al. (2020) show that simply providing workers with data on their performance has a substantial impact on their performance. Our findings further support the idea that goal-setting alone can be sufficient to encourage managers to act to improve retention in the private sector. A wider literature also 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; Verbeeten, 2008). However, strong incentives can be counter-productive when agents are pro-socially motivated (Ellingsen and Johannesson, 2008; Bowles and Polania-Reyes, 2012), raising the question of the optimal strength of incentives in the public sector. We show that weak 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 here is best 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 effective management can be achieved without great cost in the public sector.

Moreover, we also contribute to the literature about the features of meaningful work (Kahn, 1990; Chalofsky, 2003; Bailey et al., 2019). The analysis of the activities adopted by hospital management identifies many factors in common with the aspects of work that have been shown to improve retention and job satisfaction and contribute to Cassar and

Meier (2018)'s framework of mission, autonomy, competence and relatedness. Employees have been shown to value aspects of autonomy, such as being involved in decision making processes (Böckerman et al., 2012) and having flexibility at work (Mas and Pallais, 2017); and the value of competence is illustrated by employees' responses to being recognized for their efforts and having opportunities to learn on the job (Gallus, 2017). As shown by Hoffman and Tadelis (2021), positive human resource management practices, such as having a positive attitude, setting clear expectations and internal consultations, and providing career support and coaching, may have a substantial effect on staff retention. Organizational culture also matters: Alan et al. (2023) shows that an extremely specific intervention to improve company culture reduced voluntary quits among white-collar workers in Turkey, and 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. Our results demonstrate that the changes brought about by the policy are applicable in the public sector and affect the decision to quit, an important margin in understaffed public services.

2 Institutional Background

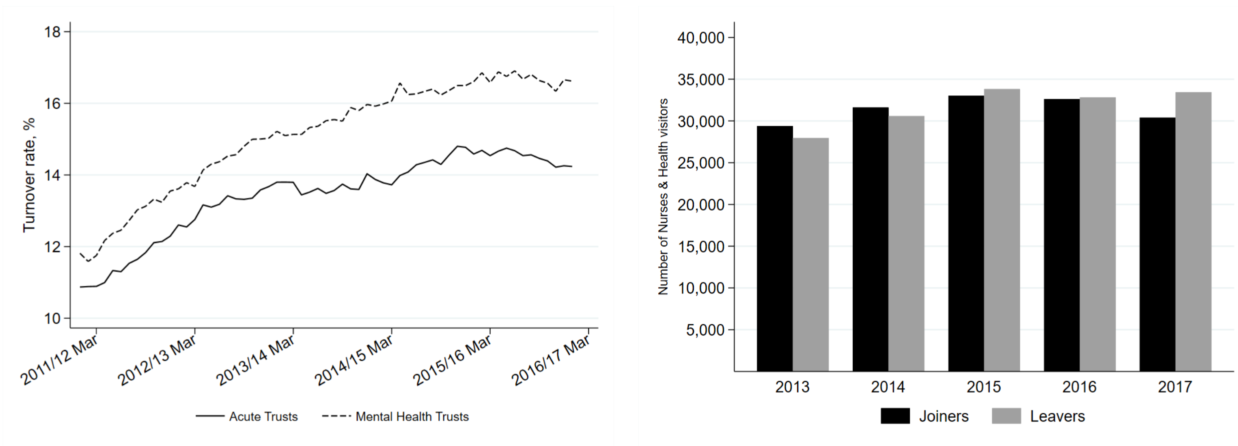
2.1 The English NHS and its nursing workforce

The NHS is publicly funded through general taxation and provides free comprehensive primary, secondary and tertiary healthcare services to over 56 million people in England[3] and a further 11 million in the other devolved nations of the UK (Scotland, Wales, and Northern Ireland). The NHS budget for England in 2017 was approximately £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

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

Figure 1: Trends in NHS Nursing Workforce

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



Notes: Panel (a): Authors' calculation from the Electronic Staff Records 2009-2020. Turnover is measured for each month from one year to the following. The Figure produced using the ESR might differ from the official Workforce Statistics released by NHS Digital. Panel (b): Headcounts of Nurses and Health Visitors in NHS Hospital and Community Health Services. Each period runs from September to the following September. Data is from NHS Digital, NHS Hospital & Community Health Service workforce statistics (NHS Digital, 2018; National Audit Office, 2020).

with the Nursing and Midwifery Council (NMC) (NMC, 2020).[4] The English NHS employs around 330,000 of these registered nurses and midwives[5], who make up almost half of the professionally qualified clinical staff.[6]

The nursing workforce has been under significant pressure from growing demand for healthcare combined with high turnover rates, and have reported high levels of work-related stress that have increased together with staff turnover rates (Perreira et al., 2018). 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. Consequently vacancy rates are high, standing at 38,000 full-time equivalent open posts in the first quarter of 2017/18 (June 2017) (NHS Digital, 2021) or 10.9%. Because of the high number of nursing vacancies, the NHS relies significantly also

[4]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.

[5]For brevity, in this work the terms "nursing staff" and "nurses" are referred to both nurses and midwives.

[6]Clinical staff includes Hospital and Community Health Service (HCHS) doctors, qualified nurses and health visitors, midwives, qualified scientific, therapeutic and technical staff, and qualified ambulance staff.

on temporary and agency staff, which cost NHS HOs approximately £1.46 billion per year (The Open University, 2018); an improvement in workforce retention is expected to reduce these labor costs as well as preventing losses in human capital.

2.2 The intervention: Retention Direct Support Programme

The Retention Direct Support Programme (RDSP) was designed by NHS Improvement to tackle the nursing supply challenge. The aim of the RDSP was to improve nursing retention in public acute care HOs and retention across the clinical workforce (i.e. nurses and doctors) in mental health HOs in England (NHS England, 2019).

The Programme was clinically-led, involving at least one member of the nursing team from the HO, and focused on factors that are 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. NHSI, the hospital monitoring body, allocated HOs into cohorts over time, starting with HOs that had above-average leaver rates.[7] The RDSP's first cohort was launched in June 2017; other cohorts started at later dates and respectively in October 2017 (Cohort 2), April 2018 (Cohort 3), and November 2018 (Cohort 4). By the end of 2018, 146 secondary care HOs had been enrolled in the Programme. In September 2019 the RDSP was extended to the other 62 Acute and Mental health HOs (Cohort 5) and all the Ambulance HOs in England until the Programme's *de facto* end in Spring 2020 due to the start of the COVID-19 pandemic.

Organization of the Retention Direct Support Programme

In the weeks following the first contact from NHSI, clinical and workforce leads from the HOs were invited to participate in a "Retention Masterclass workshop", scheduled ap-

[7] 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.

proximately with a six weeks' notice. These HOs 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 understand their retention profile and potential for improvement. The Retention Masterclass workshop was the official launch event of the RDSP; it introduced the Programme to the HOs' team and functioned as an interactive platform to review and discuss the barriers to retention in their organizations. During the workshop, NHSI also presented the domains that HOs might focus on to reduce their leaver rates, showcased best practices, and demonstrated how data can be used to inform decision-making. HOs were also given guidelines on ways to develop action plans.[8]

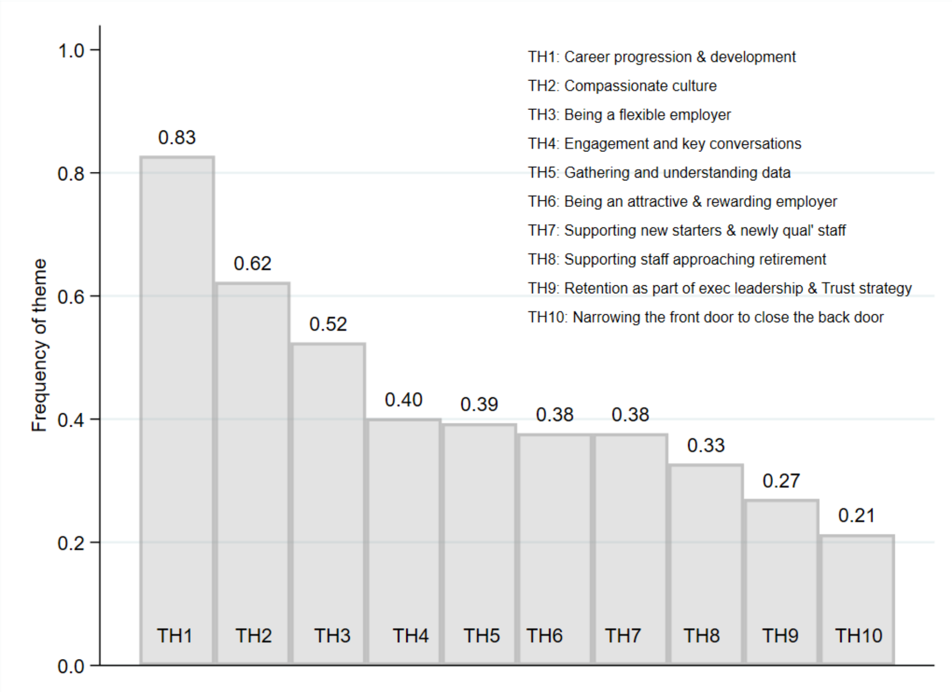
Each HO was matched with an NHSI officer acting as lead and collaborating with the HO for the whole duration of the RDSP. HOs were given 90 days to develop and submit a Retention Improvement Action Plan, and expected to use this period to review their data, identify areas of improvement, and set clear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

An important difference, setting the RDSP apart 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, was that the Programme 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, RDSP encouraged HOs to focus on retention challenges endemic to their workforce and to set their turnover goals accordingly. In this way, the Programme also enabled HOs to incorporate existing and planned workforce governance initiatives into their action plans.

NHSI identified 10 recurring themes from the submitted Action Plans. Figure 2 shows

[8] For more details on the workshop, the presentation for Cohort 1 is available at `https://www.networks.nhs.uk/nhs-networks/nhsi-retention-support-cohort-1/documents/14-july-workshop-slides`.

Figure 2: Frequency of Themes in RDSP Action Plans



Notes: Authors' calculations from NHSI thematic coding matrix from Cohorts 1 to 4. The themes are categorized by NHSI using action plans submitted by 122 Hospital Organizations.

the frequency of recurring themes in all Action Plans. At least three-quarters of the Trusts in the first four cohorts focused on career progression and development (TH1) of their nursing workforce, including strategies to develop clear and attainable career paths, re-design appraisal processes, and career coaching.[9]

The second most frequent theme arising from the Action Plans was establishing a compassionate culture (TH2). This theme covered action points and initiatives on different aspects of nurses and midwives' experiences at their organization, ranging from focusing on mental health and wellbeing needs to managing workloads and preventing burn-out. The emphasis on these aspects is in line with the high levels of stress identified in the NHS Staff Survey. Some HOs also mentioned their aim to work on recognition of good work and valuing staff, as well as strategies to reduce negative workplace experiences such as bullying and harassment.

[9]This is explained by 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, 2019), also due to the cuts to 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).

Slightly more than half of the HOs in the first four cohorts identified strategies to improve flexibility at work, another aspect emphasized by House of Commons Health Committee (2018) and the 2019 NHS Long-term Plan (NHS, 2019) as necessary for improved retention. The "Being a flexible employer" theme contained strategies offering flexibility in rotations, improving online shift-scheduling, and facilitating transfer schemes.

While pay is a contentious topic among nurses and midwives (Mirror, 2020; The Guardian, 2021; The New Statesman, 2021), it was not a recurring theme in the retention improvement plans. Only 13 of 122 action plans mentioned pay and it was classified under the promotion of the "rewards and pay" sub-theme of Theme 6, "Being an attractive employer". This is perhaps not surprising, as NHS nurses' and midwives' wages are negotiated and determined at the national level, with no scope for individual bargaining.

3 Data and Methods

To understand the impact of the RDSP on nursing retention in English secondary care, we construct a monthly panel of NHS HOs in England using various micro-level datasets.

We construct measures of retention from the monthly Electronic Staff Records (ESR) 2009-2020. 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 NHS in England. The information on the RDSP comes from NHS Improvement (NHSI), which was the NHS monitoring body responsible for the development and implementation of the intervention.[10] These data contain information about the timing of the Programme's roll-out.

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 wellbeing at

[10] NHSI works with the Department of Health and Social Care, and, together with NHS England, it monitors, oversees and provides support to NHS HOs to improve the provision of healthcare services to patients.

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 structure of our data and its setup.

Finally, to investigate the potential 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 information on admissions to acute care English NHS hospitals. For patient outcomes, 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, and to measure the hospital-level productivity as the number of emergency and planned admissions to acute care hospitals.

3.1 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. 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.In other words, we measure how many nurses and midwives are retained in a HO on a year-on-year basis for each month, which accounts for seasonality by comparing the same month a year apart.[11] By definition, the stability rate in calendar time t, S_{ht}, reflects the leaving decisions that occurred between $t-12$ and t; this may have implications for the impact evaluation of the Programme, as we further discuss in Section 4.

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), which is the rate of nurses and midwives who leave the NHS. While the RDSP did not directly aim to reduce the number of nurses and midwives who leave the NHS, some organizational changes instigated by the RDSP may also discourage nurses from leaving the NHS. Thus, 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.

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

We limit the sample for NHS leaver rates to nurses and midwives below the age of 65, which is the standard retirement age for nursing staff. Nevertheless, our measure still captures nursing staff who retire early, which is an important (and possibly preventable) source of exits from the NHS.

3.2 Empirical Strategy

We employ a difference-in-differences (DiD) strategy to assess how effective the RDSP has been in improving nursing retention in acute and mental health HOs in the English NHS. A naive estimation of the impact of the RDSP on retention, S_{ht}, and on leaver rates, L_{ht},

[11] 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%.

can be achieved with the following (static) two-way fixed effects (TWFE) baseline model:

$$Y_{ht} = \mu_h + \lambda_t + \beta^{TWFE} D_{ht} + \varepsilon_{ht} \quad , \tag{1}$$

where $Y_{ht} = \{S_{ht}, L_{ht}\}$ are the retention outcomes in HO h at calendar time t. D_{ht} is the treatment indicator, and takes the value 1 for all periods when HOs launched the RDSP. μ_h and λ_t are HO and calendar time fixed effects, respectively. The parameter of interest in this specification is β, which, under the parallel trends assumption, identifies the overall average treatment effect on the treated (ATT).

A common way to analyze the dynamics of treatment effects is through an event-study TWFE specification:

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

where D_{ht}^k is the event-time indicator for the relative time (in months) to/from RDSP, k. The lag parameters, $\delta_{k\geq 0}$, are the estimates for the treatment effects at k, and $\delta_{k<-1}$ are pre-treatment estimates, which are conventionally used for testing the parallel trends assumption.[12]

Recent methodological advances in the DiD literature have shown that β in Eq. (1) and δ_s in Eq. (2) may be biased when there is a staggered treatment adoption with multiple periods and heterogenous treatment effects (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; de Chaisemartin and D'Haultfœuille, 2020; Borusyak et al., 2021). The bias in β^{TWFE} stems from the variance weighting of the OLS, and, more importantly, from using the early-treated units as controls for later-treated units, i.e. making "forbidden" comparisons (Goodman-Bacon, 2021; Baker et al., 2022). Compared to static TWFE, the

[12]As it is standard in the literature, we exclude the month before the RDSP launched, $k = -1$, as the reference period.It is also common practice to bin or trim the relative time periods that are too distant from the treatment. In our estimation of equation (2), we bin the periods before and after 12 months to/from the RDSP.

differential treatment timing becomes less concerning in event-study TWFE approach as the length of exposure to treatment (start of the RDSP in our context) is taken into account explicitly. Yet, the lead and lag estimates might still be biased due to treatment effect heterogeneity, and due to treatment effects from other relative time periods (Sun and Abraham, 2021). To avoid the bad comparisons of the TWFE, the alternative heterogeneity-robust DiD estimators either provide flexible specifications by adding interaction terms for cohorts (e.g. Sun and Abraham, 2021; Wooldridge, 2021) or by transforming the comparisons into a conventional two groups - two periods setting (i.e treatment and control, before and after) and aggregating treatment effects (Callaway and Sant'Anna, 2021).[13]

In our analysis, we evaluate the RDSP's impact on nursing 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.[14] 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.[15]

Staggered-DiD and heterogeneity-robust estimation

The staggered-DiD approach by Callaway and Sant'Anna (2021) is based on a series of average treatment effect 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. (2022), under parallel trends and no anticipation, the average treatment effects in post-treatment

[13]For a review of TWFE and recent DiD estimation methods see Roth et al. (2022) and de Chaisemartin and D'Haultfœuille (2022).

[14]We use did package version 2.0.0 (Callaway and Sant'Anna, 2020) and csdid package (Rios-Avila et al., 2021) to estimate CSA models in R and Stata 16, respectively.

[15]SA interact relative time periods with cohort indicators, excluding indicators for never-treated group (or last treated cohort) in a linear TWFE framework. We use eventstudyinteract command in Stata 16 for estimation (Sun, 2021). The CSA and SA estimates are comparable when the control group consists of never-treated HOs and without covariates. The inclusion of time-varying covariates (linearly) in TWFE framework requires additional assumptions (Sant'Anna and Zhao, 2020) as treatment may have different effects across subgroups of the treated HOs (Baker et al., 2022; Roth et al., 2022).

period when $c \leq t$:

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

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 can consists of HOs that are either *never-treated* or *not-yet-treated* by t. Likewise the pre-treatment effects for the period when $c > t$ are

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

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 use the last period before the treatment as the reference period for all differences.[16]

CSA also allows for parallel trends to hold after conditioning on covariates through a doubly robust estimation method.[17] Conditioning on trends on observed covariates can be considered as a more credible approach compared to unconditional parallel trends assumption if retention trajectories depend on factors that would determine HOs' allocation into cohorts. As the HOs allocation into RDSP cohorts was not exactly random, we provide an additional robustness check by assuming under parallel trends conditional on past retention values and other covariates.

A challenge in estimating models with conditional parallel trends is to find covariates that provide a common support for the propensity score for Callaway and Sant'Anna (2021)'s doubly-robust estimator. We use the difference between the average past retention of each HO and its allocated cohort to capture variations in past retention. As shown in Figure 3,

[16] For a discussion on varying reference periods, see Callaway (2021).

[17] 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).

there was some variation in retention across HOs in each cohort, providing common support for *treated* and *never-treated* HOs. To capture baseline differences in organizational and workforce practices, we also control for the age of creation of the HO[18] and nurses and midwives' sickness absence rate. In specifications using stability rates as outcomes, we also include items from the NHS Staff Surveys to characterize the workplace environment before the launch of the RDSP: the share of nurses and midwives among the NSS respondents of the HO; and the shares of nurses and midwives who are satisfied with their recognition of good work, and with the support from their immediate managers and co-workers, which has been found to influence nursing retention (Marufu et al., 2021). All these covariates are set to their pre-treatment values, i.e. the month before the RDSP was launched in a cohort, and employed only to estimate the propensity score used to re-weight the effects, thus they do not influence the RDSP retention outcome directly.

Policy makers are often interested in an overall effect as well as how the effects evolve over time. An additional advantage of the CSA approach is the option to aggregate average treatment effects into a single policy effect. We can aggregate $ATT(c,t)$s to understand the overall impact of taking part in the RDSP, across cohorts and over time:

$$ATE = \sum_{c \in C} \sum_{t} \omega(c,t) \cdot ATT(c,t) \quad ,$$

where we choose relevant weights, $\omega(c,t)$, e.g. by different lengths of treatment exposure (event-study ATEs), or by the time each cohort spends under treatment (cohort-specific ATEs).

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,

[18]Older HOs might have some established workforce retention practices.

and HOs in Cohort 5 as controls, i.e. *never-treated* HOs.[19] One advantage of using Cohort 5 as the never-treated comparison group is that it allows us to estimate the RSDP's impact on each cohort for a period of at least 12 months. In particular, this affects HOs in Cohort 4, whose 12$^{\text{th}}$ month into RDSP falls on November 2019. As such, we restrict our estimation 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 changing the definition of the control group to be *not-yet-treated* HOs, i.e. all the HOs belonging to cohorts that have not yet started the intervention.

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 weeks in advance, with only limited information about the scope and the extent of the Programme as this was delivered by the initial Retention Masterclass workshop. The short notice period minimizes the risk of potential anticipation for individual HOs.

Effects on quality

NHS hospital Trusts are frequently subject to localized policy interventions aimed at improving patient care. These policies often prioritize hospital organizations falling short in terms of minimum care quality standards, which is known to be negatively associated with high staff turnover. Hence, the Callaway and Sant'Anna (2021) strategy may be inadequate to infer about the indirect effect of RDSP on hospital quality measures, as the pre-intervention trends in hospital quality performance might be affected by target interventions aimed at raising the minimum care quality standards and different from the RDSP.

For this reason, to study the potential effects of the RDSP on patient care quality we

[19]It is not possible to capture the impact of the Programme for Cohort 5 in such a limited time frame.

rely on the Arkhangelsky et al. (2021) synthetic difference-in-difference (SDiD) approach. Intuitively, the SDiD estimator allows us to compare the change in quality performance of treated hospital organizations over time with that of a re-weighted set of never-treated hospital organizations, where weights are chosen with the explicit objective of satisfying the parallel trend assumption prior to the RDSP start. 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, and hospital waiting times for planned procedures. The mortality and readmission hospital quality measures have been computed through the standardization method used by NHS Digital, the official data provider and statistics office of the NHS.[20]

$RDSP_{h,t}$ is our 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.

[20]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.

Sample restrictions

We restrict our attention to nursing and midwifery staff working in acute (including Community and Specialist care) and mental health care HOs. We also exclude a small number of HOs that have undergone organizational changes, e.g. mergers and acquisitions, from July 2016 onward, because in this case it is not possible to univocally assign a HO to a certain cohort. We also exclude one HO with very few nursing staff[21], one HO that did not keep workforce information in ESR and one HO which have started using the ESR towards the end of the sample period.

Our analysis 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).

3.3 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).

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 under the age of 65 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 3 shows that the

[21] The HO had on average only 20 nurses in each month, with the second smallest HO having a nursing staff eight times larger.

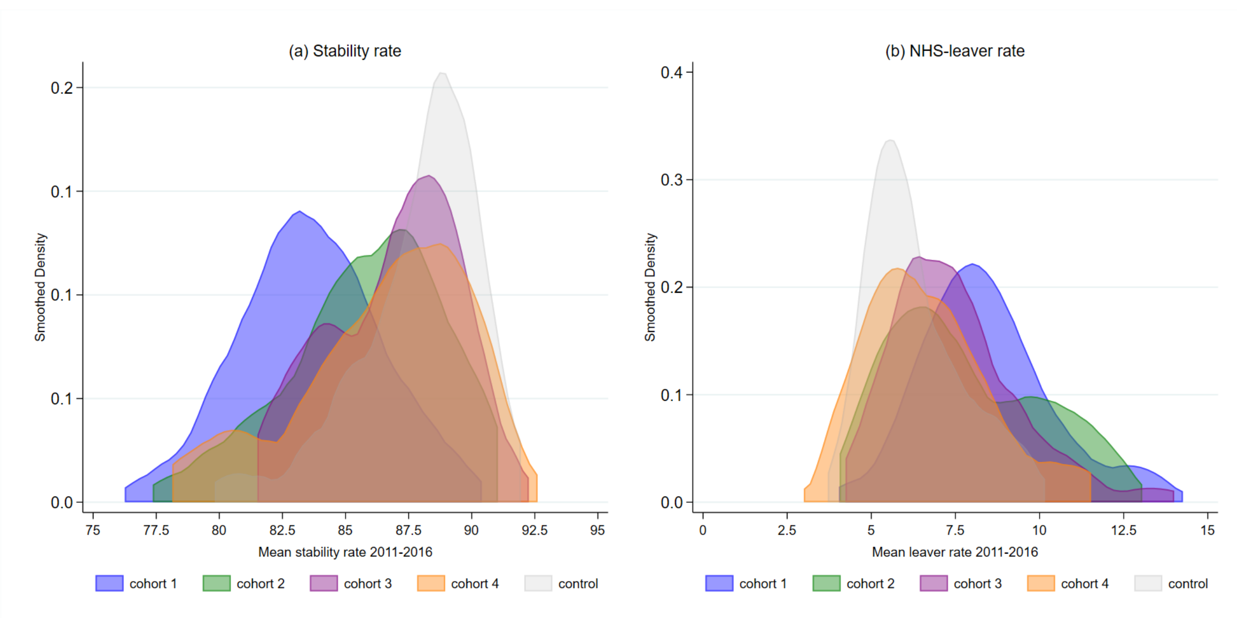
Table 1: Overview of Cohorts in the RSDP

	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	8.45% (1.98)	7.79% (2.16)	7.50% (1.85)	6.68% (1.77)	6.28% (1.37)
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.

distributions of pre-RDSP retention measures substantially overlap among the treated and the control group.

Figure 3: Distribution of average monthly retention measures between 2011/12 - 2015/6



Notes: Smooth histograms are calculated using a kernel density smoother.

Despite differences in their retention levels, HOs across all treated cohorts experienced similar trends with decreasing (increasing) stability rates (NHS leaver rates) from early

2011/12 to 2016/17 until the RDSP was launched (see Figure A2 panels (a) and (b) in Appendix). Similarly, from a closer visual inspection of the pre-trends in Figure 4, we see that all treated cohorts exhibit retention trends similar to the control cohort in the months leading up to the RDSP.[22] This finding is not surprising, considering that the allocation of HOs into RDSP cohorts was primarily based on the retention *levels* in the 5 years preceding the start of the Programme. As this is quite a long period, selection into the Programme was not correlated with retention trends.

In Table A1, we test the mean differences in retention outcomes between *treated* and *control* cohorts for the month before the RDSP launched (pre-RDSP) and at the end of the sample period: 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).

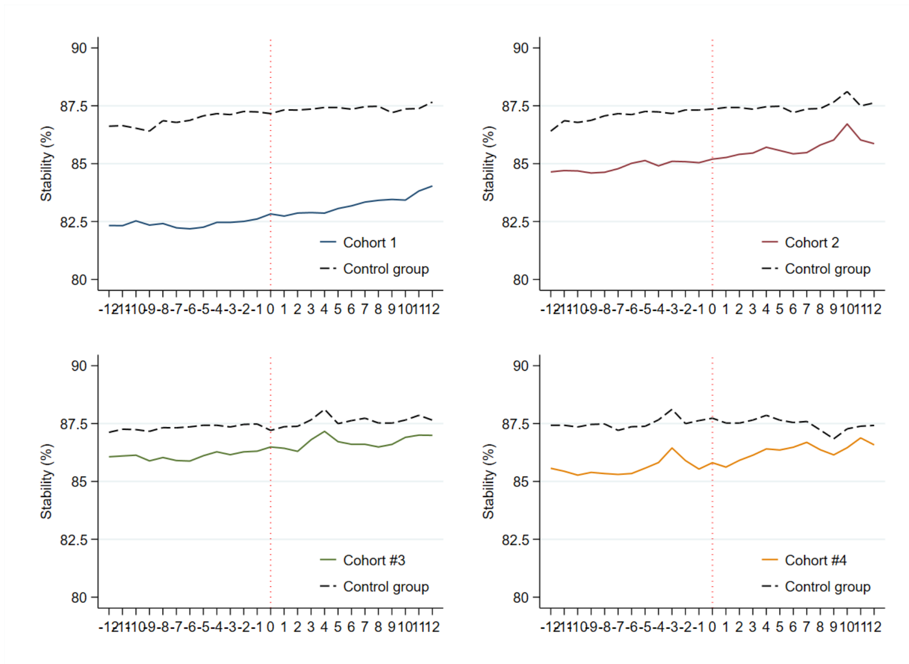
Finally, Appendix Table A2 presents the summary statistics for selected characteristics from June 2017, i.e. the month before the RDSP was first launched.[23]

[22] In the next section, we show that parallel trends hold unconditionally under multiple hypothesis testing.

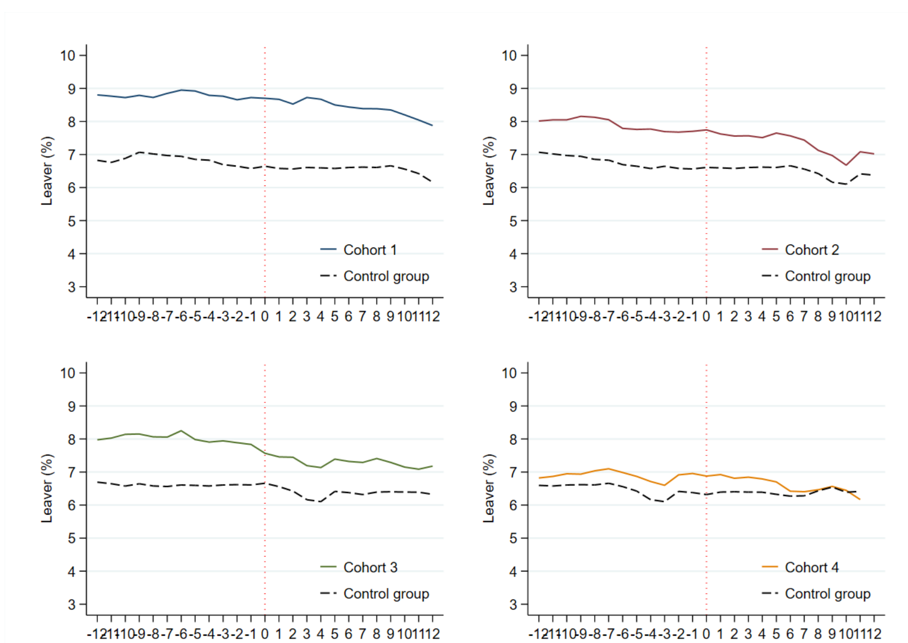
[23] The majority of nursing staff is female in all cohorts, and one-in-eight nursing staff is from overseas in the first two cohorts, whereas this rate reduces to 1 in 15 in the control cohort. There are slightly more nurses and midwives from an ethnic minority background in treated cohorts compared to in the control group. Many of these differences are likely to reflect regional labor market differences, which are not directly under the control of the HOs.Retention also varies across regions, and this is partially reflected in the composition of the cohorts. More than a quarter of the HOs in London and East of England were in Cohort 1, followed by one fifth of secondary care HOs in South East. On the other hand, more than half of the HOs in South West were in the control group. The average monthly stability rate of nurses and midwives between 2011/12 and 2015/16 ranged from 82.64% in London to 89.20% in North East and Yorkshire. Before the RDSP's roll-out, the overall nursing staff engagement stood on average at 7.15 on a scale of 0 (lowest) to 10 (highest), and only Cohort 3 had lower average engagement score than the control group in 2015 NSS.In all cohorts, more than half of the nursing staff agreed that they receive recognition for their good work. In terms of work organization, the nursing staff in control HOs is less likely to work more than 11 additional unpaid hours than treated HOs with around one in 16 nurses in the first cohorts working additional unpaid hours. The average monthly sickness rate is also similar across cohorts and control group in the month preceding the RDSP's first launch at around 4%.

Figure 4: Common trends between treated and control cohorts

(a) Stability rate



(b) NHS leaver rate



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.

4 Results

4.1 RDSP effects on nursing workforce retention outcomes

At cohort level, the RDSP showed positive treatment effects in the months after its launch. Figure A3 panel (a) shows that the Programme had took some time to yield positive effects on nurses' and midwives' stability rates in treated hospitals. The first group of treated HOs (Cohort 1) started improving their nursing staff's stability significantly towards the end of the observation period, whereas the last Cohort (Cohort 4) experienced stronger effects after a shorter exposure to the RDSP. As our sample has a large number of pre- and post-treatment periods which render the individual cohort-time effects not very informative to evaluate the effectiveness of the Programme. Thus in the remainder of this section, we primarily focus on the ATT estimates aggregated across the full set of HOs, which are more policy relevant, and also briefly discuss the overall impact of the programme at cohort-level.

The first panel of Figure 5 is an event-study plot presenting aggregated treatment effects over time across treated HOs in the first year of the RDSP and compares estimates from alternative difference-in-differences strategies.[24] The identifying assumption for our analysis, i.e. the parallel trends assumption, is supported by the pre-test proposed by CSA[25], and they are not sensitive to potential linear extrapolation of parallel trends based on the formal sensitivity test proposed by Rambachan and Roth (2022).[26]

On average, the impact of the RDSP took off after the first three months. This is likely because during this period HOs worked on designing their Retention Improvement Action

[24]Figure A4 also illustrates the dynamic treatment effects based on bootstrapped standard errors with uniform 95% confidence intervals.

[25]The pre-test is based on pre-treatment cohort-time estimates, which are also robust to selective treatment timing. We test the hypothesis whether the pre-treatment estimates for the 6 (12) months leading to the treatment are jointly equal to zero using an augmented Wald test. We fail to reject with a p-value of 0.273 (0.080).

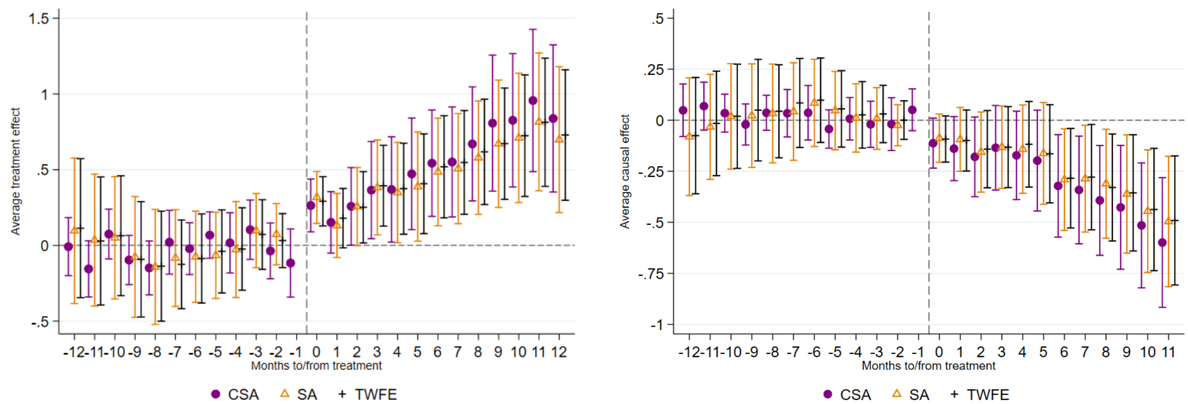
[26]We test the sensitivity of the slope (smoothness) restriction that the post-treatment violations in parallel trends cannot deviate much from a linear extrapolation of pre-treatment trends by a factor $M.M = 0$ indicates that counterfactual difference in trends is linear and $M > 0$ means non-linearity.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

Plans, but had not implemented them yet. We find a stronger increasing impact of the RDSP over time and we explore this increasing pattern further in Section 4.2.

Figure 5: Event study estimates of the RDSP on nurses' and midwives' retention

(a) Stability rate

(b) NHS Leaver rates



Notes: Event study estimates are based on aggregated cohort-time ATTs under unconditional parallel trends assumption. To facilitate comparisons across different estimators, the figure shows point-wise 95% confidence intervals from asymptotic standard errors clustered at HO level. Both in Sun and Abraham (2021) and the TWFE specifications use universal reference period, i.e omits the month before the RDSP, $k = -1$ and we bin the periods for $k < -12$ and $k > 12$ rather than trimming the sample.

We find similar results using both Sun and Abraham's (2021) interaction-weighted estimator and dynamic TWFE. A potential explanation for the similarity between the dynamic TWFE estimates and the robust estimators is the relatively large number of *never-treated* HOs compared to treated HOs at each time period. The large size of the control group helps reducing the importance of negative weights at each length of exposure to treatment and minimizing the bias from using early-treated units as controls (Jakiela, 2021; Baker et al., 2022).

Overall and cohort-specific impact of RDSP on retention under unconditional (columns 1 and 3) and conditional (columns 2 and 4) parallel trends assumption are presented in Table 2. Overall, the Programme improved nurses' and midwives stability by 0.78ppt, which translates into a 0.92% increase in retention based on the average pre-treatment stability rate and comparable to the doubly-robust estimate of the overall ATT at 0.86ppt.[27] Compared to

[27]We also estimated alternative specifications including the share of nurses who felt unwell due to work

Table 2: Average treatment effects of RDSP on retention outcomes

	Stability rate		NHS-leaver rate	
β^{TWFE}	0.472		-0.249	
	(0.170)***		(0.125)**	
Callaway and Sant'Anna (2021)				
Overall ATT	0.775	0.862	-0.408	-0.389
	(0.188)***	(0.199)***	(0.131)***	(0.137)***
	[0.196]§	[0.195]§	[0.128]§	[0.135]§
Cohort-specific ATTs				
Cohort 1	0.950	1.017	-0.498	-0.448
	(0.352)***	(0.366)***	(0.254)**	(0.261)*
	[0.360]§	[0.388]§	[0.233]	[0.278]
Cohort 2	0.677	0.616	-0.416	-0.377
	(0.303)**	(0.361)*	(0.202)**	(0.236)
	[0.309]	[0.348]	[0.209]	[0.231]
Cohort 3	0.557	0.843	-0.394	-0.386
	(0.336)*	(0.403)**	(0.248)	(0.253)
	[0.333]	[0.394]	[0.250]	[0.251]
Cohort 4	0.912	0.944	-0.341	-0.351
	(0.267)***	(0.265)***	(0.183)*	(0.183)*
	[0.269]§	[0.259]§	[0.181]	[0.183]
Conditional parallel trends (PTA)	no	yes	no	yes
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 clustered at HO level. Asymptotic standard errors are in parentheses, and estimated using csdid package in Stata 16 (Rios-Avila et al., 2021), p-values *p<0.1; **p<0.05; ***p<0.01. Clustered bootstrapped standard errors are in brackets, and estimated using did package in R (Callaway and Sant'Anna, 2020)). § indicates that the 95% simultaneous confidence band does not cover 0. Estimation period ends in November 2019 for stability rate and October 2019 for leaving the NHS rate.

conventional DiD estimation, the overall effect under unconditional parallel trends assumption is twice the magnitude of the conventional TWFE estimate (0.47ppt) as β^{TWFE} is an average difference between the treated and control group which does not take into account the differential treatment timing.

We also find that the RDSP increased the stability of nursing staff in all cohort, particularly in the first and last cohorts. On average, Cohort 1 increased its nursing stability by

stress, and the share of those who were bullied at work by their managers or co-workers in the last 12 months. The results are similar to the baseline model presented in Table 2, and are available upon request.

1.15% in 28 months. This translates to a one-third of a standard deviation based on their pre-RDSP stability rate. Despite spending half of the time under RDSP, HOs in Cohort 4 improved their nurses' stability by 1.06% in 12 months. We find similar results under conditional parallel trends assumption, with a slight, but statistically insignificant, increase in the size of ATTs, particularly for Cohort 3.

The second set of results relate to the loss of human capital from the NHS, i.e. the NHS leaver rates of nurses and midwives. As shown in panel (b) of Figure 5, the treatment effect profiles are similar across cohorts. The RDSP led to a gradual decrease in NHS-leaver rates of nurses and midwives in treated cohorts.[28] The NHS leaver rate drops on average by 0.32ppt at the 6 month and 0.60ppt in the 11 month of the Programme.[29] We also find that our results are robust to alternative dynamic DID estimation methods.

By aggregating the cohort-specific ATTs across all cohorts, we find that the RDSP decreased the NHS leaver rates of nursing staff on average by 0.41ppt, which is a 5.38% reduction in the pre-RDSP NHS leaver rates. We find a similar result under conditional parallel trend assumption with 5% reduction in NHS-leaver rates of nurses and midwives compared to the pre-treatment average.[30] The impact of the Programme was less precisely estimated at cohort level, but has the expected negative signs.

4.2 Robustness Checks

Partial treatment and the impact of the RDSP over time

As described in Section 3.1, 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

[28] As with the stability rates, the pre-test for parallel trends provides statistical evidence that the nurses and midwives' NHS leaver rates follow common trends in pre-treatment periods. The associated p-values for the test are presented in the last two rows of columns 3 and 4 of in Table 2.

[29] As described in Section 3.1, to distinguish NHS leavers from churns, the outcome measure is constructed by following up nurses and midwives in the payroll data for 5 consecutive months, until $t+5$. Therefore, the last non-missing value for the NHS leaver rates is on October 2019.

[30] Through the propensity score estimation, we indirectly control for the mean difference between the past 5 years' average NHS leaver rates of HO and the allocated cohort, the HO-level sickness absence rate of nurses and midwives and their pay satisfaction. These findings are robust to alternative specifications including work stress, bullying and fair career progression.

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 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.

We expect that the nature of the outcome variables leads to an underestimation of the ATT. To check its implications, we compare the overall and cohort-specific ATTs in Table 2 with the censored ATTs, which we obtain by averaging the cohort-specific ATTs from 12$^{\text{th}}$ to 19$^{\text{th}}$ months into the Programme, and also from 12$^{\text{th}}$ to the final observations of our sample.

Table 3: Breakdown of RDSP impact over time

(a) Stability rates

	ATT [0,11]		**ATT [12,19]**		**ATT [12,τ]**	
Overall	0.519	[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]	0.853	[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-leavers' rates

	ATT [0,11]		**ATT [12,18]**		**ATT [12,$\tau-1$]**	
Overall	-0.294	[0.106]§	-0.531	[0.215]§	-0.609	[0.229]§
Cohort 1	-0.270	[0.257]	-0.521	[0.308]	-0.669	[0.307]§
Cohort 2	-0.258	[0.188]	-0.507	[0.279]	-0.563	[0.312]
Cohort 3	-0.297	[0.233]	-0.560	[0.367]	-0.560	[0.367]
Cohort 4	-0.341	[0.183]				

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 set 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. $\tau=28$ for Cohort 1, $\tau=25$ for Cohort 2, $\tau=19$ for Cohort 3, and $\tau=12$ for Cohort 4.

The censored ATT estimates are presented in Table 3. Although higher (lower) in magnitude, we find that the truncated impact of the RDSP, both overall and at cohort-specific level, on stability (NHS leaver) rate is not statistically different from the main effects reported in Table 2. The largest difference between ATT estimates is for Cohort 3, but the difference is not statistically significant (0.56ppt vs 1.01ppt). From this analysis we conclude that the ATTs in Table 2 are likely to be conservative point estimates of the Programme's impact, which might be even more effective in increasing nursing workforce retention. Despite this, we favor the uncensored results as they allow us to capture the impact of Programme from its point of introduction and allow the inclusion of Cohort 4, which is only observed for 12 months post-treatment.

Choice of the reference control group

In our main analyses, we use HOs in Cohort 5 as the *never-treated* control group as they started the RDSP towards the end of the national retention Programme, and mainly consisted of HOs that have below-average turnover rates. To this extent, we set the end of our sample period to November 2019. We do not expect any large bias to arise from this, given that the launch of RDSP for Cohort 5 occurred in September 2019 and the first 90 days since the Programme enrollment were devoted to developing and submitting the Retention Improvement Action Plan, thus without any active implementation happening before the formal agreement with the NHSI monitoring body.

To check whether our findings are sensitive to the definition of the comparison group, we re-estimate our models where our control group consists of the *not-yet-treated* HOs, rather than the *never-treated* Cohort 5. This increases the number of HOs in the control group for each time period, some of which might be more comparable to early treated HOs, but comes with a disadvantage of shorter analysis period. In practice, this means that the post-treatment period spans only until August 2019, as HOs in Cohort 5 were enrolled in the Programme from September 2019.

We present the estimation results in Table A3 for stability rates and leaving the NHS rates. Columns (II) show the results from the model 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). We find that the RDSP increased nursing retention on average by 0.68ppt, which is slightly, but not significantly, lower than our baseline estimates. The difference is mainly because the post-treatment period spans only until September 2019, when the last Cohort was enrolled in the RDSP. As RDSP's impact increased over time, reducing the post-treatment period leads to a lower overall effect of RDSP on nursing retention; this occurrence may particularly affect Cohort 4, which ends prematurely at 9 months rather than 12 months and is therefore more subject to the partial treatment issue mentioned above. Our results hold for nurses and midwives' NHS leaver rates when we re-define our comparison group as not-yet treated HOs: we find the RDSP has decreased the NHS leaver rates on average by 0.37ppt. Thus, our results are robust using Cohort 5 as the never-treated control group with the additional advantage of observing treatment effects in later periods by increasing the length of post-treatment period.

Falsification tests

We test the validity of our results for stability of nurses and midwives by performing two placebo tests by (i) changing the start time of the Programme and (ii) randomizing the cohort compositions.

In addition to the suggestive statistical evidence and formal sensitivity tests on common trends à la Callaway and Sant'Anna (2021) and Rambachan and Roth (2022), we perform a conventional placebo test for pre-trends by changing the time period the RDSP was launched. We use three different time periods before July 2017 and assume that the RDSP was introduced in the same month but in earlier years, i.e. June 2010 - December 2013 (Placebo 1), June 2013 - December 2016 (Placebo 2) and June 2011 - December 2014 (Placebo 3). As the placebo periods date back the original time frame, some HOs were not open during some placebo periods. This means there are fewer HOs in some periods, but as shown in Table A4, the relative size of cohorts are similar to the original distribution. Table A5 presents the results for the first falsification test together with the re-estimated model for the original time period using the same set of HOs as in placebo samples. In none of the placebo periods, we find a positive effect of the programme, which supports our identifying assumption for the main models.

We perform also another falsification test to evaluate the likelihood that our estimates on retention outcomes are false positives, by randomly change the allocation of HOs to cohort. We use the original observation period, from June 2016 to November 2019, randomly allocate HOs into cohorts but keeping the cohort sizes equal to the original allocation, and estimate the effects of interest. We repeat this estimation exercise for 500 replications. We should expect to find high chances of significant effects of the Programme on retention outcomes even under a random allocation of HOs to cohorts, if our Table 2 estimates are highly falsifiable. Figure A6 presents the overall ATT estimates from random cohort allocations, for all replications (panel a) or just for the random allocations that do not violate the parallel pre-trends test at 12 months (panel b). Compared to our baseline overall ATT, the point estimates from the replicated samples are smaller and none of them are significantly different than zero, which provides additional reassurance about the robustness of our estimates.

Spillover effects and workers' mobility

Another potential concern in our identification setting is the existence of potential spillover effects across HOs from different cohorts. These could arise among HOs that are geographically close to each other or have some sort of cooperation. This possibility should be discounted for at least two reasons: first, the customized nature of the retention strategies adopted by each HO means that the strategy adopted by one may not be suitable elsewhere; additionally, HOs not included in the RDSP would not have access to the bespoke data and support from their supervising lead officer at NHSI. Nevertheless, we conduct some empirically tests. An immediate approach to check whether nursing staff migrate to HOs where work standards are improved by the RDSP adoption is to look at the transition frequencies of nursing staff before and after the RDSP was launched in their origin hospital, by focusing on workers who changed their HO at least once between June 2016 and November 2019. If employees' churn across HOs was motivated by the RDSP introduction, we should find higher transition rates from untreated hospitals to treated hospitals. Each row in Figure A7 illustrates the transition rates of nursing staff who switch hospitals before (panel a) and after (panel b) the RDSP was launched in their HO (in y-axis) taking into account the staggered adoption of the Programme. The green bars indicate transitions rates to not-yet treated hospitals, 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 for before and after RDSP was launched. This provides support that our results are not driven by such movements.

Next, we look at the potential impact of the RDSP on the share of overall new joiners, joiners to the NHS and joiners from other NHS HOs (churn joiners). As shown in Table A6, the overall impact of RDSP on joiners is small and the estimates are not significantly different from zero, moreover with negative signs. This is not completely surprising: as the RDSP increased the retention rates on average, more nursing staff were retained within their HO, thus likely leading to fewer vacancies and less need of new joiners to be hired as nurses and midwives.

4.3 Which workers benefited the most? Heterogeneous effects by workers' and jobs' characteristics

Our main analysis is based on the whole population of nurses and midwives actively employed by English NHS HOs. Here 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 seniority, ethnicity minority status and nationality (Non-British vs British).

We define seniority of nurses and midwives by their NHS Agenda for Change pay bands and by their age, separately. Pay bands broadly defines two skill groups, with senior nurses in pay bands 6 and higher.[31] As experience is highly correlated with age, we also assume

[31] 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 pre-treatment period.

that nursing staff who are 41 and older are in more senior roles than their younger colleagues. Overall, the RDSP improved both junior and senior nursing staff's stability by 1 ppt and 0.57ppts, respectively. We find similar results when we use the age criterion to assign seniorty with younger (older) nurses increasing their stability due to the Programme by 0.98ppts (0.62ppts) on average. As the event-study plots in Figure A8 shows, the intervention took off quicker and with stronger impact for junior and younger nursing staff.

The Programme improved the retention of nurses and midwives from a white background by 0.66 ppts on average. However, with 1 in 5 nursing staff in treated cohorts are form an ethnic minority background (see Table A2), the Programme had some positive, albeit not statistically significant, effects on the stability of Black, Asian and other minority ethnic groups. We also find that the programme was effective for retaining British nursing staff and had very little effect, if any, on non-British nurses and midwives working in treated NHS HOs. Figure A9 presents the event-study plots by ethnicity and nationality of the nursing staff.

Additionally, we examine the impact of the Programme on the retention of several sub-groups of nurses and midwives: (i) those *not* on zero-hour, temporary contracts that are not shift-based (called "*Bank* staff" in the NHS), (ii) those on permanent contracts (as opposed to workers on fixed-term contracts), (iii) those employed by acute care HOs (as opposed to nurses employed by mental health care HOs), and finally, just for the stability rates, nursing staff below the legal retirement age of 65 years.[32] Table B1 presents the aggregated average treatment effects on retention outcomes for aforementioned different sub-samples, following the same CSA specifications under conditional parallel trends assumption in Section 4.1. The overall impact of the RDSP are very similar to the retention outcomes defined over a broader group of nursing staff in the baseline models. The cohort-level ATTs also exhibit similar patterns with the first and last cohorts gaining higher returns from the RDSP. The only exemption is for Cohort 3 HOs, which significantly improved the stability of nursing

[32]This is the same group of nurses used for the NHS leaver rates analyses.

staff who do not work only as Bank staff (column 2 in Table B1).[33]

4.4 Heterogenous effects on retention outcomes by RDSP themes

The RDSP policy has improved nurses' and midwives' retention overall. From a policy point of view, we wish to understand which aspects of the policy and hospital strategies were most effective. To do so, we exploit the information that we have about the broad themes that appeared in the action plans for a large sub-sample of hospital organizations from Cohort 1 to 4 (112 out of 132 treated HOs).[34]

As discussed in Section 2, the categorization of the broad themes was done by the Programme's monitoring body, NHSI, and derived from an ex post grouping of the approved RDSP plans. Table A7 presents percentage share of the themes adopted by hospitals within cohorts. As each HO is encouraged to create their bespoke action plans tailored to their specific retention challenges, we observe some differences in broad themes they tackled. We use this information to understand what strategies might have worked better to improve retention across HOs in different cohorts.

We estimate the impact of the specific areas of the intervention on nursing retention, and run staggered difference-and-differences models for each theme, separately. In this case, the treatment defined $d_{hca} = 1$ for an HO h in cohort c adopting an action plan under theme a, and $d_{hca} = 0$ for control group, i.e. Cohort 5. As most of the strategies are multifaceted and might have spillover effects, we exclude treated HOs that did not have any strategies under theme a.

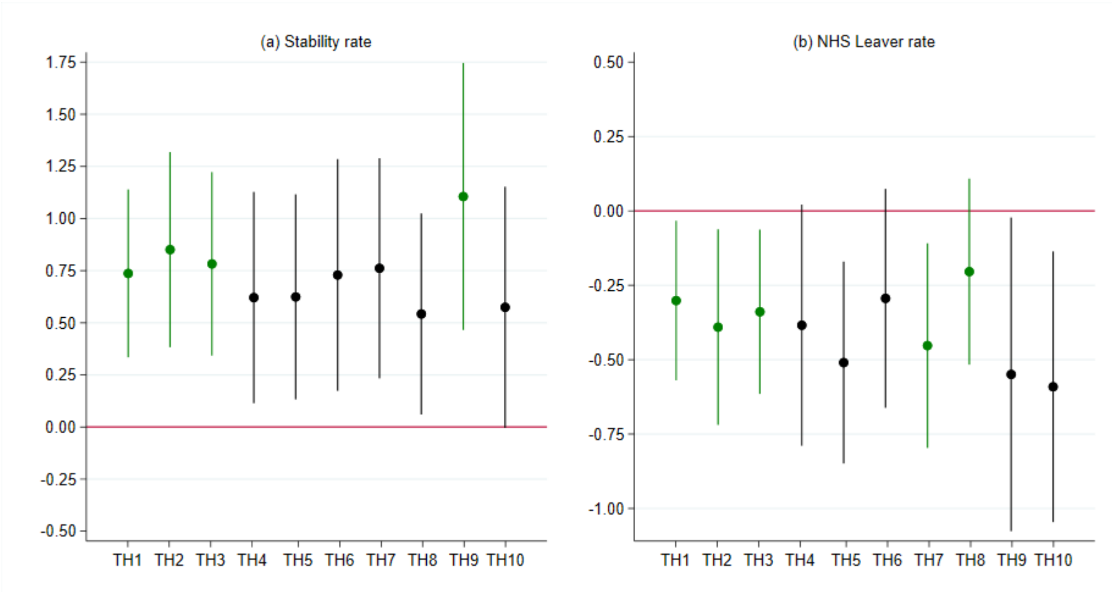
Figure 6 presents the overall impact of the themes on nursing staff's stability (panel a) and NHS leaver rates (panel b).[35] Overall all themes had some positive effect on retention of nurses and midwives in treated HOs. The parallel trends assumption holds for some

[33] The full set of results from the heterogeneity analysis are discussed in detail in Appendix B.

[34] We also check the robustness of our main set of results by using only 112 treated hospitals in CSA estimation. Under unconditional parallel trends, the overall impact of the Programme is comparable with those presented in Table 2 with 0.81ppt increase in stability and 0.44ppt reduction in NHS leaver rates of nurses and midwives.

[35] Table A8 shows the overall treatment effects and cohort-specific effects of each theme in detail.

Figure 6: Effects of RDSP themes on Stability rates for Cohorts 1-4



Notes: The effects of the RDSP action plan themes on nurses and midwives' stability (panel a) and NHS leaver rates (panel b) under unconditional parallel trends assumption. Those that do not violate the parallel trends assumption for the 6 pre-treatment months are shown in green. The point estimates are shown with 95% confidence bands from clustered asymptotic standard errors clustered at HO level.

models (depicted in green), for which we can reliable attribute a causal effect of the theme on retention. Despite not being central to many HOs, having retention as a part of the executive leadership and HO strategy (TH9) helped improving the stability of nurses the most by 1.1ppts on average. This broad theme involved plans such as including retention into quality improvement strategy, commitment to diversity & inclusion, managing capacity and intensity of workload, and support from senior management for retention initiatives.

The other themes that had significant impact on nursing staff retention, both in terms of stability and NHS leaver rates, were career progression & development (TH1), creating compassionate culture (TH2), and being a flexible employer (TH3).[36] These themes highlight the most commonly adopted strategies among treated hospitals and based on improving working conditions and culture in the workplace.

[36]The strategies under these themes involved developing clear career pathways, career coaching and additional training (TH1), investing in staff's physical and mental health needs at work, providing pastoral support and taking action to reduce negative experiences such as bullying and harassment at work (TH2), revisiting existing flexibility offers, supporting staff to create work-life balance and using electronic rosters for shift management (TH3).

In addition to these three themes, supporting new starters/newly qualified staff (TH7) and those who are approaching retirement (TH8) reduced the NHS leaver rates. Strategies such as ensuring the new recruits are confident in their posts, having more informative induction sessions and providing continued support to new joiners reduced the share of nurses and midwives who would have otherwise left not only their hospital but the NHS as a whole. Despite positive impact of strategies centered around retirement staff had a weaker effect in reducing NHS leaver rate.

4.5 RDSP effects on labor supply at the intensive margins

So far our analysis has considered nurses' and midwives' labor supply at the extensive margin as proxied by stability and NHS leaver rates, because the explicit aim of the RDSP was to reduce turnover rates and improve retention within and across HOs. Nevertheless, some strategies outlined in action plans, such as e-Rostering[37], might have led to a re-allocation of working hours and encouraged nurses to work more.

We focus on the average monthly hours worked by nurses and midwives who work full-time.[38]

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.[39] There was very little variation across full-time hours worked across cohorts in the last 3 years leading up to the RDSP (Figure A10 panel a). Yet, there were significantly fewer full-time nursing staff in the control group than there was in treated cohorts, particularly in the first two cohorts (Figure A10 panel b).

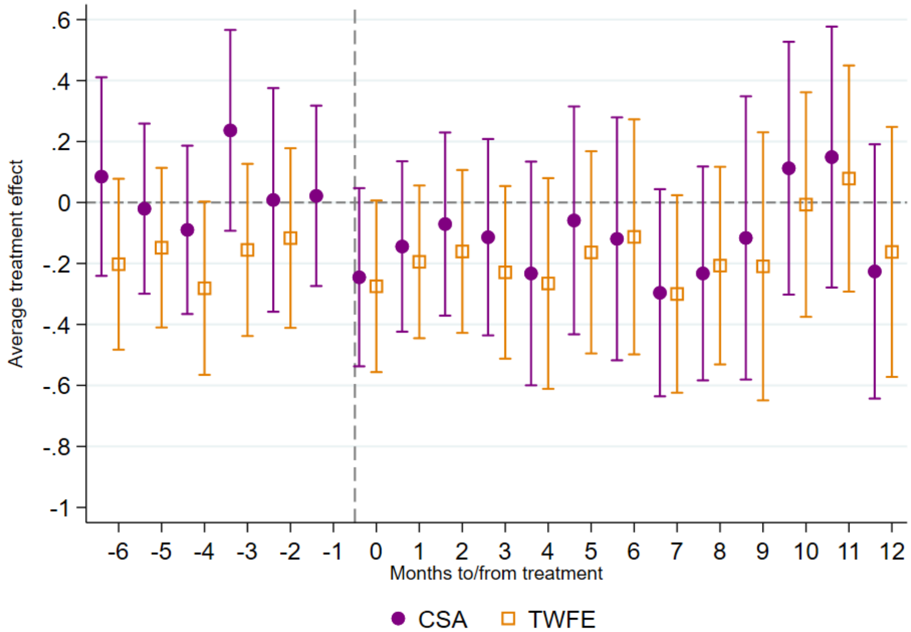
The RDSP did not have any impact on the average hours worked by full-time nurses

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

[38]We exclude negative and zero hours from the sample, and define full-time job by the total monthly work-time equivalent (WTE) of at least 0.95. 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.

[39]The ESR is a payroll data, thus it does not have information on unpaid hours. Nurses and midwives are likely to work additional unpaid hours to cover shifts and provide quality patient care.

Figure 7: Dynamic average treatment effects on full-time hours worked



Notes: Hours worked is calculated using nurses and midwives who worked full-time with total WTE $\geq$ 0.95. Figure is from the model estimated under unconditional PTA, and shows estimates with 95% point-wise confidence intervals based on asymptotic standard errors clustered at HO level. The p-value for the augmented Wald test for that CSA pre-treatment estimates in the 6 months leading up to the RDSP is 0.153. For dynamic TWFE, the periods more than 6 months before and 12 months after the RDSP are binned and not shown in the figure.

and midwives in treated HOs, as shown in Figure 7. This is not surprising, as the RDSP's primary target was to improve working conditions affecting staff retention, but the same conditions do not provide any strong incentive to work longer hours.

The Programme might have improved flexibility offers through additional Bank work. In 2016, the year before the RDSP was launched, the average Bank hours made up 1.5% of nurses and midwives monthly hours, and conditional on being registered as a Bank nurse or midwife the average increased to 12.3%. The share of Bank hours within nursing staff's working hours converged across cohorts in 2017, and has been on an increasing trend since then (Figure A11). We do not find supporting evidence that the retention Programme had an impact on the share of Bank hours (see Figure A12).

4.6 Effects on hospital quality and waiting times

There is limited empirical evidence on the impact of non-monetary policies aimed at improving staff retention on patients' health outcomes and hospital performance measures. 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, it is also possible that the interventions within RDSP to improve nurses' working conditions redirected some effort that might have been dedicated to patients. Moreover, as no benchmark on patients health outcomes was ever defined for the RDSP, any indirect RDSP effect on quality would have to be mediated by the direct effect on nurses' retention.

To understand the impact of this non-monetary intervention on patient outcomes and hospital performance, we focus on acute care hospitals only and use 30-day standardized hospital mortality indicators (SHMI) for all hospital admissions, 30-day emergency readmission rates of patients previously admitted for planned treatment, and hospital waiting times for planned patients. The estimated ATTs, based on the synthetic DiD method, are reported in Table 4.[40] The RDSP effects on mortality within 30 days from admission is negative and significant at 5% level; the estimate of -0.132 implies a 3.45% decrease in mortality in the treated NHS hospitals versus the control group, with respect to the baseline unconditional mean mortality of 3.82%. While the magnitude of the effect might appear tiny, it is quite

Table 4: Effect of the RDSP in health outcomes of patients admitted to Acute hospitals

	(1)	(2)	(3)	(4)	(5)
	30-day in/outside hospital mortality (all patients)	30-day emergency readmission (planned patients)	Waiting times (in days; planned patients)	Stability Index	NHS Leaving rate
ATT	-0.132**	0.100	-4.147*	0.622***	-0.321**
	(0.065)	(0.133)	(2.188)	(0.204)	(0.140)
Mean (outcome)	3.820	6.800	69.249	0.8615	0.734
Std (outcome)	1.302	1.840	45.303	2.73	1.83

Notes: Bootstrapped standard errors (500 replications) in parentheses. * $p < 0.1$, ** $p < 0.05$ *** $p < 0.01$.

[40] Figure 8 shows plots of the trends in 30-day SHMI, based on cohort-level synthetic DiD.

policy relevant if we consider that the population at risk is given by all patients admitted to hospital care in England in a given year. For example, based on the 8,681,000 (i.e. 5,222,548 planned and 3,458,452 emergency) patient admissions occurred in RDSP-treated NHS organizations during financial year 2019/20, i.e. the period when all four courts had received the RDSP intervention, our results implies that about 11,441 fewer patients died.[41] If we assume the value of £60,000 for one year in full health, as considered by the English Department for Health and Social Care (DHSC), and, quite conservatively, only one year of residual life after discharge for each patient, the indirect effect of RDSP on mortality is worth £686.460 ($851.210) millions.

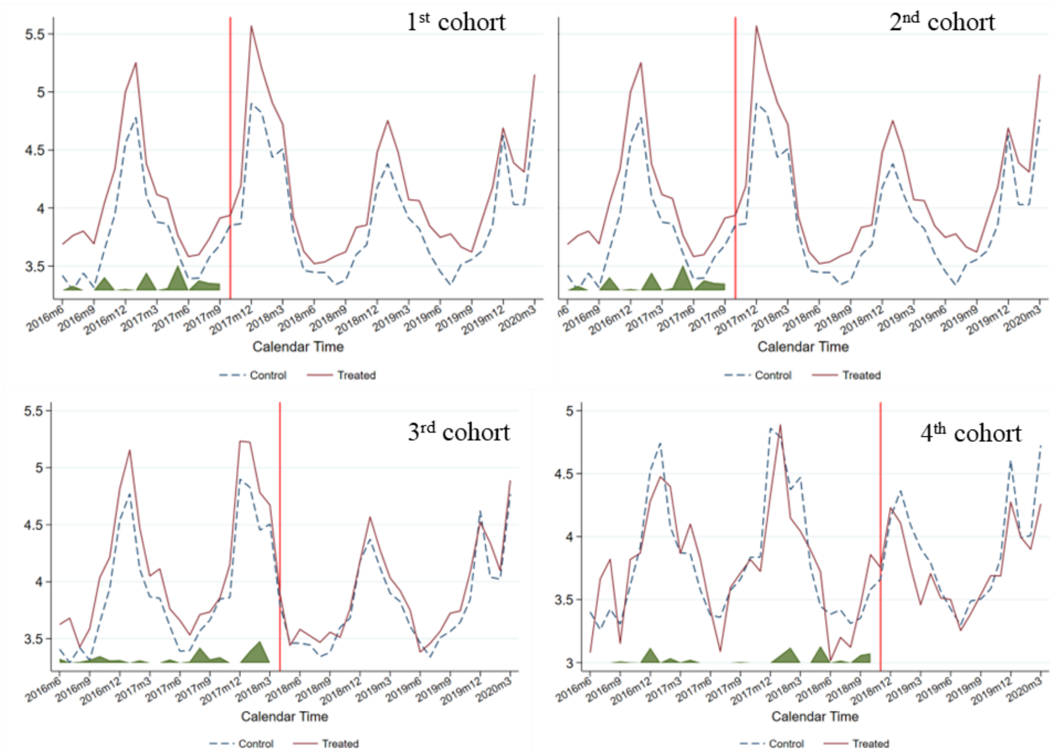
We also find a reduction of waiting times for planned care, although significant only at 10% level, and instead we do not find a significant effect on emergency readmissions after discharge for planned patients. Finally, to double check that our estimates of the RDSP effects on patient outcomes and waiting times are consistent with the direct effects of the Programme estimated with the Callaway and Sant'Anna (2021) approach, we re-estimate the RDSP effects on stability index and NHS leaving rates using Arkhangelsky et al. (2021) synthetic DiD approach. The results, in the last two columns of Table 4, show statistically significant estimates whose point values fall within the 95% uniform confidence bands of the Callaway and Sant'Anna (2021) estimates.

5 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 is the first to examine the impact of a large scale, national-level intervention, the Retention Direct Support Programme (RDSP), which aimed to increase nurse retention in NHS hospital providers by improving

[41] 8,681,000 $\times$ 3.82% $\times$ 3.45%.

Figure 8: Effects of RDSP on 30-day mortality, acute care NHS hospitals



Notes: Trends in all patients 30-day mortality from hospital admission. Sample: acute care NHS Trusts.

nurses' non-financial conditions.

We find that the RDSP achieved its objective in terms of reducing nurses' turnover. 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 exits from the NHS overall. 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.

While the RDSP succeeded in improving retention, it was insufficient to resolve the retention problem in NHS hospitals. Nevertheless, in interpreting these results 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 and potentially complementary to other policies designed to alleviate workforce pressures.

A policy achieving its targets does not imply it is also cost effective. NHSI estimates that it costs £11,000 to replace a nurse (NHS Improvement, 2018), which implies the Programme saved £18,667,000 (i.e. $11,000 \times 1,697$ nurses who did not change hospital organization or leave the NHS) from the NHS budget. However, it is hard to compare this with the cost of the Programme: while no additional funding was made available to the treated NHS providers, it is clear that staff time was used, both in hospital organizations and at the NHSI monitoring body. We know little about the amount of staff time spent and even less about its opportunity cost, although we have indirect evidence that the opportunity cost was not too large - as it is not visible as reduced admissions numbers or higher mortality rates. Thus, it seems likely that the value of staff time spent per hospital organization was less than £141,500 per HO on average (i.e. £18,667,000 divided by the 132 treated NHS hospital organizations), implying a probable cost-effectiveness of the RDSP. We also find that the RDSP had a positive effect on hospital performance, through a reduction of both 30-day mortality for all admitted patients and waiting times for planned treatments. About 11,441 fewer patients died in RDSP-treated hospital organizations, for an estimated monetary value of £686.5 millions of life gained, which adds to the evidence of the cost-effectiveness of the policy.

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. This contrasts with the paradigm that fully decentralized systems are usually more efficient (Alonso et al., 2008; Besley and Coate, 2003; Dewatripont and Maskin, 1995). 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.

Bailey, C., Yeoman, R., Madden, A., Thompson, M. and Kerridge, G. (2019), 'A review of the empirical literature on meaningful work: Progress and research agenda', *Human Resource Development Review* **18**(1), 83–113.

Baker, A. C., Larcker, D. F. and Wang, C. C. (2022), 'How much should we trust staggered difference-in-differences estimates?', *Journal of Financial Economics* **144**(2), 370–395.

Bakker, A. B. and Demerouti, E. (2007), 'The job demands-resources model: State of the art', *Journal of managerial psychology* **22**(3), 309–328.

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. Special Issue: The Impact of Nursing.

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.

Besley, T. and Coate, S. (2003), 'Centralized versus decentralized provision of local public goods: a political economy approach', *Journal of public economics* **87**(12), 2611–2637.

Besley, T. and Ghatak, M. (2005), 'Competition and incentives with motivated agents', *American Economic Review* **95**(3), 616–636.

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.

Böckerman, P., Bryson, A. and Ilmakunnas, P. (2012), 'Does high involvement management improve worker wellbeing?', *Journal of Economic Behavior & Organization* **84**(2), 660–680.

Borusyak, K., Jaravel, X. and Spiess, J. (2021), 'Revisiting event study designs: Robust and efficient estimation'. arXiv preprint:2108.12419, available at `https://arxiv.org/abs/2108.12419`.

Bowles, S. and Polania-Reyes, S. (2012), 'Economic incentives and social preferences: substitutes or complements?', *Journal of Economic Literature* **50**(2), 368–425.

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. (2002), 'Nursing shortages and evidence-based interventions: a case study from Scotland', *International Nursing Review* **49**(4), 209–218.

Buchan, J., Ball, J., Shembavnekar, N. and Charlesworth, A. (2020), 'Building the NHS nursing workforce in England', *London: The Health Foundation* . `https://doi.org/10.37829/HF-2020-RC14`.

Buchan, J., Charlesworth, A., Gershlick, B. and Seccombe, I. (2019), 'A critical moment: NHS staffing trends, retention and attrition', *London: Health Foundation* . `https://www.health.org.uk/publications/reports/a-critical-moment`.

Bungeroth, L., Fennell, E. and S., A. (2018), 'Investing in a Safe and Effective Workforce: Continuing professional development for nurses in the UK'. Royal College of Nursing. Accessed on 2/11/2021 at `https://www.rcn.org.uk/-/media/royal-college-of-nursing/documents/publications/2018/may/pdf-007028.pdf?la=en`.

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', *The Economic Journal* **127**(605), F117–F141.

Callaway, B. (2021), 'Universal vs. varying base period in event studies'. November 20201. `https://bcallaway11.github.io/posts/event-study-universal-v-varying-base-period`.

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. Themed Issue: Treatment Effect 1.

Campbell, D. (2023), 'Health unions urge sunak to resolve pay dispute before unprecedented strike'.

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.

Chalofsky, N. (2003), 'An emerging construct for meaningful work', *Human Resource Development International* **6**(1), 69–83.

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.

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, available at http://www.nber.org/papers/w29691.

Dewatripont, M. and Maskin, E. (1995), 'Credit and efficiency in centralized and decentralized economies', *The Review of Economic Studies* **62**(4), 541–555.

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.

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

Forbes (2015), 'The World's Biggest Employers [Infographic]', *Forbes* . By McCarthy, Niall. Published online on June 23, 2015 at https://www.forbes.com/sites/niallmccarthy/2015/06/23/the-worlds-biggest-employers-infographic/?sh=1a5b9711686b.

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.

Gallus, J. (2017), 'Fostering public good contributions with symbolic awards: A large-scale natural field experiment at Wikipedia', *Management Science* **63**(12), 3999–4015.

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.

Goodman-Bacon, A. (2021), 'Difference-in-differences with variation in treatment timing', *Journal of Econometrics* **225**(2), 254–277. Themed Issue: Treatment Effect 1.

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.

Helm, C. and Bungeroth, L. (2017), 'Safe and effective staffing: the real picture'. Royal College of Nursing London. https://www.rcn.org.uk/professional-development/publications/pub-006195.

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', *JL Econ. & Org.* **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.

Jakiela, P. (2021), 'Simple diagnostics for two-way fixed effects'. arXiv.2103.13229, available at https://arxiv.org/abs/2103.13229.

Kahn, W. A. (1990), 'Psychological conditions of personal engagement and disengagement at work', *Academy of management journal* **33**(4), 692–724.

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

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* .

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

Marufu, T. C., Collins, A., Vargas, L., Gillespie, L. and Almghairbi, D. (2021), 'Factors influencing retention among hospital nurses: systematic review', *British Journal of Nursing* **30**(5), 302–308.

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?'. Published on March 14th, 2002, available at https://blogs.lse.ac.uk/businessreview/2022/03/14/do-management-practices-matter-in-further-education/.

Mirror (2020), 'NHS faces fresh crisis as 250,000 nurses consider quitting because of low pay', *Mirror* . By Nelson, Nigel and McLaughlin, Chris. Published online on July 25, 2020 at https://www.mirror.co.uk/news/uk-news/nhs-faces-fresh-crisis-250000-22416470. Accessed on 19/05/2022.

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/.

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

NHS Digital (2018), 'Nurse and Health Visitor joiners and leavers from the NHS'. Accessed on 15/09/2021 at 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.

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

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

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 Improvement (2017), 'Retention Support Programme'. Accessed on 05/11/2020 at https://www.networks.nhs.uk/nhs-networks/nhsi-retention-support-cohort-1/documents/programme-outline.

NHS Improvement (2018), 'Treasuring the NHS's greatest asset - retaining our workforce'. Presented by Mark Radford, Director of Nursing, NHSI on 7th June 2018. Accessed on 16/06/2021 at https://www.hfma.org.uk/docs/default-source/educationandevents/national-events/brighter-together-2018/(2)-mark-radford.pdf?sfvrsn=2.

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.

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.

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. (2022), 'An honest approach to parallel trends', *Working Paper* . October 2022. https://www.jonathandroth.com/assets/files/HonestParallelTrends_Main.pdf. Forthcoming, 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'. "https://EconPapers.repec.org/RePEc:boc:bocode:s458976".

Roth, J., Sant'Anna, P. H. C., Bilinski, A. and Poe, J. (2022), 'What's trending in difference-in-differences? a synthesis of the recent econometrics literature'. Available at arXiv: https://arxiv.org/abs/2201.01194.

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.

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. (2021), 'Eventstudyinteract: Stata module to implement the interaction weighted estimator for an event study'.

Sun, L. and Abraham, S. (2021), 'Estimating dynamic treatment effects in event studies with heterogeneous treatment effects', *Journal of Econometrics* **225**(2), 175–199. Themed Issue: Treatment Effect 1.

The Guardian (2021), 'Threat of nurses' strike over 3% pay award for NHS staff', *The Guardian* . By Campbell, Denis and Allegretti, Aubrey. Published online on July 21, 2021 at https://www.theguardian.com/society/2021/jul/21/government-offers-nhs-staff-in-england-3-pay-rise. Accessed on 19/05/2022.

The New Statesman (2021), 'Exclusive polling: Majority say NHS nurses and doctors are paid too little', *The New Statesman UK Edition* . By Chakelian, Anoosh. Published online on August 31, 2021 at https://www.newstatesman.com/politics/2021/08/exclusive-polling-majority-say-nhs-nurses-and-doctors-are-paid-too-little. Accessed on 19/05/2022.

The Open University (2018), Tackling the nursing shortage, Technical report, The Open University. 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'. Accessed on 22/10/2021 at https://www.commonwealthfund.org/international-health-policy-center/countries/england.

Verbeeten, F. H. (2008), 'Performance management practices in public sector organizations: Impact on performance', *Accounting, Auditing & Accountability Journal* .

Wooldridge, J. M. (2021), Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. (August 17, 2021). Available at SSRN:http://dx.doi.org/10.2139/ssrn.3906345.

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

	Stability rate		NHS-leavers 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, ¿=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: Stability rates: never-treated vs. not-yet-treated comparison groups

	Stability rate			Leaving the NHS rate		
	(I)	**(II)**	**(III)**	**(I)**	**(II)**	**(III)**
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	-0.408	-0.361	-0.371
	(0.188)***	(0.180)***	(0.176)***	(0.131)***	(0.126)***	(0.125)***
	[0.196]§	[0.182]§	[0.174]§	[0.128]§	[0.125]§	[0.120]§
partially aggregated						
Cohort 1	0.950	0.851	0.971	-0.498	-0.454	-0.481
	(0.352)***	(0.340)**	(0.324)***	(0.254)**	(0.251)*	(0.244)**
	[0.360]§	[0.329]§	[0.349]§	[0.233]	[0.261]	[0.262]
Cohort 2	0.677	0.579	0.589	-0.416	-0.388	-0.394
	(0.303)**	(0.289)***	(0.278)**	(0.202)**	(0.196)**	(0.189)**
	[0.309]	[0.273]	[0.280]	[0.209]	[0.199]	[0.193]
Cohort 3	0.557	0.424	0.378	-0.394	-0.348	-0.358
	(0.336)*	(0.327)	(0.319)	-0.248	(0.242)	(0.234)
	[0.333]	[0.351]	[0.323]	[0.250]	[0.241]	[0.234]
Cohort 4	0.912	0.773	0.773	-0.341	-0.274	-0.274
	(0.267)***	(0.258)***	(0.258)***	(0.183)*	(0.169)	(0.169)
	[0.269]§	[0.255]§	[0.266]§	[0.181]	[0.166]	[0.166]
pre-trend Wald test (df = 48) p-value	0.080	0.080	0.236	0.135	0.135	0.083

Notes: Aggregated treatment effect parameters under the Unconditional DiD Assumption of CSA with never-treated HOs as control group. Asymptotic standard errors are in parentheses, and estimated using csdid package in Stata (Rios-Avila et al., 2021), p-values *p<0.1; **p<0.05; ***p<0.01. Clustered bootstrapped standard errors are in brackets, and estimated using did package in R (Callaway and Sant'Anna, 2020). The cohort-specific effects take into account selective treatment timing. § indicates that the 95% simultaneous confidence band does not cover 0.

Table A4: Placebo Testing: Distribution of HOs in Cohorts

	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 A5: Placebo Testing: Different timings for the RDSP

	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 A6: Average treatment effects of RDSP on new joiners

	Joiner	**NHS joiner**	**Churn joiner**
Overall ATT	-0.067 [0.394]	-0.031 [0.33]	-0.039 [0.151]
Cohort 1	-0.778 [1.366]	-0.48 [1.283]	-0.303 [0.388]
Cohort 2	0.427 [0.465]	0.604 [0.394]	-0.186 [0.235]
Cohort 3	0.062 [0.506]	-0.157 [0.393]	0.216 [0.23]
Cohort 4	0.019 [0.456]	-0.032 [0.328]	0.057 [0.227]
PTA p-value (12 months)	0.0011	0.0000	0.0669
PTA p-value (6 months)	0.1011	0.0107	0.4322
Avr. rate, treated	16.16%	10.74%	5.45%
Avr. rate, control	14.17%	9.19%	5.00%

Notes: Estimated under unconditional parallel trends. Bootstrapped standard errors are in parentheses. § indicates that the 95% simultaneous confidence band does not cover 0.

Table A7: RDSP Action Plans by cohorts, theme frequencies (%)

	Cohort 1	**Cohort 2**	**Cohort 3**	**Cohort 4**	**Total**
TH1: Career progression & development	81.48	88.89	76.47	83.33	82.14
TH2: Compassionate culture	66.67	66.67	64.71	58.33	64.29
TH3: Being a flexible employer	44.44	55.56	41.18	70.83	51.79
TH4: Engagement and key conversations	22.22	37.04	52.94	37.50	38.39
TH5: Gathering and understanding data	40.74	33.33	44.12	37.50	39.29
TH6: Being an attractive & rewarding employer	44.44	51.85	29.41	20.83	36.61
TH7: Supporting new starters & newly qualified staff	40.74	33.33	35.29	41.67	37.50
TH8: Supporting staff approaching retirement	29.63	25.93	26.47	50.00	32.14
TH9: Retention as part of executive leadership & Trust strategy	37.04	29.63	20.59	25.00	27.68
TH10: Narrowing the front door to close the back door	33.33	14.81	26.47	12.50	22.32
Number of action plans	27	27	34	24	112

Notes: Based on 112 treated HOs (out of 132). The breakdown of excluded HOs by cohort are: 4 HOs in Cohort 1, 1 in Cohorts 2 and 3, and 8 HOs in Cohort 4. There is also 6 HOs that were not recorded in NHSI's action plan data.

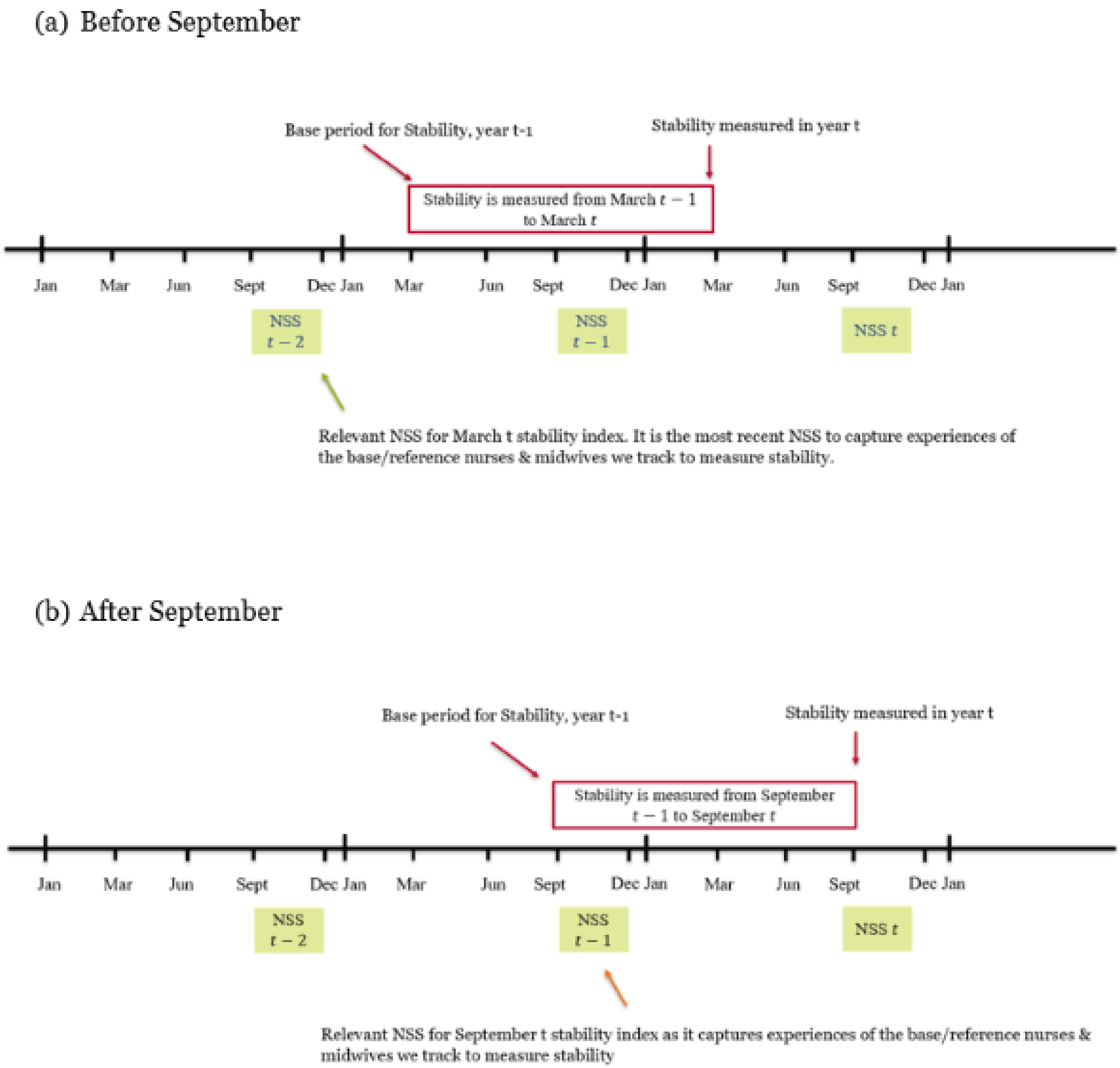
Table A8: Effects of RDSP action plans (themes) on the retention of nurses and midwives

(a) Stability rate	**TH1**	**TH2**	**TH3**	**TH4**	**TH5**	**TH6**	**TH7**	**TH8**	**TH9**	**TH10**
Overall ATT	0.737***	0.851***	0.782***	0.621**	0.624**	0.729**	0.761***	0.542**	1.106***	0.574*
	(0.205)	(0.239)	(0.225)	(0.258)	(0.251)	(0.284)	(0.269)	(0.246)	(0.327)	(0.295)
Cohort 1	0.887**	1.060**	0.231	0.899	0.400	0.393	0.388	0.516	1.114*	0.344
	(0.386)	(0.478)	(0.382)	(0.956)	(0.556)	(0.399)	(0.484)	(0.640)	(0.615)	(0.526)
Cohort 2	0.771**	0.832**	0.817**	0.216	0.522	0.987**	0.192	0.262	0.836	0.750
	(0.342)	(0.397)	(0.357)	(0.304)	(0.393)	(0.461)	(0.551)	(0.459)	(0.610)	(0.730)
Cohort 3	0.391	0.580	0.972*	0.655	0.503	0.600	1.269***	-0.018	0.964	0.507
	(0.347)	(0.403)	(0.519)	(0.406)	(0.420)	(0.651)	(0.462)	(0.390)	(0.686)	(0.391)
Cohort 4	0.978***	1.032***	0.985***	0.816***	1.203***	1.073**	1.075**	1.144***	1.617***	1.233*
	(0.315)	(0.400)	(0.335)	(0.312)	(0.353)	(0.536)	(0.497)	(0.343)	(0.437)	(0.637)

(b) NHS-leaver rate	**TH1**	**TH2**	**TH3**	**TH4**	**TH5**	**TH6**	**TH7**	**TH8**	**TH9**	**TH10**
Overall ATT	-0.301**	-0.391**	-0.339**	-0.384*	-0.510***	-0.294	-0.452***	-0.204	-0.549**	-0.591**
	(0.137)	(0.168)	(0.141)	(0.207)	(0.173)	(0.188)	(0.175)	(0.159)	(0.269)	(0.232)
Cohort 1	-0.361	-0.635*	-0.115	-0.540	-0.243	-0.175	-0.320	-0.203	-0.663	-0.563
	(0.276)	(0.366)	(0.235)	(0.556)	(0.363)	(0.274)	(0.380)	(0.362)	(0.538)	(0.489)
Cohort 2	-0.415**	-0.350	-0.453*	-0.338	-0.727**	-0.426	-0.630*	-0.336	-0.269	-1.198***
	(0.209)	(0.231)	(0.261)	(0.308)	(0.321)	(0.280)	(0.325)	(0.332)	(0.343)	(0.297)
Cohort 3	-0.173	-0.342	-0.559*	-0.545	-0.536**	-0.480	-0.402	0.140	-0.658	-0.481
	(0.244)	(0.316)	(0.303)	(0.372)	(0.272)	(0.477)	(0.306)	(0.296)	(0.708)	(0.335)
Cohort 4	-0.265	-0.205	-0.215	-0.008	-0.574*	0.162	-0.499**	-0.386*	-0.607*	-0.196
	(0.188)	(0.233)	(0.186)	(0.227)	(0.304)	(0.245)	(0.253)	(0.211)	(0.329)	(0.290)

Notes: Estimates comes from separate models under the unconditional parallel trends assumption. Asymptotic standard errors clustered at HO level are in parentheses. *p<0.1; **p<0.05; ***p<0.01.

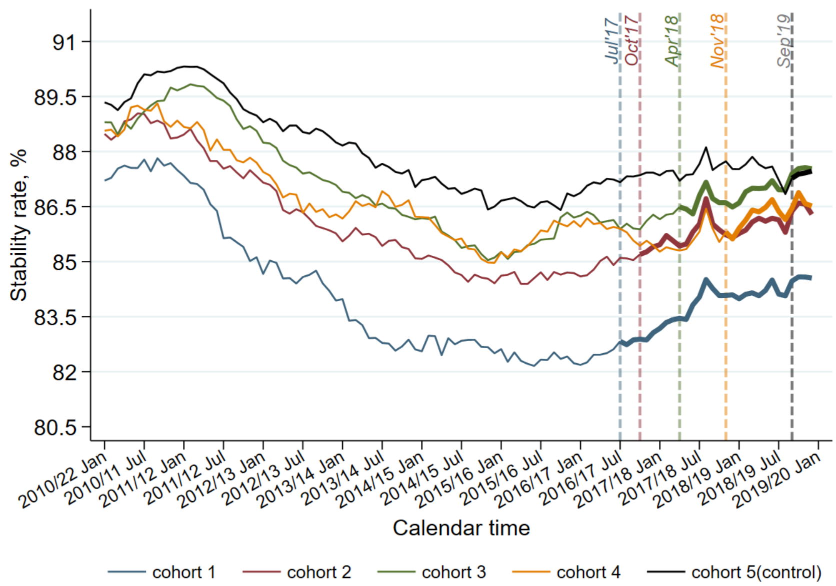
Figure A1: Data setup



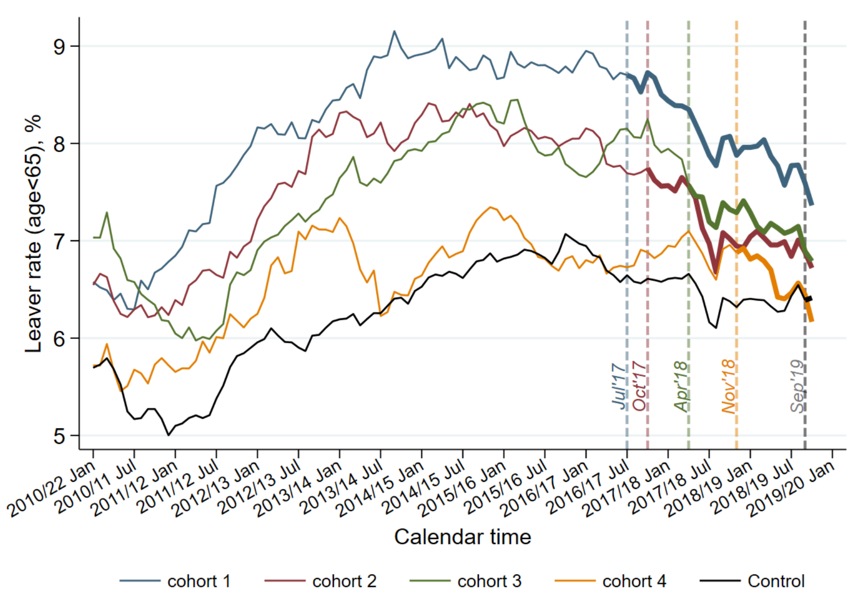
Notes: The same holds for NHS leaver rates. t refers to the analysis year, $t-1$ is the base year. NSS refers to the NHS Staff Survey which is conducted every year in autumn since 2003. Staff working in HOs in 1st September are eligible to respond to the NSS. The NSS runs from the mid-September and remains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

Figure A2: Each cohort's retention profile and RDSP launch dates

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



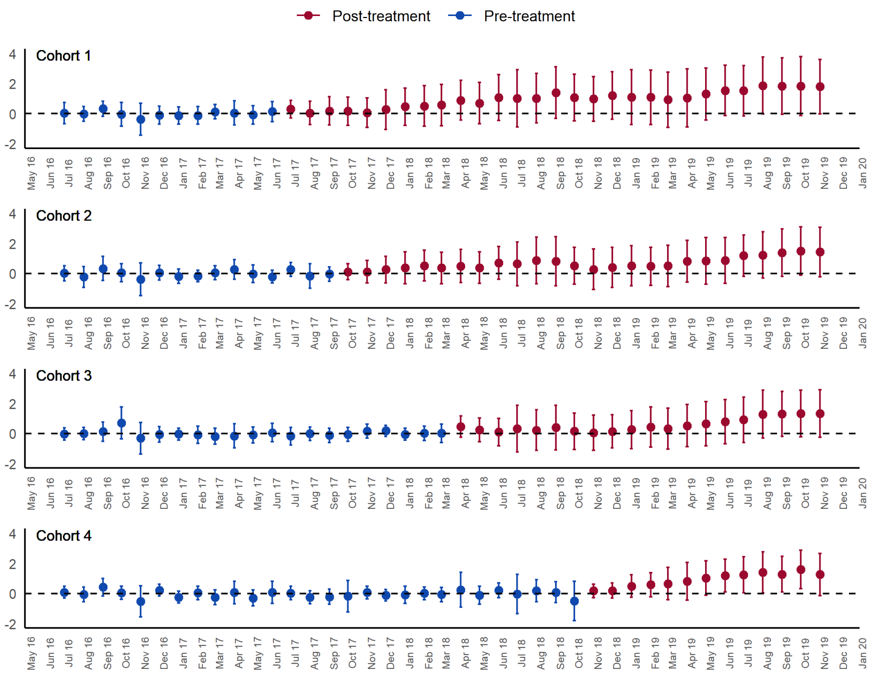
(b) NHS Leaver rates from 2010/11 to 2019/20



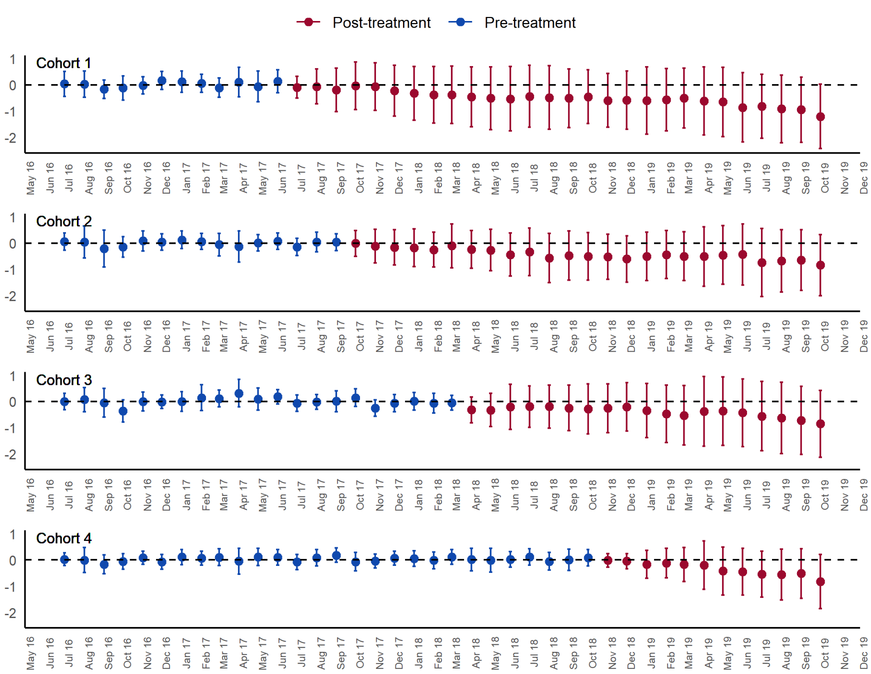
Notes: Cohort 5 includes 2 additional HOs that were not in the NHSI allocation. Vertical lines show the RDSP start dates for each cohort, and the thicker horizontal lines indicate post-RDSP period in each cohort.

Figure A3: Each cohort's retention profile and RDSP launch dates

(a) Stability rates



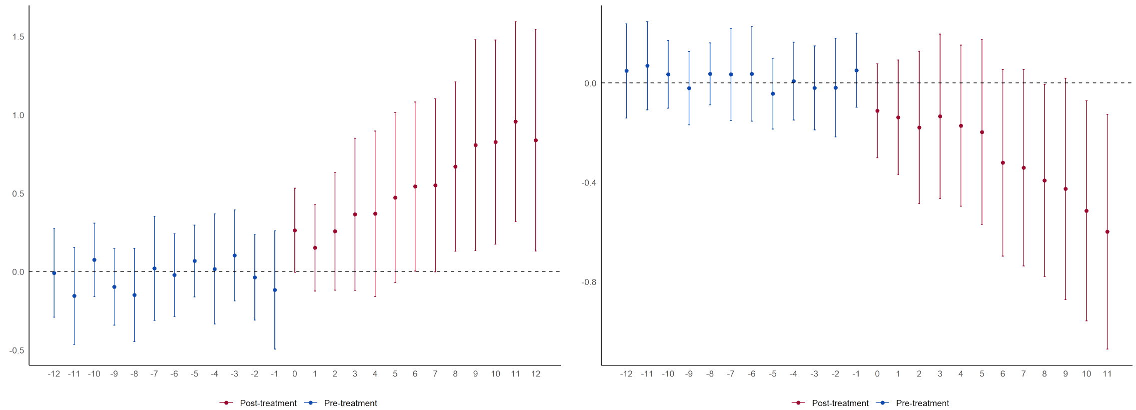
(b) NHS Leaver rates



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.

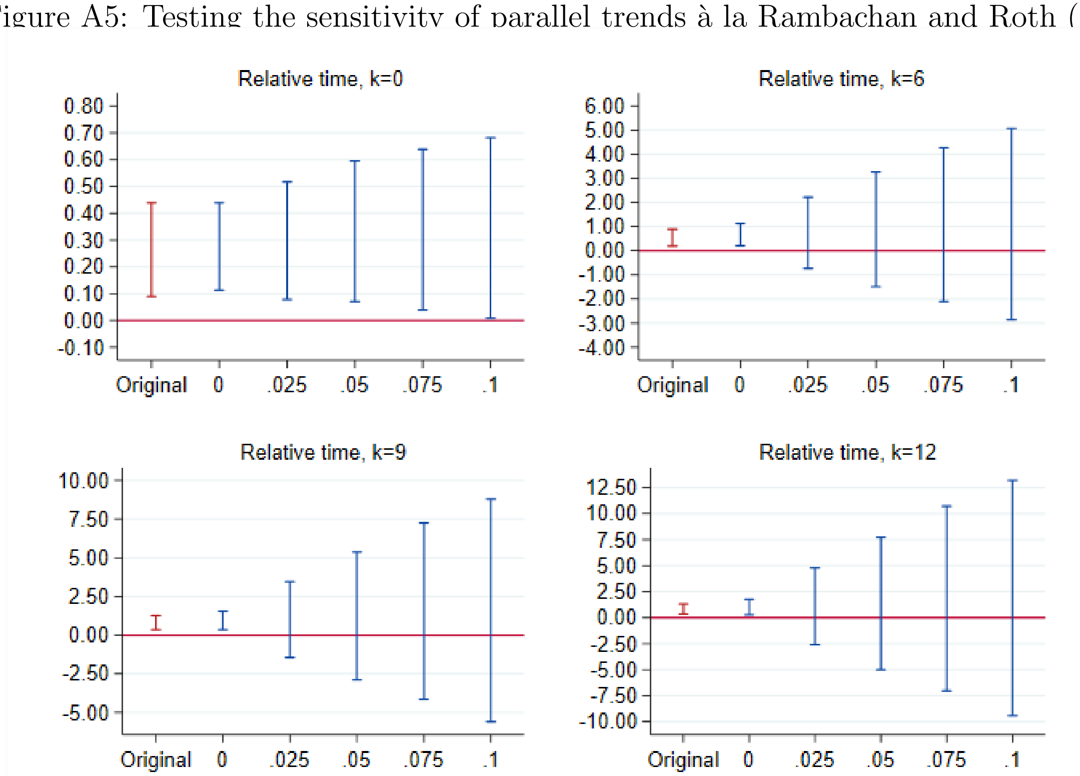
Figure A4: Dynamic treatment effects of the RDSP on nursing retention

(a) Event study plot of stability rate (b) Event study plot of NHS Leaver rates



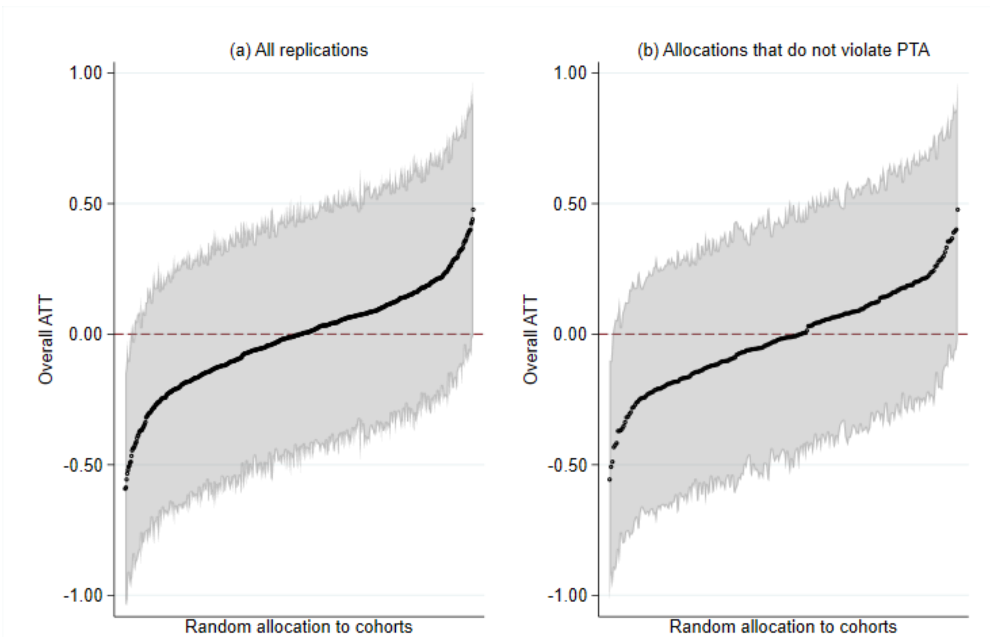
Notes: Both event study estimates are computed under the unconditional parallel trends assumption. The figure shows estimates with uniform 95% confidence intervals using bootstrapped standard errors clustered at HO level. The analysis is performed using the "did" package (Callaway and Sant'Anna, 2020) in R.

Figure A5: Testing the sensitivity of parallel trends à la Rambachan and Roth (2022)



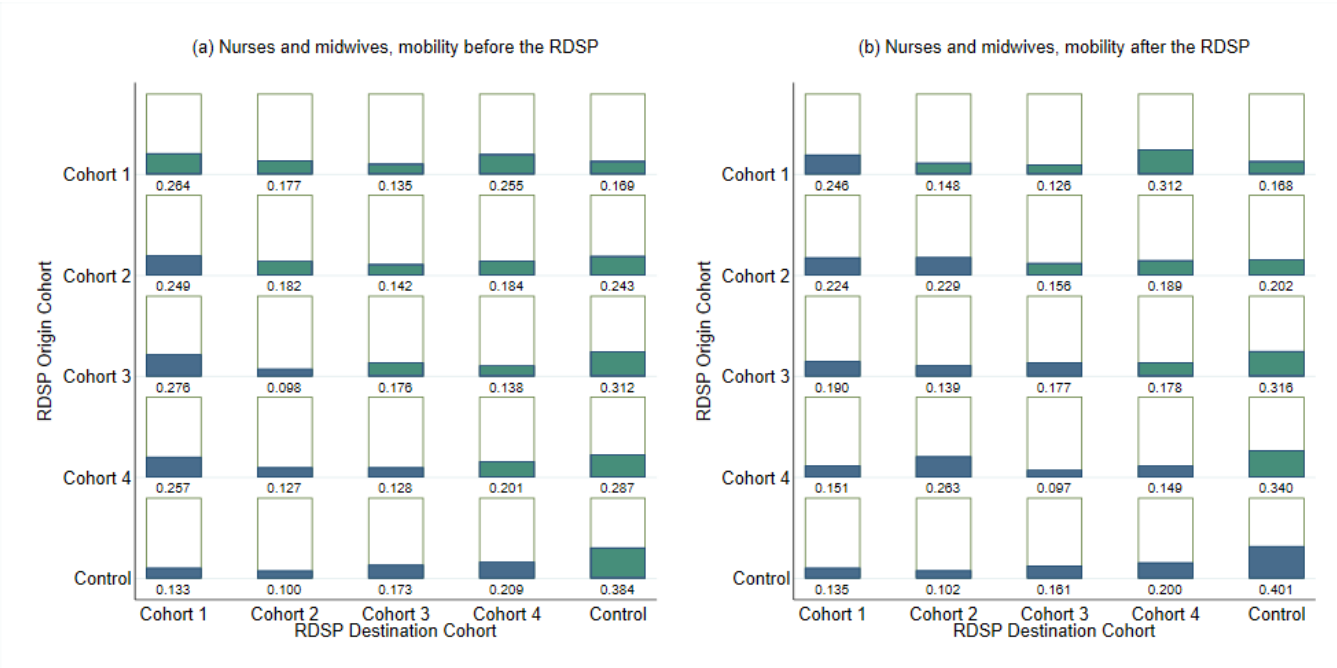
Notes: The x-axis shows different values for M, the smoothness restriction, ranging from 0 to 0.1, and the y-axis shows the 95% robust confidence interval. The original indicates the confidence interval for ATT at exposure month k using asymptotic standard errors (as in panel (b) of Figure 5).

Figure A6: Placebo test: Random allocation of HOs into Cohorts



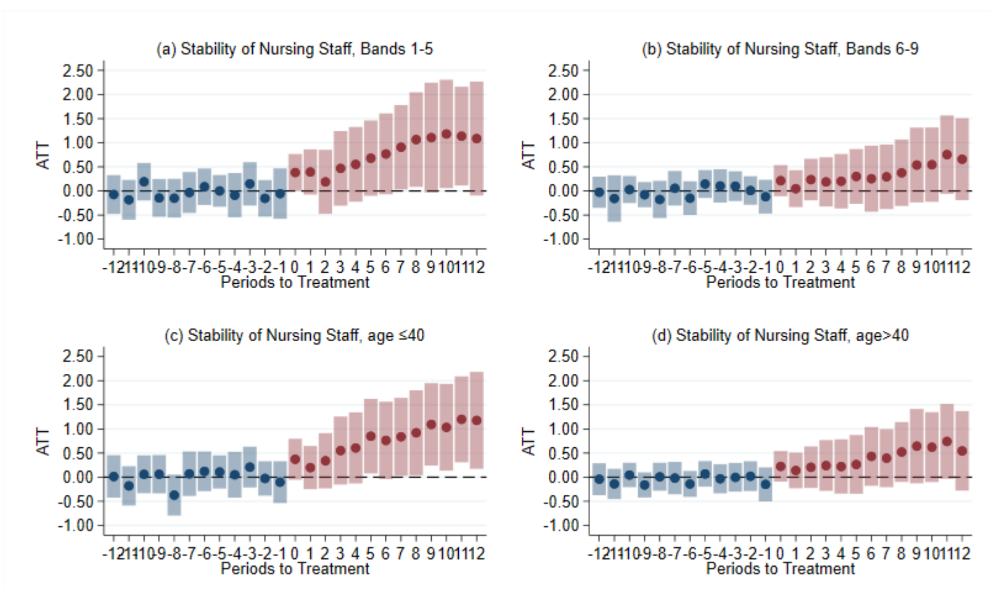
Notes: Overall ATTs from replications under unconditional parallel trend assumption. Uniform 95% confidence intervals using bootstrapped standard errors clustered at the HO level. All replication results are presented in panel (a) with 500 replications, and panel (b) shows results only from those that do not violate the unconditional parallel trends assumption at 12 months (248 of 500 replications).

Figure A7: Mobility of nurses and midwives before and after the RDSP



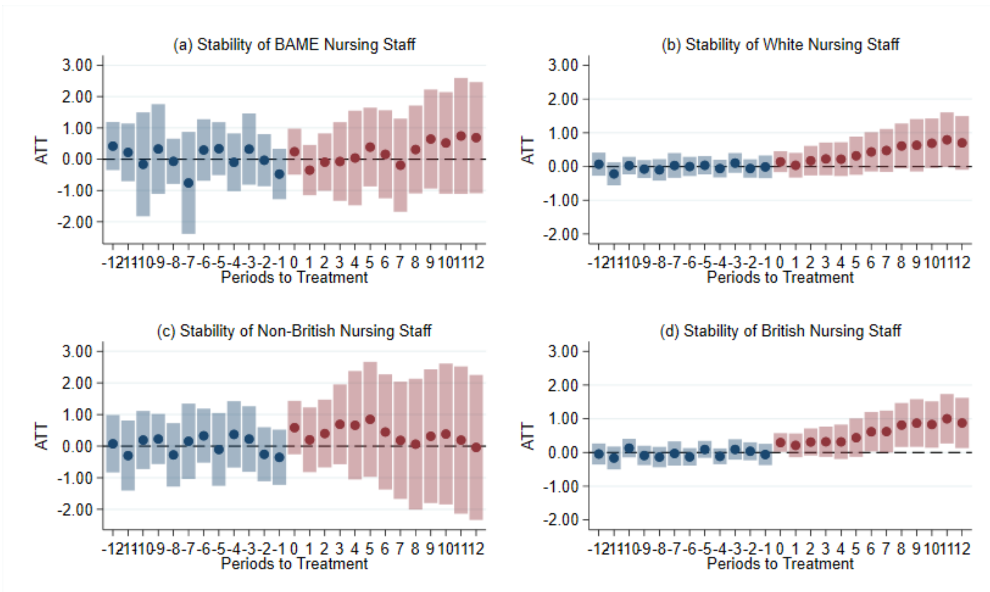
Notes: Overall, 28,558 nurses and midwives switched hospitals at least once (6.7%) from June 2016 to November 2019 (34,472 switches in total). The period before/after the RDPS in panels (a) and (b) varies for each cohort depending on when they launched the programme. Green bars indicate the transitions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

Figure A8: 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).

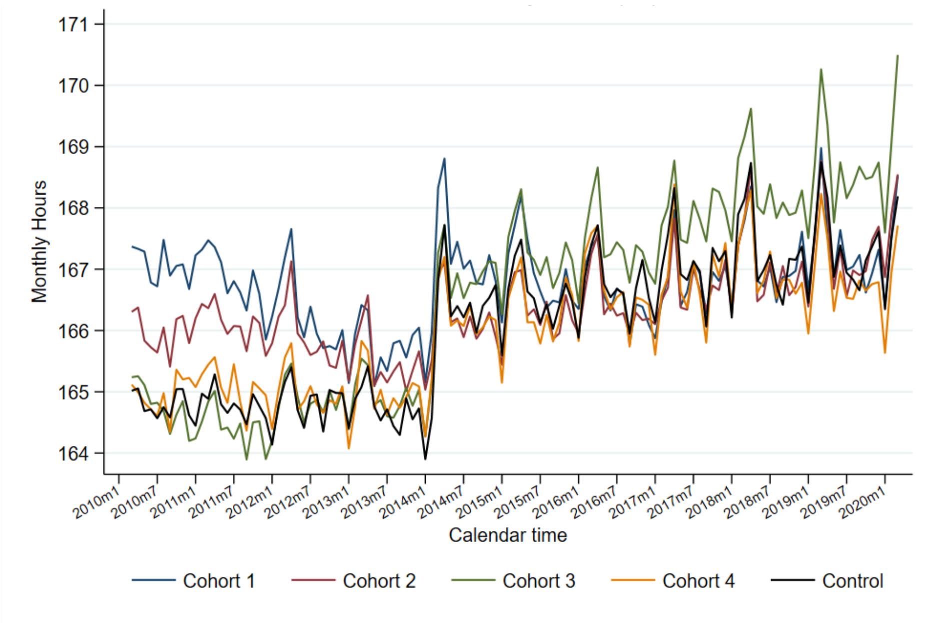
Figure A9: Ethnicity and nationality of nursing staff - an event study approach



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 A10: Full-time nursing staff and monthly hours worked

(a) Average monthly hours worked by full-time nursing workforce



(b) Share of full-time nursing staff

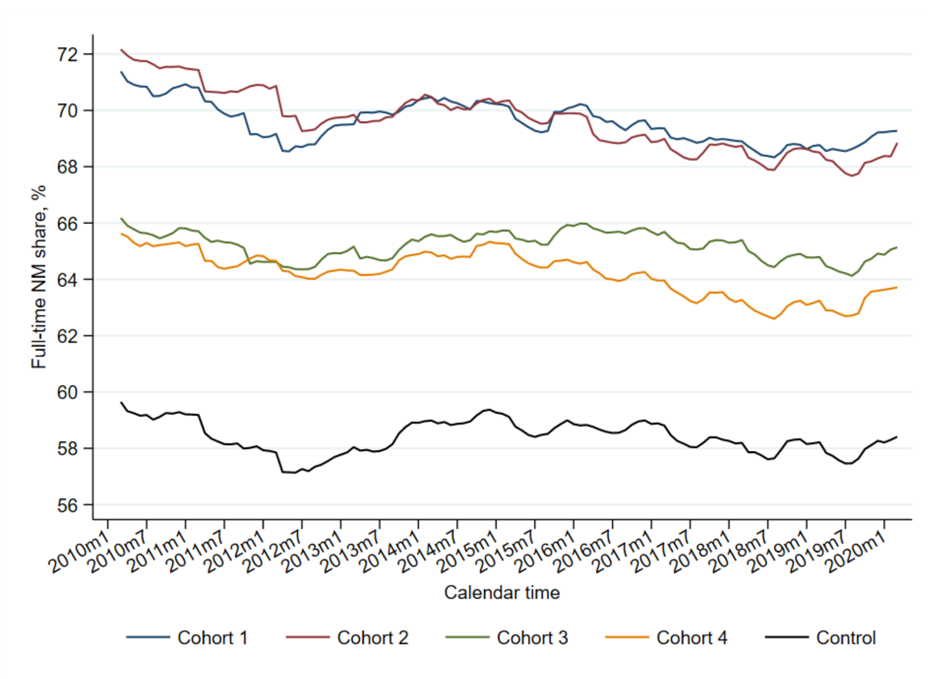
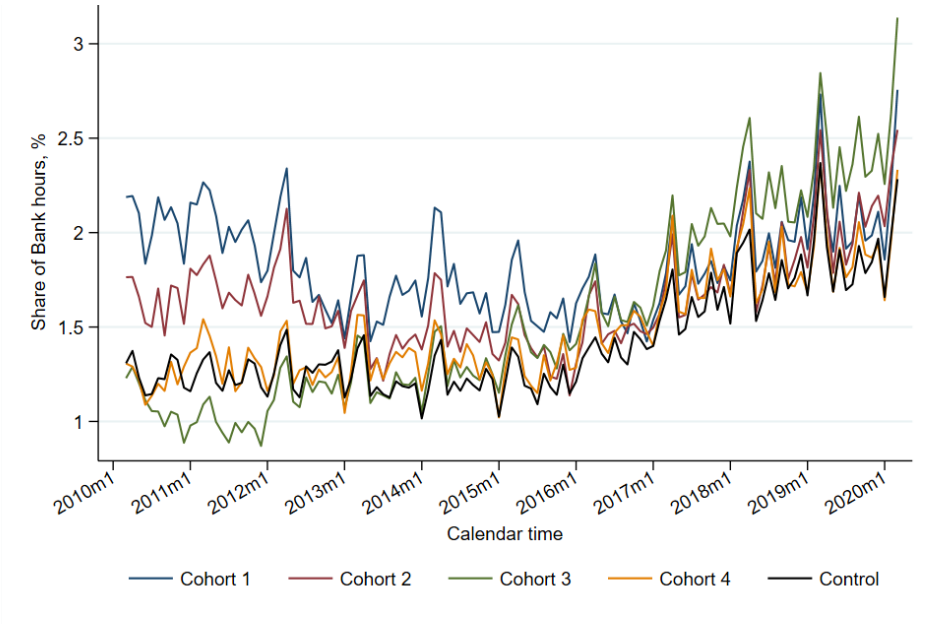
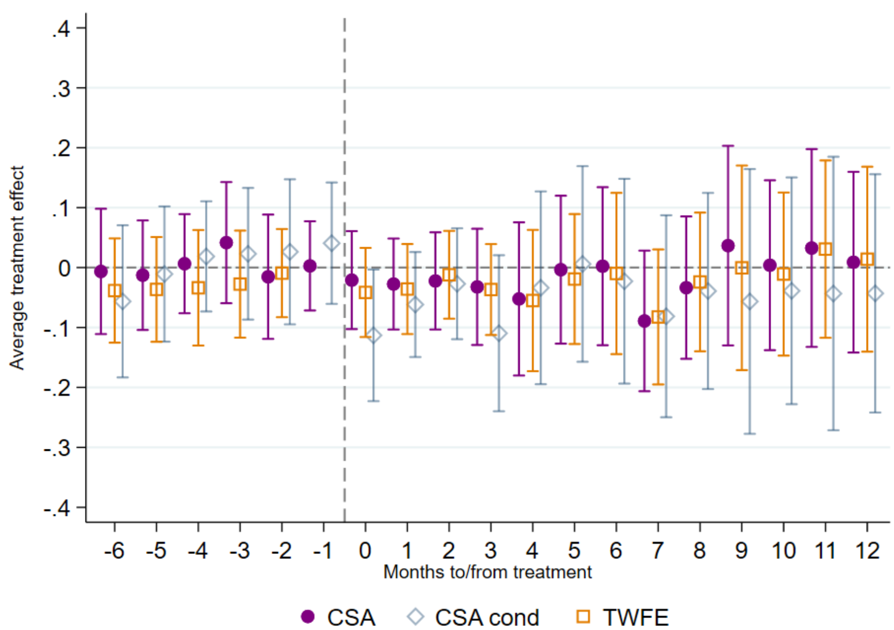


Figure A11: Share of Bank Hours within monthly hours worked by nursing staff



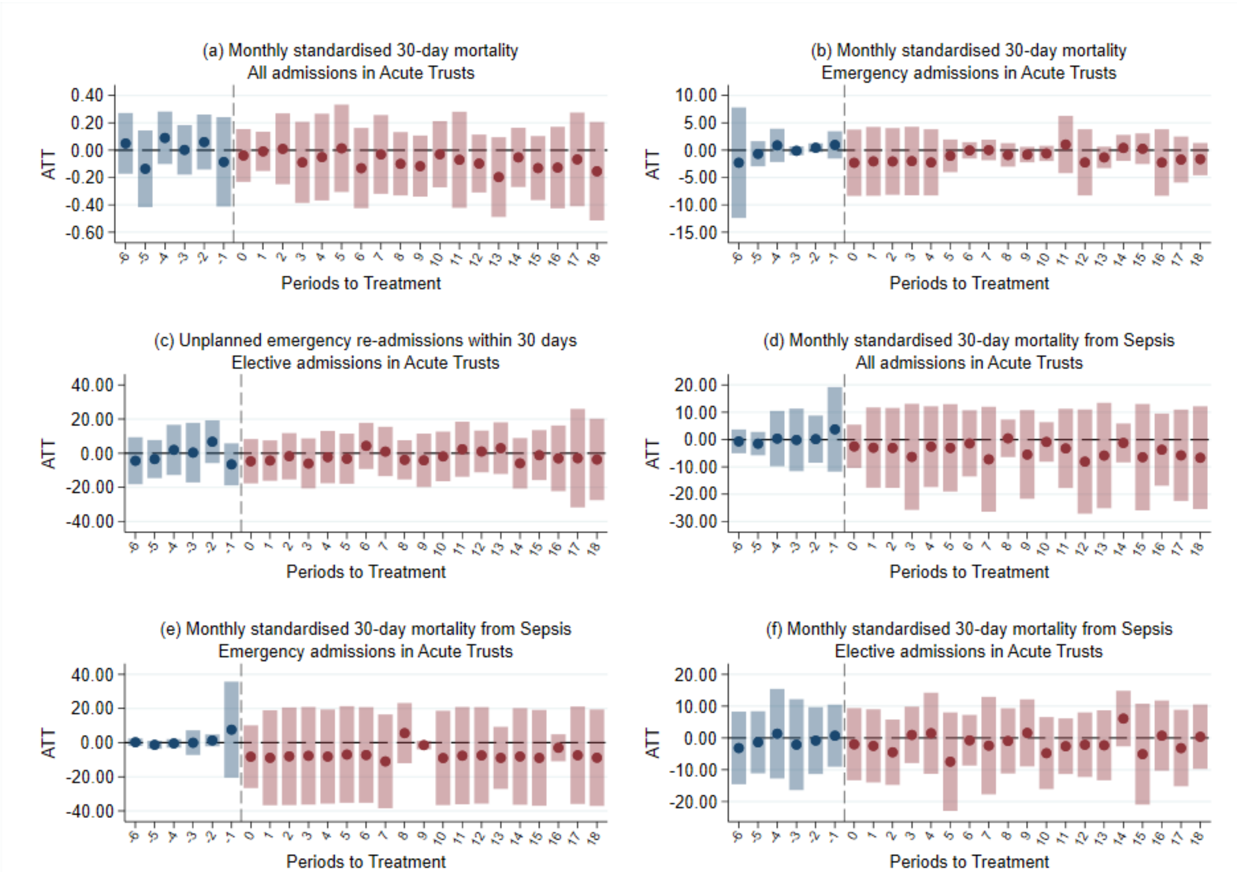
Notes: Average hours excludes negative and zero hours and includes both full-time and part-time working nurses and midwives.

Figure A12: Dynamic average treatment effects on share of Bank hours



Notes: CSA is under unconditional PTA, and "CSA cond" is the CSA estimation under conditional PTA. The set of covariates include nurses and midwives' absence rates, support from co-workers, and share of full-time nurses and midwives except for Bank work. Asymptotic standard errors are clustered at HO level and shown with point-wise 95% confidence intervals. The p-value for the augmented Wald test for CSA that pre-treatment estimates for the 6 months preceding the treatment is 0.316 under unconditional PTA and 0.383 under conditional PTA. For the TWFE, the periods more than 6 months before and 12 months after the RDSP are binned and not shown in the figure.

Figure A13: Effects of RDSP on risk-adjusted mortality and unplanned emergency re-admissions for patients in acute hospitals



Notes: Estimated under unconditional parallel trends assumption. The estimates are presented with uniform 95% confidence intervals with bootstrapped standard errors clustered at HO level. The pre-treatment parallel trends hold for the 6 months preceding the RDSP in all models.

Appendix B: Heterogeneity analysis

This Appendix provides further discussion on the results from heterogeneity analysis on different sub-samples of nursing staff as described in Section 4.2.

The stability measure in the main analysis uses all nurses and midwives employed at a Trust between two time periods, and includes nursing staff of all ages under different contractual agreements. We have two motivations to include nursing staff who are close to or beyond the retirement age in our main stability measure: (i) The RDSP actions include strategies targeting staff close to retirement (e.g. Action Plan theme 8) such as introducing flexible retirement options and "retirement & return" policies. (ii) The rate of retirement among nurses and midwives was constant over time, and the difference is an intercept shift in stability rates. It is possible, however, that nurses nearing retirement age may be more likely to leave regardless. As a robustness check, we re-compute the stability rates by restricting our attention to nurses and midwives who are younger than 65. The estimated overall average treatment effect under this age restriction is slightly lower in magnitude, but both cohort-specific ATTs and the overall ATT presented in Column 1 in Table B1 present similar patterns to those of column 2 in Table 2 indicating that the impact of the RDSP does not work primarily through the prevention of retirements.

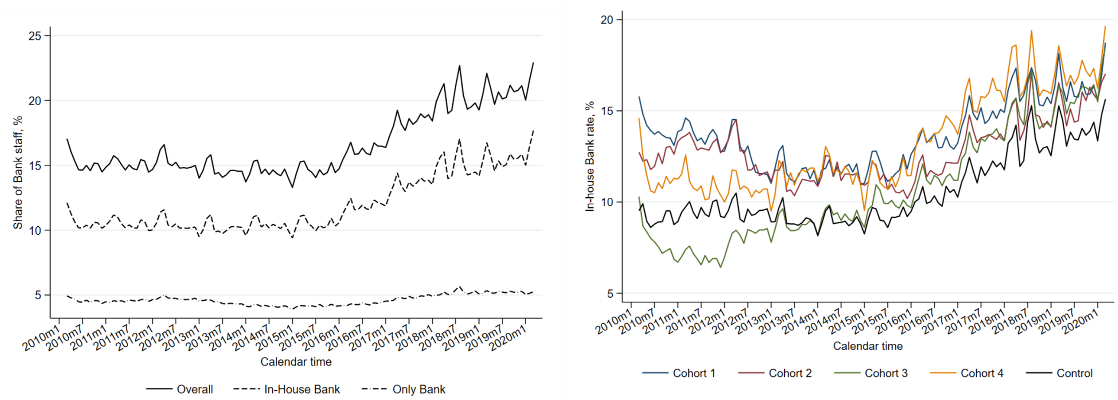
The main retention measures also include staff who employed as "staff Bank" workers by the hospital Trusts. "Staff Bank" work is carried out by employees who are registered to provide shifts on a temporary basis, mostly on a zero-hours contract, with no further obligation for regular work at hospital 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 employees of a hospital Trust (in-house bank) or from employees of an outside organization who are only contracted as bank workers in the Trust (bank-only). The difference between Bank-only and in-house Bank staff is that Bank-only staff may leave the Trust once their period of temporary employment terminates, whereas in-house bank staff are nurses already employed within the Trust and providing additional labor as staff bank workers.

Starting from 2016, the share of nurses and midwives who registered as Bank has increased across all cohorts, some more than others, and the increase in the Bank workforce was driven by the in-House Bank registrations for all cohorts as shown in Figures B1. While the extent of the use of bank staff may signal staffing difficulties such as increasing need to cover staff, bank work also provides flexibility to nursing staff as they can choose which shifts to work (Buchan, 2002; Buchan et al., 2019).

Figure B1: Share of Bank staff over time

(a) Overall, in-house Bank and Bank-only nurses and midwives

(b) Share of in-house Bank work, by cohort



To understand whether the main estimates are affected by the Bank workforce, we re-compute our retention outcomes by excluding the nurses and midwives who are employed as Bank-only staff, as their inclusion might contaminate the retention outcome and bias the effects of interest, and instead we keep in-house Bank staff. B1 column 2 shows a very similar overall impact of the programme on nursing retention. The key difference is between the cohort-specific effects, where in the absence of Bank-only staff, the RDSP lead a significant increase in retention of nursing staff in Cohort 3 Trusts. This effect is slightly higher than

Cohort 4 Trusts. The impact of the retention programme is slightly limited, yet we find significant reduction in the nurses and midwives leavers' rates due to RDSP participation.

We introduce a further restriction on the outcomes and exclude nurses and midwives who are on any temporary and fixed-term contracts. The results are similar to our baseline estimates, which is expected as the majority (97%) of the nurses and midwives work under permanent contracts. The overall average impact of the RDSP on nurses and midwives' retention, both for stability and NHS leavers' rate.

Lastly, we focus on nurses and midwives who work in acute care NHS Trusts. The health care services provided in acute care hospitals are different than mental health Trusts. As 91% of the mental health Trusts are treated in the first three cohorts, and the majority of control group consists of acute care Trusts, there might be some limits in the comparability across treatment and control groups for the mental health Trusts. However, our main results are robust to restricting the sample to acute Trusts: we find that overall the RDSP improved retention by 0.87 ppt on average for all treated Trusts, and reduced the leaver rates by 0.42 ppt.

Table B1: Robustness checks: Average treatment effects of RDSP on retention outcomes for sub-samples of nurses and midwives

	Stability rates				**NHS leaver rates**		
	Age < 65 y.o.	**No bank-only**	**Permanent**	**Acute Trusts**	**No bank-only**	**Permanent**	**Acute Trusts**
Overall ATT	0.798	0.839	0.789	0.868	-0.380	-0.356	-0.415
	(0.196)***	(0.201)***	(0.199)***	(0.200)***	(0.137)***	(0.137)***	(0.136)***
	[0.208]§	[0.201]§	[0.195]§	[0.200]§	[0.145]§	[0.146]§	[0.139]§
Cohort-specific ATTs							
Cohort 1	0.968	1.148	1.108	1.390	-0.533	-0.510	-0.561
	(0.371)***	(0.382)***	(0.387)***	(0.436)***	(0.258)**	(0.255)**	(0.311)*
	[0.372]§	[0.386]§	[0.386]§	[0.427]§	[0.283]	[0.264]	[0.313]
Cohort 2	0.520	0.417	0.391	0.429	-0.301	-0.261	-0.159
	(0.340)	(0.345)	(0.343)	(0.263)	(0.250)	(0.247)	(0.176)
	[0.329]	[0.348]	[0.345]	[0.264]	[0.247]	[0.240]	[0.186]
Cohort 3	0.750	0.942	0.901	0.511	-0.403	-0.393	-0.568
	(0.394)*	(0.395)**	(0.389)**	(0.479)	(0.237)*	(0.238)*	(0.310)*
	[0.392]	[0.389]§	[0.383]	[0.511]	[0.236]	[0.242]	[0.312]
Cohort 4	0.918	0.814	0.728	0.949	-0.292	-0.267	-0.361
	(0.266)***	(0.244)***	(0.248)***	(0.273)***	(0.162)*	(0.162)*	(0.185)*
	[0.266]§	[0.249]§	[0.263]§	[0.280]§	[0.165]	[0.165]	[0.186]
Pre-trend test p-value							
12 months, df = 48	0.673	0.286	0.462	0.509	0.652	0.749	0.002
6 months, df = 24	0.743	0.916	0.966	0.539	0.707	0.591	0.830

Notes: Under conditional parallel trends assumption with the same set of controls as reported in columns 2 and 4 of Table 2. Past retention rates are adjusted for sub-samples. Standard errors are clustered at Trust level. Asymptotic standard errors are in parentheses, p-values *p<0.1; **p<0.05; ***p<0.01. Clustered bootstrapped standard errors are in brackets and § indicates that the 95% simultaneous confidence band does not cover 0. Estimation period ends in November 2019 for stability rate and October 2019 for leaving the NHS rate.