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

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

†School of Economics, University of Surrey

§School of Medicine, University of Leeds

‡IZA

May 24, 2022

Draft - Please do not cite or circulate without authors' permission

Abstract

Demand for healthcare is increasing across the world, but the supply of nurses is failing to keep up. Increasing leaver rates exacerbate workforce pressures, negatively affecting patient outcomes. We investigate the impact on nurse retention of a non-prescriptive programme (the Retention Direct Support Programme, RDSP) which encourages hospitals to increase retention through data monitoring and improving the non-pecuniary aspects of nursing jobs. Our evaluation makes use of recent methodological advances in the difference-in-difference literature for staggered treatments. Using administrative data, we find that the programme has improved nursing retention by 0.78 percentage points (ppt), and helped retain, on average, 1,697 nurses and midwives who would have left their hospital organization otherwise. The RDSP also reduced the NHS-leaver rates by 0.41ppt on average. We find no effect of the programme on hospital patient mortality. Our findings indicate that a light-touch intervention can shift management behaviour and improve hospital workforce retention.

Keywords: labour supply, workforce retention, staggered adoption, difference-in-difference, non-monetary intervention, hospital care
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, Igor Francetic, Hugh Gravelle, Elaine Kelly, Jonathan Wadsworth, 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 and University of Southampton (February 2022), Royal Economic Society Conference (April 2022). The usual disclaimer applies.

1 Introduction

Work is the engine of society. Understanding how individuals are incentivised to work is a central concern among social scientists. In the last two decades, research has widened its scope from the estimation of wage elasticities to investigate a broader range of factors shaping labour supply, including non-monetary incentives; considering their role in influencing whether, how long and how hard employees work. The responsiveness of labour supply to different types of incentives is especially pertinent in mission-driven public sector occupations, whose workers have been assumed to care less about financial rewards (Besley and Ghatak, 2005). As public sector employers struggle to fill staffing gaps with limited funding, the question of how to motivate their essential workers becomes ever more important. With this study, we aim to evaluate the effectiveness of non-monetary incentives on a large scale, by exploiting a natural experiment generated by a labour market policy aimed at decreasing the turnover of nurses working in all public hospitals in England.

Health care is a labour intensive sector and nurses are a vital part of its workforce, constituting about one third of the healthcare workers both in the United States and United Kingdom. Nurse vacancy rates are high across 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); similarly, in the US about 1.1 million new registered nurses are needed by 2030 (Bureau of Labor Statistics, U.S. Department of Labor, 2021).

High vacancy and turnover rates are particularly worrisome because 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).

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 an efficient solution to staff shortages (Shields, 2004), with the additional benefit of retaining specific human capital. 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 labour supply among nurses. Even if nurse retention was highly responsive to wages, a conspicuous pay rise across such a large sector like the UK public healthcare system[1] is unlikely to be affordable for the public finances without a significant tax rise or cuts to other public services, as shown also by recent events.[2]

An alternative approach to reducing turnover rates is to improve the non-financial elements of employment that workers value. According to a survey conducted in 2016/17 in the UK, the worsening of working conditions due to low staffing levels and high workload (Nursing and Midwifery Council, 2017) is second only to retirement as the most common reason why nurses and midwives have left their employing organization. An emerging literature discusses the importance of the non-pecuniary aspects of jobs for employee well-being, job satisfaction, labour supply and effort. In particular, Cassar and Meier (2018) emphasise the importance of specific non-financial factors to employees' behaviour, such as meaningfulness of work, mission, autonomy, competence and relatedness. For example, employees have been shown to value aspects of autonomy such as being involved in decision making processes (Böckerman et al., 2012) and having a flexible work (Mas and Pallais, 2017); and the value of competence is illustrated by employees learning on the job and being recognised for their efforts (Gallus, 2017). These factors are key features of the policy evaluated here, since they are particularly relevant for nurses. Nursing is a prime example of a mission-driven occupation, where individuals care about the outcome of their work in addition to their own remuneration (Brekke and Nyborg, 2010). A common conclusion is that nurses are likely

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

[2]In April 2022, the UK government had to increase general taxation aimed at improving NHS funding including a pay rise for NHS workers (Reuben and Eddington, 2022; Swinford et al., 2021).

to be less motivated by financial incentives than workers in other occupations (Besley and Ghatak, 2005), and may therefore be more influenced by other features of their jobs.[3]

Noticeably, very few studies look specifically at the impact of non-pecuniary occupational aspects on workforce retention (Shields and Ward, 2001; Moscelli et al., 2022). In this work we test the hypothesis of whether a low-cost non-wage intervention improves retention, i.e. labour supply at the extensive margins, within a given (healthcare) organization, but also within the whole (public hospital care) sector. Specifically, we evaluate the Retention Direct Support Programme (RDSP), which was launched in July 2017 by NHS Improvement (NHSI), the monitoring body of NHS hospital organisations. The explicit purpose of this intervention was to reduce turnover rates and improve retention among nurses in Acute care hospitals and clinical staff in Mental Health hospitals in the English National Health Service (NHS). There was no specific direction issued about how the NHS hospital organizations should achieve this goal. Instead, 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 areas of intervention considered in these plans included many of aforementioned factors that positively influence job satisfaction, in particular improving career progression and development, work-culture and staff flexibility. The provision and analysis of retention data is an especially important element 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 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. Our evaluation makes use of a unique dataset, consisting of a monthly panel of English NHS hospitals, which we constructed by combining different sources of

[3]According to Brekke and Nyborg (2010) and Heyes (2005), too strong a financial incentive might attract workers who do not care enough about their patients well-being.

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). Our identification of the impact of the programme comes from the differential timing of the RDSP across cohorts. We make use of new methodological advances in difference-in-difference (DID) estimation with staggered treatment adoption, specifically the DID estimator proposed by Callaway and Sant'Anna (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 interaction-weighted estimator of Sun and Abraham (2021) to capture dynamic treatment effects.

The intervention was designed to improve workforce retention at the hospital organization level, but we also check for effects on patient outcomes. Given that workforce shortages have been found to affect patient outcomes, it is possible that patient outcomes such as mortality would be improved 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. Alternatively, the specific focus on retention could distract hospital managers and senior clinicians from activities with more direct benefit for patients. Our investigation suggests the presence of positive (although not statistically significant) effects on mortality outcomes (i.e. a reduction in standardized mortality rates) and also on productivity, proxied by the number of admissions, in the treated organizations. There is certainly no evidence that the intervention worsened patient outcomes.

Finally we try to unpack the "black box" of the RDSP by exploiting information about the programme areas of intervention (called "themes" in the RDSP) 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 the success of the intervention was heterogeneous according to the baseline hospital retention rates and the

different areas of intervention chosen. In particular, improving career progression, development and engagement and stimulating a compassionate work culture contributed to the largest retention gains for the hospitals characterized by the worse average retention before the intervention; while improvements in staff engagement, support to new staff, selection of new joiners and the inclusion of retention in the organization strategy worked best in treated providers with the highest baseline retention.

This work contributes to several literatures. First, it adds to our knowledge of how labour supply responds to non-monetary incentives, by evaluating a large-scale intervention affecting the labour supply of the entire English NHS hospital nursing workforce. We document that the areas of improvement chosen by the hospital organizations are akin to, or the same as, those shown to influence worker labour supply and effort in other contexts (Bandiera et al., 2009; Beckmann et al., 2017). In addition, our setting is one where mission is especially important (Brekke and Nyborg, 2010) and workers may respond with a different intensity to non-financial, compared to financial, incentives (Besley and Ghatak, 2005). Second, our paper relates to the question of how managers and workers respond to incentives in the NHS and in the public sector more generally. It has been shown that targets and incentives can influence performance in these settings (Propper et al., 2010; Cooper et al., 2011; Gaynor et al., 2013), but the intervention we investigate is very low-powered, as no specific targets were set and the managers charged with implementation were provided with only weak incentives for success. This suggests that in a mission-centred workplace low-powered incentives combined with information (see below) can be a powerful tool for improvement. Third, and relatedly, this paper links to the emerging literature on the impact of management on organisational performance (Bloom and Van Reenen, 2007). The policy that is being evaluated is not prescriptive. Instead, it involves the provision of information and a "nudge" towards adopting best practice, particularly with regards to human resource management. Data, monitoring and people management are emphasised as especially important for good management and performance in public sector organisations (Bloom et al., 2014; McNally et al., 2022) and

there is evidence of room for improvement in management practices in the NHS (Bloom et al., 2012). The management of NHS organizations can improve in response to incentives, even in the fairly short run (Bloom et al., 2015) and here we show that light-touch interventions can shift organisations' management practices in ways that are beneficial for retention. Most research on the relationship between management quality and performance is not causal, and the staggered implementation of the RDSP policy allows us proceed further than usual in this direction. By focusing on the effects on workforce retention rather than productivity, our work has a slightly different focus from other studies on effective management practices, but tackling low retention is of great importance to the performance of healthcare systems under increasing pressure for the demand of healthcare, such as the NHS.

In Section 2 we present the institutional settings of the English NHS, its nursing workforce and the RDSP policy. In Sections 3 and 4 we respectively describe the data and empirical strategy used, and the results of the analysis. Section 5 concludes.

2 Institutional Background

2.1 The English NHS and its nursing workforce

The English NHS is publicly funded through general taxation and provides free comprehensive primary, secondary and tertiary healthcare services to over 56 million people in England [4]. With a budget of approximately £110bn in 2017 for England, the NHS dominates provision of healthcare in England and the other devolved nations of the UK (Scotland, Wales and Northern Ireland). Public hospitals providing secondary care are run by organizations called NHS hospital Trusts, or simply NHS Trusts.

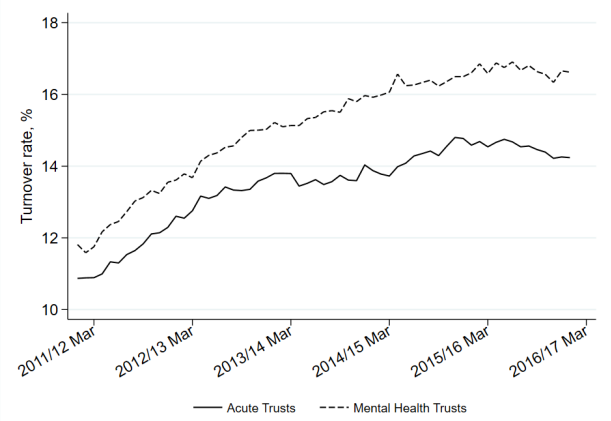
As of March 2020, about 564 thousand nurses, midwives and nursing associates living in England were registered with the Nursing and Midwifery Council (NMC, 2020).[5] The

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

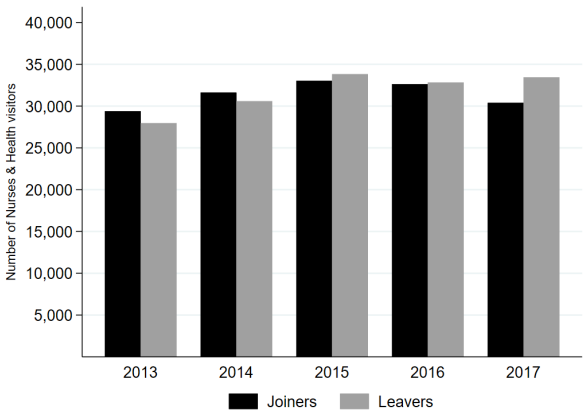
[5]The NMC is the professional body for nurses and midwives in the UK. To practice their profession, nurses and midwives need to register with the NMC and qualify to the NMC's standards.

Figure 1: Trends in NHS Nursing Workforce

(a) Nurses' 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 the 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).

English NHS employs around 330,000 nurses and midwives.[6], who make almost half of the professionally qualified clinical staff[7]

Nursing has been under significant pressure from rising staff shortages due to high turnover and vacancy rates, against the backdrop of growing demand for healthcare. The number of nurses and health visitors joining the NHS has been offset by the number of leavers in 2016/17 (Figure 1(a)), and 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 the need in the health services. The same period has also seen increasing turnover rates from nursing staff (Figure 1(b)), and increasing vacancy rates which stood at 10.9% with over 38,000 full-time equivalent open posts in the first quarter of 2017/18 (June 2017) (NHS Digital, 2021). The high number of nursing vacancies and staff shortages means that the NHS relies significantly on temporary and agency staff, which is estimated to cost NHS Trusts approximately £1.46 billion (The Open University, 2018). A potential way to reduce

[6]For brevity, in this work the terms "nursing staff" and "nurses" are referred to both nurses and midwives.
[7]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.

temporary staffing costs for NHS hospital providers is to invest in existing human capital by increasing the retention of the nursing workforce.

Staff shortages also affect existing staff's morale and wellbeing at work (Perreira et al., 2018). High turnover rates mean increased workload for already thinly stretched nursing staff and contribute to work related stress. The results from the NHS Staff Surveys show that the share of nurses and midwives who experienced work related stress in the last 12 months has increased from 40.87% in 2015 to 43.31% in 2019. Alongside stress at work, House of Commons Health Committee (2018) and the 2019 NHS Long-term Plan (NHS, 2019) both emphasised that lack of flexibility and limited continuous learning and development opportunities pose barriers to improve nursing retention. Some macroeconomic and social factors such as (national) pay regulations, changing workforce due to Brexit and cuts to central funding for Continuing Professional Development (CPD) are clearly beyond the control and influence of individual NHS Trusts. On the contrary, the non-pecuniary aspects of work[8] can be acted upon to redress burnout and work-life imbalance through adequate managerial strategies and workforce planning, which in turn would improve workforce retention.

2.2 Retention Direct Support Programme

The Retention Direct Support Programme (RDSP) was designed by NHS Improvement as part of a three-year National Retention Programme to tackle the nursing supply challenge. The aim of the RDSP was to improve nursing retention in public Acute-care Trusts and retention right across the clinical workforce in Mental Health Trusts in England (NHS England, 2019) by providing targeted information and support to Trusts to reduce turnover rates.

The programme was clinically-led, involving at least one member of the nursing team from the Trust, and focused on factors that are under Trusts' control (NHS Improvement, 2017). The RDSP took a unique form for each Trust. It started by identifying potential

[8]The non-monetary aspects of work are, but not limited to: creating supportive work environment, providing clear career paths and career development opportunities, and improving staff engagement.

areas of improvement before developing and carrying out a unique action plan to improve retention over the next 12 months.

The RDSP was rolled out in a staggered fashion over 5 cohorts from June 2017 until 2020. NHSI allocated Trusts into cohorts over time, starting with Trusts that had above-average leaver rates.[9] The RDSP was launched in June 2017 for the first cohort of Acute and Mental Health Trusts. Additional cohorts started the RDSP at later dates: Cohort 2 in October 2017, Cohort 3 in April 2018, and Cohort 4 in November 2018. By the end of 2018, the RDSP reached 146 secondary-care Trusts. In September 2019, shortly before the national programme's end, RDSP was extended to the other 62 Acute and Mental health Trusts (Cohort 5) in England, plus Ambulance Trusts.

Structure of the RDSP interventions

In the weeks following the first contact from NHSI, clinical and workforce leads from the Trusts were invited to participate a "Retention Masterclass workshop", scheduled around six weeks in advance. These Trusts also received data packs from the NHSI a few days before the workshop, which contained Trust-specific retention measures with regional benchmarks (NHS Improvement, 2017), to help the Trust understand their retention profile and its potential for improvement. The Retention Masterclass workshop was the official launch event of the RDSP; it introduced the programme to the Trust's team, and functioned as an interactive platform to review and discuss the barriers to retention in their Trusts. During the workshop, NHSI also presented the domains that Trusts might focus on to reduce their leaver rates, showcased best practices and demonstrated how can be used to inform decision making. Trusts were also given guidelines on ways to develop action plans.[10]

Each Trust was matched with an NHSI lead, who collaborated with the Trust for the duration of the RDSP. After introductory calls and visits by the NHSI, Trusts were given

[9]The selection was based on several factors, but higher weight was given to Trust's turnover rates and trends in the five years preceding the RDSP. Trusts did not know which Cohort they were allocated into until only a few weeks before they were contacted about the programme.

[10]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`.

90 days to develop and submit a Retention Improvement Action Plan. Trusts were 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 monitored the progress of the Trusts, provided quarterly data packs and supported Trusts that lagged behind their targets.

Retention Improvement Action Plans

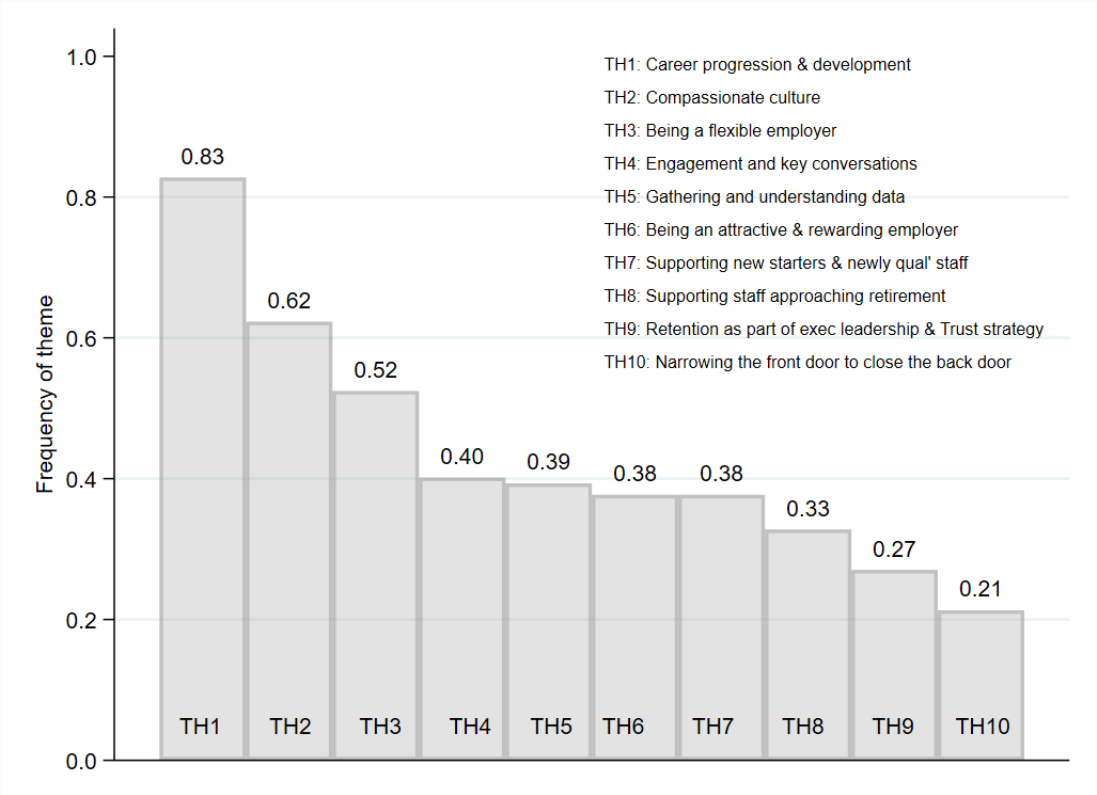
An important difference that sets RDSP apart from other national interventions was that the programme did not set any specific turnover rate targets for Trusts to achieve. NHSI's expectation was to see an *improvement in turnover rates* in the 12 months following the start of the programme. Additionally, rather than a one-fits-all approach, RDSP encouraged Trusts to focus on retention challenges endemic to their workforce, and to set their turnover goals accordingly. In this way, the programme also enabled Trusts to incorporate existing and planned workforce governance initiatives into their action plans.

The NHSI identified 10 recurring themes from the submitted action plans. Figure 2 shows 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. With cuts to CPD from the Health Education England's budget[11], Trusts are responsible for supporting developmental training and access to potential career opportunities. Some of the strategies under this theme included developing clear and attainable career paths, re-designing appraisal processes, and career coaching.

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. Some Trusts 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.

[11]The workforce development budget is mostly used for training of nurses and there has been a 60% cut in CPD budget from 2015/16 to 2017/18 (Bungeroth et al., 2018).

Figure 2: Frequency of Themes in Retention Improvement Action Plans



Notes: Authors calculations from NHSI thematic coding matrix from Cohorts 1 to 4. The themes are categorised by NHSI using action plans submitted by 122 Trusts.

Slightly more than half of the Trusts in the first four cohorts identified strategies to improve flexibility at work. 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 (Nelson and McLaughlin, 2020; Campbell and Allegretti, 2021; Chakelian, 2021), it was not a recurring theme in retention improvement plans. Only 13 of 122 action plans mentioned pay and it was classified under the promotion of "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 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 Trusts in England using various micro-level datasets.

We construct measures of retention from the monthly Electronic Staff Records (ESR) 2009-2020, 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.[12] The data contain information about programme's roll-out such as cohort allocation at hospital Trust level.

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 Trust level. The NSS are annual staff surveys commissioned by the NHS to collect information on NHS employees' experiences and wellbeing at work (NHS England, 2022). The NSS serves as a valuable resource to understand the differences in nursing staff's beliefs and perceptions about their workplace, which might have an impact on Trust's retention even before the RDSP was launched. 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 Trust-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 patient level to calculate 30-day monthly standardized hospital mortality indicators (SHMI)[13], and measure the

[12] NHSI works with the Department of Health and Social Care, and together with NHS England, monitors, oversees and provides support to NHS trusts to improve patient care.

[13] ONS linked data allows for the computation of mortality inside and outside the hospital, within 30 days of discharge. We follow the official SHMI specification by NHS Digital (2022) to create monthly SHMI, with minor differences in risk modelling, e.g. we use monthly indicators rather than yearly ones, to be consistent with the monthly evaluation of the RDSP.

hospital-level productivity as the number of emergency and elective admissions to acute care hospitals.

3.1 Measures of retention

We measure nursing retention in two ways: with stability rate (index) and with NHS-leaver rate, which are computed for each month on a year-on-year basis for all hospital Trusts. More specifically, we define the stability rate for nurses and midwives' in Trust h at calendar time measured in month t, S_{ht}, as

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

The stability rate indicates the percentage of the nurses and midwives who were actively employed in Trust h at $t-12$ and were still employed in the same Trust at t.In other words, we measure how many nurses and midwives are retained in a Trust on a year-on-year basis for each month, which accounts for seasonality by comparing the same month a year apart.[14] By definition, the stability rate at month t, S_{ht}, reflects the leaving decisions that occurred between $t-12$ and t, which may have implications on the programme evaluation that 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 Trusts, 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 organisational 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 organisations at t and do not reappear in the NHS payroll within the following six

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

months, $t+6$, i.e.

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

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 Trusts in the English NHS. A naive estimation of the impact of the RDSP on retention, S_{ht}, and on leaver rates, L_{ht}, can be achieved with the following (static) two-way fixed effects (TWFE) baseline model:

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

where $Y_{ht} = \{S_{ht}, L_{ht}\}$ are the retention outcomes in Trust h at calendar time t. D_{ht} is the treatment indicator, and takes the value 1 for all periods when Trusts launched the RDSP. μ_h and λ_t are Trust 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 analyse the dynamics of treatment effects is through an event-study TWFE specification:

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

where D_{ht}^k is the event-time indicator for relative time 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. As it is standard in the literature, we exclude the month before the RDSP launched, $k = -1$, as the reference period.[15]

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 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 interactions terms for cohorts (e.g. Sun and Abraham, 2021; Wooldridge, 2021) or by transforming the comparisons into a conventional two groups - two periods setting and aggregating treatment effects (Callaway and Sant'Anna, 2021).[16]

In our analysis, we evalute 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 Trusts for identification.[17] We also

[15] 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.

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

[17] 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.

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.[18]

Callaway and Sant'Anna (2021) heterogeneity-robust estimator

The DID approach by CSA 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, $ATT(c,t)$s, are the average treatment effects at calendar time t for hospital Trusts 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 period when $c \leq t$:

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

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 Trusts in cohort c and the control group, NT. The control group can consists of Trusts 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.[19]

CSA also allows for parallel trends to hold after conditioning on covariates through a

[18]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 Trusts and without covariates. The inclusion of time-varying covariates (linearly) in TWFE framework requires additional assumptions as treatment may have different effects across subgroups of the treated Trusts (Baker et al., 2022; Roth et al., 2022), and introduce bias to the estimates (see Sant'Anna and Zhao (2020) for details on inclusion of covariates in DID settings).

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

doubly robust estimation method.[20] As the cohort allocation of RDSP was not exactly random, imposing parallel trends conditional on past retention values (along with other conditions) provide further robustness to our results.

Policy makers are often interested in an overall effect as well as how the effects evolve over time. The flexibility in aggregating average treatment effects into a single policy relevant estimate also contributes to our choice of CSA estimator. We can aggregate $ATT(c,t)$s to understand the overall impact of taking part in the RDSP, across cohorts and over time:

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

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).

Trusts 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 Trusts allocated to the first four cohorts as treated, and Trusts in Cohort 5 as controls, i.e. *never-treated* Trusts.[21] One advantage of using Cohort 5 as the never-treated comparison group is that it allows us to estimate the RSDP's impact on every cohort for at least over a 12 month period. This particularly affects Cohort 4 Trusts, whose 12-month under 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* Trusts, i.e. all the Trusts belonging to cohorts that have not yet started the intervention.

[20] This method combines inverse probability weighting (matching) and outcome regression method to minimise 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).

[21] Under any circumstances it would be hard to capture the impact of the programme for Cohort 5 in such a limited time frame.

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 Trusts, the policies remained in place for the remainder of the sample period. There is no (expected) anticipation effects of the treatment for the treated organizations, as hospital Trusts were informed about their involvement in the RDSP only 6 weeks in advance without further information about the scope and the extent of the programme. This short notice period minimises the risk of potential anticipation for individual hospital Trusts. Another potential concern in this identification setting is the existence of potential spillover effects across Trust from different cohorts. These could arise if Trusts are in networks, and might be more likely among Trusts that are geographically close to each other. We discount this possibility for a number of reasons. First, the customised nature of the retention strategies adopted by each Trust means that the strategy adopted by one may not be suitable in elsewhere. Additionally, Trusts not included in the RDSP would not have access to bespoke data and support from NHSI. Moreover, the CSA estimator is suitable to our study as it allows to aggregate event-study ATTs estimates into overall ATT and cohort-specific ATTs, and it provides also the flexibility with the choice among two estimators, one unconditional and the other conditional on pre-determined covariates and doubly-robust.

Sample restrictions

We restrict our attention to nursing and midwifery staff working in acute (including Community and Specialist care) and mental health care Trusts, and exclude a small number of Trusts that have undergone organisational changes, e.g. mergers and acquisitions, from July 2016 onward.

We exclude these Trusts for two reasons: (i) the information on cohort allocation of at least one Trust that has undergone such an organisational change is missing in NHSI data collection; (ii) when the information is available, which is only a few cases, the Trusts that

were involved in a merger/acquisition operation have been allocated to different cohorts at different times, e.g. one before the merger one after. This makes the allocation of treatment status impossible. Moreover, the need for organisational change might be driven by or correlated with Trusts retention profiles, hence including these Trusts into our analysis may mask the true impact of the RDSP. We also exclude one Trust with very few nursing staff, Trusts that do not keep workforce information in ESR and have started using the ESR towards the end of the sample period.

Our analysis sample includes 193 NHS Trusts that we observe from June 2016 to November 2019. This time window allows us to observe all Trusts in the treated group for at least 12 months before and after their enrolment 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 Trusts in each cohort. Table 1 presents an overview of the NHS hospital Trusts in each RDSP cohort. The estimation sample is a panel balanced 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).

Between 2011/12 and 2015/16, the five-year period before the RDSP first took place, 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 at 7.15%. The first Cohort of the RDSP had the lowest average stability rate over the 5 years before the RDSP took place, with more than two-thirds of its Trusts at the bottom quartile of the pre-RDSP stability distribution. Compared to the first Cohort, only 27.59% of Trusts in Cohort 2 were performing poorly in terms of stability (Table 1), and more than half of the Trusts in Cohorts 3 and 4 were from the top two quartiles. Nevertheless, Figure 3 shows that the distributions of pre-RDSP retention measures substantially overlap among the treated and the control group.

Despite differences in their retention levels, hospital Trusts across all cohorts experienced

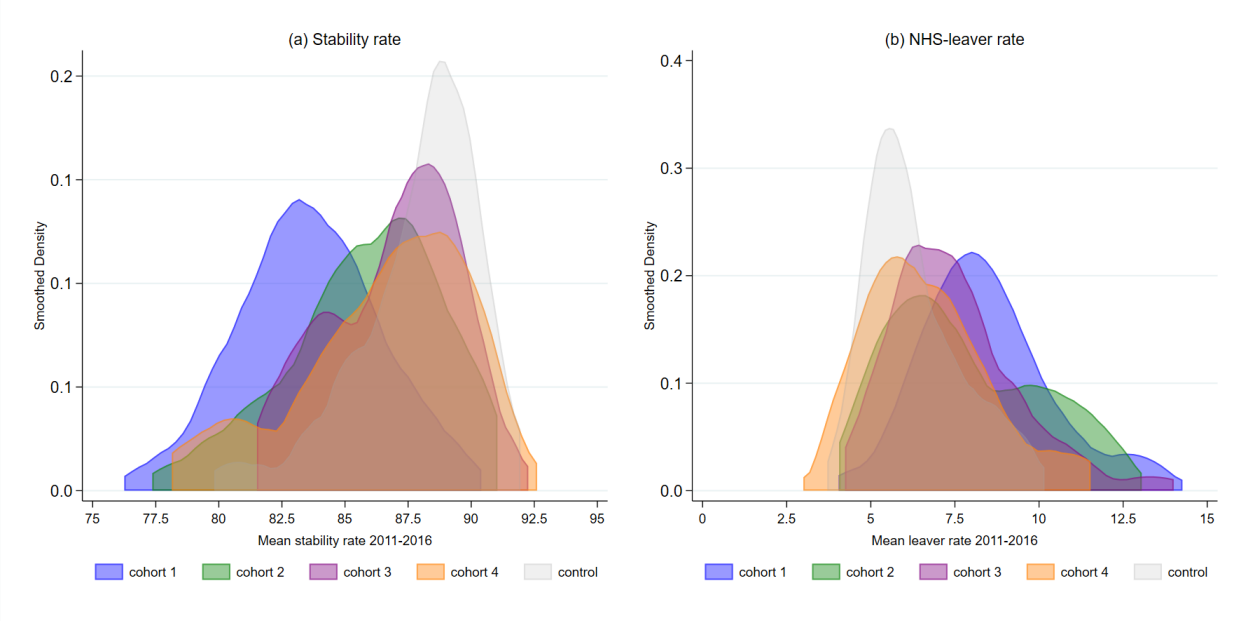
Table 1: Overview of Cohorts in the RSDP

	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 Trusts in Cohort 5 and two additional Trusts 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 Trusts in each quartile within cohorts.

similar trends with 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, a visual inspection of the pre-trends in Figure A3 reveals that all treated cohorts exhibit retention trends similar to the control cohort in the months leading

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



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

Table 2: Average retention before and after 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 Trusts, i.e. the timings for each cohort are June 2017, September 2017, March 2018, and October 2018, respectively. The end of the RDSP columns is based on the stability rates in November 2019, and 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.

to the RDSP.[22]

Table 2 presents 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). As expected, the largest differences in retention measures are between Cohort 1 and the control Cohort. The mean differences shows that the retention gap between these two groups almost halved, yet

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

a significant difference in retention remained in November 2019. On the other hand, the gap in stability rates among the other treated cohorts narrowed, and they caught up with the control group with no significant differences in retention outcomes by the end of sample period. Table A1 presents the summary statistics for selected characteristics from June 2017, i.e. the month before the RDSP was first launched. While the majority of nursing staff is female in all cohorts, only 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. Likewise, there are slightly more nurses and midwives from an ethnic minority background in treated cohorts compared to in the control group. As with the location of the hospital Trusts, these differences are likely to reflect regional labour market differences, which are not directly under the control of the Trusts.[23] 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 Trusts is less likely to work more than 11 additional unpaid hours than treated Trusts with around one in 16 nurses in the first cohorts working additional unpaid hours. Average monthly sickness rate is also similar across cohorts and control group in the month preceding the RDSP's first launch at around 4%.

4 Results

4.1 Retention outcomes

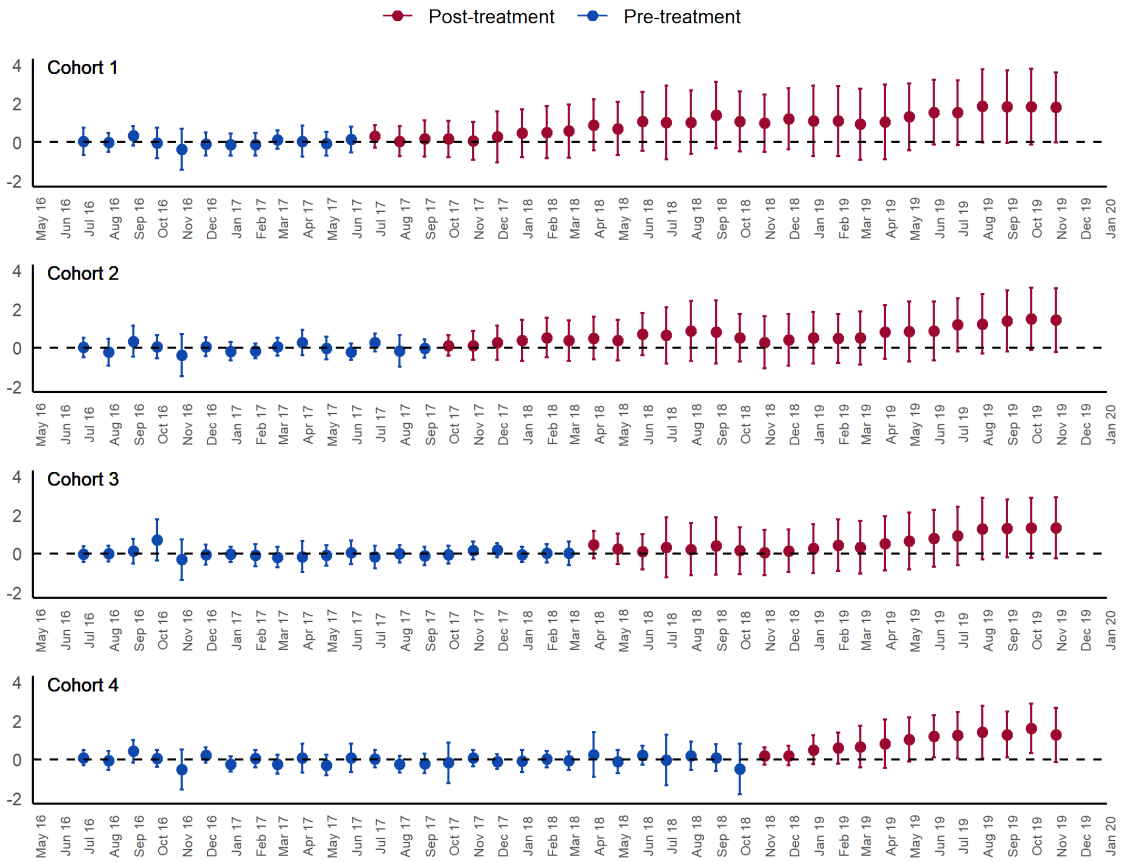
We first discuss the RDSP's impact on stability rates. Figure 4 presents the $\widehat{ATT}(c,t)$ for each Cohort and time period with simultaneous 95% confidence intervals under unconditional parallel trends. Using the augmented Wald-test proposed by CSA, we test whether the pre-

[23]Retention also varies across regions, and this is partially reflected in the composition of the cohorts. More than a quarter of the Trusts in London and East of England were in Cohort 1, followed by one fifth of secondary care Trusts in South East. On the other hand, more than half of the Trusts 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.

treatment estimates for the 6 (12) months leading to the treatment are jointly equal to zero, which we fail to reject with a p-value for 0.273 (0.080).

The impact of the RDSP on nurses and midwives' stability rate exhibits an increasing pattern in all cohorts; yet, there are also some heterogeneous treatment effects across cohorts as shown in Figure 4.

Figure 4: RDSP cohort-time average treatment effects on stability rate



Notes: The effect of RDSP on nurses and midwives' stability rates under the unconditional parallel trends assumption. The point estimates are shown with simultaneous 95% confidence bands from bootstrapped standard errors allowing for clustering at Trust level.

For the first group of Trusts (Cohort 1), which started the RDSP in July 2017, the intervention started yielding some (positive) effects from 7 months into the cohort's enrolment, with the effects becoming statistically significant at 5% level only towards the end of our sample period, i.e. after 24 months in the programme. On the other hand, as shown in the last panel of Figure 4, we estimate that Cohort 4 Trusts, which joined the RDSP the latest, experienced stronger effects after a shorter exposure to the RDSP compared to other

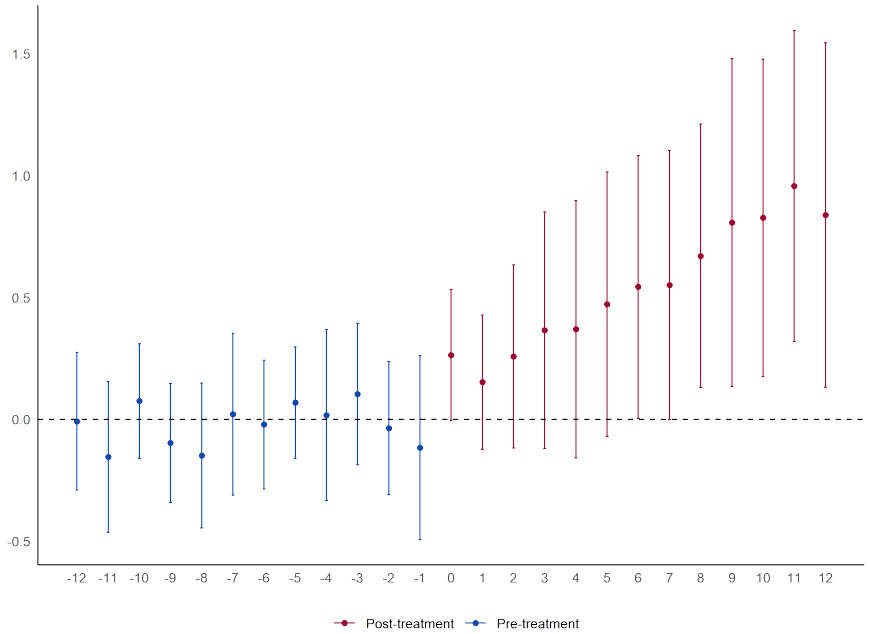
cohorts. However, as our sample has a large number of pre- and post-treatment periods, the cohort-by-month ATTs are mostly imprecisely estimated. Thus, in the remainder of the paper, for stability rate, we focus our attention mostly on ATT estimates aggregated across the full set of Trusts or by Trusts in the same cohort, which are not only more precisely estimated, but also more policy relevant.

Figure 5 is an event-study plot of the average treatment effects for the first 12 months after the exposure to the RDSP. The average treatment effects come from a balanced sample in terms of treatment timing, i.e. all cohorts have at least 12 months post-treatment period. We find a stronger increasing impact of the RDSP over time when compared to cohort-time ATTs shown in Figure 4. On average, there is no significant impact of RDSP on stability rate in the first three months, which is expected as Trusts worked on designing the retention improvement action plan to be signed off by NHSI officers without any strategies being implemented yet. Besides the progression of the RDSP over time, it is also likely that the increasing event-time estimates over the first 12 month period might be due to the definition of our outcome variables, which we anticipate to underestimate the true impact of the RDSP in its first year. As described in Section 3.1, retention measures reflect nurses and midwives' decision to leave their Trust (or the NHS) in the last 12 months; thus, retention measures for the first 11 months of the RDSP contain some pre-treatment periods to account for the leaving decisions.[24] In the next sub-section we come back to this point and evaluate the impact of this on overall and cohort-level estimates.

We find similar results using both the dynamic TWFE and Sun and Abraham's (2020) interaction-weighted estimators. The comparison of event-study estimates are presented in Figure 6, under unconditional parallel trends. A potential explanation for the dynamic TWFE estimates being close to the robust estimators is the existence of a relatively large number of *never-treated* Trusts compared to treated Trusts at each time period. The large

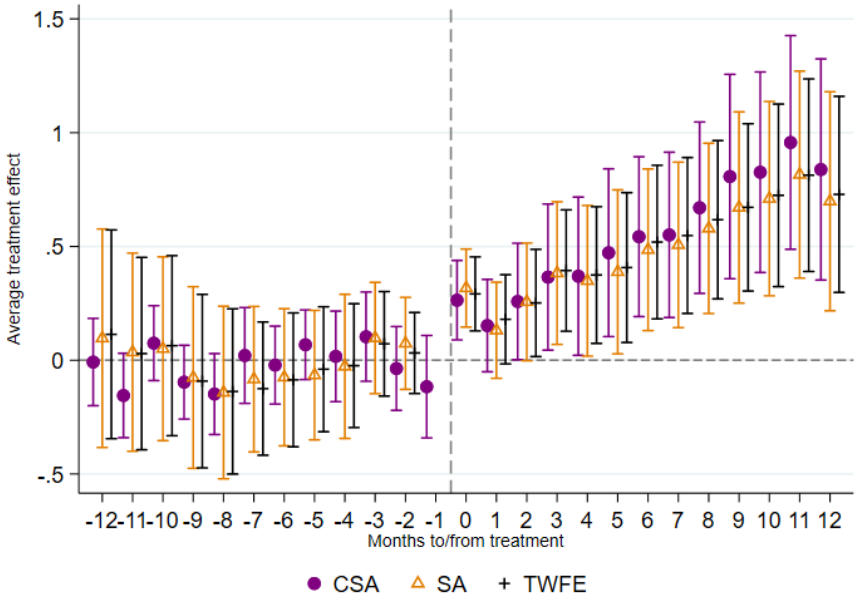
[24]For instance, the stability index for the first month of the RDSP looks at 11 months pre-RDSP and 1 month post-RDSP, the second months looks at 10 months pre-RDSP and 2 months post RDSP, and so on until 12 months into RDSP when all periods used to compute the stability index comes from post-RDSP months.

Figure 5: Dynamic treatment effects on stability rate



Notes: Event-study estimates under unconditional parallel trend assumption with uniform 95% confidence intervals and bootstrapped clustered standard errors allowing for clustering at Trust level using did package in R (Callaway and Sant'Anna, 2020). We use balanced cohorts with relative time periods spanning from 12 months pre- and post-treatment.

Figure 6: Dynamic treatment effects on stability rate, alternative estimators



Notes: Event-study estimates under unconditional parallel trend assumption. To facilitate comparisons across different estimators, the figure shows asymptotic standard errors clustered at Trust level with point-wise 95% confidence intervals. In Sun and Abraham (2021) and TWFE specifications, the month before the RDSP, $k = -1$, is omitted to avoid multicollinearity; and periods $k < -12$ and $k > 12$ are binned instead of trimming the sample.

size of the control group helps reducing the importance of negative weights and minimising the bias from using early-treated units as controls (Jakiela, 2021; Baker et al., 2022).

Table 3: Average treatment effects of RDSP on retention outcomes

	Stability rate		**NHS-leaver rate**	
β^{TWFE}	0.472 (0.170)***		-0.249 (0.125)**	
Callaway and Sant'Anna (2021)				
Overall ATT	0.775 (0.188)*** [0.196]§	0.862 (0.199)*** [0.195]§	-0.408 (0.131)*** [0.128]§	-0.389 (0.137)*** [0.135]§
Cohort-specific ATTs				
Cohort 1	0.950 (0.352)*** [0.360]§	1.017 (0.366)*** [0.388]§	-0.498 (0.254)** [0.233]	-0.448 (0.261)* [0.278]
Cohort 2	0.677 (0.303)** [0.309]	0.616 (0.361)* [0.348]	-0.416 (0.202)** [0.209]	-0.377 (0.236) [0.231]
Cohort 3	0.557 (0.336)* [0.333]	0.843 (0.403)** [0.394]	-0.394 (0.248) [0.250]	-0.386 (0.253) [0.251]
Cohort 4	0.912 (0.267)*** [0.269]§	0.944 (0.265)*** [0.259]§	-0.341 (0.183)* [0.181]	-0.351 (0.183)* [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 Trust 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.

Table 3 reports the ATTs aggregated by cohort and across all cohorts (overall ATT). Under unconditional parallel trends (column 1), the highest average treatment effect is found

for Cohort 1 Trusts, which on average corresponds to one-third of a standard deviation increase in nurses' stability rate. This is perhaps not surprising as these Trusts were the first to be enrolled into RDSP, spending 28 months in the treatment. Despite spending less than half of the time Cohort 1 spent in RDSP, Trusts in Cohort 4 increased their retention on average by 0.91ppt in 12 months, which is equal to 28% of the standard deviation in the month before the RDSP was launched in Cohort 4. On average, the programme had some positive, but imprecisely estimated effects on Trusts in Cohorts 2 and 3.

The cohort-specific effects provide the average treatment effect for each cohort, i.e. the impact of being enrolled to the RDSP averaged across the time the cohort spends in the programme. We compute the overall impact of the programme by taking a weighted average of the cohort-specific ATTs. The overall impact of participating to the RDSP is on average a 0.78ppt increase in stability of nursing workforce in treated Trusts. The conventional counterpart of the overall ATT is DID estimate from the TWFE, β^{TWFE}, which at 0.47ppt is almost half the size of the overall ATT under unconditional parallel trends.

We present the average treatment effects of the RDSP on stability of nursing staff under conditional parallel trends in the second column of Table 3. Conditioning trends on observed covariates might be considered as a more credible approach compared to unconditional parallel trends assumption if retention trajectories depend on factors that would determine Trusts allocation into cohorts. Given the nature of allocation into RDSP cohorts, this approach gives rise to an additional challenge of finding a set of covariates that provide a common support for the propensity score for the doubly-robust estimator Callaway and Sant'Anna (2021). As conditions, we use the difference between the average past retention of the Trust and its allocated Cohort to capture variations in past retention. As shown in Figure 3, there was some variation in retention across Trusts in each cohort, providing common support for *treated* and *never-treated* Trusts. We also control for the age of creation of the Trust[25] and nurses and midwives' sickness absence rate, to capture organizational and workforce practices. In specifications using stability rates as outcomes, we also include items

[25]Older Trust might have some established workforce retention practices.

from the NHS Staff Surveys to characterise the workplace environment before the launch of RDSP: the share of nurses and midwives among the NSS respondents of the Trust; 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).[26] All these covariates are set to their pre-treatment values, i.e. the month before the RDSP was launched, and used only for the estimation of the propensity score used to re-weight the effects, thus they do not influence the RDSP retention outcome directly. The doubly-robust ATTs estimates are of comparable magnitude to those estimated under unconditional parallel trends. We find that the RDSP helped increasing nurses and midwives' retention of treated NHS hospital Trusts on average by 0.86ppt. The cohort-specific ATTs are also similar, with a slight increase in magnitude particularly for Cohort 3.

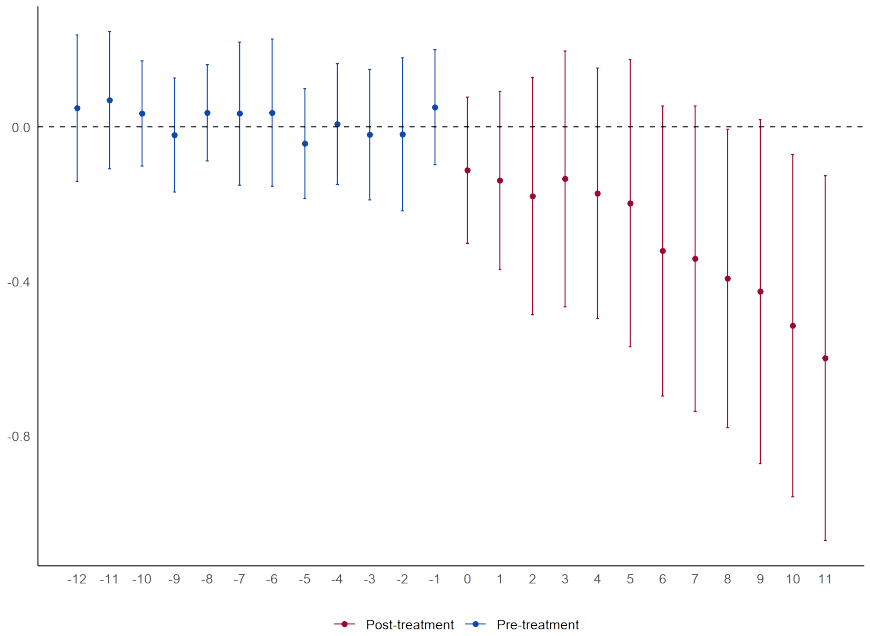
Our second set of main results uses the NHS leaver rates of nurses and midwives as the retention outcome of interest, to investigate whether the RDSP reduced the loss of human capital to the NHS overall. We first estimate the impact of the RDSP on NHS leaver rates under unconditional parallel trends. The augmented Wald test on the pre-treatment estimates provides statistical evidence that the nurses and midwives' NHS leaver rates follow common trends in pre-treatment periods (see column 3 in Table 3). The RDSP had some positive impact in lowering the leaver rates in each cohort, but the cohort effects are less precisely estimated.[27] The RDSP had decreasing dynamic treatment effects on nursing staff NHS-leaver rates in the first 11 months of RDSP as illustrated in Figure 7.[28] The average impact of the RDSP goes from 0.39ppt reduction in NHS-leavers rate at 8 months to 0.60ppt

[26] We also estimated alternative specifications including the share of nurses who felt unwell due to work 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 3, and are available upon request.

[27] Figure A4 presents the cohort-time ATTs.

[28] We focus on 11 months after the RDSP kicked off in Trusts because NHS leaver rates, as described in Section 3.1, are constructed by following up nurses and midwives in the ESR for $t + 5$ months to differentiate between churns and NHS leavers. Thus the last non-missing value for NHS leaver rates is the one for October 2019.

Figure 7: Dynamic treatment effects on leaving the NHS rate



Notes: Event-study estimates under unconditional parallel trend assumption with uniform 95% confidence intervals and bootstrapped clustered standard errors allowing for clustering at Trust level using did package in R (Callaway and Sant'Anna, 2020). We use balanced cohorts with relative time periods spanning from 12 months pre- and 11 months post-treatment.

decrease at the 11^{th} month of RDSP. As discussed above for dynamic effects on stability rate, another potential explanation for this decreasing pattern could be the definition of NHS-leavers, and we return to this point in robustness checks.

Similar to stability rates, alternative estimation methods provide similar estimates for the dynamic treatment effects (see Figure A5). When we aggregate the cohort-specific ATTs, we find that the RDSP reduced nurses and midwives' NHS leaver rates by 0.41ppt in treated Trusts. This amounts to one-fifth of a standard deviation decrease in the NHS leaver rates of nurses and midwives (where the SD is calculated for the month before Trusts' RDSP launch). The last column in Table 3 presents the results under conditional parallel trends[29], which are quite similar in magnitude to the ones obtained under unconditional parallel trends.

[29]Through the propensity score estimation, we indirectly control for the mean difference between the past 5 years' average NHS leaver rates of Trust and the allocated cohort, the Trust-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.

4.2 Robustness Checks

"Mechanical contamination" and breakdown of RDSP impact over time

As described in Section 3.1, the retention measures are calculated by comparing in each month and Trust the number of nurses to the number of same nurses in the following 12 months. The number of workers leaving before the RDSP started might introduce a "mechanical contamination", leading to a downward bias (i.e. towards zero) in the estimates of interest. In order to check the direction and the magnitude of this bias, we compare the overall and cohort-specific ATTs in Table 3 with the censored ATTs, which we obtain by averaging the cohort-specific ATTs from 12^{th} to 19^{th} months into the programme, and alternatively from 12^{th} months to the final observations of our sample. The censored ATT estimates are presented in Table 4.[30]

Although higher (lower) in magnitude, we find that the truncated impact of the RDSP, both overall and at cohort-specific level, on stability (NHS leavers rate) is not statistically different from the main effects reported in Table 3. The largest difference between ATT estimates is for Cohort 3, but the difference is not statistically significant (0.56 vs 1.01). From this analysis we conclude that the definition of the retention measures do not introduce any significant bias to our estimates, and that the ATTs in 3 are likely to be conservative point estimates of programme's impact, which might be even more effective in increasing nursing workforce retention.

[30]The censored ATTs are obtained using the same unconditional PTA model, but instead of aggregating all post-treatment periods as presented in Table 3, we aggregate the ATTs only for a set of post-treatment periods on and after 12 months.

Table 4: 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: Bootstrapped standard errors are clustered at Trust level.§ indicates that the 95% simultaneous confidence band does not cover 0. τ indicates the elapsed time of RDSP, which corresponds to November 2019 for stability rate and October 2019 for the NHS-leavers rate. $\tau = 28$ for Cohort 1, $\tau = 25$ for Cohort 2, $\tau = 19$ for Cohort 3, and $\tau = 12$ for Cohort 4.

Choice of the reference control group

In our main analyses, we use Trusts 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 Trusts that have above-average turnover rates. An advantage of using this cohort as the control group was to be able to estimate the impact of the RDSP on Cohort 4 for the full 12 month period. To do this we adjusted our sample period to end in November 2019 to limit the potential bias that the launch of RDSP for Cohort 5 in September 2019 might have introduced.

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* Trusts, rather than the *never-treated* Cohort 5. This increases the number of Trusts in the control group for each time period, some of which might be more comparable to early treated Trusts, but comes with a disadvantage of shorter analysis period for treated cohorts as Cohort 5 started

RDSP in September 2019. In practice, this means that the post-treatment period spans only until August 2019, instead of November 2019 when our sample period ends.

We present the estimation results in Tables A2 and A3 for stability rates and leaving the NHS rates, respectively. Columns (II) show the results from the model using Cohort 5 Trusts 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 launched the RDSP. As RDSP's impact increased over time, reducing the post-treatment period, particularly for Cohort 4 which ends prematurely at 9 months rather than 12 months, leads to a lower overall effect of RDSP on nursing retention. Likewise, our results hold for nurses and midwives' NHS leaver rates when we re-define our comparison group as not-yet treated Trusts. 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.

Heterogeneity analysis based on different sub-samples

Our main analysis is based on the whole population of nurses and midwives actively employed in English NHS hospital Trusts. We further assess whether the impact of the policy differed for particular groups of nurses and midwives. We refine our retention measures for nursing staff who are not on shift-based, mostly zero-hour, temporary contracts (Bank work), those on permanent contracts only, and nurses in Acute Trusts. We also re-compute the stability rates for nursing staff who are below the retirement age, which is the group we already focus on for the NHS-leaver outcome.

Table B1 presents the aggregated average treatment effects on retention outcomes for different sub-samples of nursing staff 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 Trusts, which significantly improved the stability of nursing staff who do not work only as Bank staff (column 2 in Table B1.[31]

4.3 The RDSP impact on labour supply at intensive margins

So far our analysis took into account the nurses and midwives' labour supply at the extensive margin in terms of stability and leaver rates. as the aim of the RDSP was to reduce turnover rates and improve retention within and across Trusts. Nevertheless, some strategies outlined in action plans such as e-Rostering[32] might have led to a re-allocation of working hours and might have alleviated the working conditions of the overstretched nursing workforce.

We focus on the average monthly hours worked by nurses and midwives who work full-time.[33] 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.[34] There was very little variation across full-time hours worked across cohorts in the last 3 years leading up to the RDSP (Figure A6 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 A6 panel b).

The RDSP did not have any impact on the average hours worked by full-time nurses and midwives in treated Trusts, as shown in Figure 8. This is not surprising as the RDSP's primary target was to improve working conditions through managerial and organisational

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

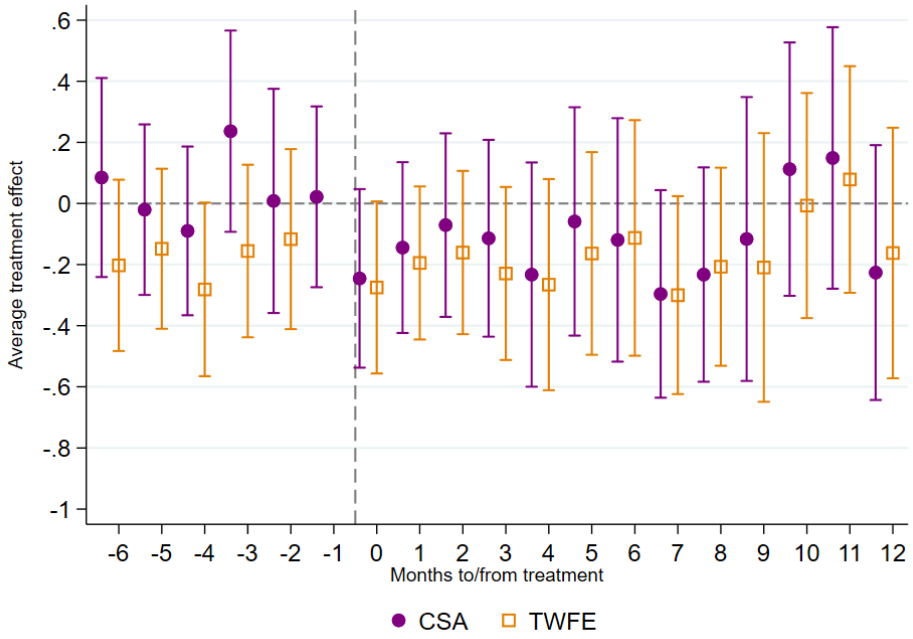
[32] E-Rostering is an electronic shift management system that provides real time information on staffing levels to meet healthcare demands and also facilitate workforce flexibility.

[33] 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 Trust 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.

[34] 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.

improvements.

Figure 8: 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 Trust 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..

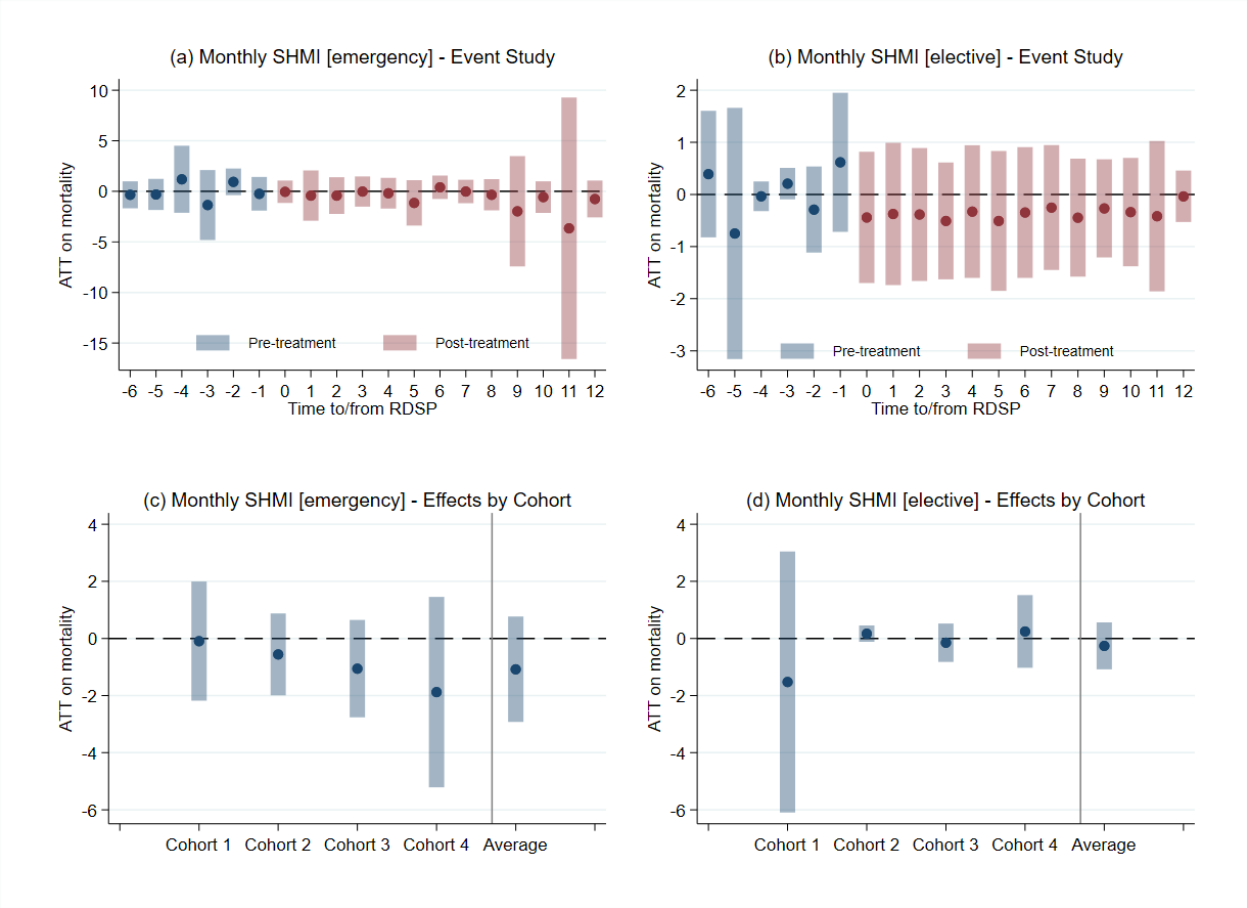
The RDSP might improved flexibility offers through additional Bank work. The share of Bank hours within nursing staff's working hours converged across cohorts in 2017, and has been on an increasing trend since then (Figure A7). In 2016, the year before the RDSP was launched, the average bank hours made 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%. We do not find supporting evidence that the retention programme had an impact on the share of Bank hours (see Figure A8).

4.4 RDSP impacts on Mortality and Productivity

We have speculated that the RDSP may also have impacts on patient outcomes, either indirectly as improved retention benefits patient care or directly through either improved

management (leading to an improvement for patients) or a diversion of attention and resources (leading to negative effects for patients). In Figure 9, we show the impact of the RDSP on standardized hospital mortality indicators (SHMI) for emergency and elective patients admitted for treatment to acute NHS hospitals in our sample. The effects are shown to be negative for emergency patients in all cohorts, and also for elective patients in Cohort 1 Trusts, but they are not statistically significant at conventional confidence levels.

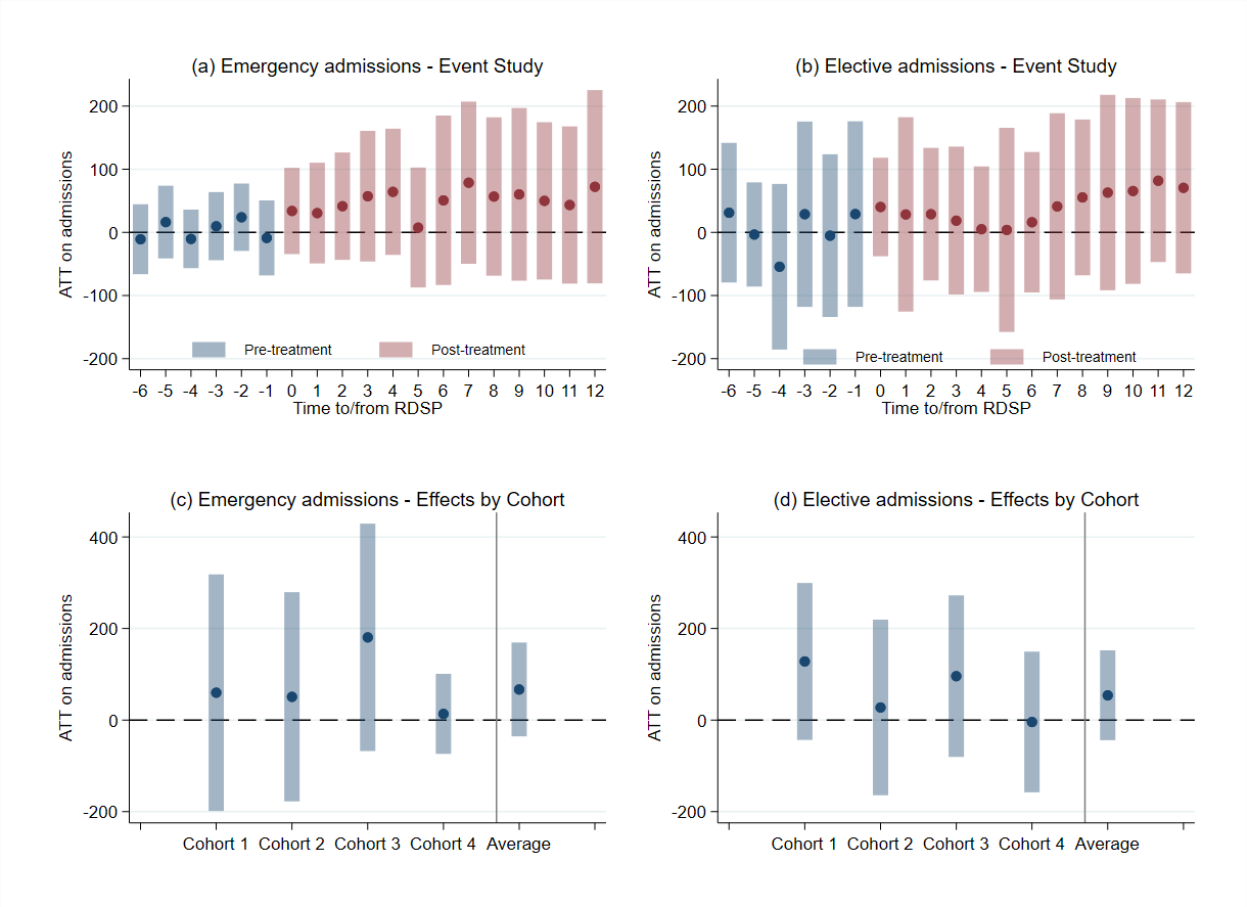
Figure 9: The Effects of RDSP on SHMI of patients aged below 75 in Acute Trusts



Notes: Average treatment effects are estimated using CSA (2020) under unconditional parallel trends assumption. The estimates are presented with uniform confidence intervals with bootstrapped standard errors clustered at Trust level. The pre-treatment parallel trends hold for the 6 months preceding the RDSP with p-values 0.520 for emergency SHMI and 0.069 for elective admissions' SHMI.

Figure 10 reports the impact of the RDSP on the number of emergency and elective patients admitted for treatment to acute NHS hospitals in our sample. We can think of this as a proxy for productivity at organization level. The RDSP effects are shown to be positive, although never statistically significant at the 5% level, for emergency and elective patients

Figure 10: The Effects of RDSP on level of admissions in Acute Trusts



Notes: Average treatment effects are estimated using CSA (2020) under unconditional parallel trends assumption. The estimates are presented with uniform confidence intervals with bootstrapped standard errors clustered at Trust level. The pre-treatment parallel trends hold for the 6 months preceding the RDSP with p-values 0.159 for emergency admissions and 0.403 for elective admissions.

across all cohorts, with the only exception of a small, negative and non-significant effect on the elective patients' admissions in Cohort 4 Trusts.

4.5 What works? Cohort-specific effects of the RDSP themes

Even though we show the RDSP policy has improved nurses and midwives' retention, there is evidence of heterogeneity in the RDSP's effect across cohorts. This can be due to several reasons: differences in treatment timing; a different responsiveness of the labour supply across Trusts of different cohorts; different choices regarding the RDSP policy implementation, which may be captured by the RDSP themes and that we analyze in this

subsection. We do not have a detailed account of which strategies each Trust has developed, acted upon or successfully completed within the RDSP, but we have information about the broad themes that appeared in the action plans for a sub-sample of Trusts from Cohort 1 to 4 (112 Trusts out of 132 treated Trusts in our estimation sample). As briefly discussed in Section 2.2, NHSI grouped ex post the approved RDSP plans into broader themes. We use this information to understand what strategies might have worked better to improve retention across Trusts in different cohorts.

Table A4 presents the frequency of themes that appeared in retention improvement plans by each cohort. On average, action plans contained 4 themes. Career progression and development (CPD) has been on top of the themes in all Cohorts, and more than three quarters of the Trusts submitted a strategy on CPD provision. Similarly, initiatives on creating a compassionate culture within the Trust dominated the action plans. The aim of the RDSP was to help Trusts to identify and improve the areas that hinder nursing retention. Thus, it is not surprising that different themes prevailed across cohorts. Earlier cohorts concentrated more on being an attractive and rewarding employer and included retention as part of executive strategies, while Trusts in Cohort 4 prioritised being a flexible employer and supporting staff approaching retirement.

We focus on treated hospital organisations, and estimate the following constrained linear regression model

$$S_{ht} = \alpha + \sum_{c=1}^{4} \sum_{a=1}^{10} \beta_c^a I_{c,h} A_{a,h} D_t + \mu_h + \tau_t + \varepsilon_{ht} \tag{3}$$

with time and hospital Trust fixed effects, and where $I_{c,h}$ is the cohort identifier for Trust h in cohort c, $A_{a,h}$ is a dummy variable for theme a adopted by Trust h during RDSP, and as before D_t takes the value one for the post-treatment period. We are interested in the parameter of the interaction term, β_c^a, which provides the difference in retention within cohorts for adopting a theme. The constraints are placed on the parameter of interests, β_c^as, where we set the sum of the estimated effects of the themes $\sum_{a=1}^{10} \beta_c^a$ for each treated Cohort, $c = \{1, 2, 3, 4\}$, to be equal to the cohort-specific ATTs estimated under unconditional

parallel trends assumption. Compared to the analyses reported in our previous sections, the rigorousness of this investigation is more limited due to missing data on the action plans for about 15% of treated Trusts, and the discretion of each Trust in choosing the main themes to act upon in their action plans, which is potentially endogenous. However, the aim of this exercise is also rather different from our main analyses which is to understand the mechanisms behind the RDSP's success in improving retention in each cohort, we use the constrained regression in Eq. 3 to decompose the estimated ATTs of each cohort according to the themes chosen by the Trusts.[35] While the results from this decomposition do not have a causal interpretation, they can provide suggestive insights about which areas of RDSP intervention were associated with larger gains in nursing workforce retention.

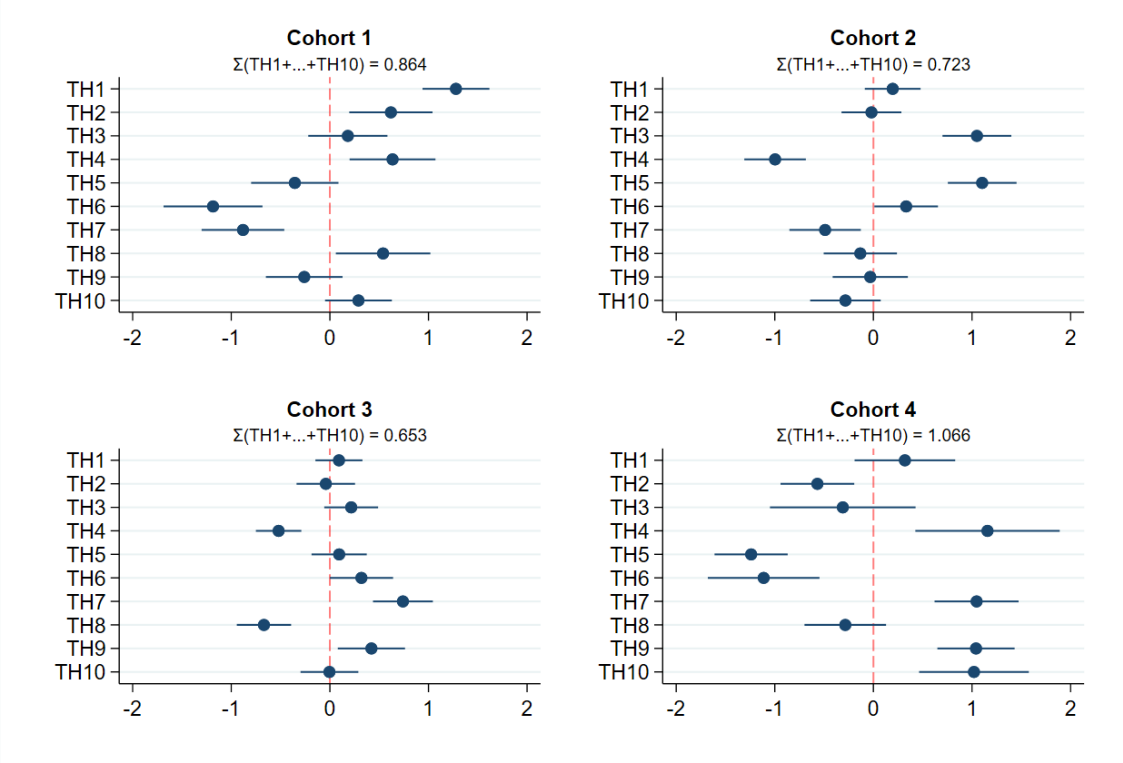
Figure 11 presents the estimated associations between action plan themes and Trusts' stability rates for each cohort.[36] In all cohorts, we find a number themes, ranging from 3 to 5, with a positive and significant impact on the stability of nurses and midwives. Retention gains are associated with different strategies (themes) in different cohorts. This is not unexpected, as the RDSP was not a prescriptive one-size-fits-all intervention, but one which allowed Trusts enough flexibility what to prioritize to improve workers' retention.

In Cohort 1, the highest positive associations are found with themes addressing the personal and professional needs of nursing staff, such as *career progression and development* (TH1), promoting *engagement and key conversations* (TH4), fostering *compassionate culture* (TH2) and *supporting staff approaching retirement* (TH8). Despite a positive but statistically insignificant impact of the RDSP on Cohorts 2 and 3 stability rates, we find some action points worked better than others in these Trusts: for Cohort 2 Trusts, the action points were centred around organizational matters such as being a*flexible* (TH3) and *an attractive*

[35]The estimated cohort-level ATTs imposed as constraints are unbiased as they depend just on the dynamic treatment assignment of RDSP cohorts. We expect that the imposition of these unbiased constraints reduces issue due to the endogeneity of the themes in our regression-based decomposition.

[36]The sum of theme parameters is equal to cohort-specific ATTs from a model estimated using the 112 Trusts for which an action plan with themes as treated unites, and aggregated in the same fashion as in Table 3. The results are available upon request.

Figure 11: The Effects of Themes on Stability rates for Cohorts 1-4



Notes: Results from estimating the constrained regression equation 3 using Trusts in Cohorts 1-4 with 95% confidence intervals from robust standard errors. The sum of theme parameters equals cohort-specific ATTs from a model estimated using the 112 treated Trusts that submitted an action plan with themes, and aggregated in the same fashion as in Table 3. For details on the themes within Cohorts, see Table A4.

& rewarding employer (TH6); in Cohort 3 Trusts, the retention gains are associated with *supporting new staff* (TH7) and having retention on the agenda of a Trust managers (TH9). Interestingly, the highest positive theme association on nursing stability for Cohort 2 Trusts were from the *gathering and understanding data* theme (TH5), which had been introduced as a step to identify and address barriers against retaining staff in the Trusts during RDSP retention workshops (see footnote 10). Trusts in the last treated cohort, Cohort 4, spent 12 months in RDSP until the end of our observation period. During this time, we find that a mix of staffing policies such as fostering engagement, and managerial strategies to improve retention are associated with gains in nursing stability.

Finally, the results from constrained linear regression for nursing NHS-leaver rates from equation 3 are presented in Figure A9, and show patterns of associations between themes

and drops in NHS leavers are very similar, although specular, to those with stability rates.

5 Conclusions

Staffing pressures are intense in the public sector as demand continues to grow while turnover rates increase. We speculate that public sector workers are strongly motivated by the non-financial aspects of their jobs, but we know relatively little about how this can be harnessed to improve employee retention. This paper has examined the impact of the Retention Direct Support Programme (RDSP), which exploited the amelioration of hospital job's non-financial conditions to raise nursing retention in NHS hospital providers (Trusts) in England as a part of a 3-year national retention improvement programme.

The RDSP was rolled out in 5 cohorts in a staggered fashion, with potential differences in treatment effects across cohorts and over time. As the recent literature on difference-in-differences has shown, estimating the impact of a policy with differential timing such as the RDSP with the two-way fixed effects model may lead to biased findings due to early-late treated comparisons and heterogeneous treatment effects. Thus, we use the approach proposed by Callaway and Sant'Anna (2021), whose identifying assumptions fit with the RDSP setup, and allow us to aggregate the treatment effects by treatment cohorts and over time.

Our most conservative estimates, based on the full period under treatment for the first four cohorts, indicate that the RDSP achieved its objective to improve retention among treated Trusts. Overall, the programme improved the stability rates of nurses and midwives by 0.8ppt on average. This effect appears small, but is a significant improvement: the average stability rate at the start of the programme in June 2017 was 85% and the estimated impact amounts to an increase of almost a quarter of the between-Trust standard deviation in nursing retention. In terms of staffing numbers, the RDSP led to the retention of 1,697 nurses and midwives who would have otherwise left their Trusts; this is roughly equal to the average number of nursing staff in a Trust on the month before the RDSP was launched.

Perhaps not surprisingly, we find heterogeneous treatment effects across cohorts. The Trusts in Cohort 1 improved their nursing retention on average by 0.95ppt over the 28 months, while Trusts in Cohort 4 had similar returns in magnitude only after 12 months into the RDSP. There is some positive, but limited, impact of the programme in reducing exits from the NHS overall. We estimate that the NHS leaver rates for treated Trusts dropped on average by 0.41ppt, which is a one-fifth of standard deviation reduction in pre-RDSP NHS leaver rates. Moreover, these estimates might be unnecessarily conservative due to the nature of our retention outcomes, computed over 12 months. In fact, when we focus our attention to the post-treatment period beginning 12 months after the RDSP enrollment, we find even larger retention gains in terms of both stability rates and drop in NHS exits.

Overall, we find weak (insignificant) evidence that the RDSP improved patient outcomes and hospital productivity. We certainly find no indication that RDSP was detrimental to these outcomes by diverting efforts away from patient care. The limited gains in quality and productivity are largely to be expected for several reasons. First, they were outside the direct scope of the intervention. Second, patient's quality and admissions are also influenced by capacity constraints on other inputs, such as hospital doctors' capacity and the availability of hospital resources (e.g. operating theatres), which are respectively labour and capital inputs complementary to nurses' labour but not targeted by the RDSP intervention. Third, and related to the point above, the effects of a change in retention may be asymmetric, depending on whether retention increases or decreases. The NHS is considered a cost-efficient health-care system with respect to comparable systems in other countries (Dayan et al., 2018), yet it has been critically and systematically understaffed for about a decade (Fund, 2018). If hospital wards always operate close to capacity, it is likely that a fall in staff retention generates a short-term staff shortage and overburdens already proved clinicians, thus leading to worse patient health outcomes; whereas an increase in staff retention, as the one brought by the RDSP, might simply avoid the overburdening but also not being enough to raise quality, which could be achieved only by substantially increasing staff numbers.

The RDSP did not adopt a "one-size fits all" approach to supporting Trusts, instead it helped NHS hospital providers identifying the factors they could work on to develop action plans to reduce turnover rates in 12 months following the start of the programme. The monitoring body, NHS Improvement, identified ten recurring themes in action plans, which mostly align with the reasons for leaving nurses and midwives. By analysing these broader themes to understand what contributed to RDSP's *success*, we find that intervention plans that focused on increased employer flexibility improved stability rates in all cohorts, whereas other themes had heterogeneous impact on nursing retention across cohorts.

One may wish to understand whether the programme was also cost-effective. NHSI estimates that it costs £11,000 to replace a nurse (NHS Improvement, 2018). This implies that the programme saved £18,546,000 from the NHS budget. It is hard to get an estimate of the costs 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 Trusts and at NHSI. We know little about the amount of staff time spent and even less about its opportunity cost, although we have indirect evidence that both might have been contained given that we find no evidence of diverted efforts in terms of patient quality of care and number of admissions. As such, it seems likely that the value of staff time spent per Trust was less than £140,000 per Trust on average, implying a probable cost-effectiveness of RDSP.

The RDSP is a light-touch intervention. While our analyses show that there are modest returns from this local programme with respect to nursing retention in the NHS, these gains are not negligible. Given the nature of the intervention, it appears that it succeeded by filling the 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. Two obvious implications arise from these findings. At a local level, interventions such as the RDSP are not sufficient to reverse the scale of a nursing workforce crisis such as the one faced by the NHS, but such initiatives are sustainable and can complement other policies designed to alleviate workforce pressures. At a more general level, our results are informative about the trade-off between centralization and decentralization in the management of organizations

providing public services and goods (Marschak, 1959; Sah and Stiglitz, 1991; Alonso et al., 2008). It appears that preserving a certain level of centralization, at the very least in terms of disseminating information and providing guidance on best practices, may help decentralized units to overcome information asymmetries and the consequent economic inefficiencies that these organizations would be unable to address with the same success, if left on their own. This is partially in contrast with the paradigm that fully decentralized systems are usually more efficient (Alonso et al., 2008; Besley and Coate, 2003; Dewatripont and Maskin, 1995), although it is not unlikely since the case we consider does not require coordination of efforts across the decentralized units as in Alonso et al. (2008). Finally, our findings suggests that an effective configuration of the delivery of (public) services may be achieved through an organizational structure with a convex combination of centralized and decentralized decision-making units with distinct functions and roles, e.g. one characterized by local providers of services (the "agents") subject to the constant exchange of information with and the evaluation of performance by an active centralized monitoring body (the "principal").

References

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

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.

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.

Bandiera, O., Barankay, I. and Rasul, I. (2009), 'Social connections and incentives in the workplace: Evidence from personnel data', *Econometrica* **77**(4), 1047–1094.

Beckmann, M., Cornelissen, T. and Kräkel, M. (2017), 'Self-managed working time and employee effort: Theory and evidence', *Journal of Economic Behavior & Organization* **133**, 285–302.

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., Genakos, C., Sadun, R. and Van Reenen, J. (2012), 'Management practices across firms and countries', *Academy of Management Perspectives* **26**(1), 12–33.

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`.

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.

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. and Allegretti, A. (2021), 'Threat of nurses' strike over 3% pay award for NHS staff', *The Guardian* . 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.

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.

Chakelian, A. (2021), 'Exclusive polling: Majority say NHS nurses and doctors are paid too little', *The New Statesman UK Edition* . 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.

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.

Dayan, M., Ward, D., Gardner, T. and Kelly, E. (2018), *How good is the NHS?.*, King's Fund.

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.

Fund, K. (2018), 'The health care workforce in england: make or break?'.

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.

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`.

Heyes, A. (2005), 'The economics of vocation or "why is a badly paid nurse a good nurse"?', *Journal of Health Economics* **24**(3), 561–569.

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`.

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

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.

McCarthy, N. (2015), *Forbes* . Published online on June 23, 2015 at https://www.forbes.com/sites/niallmccarthy/2015/06/23/the-worlds-biggest-employers-infographic/?sh=1a5b9711686b.

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/.

Moscelli, G., Sayli, M. and Mello, M. (2022), Staff Engagement, Job Complementarity and Labour Supply: Evidence from the English NHS Hospital Workforce, IZA Discussion Papers 15126, Institute of Labor Economics (IZA).

National Audit Office (2020), 'The NHS nursing workforce', *HC 109 Session 2019-2021. Department of Health & Social Care* . https://www.nao.org.uk/report/nhs-nursing-workforce/.

Nelson, N. and McLaughlin, C. (2020), 'NHS faces fresh crisis as 250,000 nurses consider quitting because of low pay', *Mirror* . 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.

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 Digital (2022), Summary Hospital level Mortality Indicator (SHMI): Indicator specification, Technical report, Population Health, Clinical Audit and Specialist Care, NHS Digital. https://files.digital.nhs.uk/40/C1E273/SHMI%20specification%20v1.45.pdf.

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.

Nursing and Midwifery Council (2017), 'The NMC register 2012/13 - 2016/17'. Accessed on 07/09/2021 at https://www.nmc.org.uk/globalassets/sitedocuments/other-publications/nmc-register-2013-2017.pdf.

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.

Reuben, A. and Eddington, T. (2022), *BBC News* . Published online on April 6, 2022 at https://www.bbc.co.uk/news/uk-politics-58436009.

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.

Shields, M. A. and Ward, M. (2001), 'Improving nurse retention in the National Health Service in England: the impact of job satisfaction on intentions to quit', *Journal of Health Economics* **20**(5), 677–701.

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.

Swinford, S., Smyth, C., Dathan, M. and Woolcock, N. (2021), *The Times* . Published online on July 22, 2021 at https://www.thetimes.co.uk/article/nhs-staffs-3-pay-rise-likely-to-come-out-of-social-care-tax-0w226n8fw.

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.

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: Summary statistics for selected characteristics, pre-RDSP (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, %	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†	2.424	1.772	2.036	1.387	1.964	1.558	2.500	1.351	4.176	8.610
Monthly SHMI, elective patients†	3.159	9.905	0.317	0.443	0.699	0.991	0.569	0.691	0.573	0.883
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
Bullied by manager/coworkers at work in the last 12 months	28.210	6.910	27.394	5.074	26.303	5.498	27.503	6.441	26.443	5.948
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 Trusts only, thus the sample sizes for each cohort is smaller than for other summary statistics.

Table A2: Stability rates: never-treated vs. not-yet-treated comparison groups

	Stability rate		
	(I)	**(II)**	**(III)**
Control group	Never-treated	Never-treated	Not-yet-treated
Post-treatment until	November 2019	August 2019	August 2019
overall ATT	0.775	0.656	0.677
	(0.188)***	(0.180)***	(0.176)***
	[0.196]§	[0.182]§	[0.174]§
partially aggregated			
Cohort 1	0.950	0.851	0.971
	(0.352)***	(0.340)**	(0.324)***
	[0.360]§	[0.329]§	[0.349]§
Cohort 2	0.677	0.579	0.589
	(0.303)**	(0.289)***	(0.278)**
	[0.309]	[0.273]	[0.280]
Cohort 3	0.557	0.424	0.378
	(0.336)*	(0.327)	(0.319)
	[0.333]	[0.351]	[0.323]
Cohort 4	0.912	0.773	0.773
	(0.267)***	(0.258)***	(0.258)***
	[0.269]§	[0.255]§	[0.266]§
pre-trend Wald test (df = 48) p-value	0.080	0.080	0.236

Notes: Aggregated treatment effect parameters under the Unconditional DID Assumption of CSA with never-treated Trusts 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 A3: NHS Leaver rates: never-treated vs. not-yet-treated comparison groups

	Leaving the NHS rate		
	(I)	**(II)**	**(III)**
Control group	Never-treated	Never-treated	Not-yet-treated
Post-treatment ends	October 2019	August 2019	August 2019
overall ATT	-0.408	-0.361	-0.371
	(0.131)***	(0.126)***	(0.125)***
	[0.128]§	[0.125]§	[0.120]§
partially aggregated			
Cohort 1	-0.498	-0.454	-0.481
	(0.254)**	(0.251)*	(0.244)**
	[0.233]	[0.261]	[0.262]
Cohort 2	-0.416	-0.388	-0.394
	(0.202)**	(0.196)**	(0.189)**
	[0.209]	[0.199]	[0.193]
Cohort 3	-0.394	-0.348	-0.358
	-0.248	(0.242)	(0.234)
	[0.250]	[0.241]	[0.234]
Cohort 4	-0.341	-0.274	-0.274
	(0.183)*	(0.169)	(0.169)
	[0.181]	[0.166]	[0.166]
pre-trend Wald test (df = 48) p-value	0.135	0.135	0.083

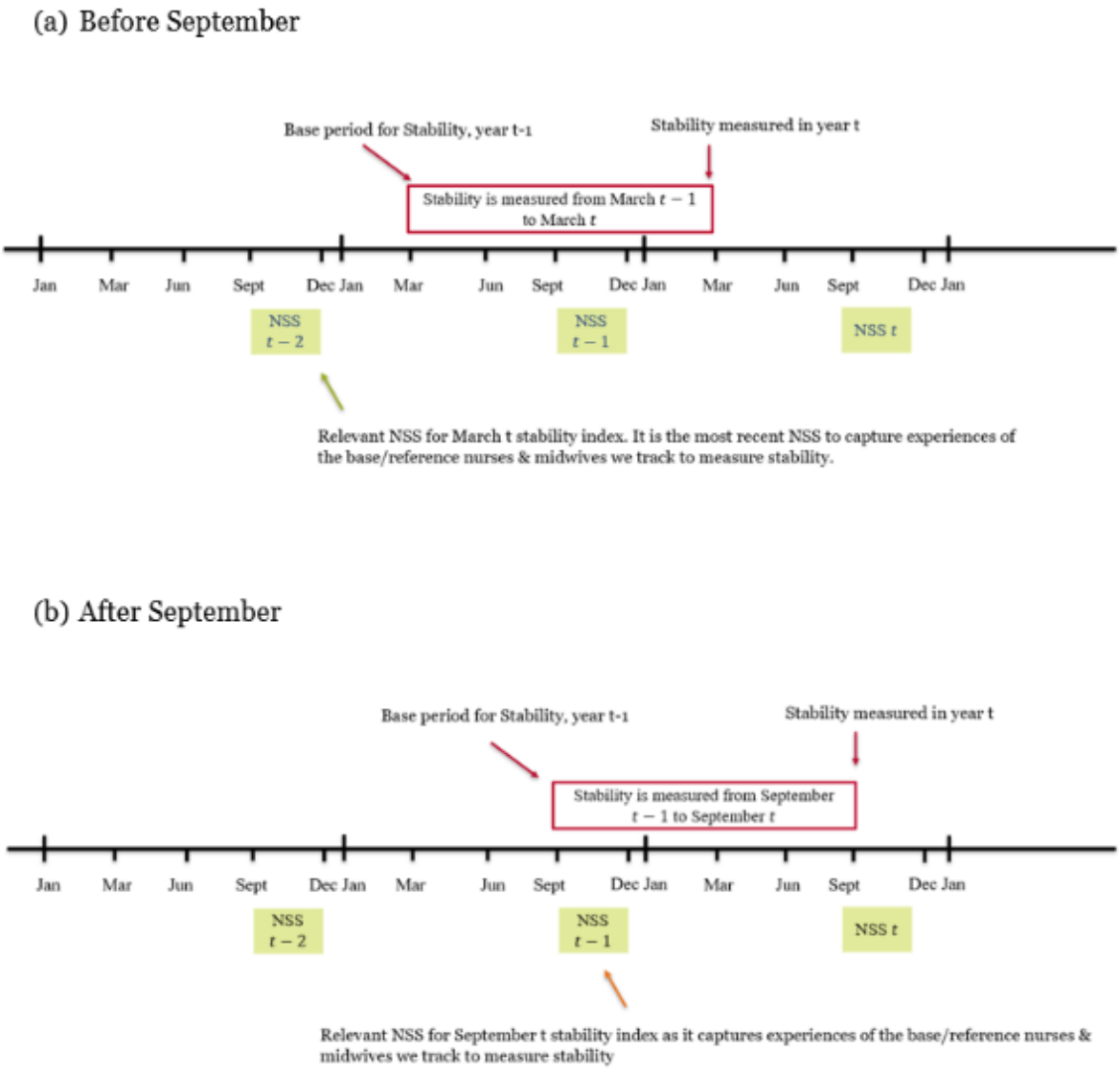
Notes: Aggregated treatment effect parameters under the Unconditional DID Assumption of CSA with never-treated Trusts as control group. Asymptotic standard errors are in parentheses, and estimated using csdid package in Stata (Rios-Avila et al., 2021), p-values *p<0.1; **p<0.05; ***p<0.01. Clustered 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: Retention Improvement Action Plans by RDSP 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: The figures in the table are from a sub-sample of Trusts from the first four cohorts with submitted action plans. Excluded are 15 Trusts did not submit or submitted their action plans later than the deadline set bu the NHSI: 4 in Cohort 1, 8 in Cohort 4 and one Trust in Cohorts 2 and 3. Also, 6 Trusts were not recorded in NHSI's action plan data (83.3% were in Cohort 4).

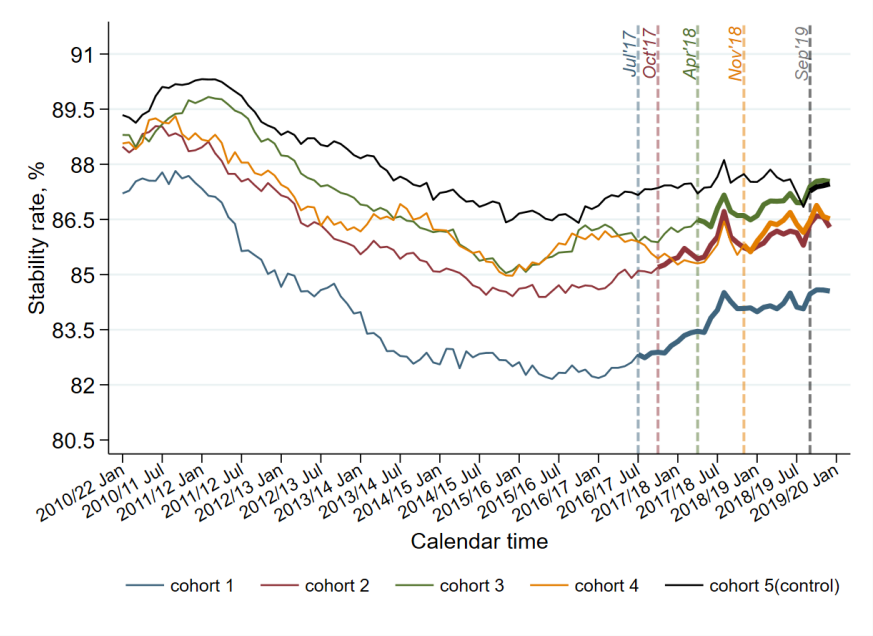
Figure A1: Data setup



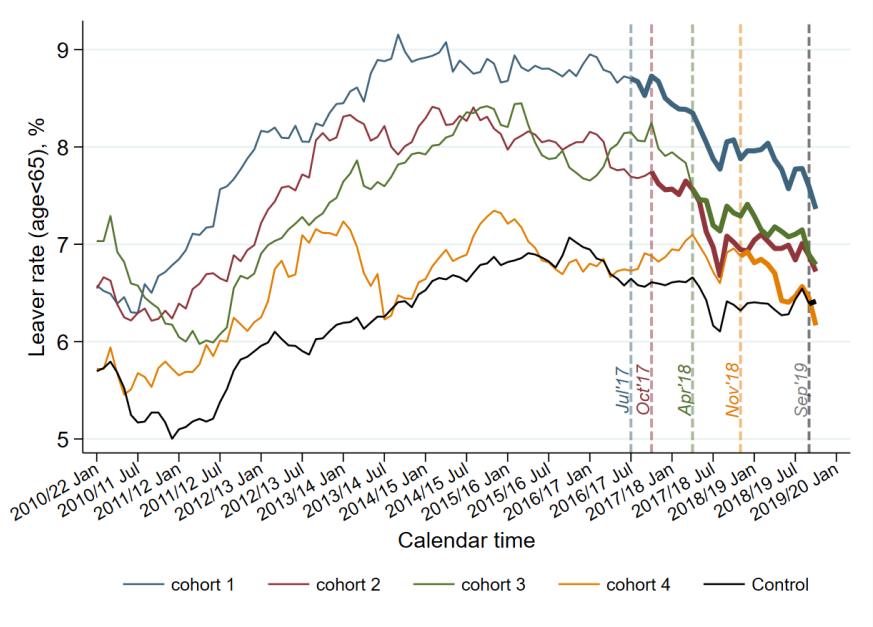
Notes: The same holds for leaver rates. t refers to the analysis year, $t-1$ is the base year. NSS refers to the NHS Staff Survey which is conducted every year in autumn since 2003. Staff working in Trusts in 1st September are eligible to respond to the NSS. The NSS runs from the mid-September and remains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



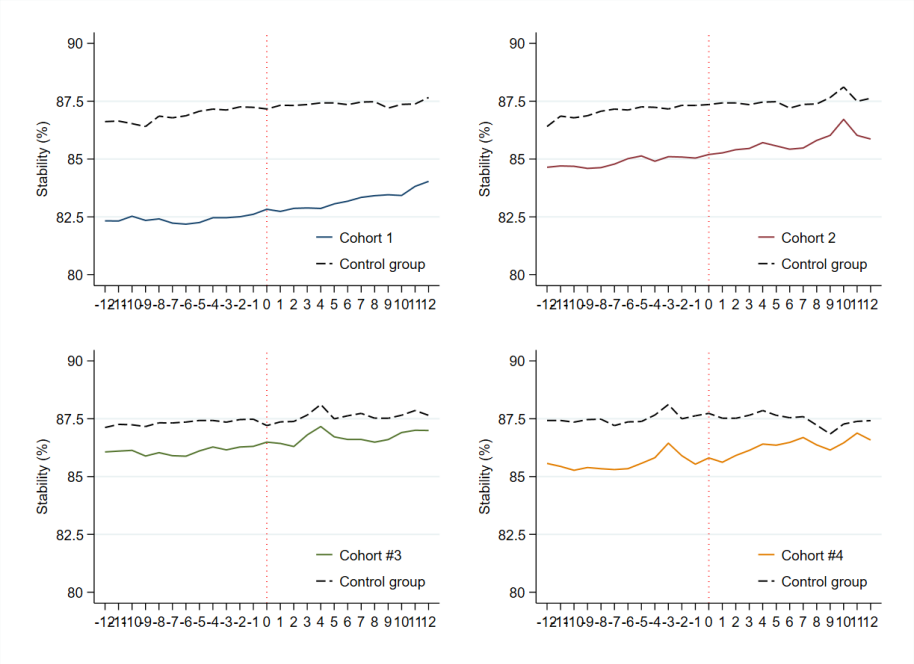
(b) NHS Leaver rates from 2010/11 to 2019/20



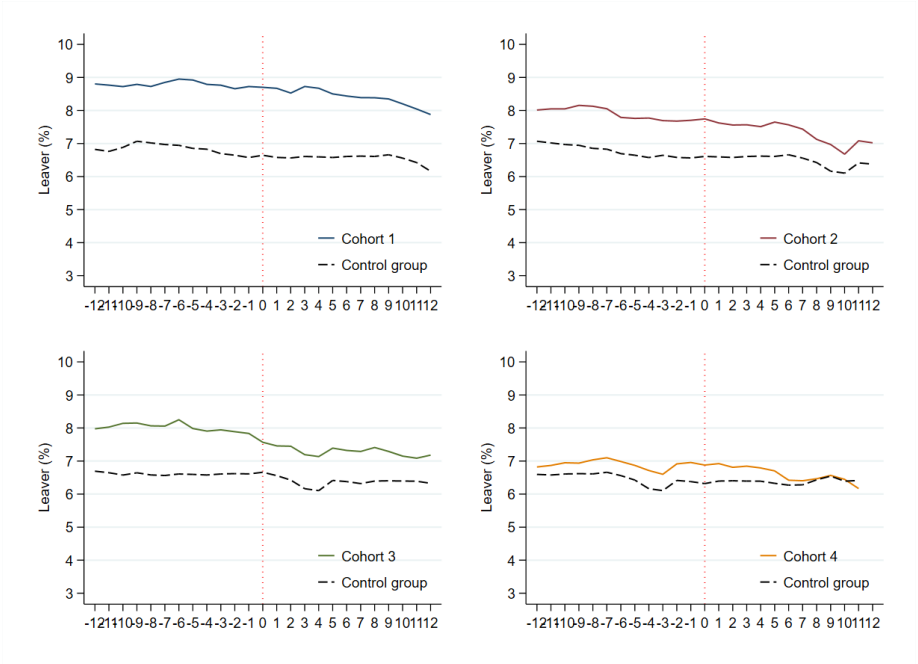
Notes: Cohort 5 includes 2 additional trusts that were not in the NHSI allocation. Vertical lines show the RDSP start dates for each cohort. The thicker lines indicate the time cohort spends in RDSP

Figure A3: Common trends between treated and control cohorts

(a) Stability rate

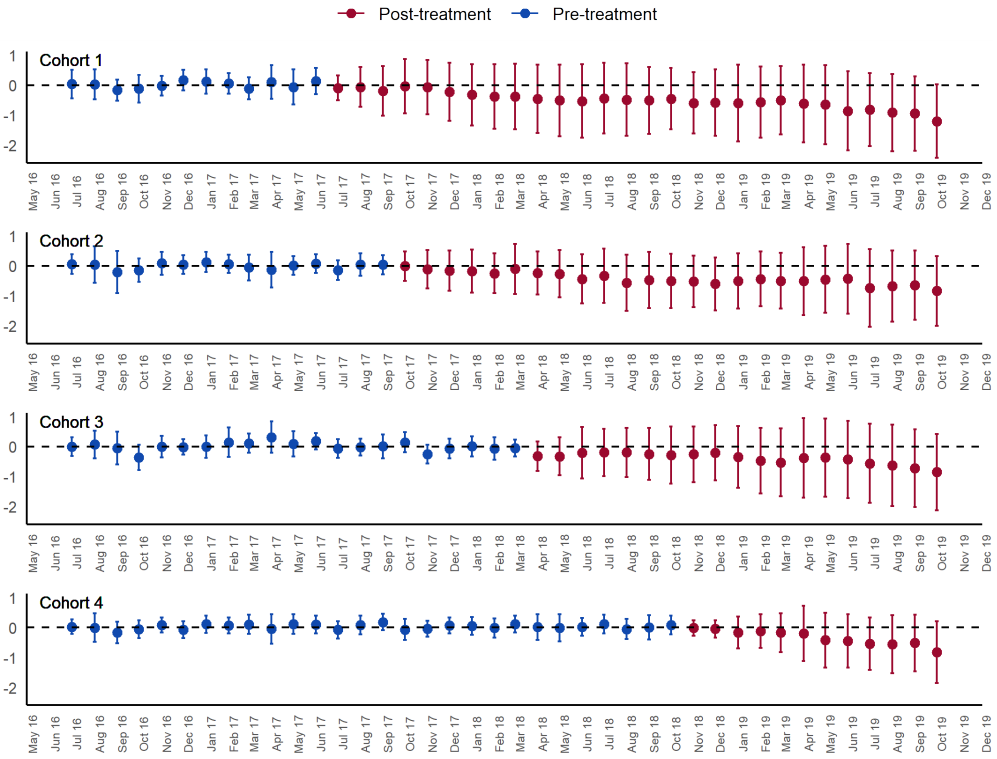


(b) NHS-leaver rate



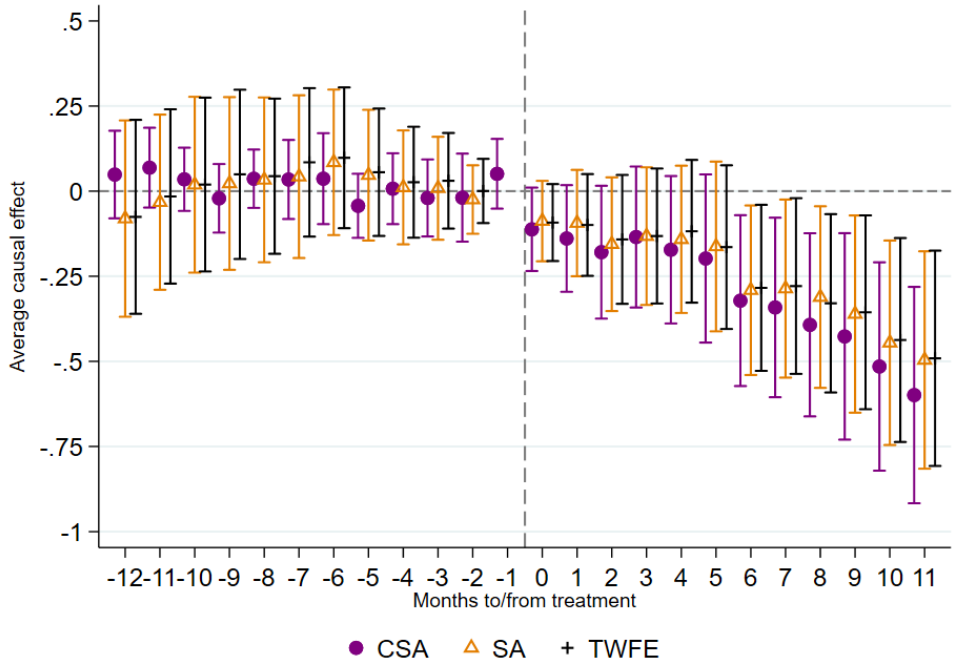
Notes: Figures are centred at the time RDSP was launched in Cohorts, Trusts, and are balanced for relative time periods. The vertical dashed line indicates the timing of the RDSP, and the figures show 12 months before and 12 months after the RDSP.

Figure A4: RDSP cohort-time average treatment effects on NHS leaver rate



Notes: The effect of RDSP on nurses and midwives' leaving the NHS rates under the unconditional parallel trends assumption. The point estimates are shown with simultaneous 95% confidence bands from bootstrapped standard errors allowing for clustering at Trust level.

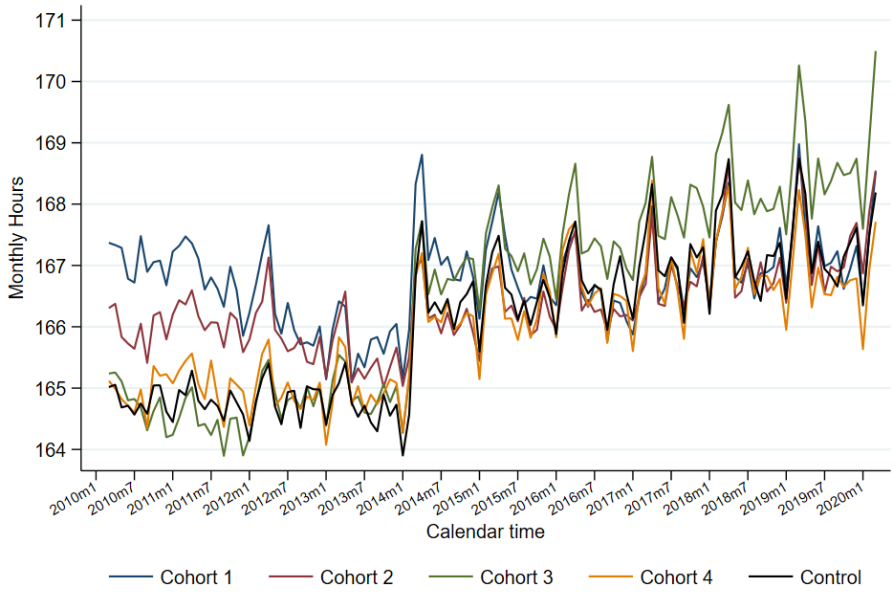
Figure A5: Dynamic treatment effects on leaving the NHS rate, alternative estimators



Notes: Event-study estimates under unconditional parallel trend assumption. To facilitate comparisons across different estimators, CSA estimates are illustrated with (asymptotic normal) clustered standard errors with 95% confidence, estimated using csdid command in Stata (Rios-Avila et al., 2021). SA are estimated using the eventstudyinteract in Stata (Sun, 2021). In Sun and Abraham (2021) and TWFE specifications, the month before the RDSP, $k = -1$, is omitted to avoid multicollinearity; and periods $k < -12$ and $k > 12$ are binned instead of trimming the sample.

Figure A6: 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

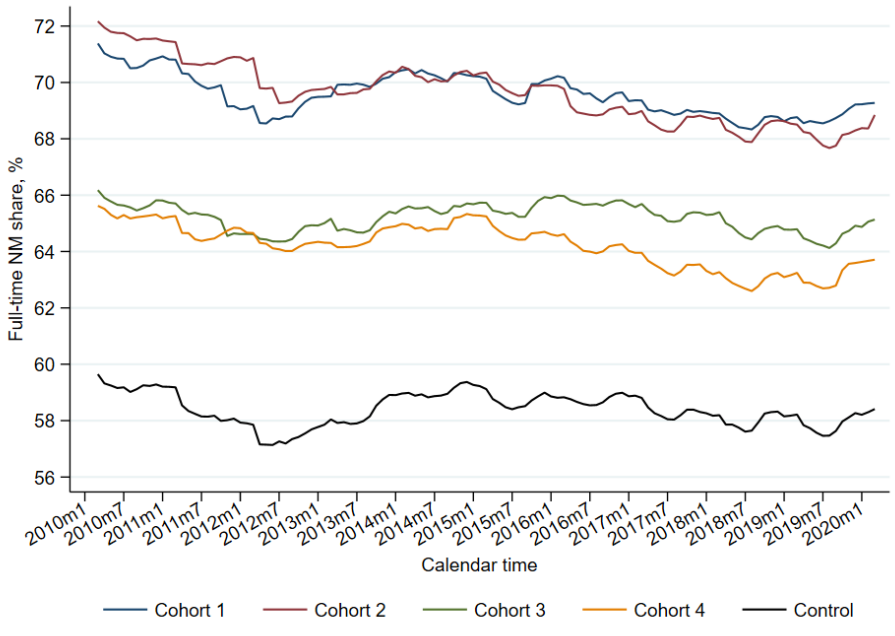
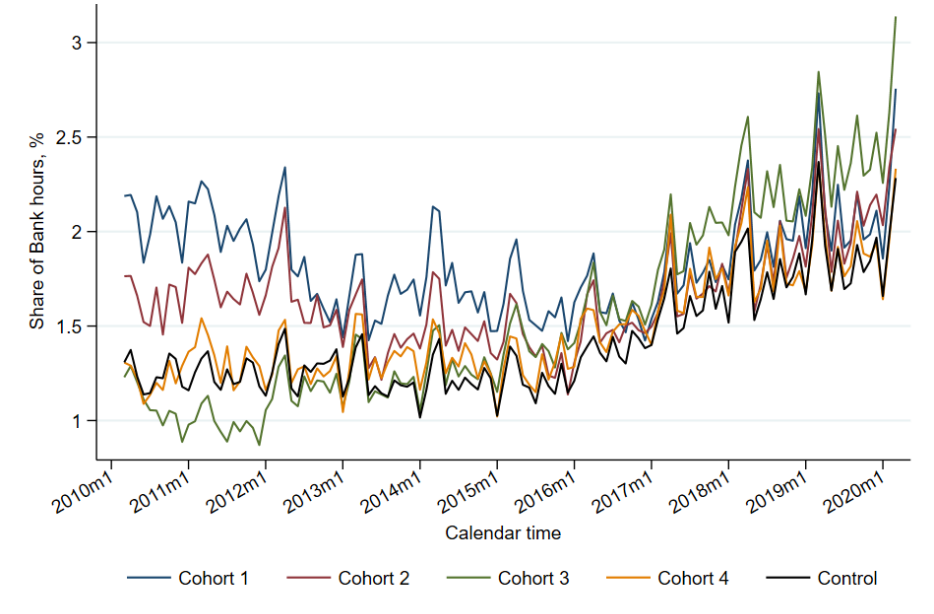
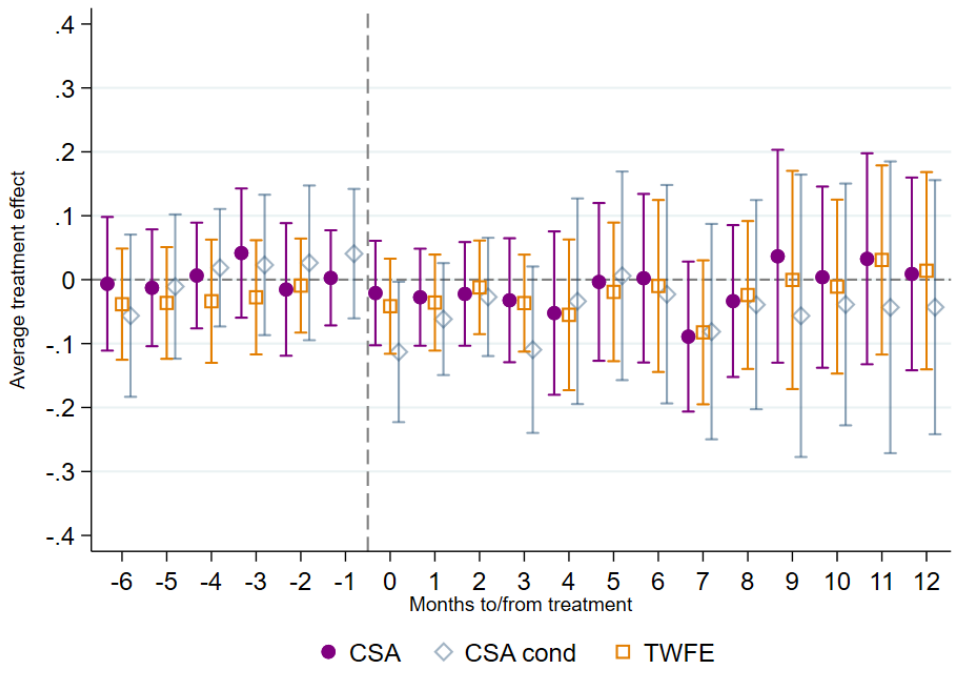


Figure A7: Share of Bank Hours within monthly hours worked by nursing staff



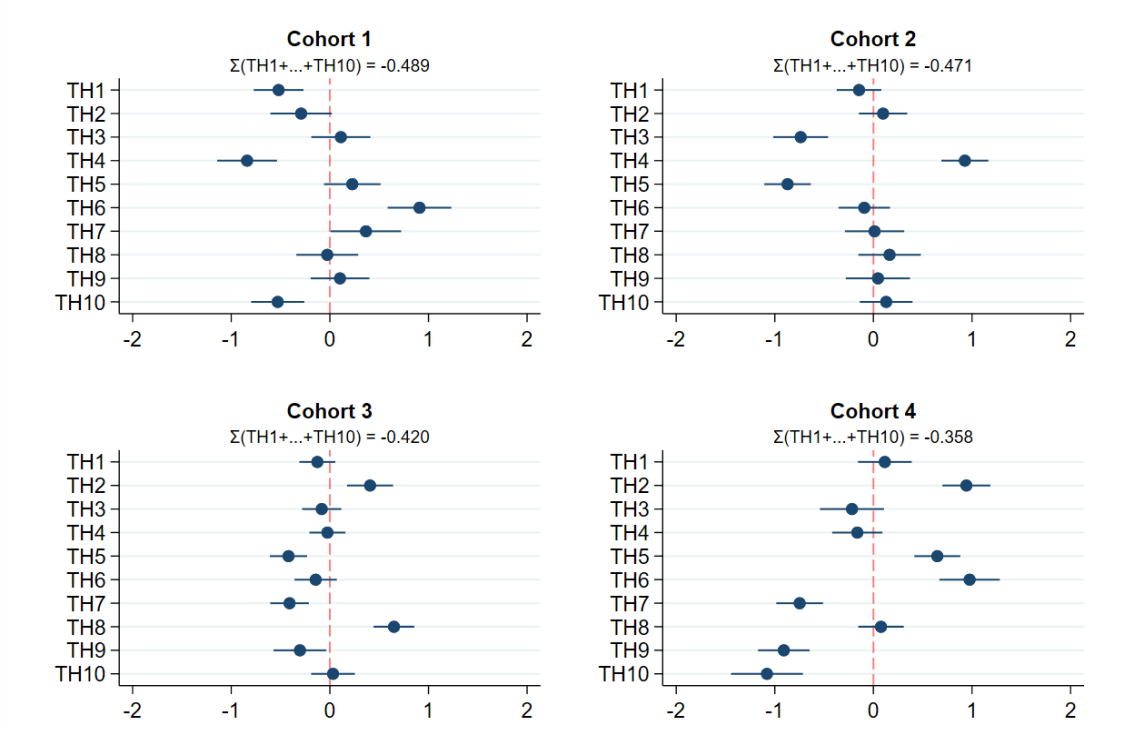
Notes: Average hours excludes negative and zero hours and includes both full-time and part-time working nurses and midwives.

Figure A8: 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 Trust 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 A9: The Effects of Themes on NHS-leaver rates for Cohorts 1-4



Notes: Results from estimating the constrained regression equation 3 using Trusts in Cohorts 1-4 with 95% confidence intervals from robust standard errors. The sum of theme parameters equal to cohort-specific ATTs from the model estimated using treated Trusts that submitted an action plan with themes, and aggregated in the same fashion as in Table 3. For details on the themes within Cohorts, see Table A4.

Appendix B

Heterogeneity analysis based on different sub-samples

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

Our 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 presents similar patterns to those of column 2 in Table 3 indicating that the impact of the RDSP does not work primarily through the prevention of retirements.

The main retention measures also include staff who are employed as Bank staff in the hospital Trusts. 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. Bank work is very common among NHS nurses and midwives with on average 16% of nurses and midwives registered as Bank in month in 2016. Bank staff may come from variety of sources, and we distinguish between the nursing staff from hospital Trust's existing employees (in-house Bank) and from an outside organization who are only contracted as Bank workers in the Trust (only Bank).

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

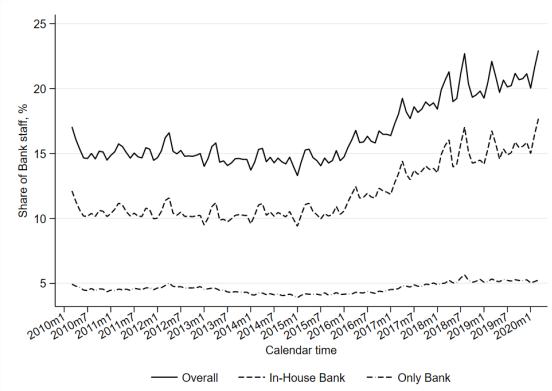
	Stability rates				NHS-leaver rates		
	Age restriction	**No Bank**	**Permanent**	**Acute Trusts**	**No Bank**	**Permanent**	**Acute Trusts**
Overall ATT	0.798	0.839	0.789	0.868	-0.380	-0.356	-0.415
	(0.196)***	(0.201)***	(0.199)***	(0.200)***	(0.137)***	(0.137)***	(0.136)***
	[0.208]§	[0.201]§	[0.195]§	[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 3. 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.

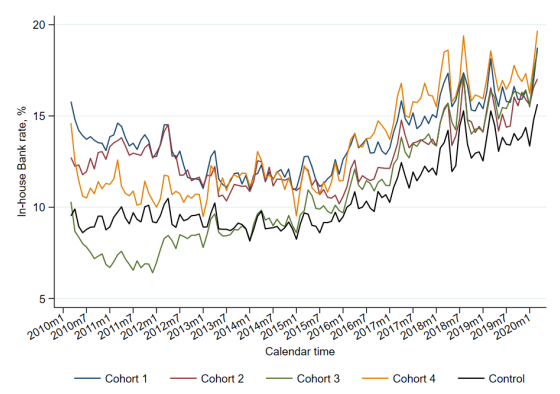
Starting from 2016, the share of nurses and midwives who registered as Bank has increased across all cohorts, some more than others, and the increase in the Bank workforce was driven by the in-House Bank registrations for all cohorts as shown in Figures B1. While the extent of the use of bank staff may 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 and Only Bank nurses and midwives



(b) Share of in-House Bank work, by Cohort



To understand whether our main estimate is affected by the Bank workforce, we re-compute our retention outcomes by excluding the nurses and midwives who are only on a Bank contract with the Trust. We keep the in-house Bank staff, as they are employed within the Trust under other tasks, so appear in nursing staff headcounts after their Bank shifts end. However, only-Bank nurses and midwives leave the Trust once their period of temporary employment, which can be as short as week, terminates. This temporary nature of only-Bank work generates lower (higher) stability rates (NHS-leaver rates), moreover, any RDSP intervention would not directly apply to this transient workforce. 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 only-Bank 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. By focusing on nursing staff on permanent contracts only, we eliminate the anticipated leaves due to fixed-term or temporary contracts, although we understand the implications of the RDSP on the total human capital in the trust.

The results are similar to the previous analysis as the majority (97%) of the nurses and midwives were on a permanent contract. The overall average impact of the RDSP on nurses and midwives' stability is slightly lower, and this seems to be driven by the lower ATT from Cohort 4. In terms of NHS-leaver rates, the estimated impact of the RDSP is smaller, but it equals to 18.9% of the standard deviation in nurses and midwives' leavers rates in treated Trusts in the month before the RDSP was launched.

Lastly, we focus on nurses and midwives who work in Acute-care NHS Trusts in England. The type of care provided in Acute-care hospitals are different than Mental Health Trusts, and we would expect different action plans arising between these two types of Trusts. 91% of the Mental Health Trusts are treated in the first three cohorts, and the majority of control group consists of Acute-care Trusts, we may therefore have some doubts about whether there is common support across treatment and control groups for the Mental Health Trusts. Our main models are robust to restricting the sample to Acute Trusts. We find that overall the RDSP improved retention by 0.87ppt on average for all treated Trusts, and reduced the leaver rates by 0.42ppt. We fail to reject the null hypothesis that there are no pre-trend in NHS-leaver rates among Acute trusts in the 12-months leading to the RDSP, but we fail to reject the null closer to programme's start.