Friday Morning Fever: Evidence from a Randomized Experiment on Sick Leave Monitoring in the Public Sector.*

Tito Boeri† Edoardo Di Porto‡ Paolo Naticchioni§ Vincenzo Scrutinio¶

Abstract

This paper offers the first population-wide controlled field experiment for home visits (HV) checking on sick leave in the Italian public sector, where absenteeism is widespread. We exploit unique administrative data on sick leaves and monitoring. Receiving a HV reduces the number of days on sick leave in the following 16 months by about 8%. Opportunism deterrence can rationalize this result, as HVs induce workers to update their perceptions of the audit risk and of the sanctions for irregular leave in presence of a double lottery (over inspections and sanctions). HVs are cost-effective: every Euro spent for the visits involves at least 6 Euros reductions in sick benefit outlays. The marginal value of public funds (MVPF) is 1.2, significantly lower than estimates of MVPF of income taxes in the US. Cost-effectiveness increases when HV target workers at the top of the earning distribution.

Keywords: Sick leave, absenteeism, randomized trial.

JEL classification: I12,J45

*We thank Thomas Le Barbanchon, Pietro Biroli, David Card, David Yanagizawa-Drott, Christian Dustmann, Nathan Hendren, Andrea Ichino, Maarten Lindeboom, John List, Francesco Manaresi, Hitoshi Shigeoka, Johannes Spinnewijn, Marco Pagano, Nicola Persico, Steve Pischke, Jeffrey M. Wooldridge, participants at the AIEL, AEHSC, Brucchi Luchino, BGSE Summer forum in Public Economics (2023), EALE-SOLE, and RES annual conferences, and seminar participants at the IEB, Bocconi University, Bologna University, CSEF, GSSI, Human Capital and Labor Market Institutions Workshop, ISPRA, Padua University, and UCFS Workshop. We are grateful to the "Coordinamento medico legale" of the Italian Social Security Administration (INPS) and all the staff of the INPS Research Department (DCSR) for their invaluable help and support. The findings and conclusions expressed are solely those of the authors and do not represent the views of INPS.
Corresponding author: Tito Boeri; Università Bocconi; Via Roentgen, 1; 20136 Milano; Italy; Email: tito.boeri@unibocconi.it

†Bocconi University, IZA, CEP-LSE

‡CSEF, INPS, University of Naples Federico II, UCFS, Uppsala University

§INPS, University of Roma Tre, IZA

¶LSE, IZA, University of Bologna

1 Introduction

Paid sick leave is a key institution protecting workers' health and providing income smoothing when they are hit by a negative health shock. The scope of sickness benefits has been considerably enhanced during the pandemic to reduce "presenteeism" and the risk of contagion: these schemes absorb more than 1% of GDP in most countries, and OECD statistics indicate that the number of compensated absences from work due to illness has substantially increased after the pandemic. Currently 18 US states (and D.C.) have mandated paid sick leave, but many others are considering introducing similar legislation, and there is support for such extensions at the federal level: the Healthy Families Act introduced in 2023 proposes to guarantee up to 7 paid sick days nationally.

As for other social insurance policies, the design of sickness benefits poses problems of moral hazard, notably in the employment protected public sector. According to the European Working Conditions Survey (EWCS), public employees stay at home for illness reasons over a one-year span about one and a half day more than their counterparts in the private sector (a 24% increase); the difference is even more marked in Mediterranean countries where public sector workers are absent for health reasons 64% more frequently than private employees. The higher incidence of sick leave in the public sector cannot be explained by differences in the characteristics of the two employment pools.[1]

Governments typically deal with moral hazard problems by adjusting the level and coverage of sickness benefits. Most reforms of sickness benefits in the last decade have been taking this route. However, the fine-tuning is difficult, especially in the presence of a highly heterogeneous population of potential beneficiaries. Too high benefits would induce over-reporting of sickness, reducing the quantity and quality of key services offered to citizens, while too low sick pay would induce presenteeism and increase the risk of contagion. Effective enforcement can be a better tool to deal with moral hazard without encouraging presenteeism than cuts in the level and duration of benefits, insofar as sickness

[1] It is true that public sector workers are slightly older than in the private sector (they are aged 43 in the public sector vs. 40 in the private sector), but they are also better educated and work fewer hours. Moreover, the incidence of sick leave is not higher in the health sector than in other public administrations, although in the health sector there could be a higher exposure to epidemiological risk.

leave monitoring involves additional checks of the actual health status of the workers involved.

While there is some literature on the design of sickness benefits, much less is known about the effectiveness of controls over their potential misuse. This paper aims at filling this gap by drawing on the first population-wide randomized experiment with home visits assessing the validity of contemporaneous sick leave certifications in the public sector. We investigate the effects of Home Visits (HV) to public employees on sick leave by doctors-inspectors working for the Italian social security administration (INPS).

Italy is an ideal case study, as there is a national administration enforcing the regulations with the same procedures over the entire country. Absenteeism in the public sector is rather widespread: one public employee out of four benefits from sick leave at least once a year, compared with one out of ten in the private sector. There are wide differences across regions in the take-up of sickness benefits, which cannot be justified in terms of composition of the workforce. There is also a large body of anecdotal evidence pointing to opportunistic behavior of public employees. For instance, on 2014 New Years's eve, 764 local police officers out of 900 claimed sick leave in Rome.[2] Data from the sickness benefit registers show that sick leave certificates are far from being uniformly distributed over week days as one would expect based on epidemiological factors. Actually, Figure 1 points to two visible peaks in the distribution, respectively on Mondays and on Fridays, de facto extending the weekend.

The generosity of sickness benefits offered to public sector workers in Italy is comparable to many other OECD countries. Austria and Germany, for instance, provide the same replacement rate and maximum benefit duration (Nelson et al., 2020). Hence, our results can be quite informative as to the effects of introducing similar monitoring mechanisms in other countries. The international scalability of our experiment is further enhanced by the fact that we test procedures that can be readily replicated by public administrations in different institutional contexts. A number of countries envisage home visits under the

[2]Click here for the link to the article in the press.

request of the employer or under the autonomous initiative of the social security agency.

We planned, performed, and analyzed a population-wide randomized control trial, run between November 22, 2017 and January 5, 2018. The national scale of the experiment implies that we can take into account of potential spillovers of the policy being investigated. The experiment allotted randomly HVs to ongoing medical certificates associated with sick leave spells while leaving unchanged all the procedures related to the organization of HVs, notably the identification of eligible certificates, the assignment of the visits to doctors, and the implementation of the visits. The use of standard procedures for the implementation of the experiment rules out the possibility that doctors or workers would react differently to the visits performed in the period of the experiment and improves the external validity of our results by excluding *Hawtorne effects*. Doctors belonging to the social security performed about 4,200 visits in the 45 days of the experiment. Our analysis is based on a unique, and so-far unexploited, dataset covering sick leaves, home visits and employment histories of Italian public sector employees.

We find that irregular leaves[3] are far from being distributed uniformly across public sector employment. Workers most likely to be found irregularly on leave at a HV are males, live in Southern regions and are more likely to be found irregularly on leave if the visit takes place on Friday.

We compare access to sick leave between individuals who received a HV during the period of the experiment and workers who did not. We find that being audited (irrespective of the outcomes of the visit) leads to a reduction in sick leave claims: workers who are audited spend about 3.6 days less on sick leave in the 16 months following the experiment, a reduction of 8% with respect to the 43.3 days on sick leave of the average worker in the control group. The reduction comes mostly from the intensive margin (duration of certificates) rather than the extensive margin (number of certificates).

We also provide suggestive evidence that the reduction in absences from work in the 16 months after a HV does not come at the cost of presenteeism. Although lack of data

[3]A worker is considered irregularly on leave if not found at home by the doctor at the time of the inspection or if considered to be fit for duty. This is discussed in greater detail in Section 2.

on coworkers and on the nature of the disease does not allow us to directly investigate the risk of contagion, we find that workers subject to HVs do not show a higher rate of relapses than the other workers. Were they coming back to work too early, they would be more likely to redevelop the same disease, and possibly be contagious for their colleagues.

Workers found irregularly absent appear to react more to a HV than workers whose GP certificate is confirmed at the visit (reducing their sick leave claims by 8 days relative to two for those found regularly on leave).

Sanctions for abuse are not automatic, but we find that workers reported to their superiors to be irregularly absent from work experience a cumulative reduction in wages close to 1.4% of their take home pay in the 16 months following the HV. This can be an implicit sanction. Notice that the actual punishment for detected non-compliance is a key policy parameter in the standard Allingham–Sandmo deterrence model (Allingham and Sandmo, 1972), but to date research has overlooked the extent of punishment (Slemrod, 2019).

Our results are robust to a large set of checks and alternative specifications, including difference in differences estimates, test of the potential bias due to selection over unobservables, as well as instrumental variables techniques dealing with it.

Finally, our findings can also be relevant in better understanding incentives in different organizational structures. By combining different data sources, we provide a classification of *street level bureaucrats* (Lipsky, 1980), that is, public employees highly exposed to interactions with the general public. These workers are not underrepresented in the pool of workers being found irregularly on leave. This does not seem to confirm the view that direct exposure to the consumers of public services operates as a disciplining device.

A simple cost benefit analysis from the standpoint of the social security administration shows that audits are highly cost-effective: we estimate that one Euro spent on HVs implies a 6.2 Euros reduction in future benefit outlays under randomized visits and 7.6 Euros if visits are targeted to workers more likely to be irregularly on leave as done in the private sector. This implies net savings of 5.2 and 6.6 Euros, respectively, for the two kinds of HVs allocation systems. These savings originate from the estimated deterrence

effects, i.e., reductions in future sick leave claims relative to the control group. Our results indicate that even audits to a worker taking a one-day sick leave, who would go back to work the day after, irrespective of the visit, could be cost-effective. Notice that expenditure savings do not seem to be eroded by program substitution: workers inspected do not have access to disability benefits or early retirement schemes more than workers in the control group.

As suggested by Slemrod (2019), it is important to characterize the distributional implications of compliance and enforcement policies. This is even more the case in our setting because income and health are strictly and positively correlated. We find that responses to HV are rather stable along the earning distribution. However, net savings originated by HVs are larger when high earners are targeted, since replacement rates offered by sickness benefits are constant along the earning distribution.

In order to assess the welfare effects of audits, we estimate the marginal value of public funds (MVPF) of HVs (Finkelstein and Hendren 2020). The MVPF is about 1.2, and this implies that HVs are more efficient than the top income taxes in the US (having, on average, a MVPF of 3) (Hendren and Sprung-Keyser, 2020). There is scant evidence on the costs and benefits of enforcement over the income distribution (Boning et al., 2023). We show that the MVPF for workers in the top quintile is lower than MVPF of workers in the bottom quintile. This suggests that targeting HVs to workers at the upper end of the wage distribution is more efficient.

The rest of the paper is structured as follows: Section 2 positions our paper in the literature, describes the institutional setting and the experiment; Section 3 describes the data and provides descriptive evidence on sick leave; Section 4 defines the empirical strategy to evaluate the effects of HVs on workers behavior; Section 5 presents the main results; Section 6 discusses evidence on the underlying mechanisms; Section 7 dwells on the cost-effectiveness and on the shadow value of the funds raised via HVs, and Section 8 concludes.

2 Background information

2.1 Our contribution to the literature

Our paper fits into two main strands of literature.

The first is the rather scant literature on sickness benefits. The focus of previous works was largely on legal rules concerning replacement rates (Ziebarth 2013; De Paola et al. 2014; Böckerman et al. 2018; Scognamiglio 2020; Marie and Vall Castelló 2022), and entitlement conditions (Markussen et al. 2012; Godøy and Dale-Olsen 2018; Hernæs 2018; Markussen et al. 2018). Recent studies, such as Pichler and Ziebarth (2017) and Pichler et al. (2020), investigated the effect of sick leave benefits on contagious presenteeism by comparing sick pay mandates across US states.[4] Enforcement is generally overlooked by these studies. Partial exceptions are Hesselius et al. (2009), Hartman et al. (2013) and Hesselius et al. (2013). In these studies, the authors exploit an experiment in Gothenburg (Sweden) in which treated workers could spend more time (from 8 to 15 days) on sick leave without providing a medical certificate. Thus, the experiment concerns the role of doctors writing the certificate while the sickness is ongoing. Our approach is different in that we deal with monitoring by external doctors directly employed by the social security to assess the validity of sick leave claims, hence not involved in the patient-doctor interaction. This makes the treatment unaffected by interpersonal relations. The assessment of the impact of this direct monitoring is unique to this study. Ferman et al. (2021) exploit a reform in a Norwegian city that removed the need for a certificate from a physician, but gave more monitoring capacity to firms. Alpino et al. (2022) analyze the impact of invitation dialogue meetings for workers on sickness leave in a field experiment in Norway and find no effect on the duration of the sickness spell of being summoned to a visit. D'Amuri (2017) investigates the response of public sector employees in Italy to changing monetary incentives and monitoring probabilities, finding a strong response to both. Unlike D'Amuri (2017) we know whether or not a worker on sick leave received

[4] Maclean et al. (2020) provide a comprehensive assessment of the introduction of mandated sick leave in the US and its welfare implications. Adams-Prassl et al. (2023) evaluate the willingness to pay for employer provided sick pay.

a home visit by an inspector, we have information on the outcomes of such visits and access to administrative data on the universe of workers taking sick leave and on their working histories. This gives us a direct measure of the treatment, randomized across individuals. Furthermore, we can look at responses along both the extensive (number of spells of sickness leave) and intensive (duration of the leave) margins and depending on the outcome of the HV.

The second avenue of research related to our work concerns the effects of auditing. This kind of studies has a prominent role in the taxation literature. In this context, randomization often concerns nudging through announcement by letter of variations in audits probability (among others, Pomeranz et al. 2014, Pomeranz 2015, Bérgolo et al. 2019, and De Neve et al. 2021). While many papers exploit natural experiments to assess the role of enforcement on tax payments (Slemrod, 2019), only a minority of studies randomize actual audits (Kleven et al., 2011; Gemmell and Ratto, 2012; Advani et al., 2017; DeBacker et al., 2018; Guyton et al., 2021; Bergeron et al., 2021). Random audits have been used in school testing (Bertoni et al., 2013), and corruption monitoring (Olken, 2007), but not in relation to sick leave.

Our contribution to this body of research is fourfold. First, we are not aware of previous studies directly investigating the enforcement of sick leave regulations through home visits. This is a crucial component of the sick leave enforcement mechanism. As in the case of the "tax system" problem (Slemrod and Gillitzer, 2013), understanding the role played by enforcement is essential to determine if a sickness benefit system is optimal. Second, we draw on our rich experimental design, which involved a population-wide randomized control trial over home visits, allowing to directly focus on the responses to actual auditing of the workers involved also taking into account the outcomes of such visits. Our experiment[5] takes place over the whole country, encompassing a wide range of different environments in terms of human and social capital endowments. Due to the

[5]It is a social experiment according to the taxonomy developed by Ferber et al. (1982) insofar as it is related to the attempt of an institutional agency to evaluate its actions by a deliberate variation in policy implementation. This definition encompasses as well natural field experiments (Harrison and List, 2004) since the subjects of the analysis (inspecting doctors and public employees claiming sick leave) are not aware of being part of the experiment.

relevant similarities in the design of sick leave benefits across countries, this experimental design strengthens the external validity and scalability of our findings (List, 2020). Third, we are able to link data on monitoring, sick leaves and workers' careers, shedding light on deterrence effects along various dimensions of workers heterogeneity and on actual workplace sanctions on workers found to be irregularly absent from work. This issue is relevant in a setting where workers do not face automatic and well-defined statutory penalties. Fourth, unlike previous literature, we target our experiment to public sector employees. This is particularly attractive as we have little evidence on productivity of public employees (Gittleman and Pierce, 2012; Colonnelli et al., 2020; Fenizia, 2022) and use of sick leave is a relevant issue in this setting across the two sides of the Atlantic (Cronin et al., 2022). An additional advantage of our focus on public employees is that public employment is characterized by 30 to 50% lower turnover rates, and most notably lower quits, than the private sector, and flows from public to private employment are extremely rare (Gomes, 2018; INPS, 2023). Hence, we can meaningfully use information on separations and wage reductions following home visits as a measure of implicit sanctions.

2.2 Institutional setting

All private and public sector employees in Italy are eligible for sickness benefits. The reason for the absence from work has to be certified by a GP or a specialist doctor, and the certificate, reporting the days in which the worker will be absent from work, must be notified to both the employer and INPS. The latter provides payments for the sick leave benefits. In the public sector, workers are entitled to up to 18 months of paid sick leave, and to an additional 18 months of unpaid leave over a 4-year period. The amount of the benefit is determined as follows: from the first day up to the 10^{th} day of the leave, the benefit covers the full contractual wage except the accessory (fixed and variable) components.[6] From the 10^{th} day up to 9^{th} months of absence from work, sick leave

[6] In Italy, wages in the public sector are set by collective agreements, which identify minimum levels of pay ("minimo tabellare") and other fixed accessory pay allowances related to the particular job (e.g., compensation for injury or epidemiological risk, required availability outside working hours, "indennità di turno" for those employees working during nighttime shifts). In addition, workers can receive variable accessory pay, related for instance to overtime work, and performance related pay. The latter component is more common for managers and highly skilled workers.

replaces 100% of the contractual wage including fixed accessory pay, but still excluding all variable components of pay (either accessory or performance-related). The replacement rate declines to 90% from the 9th month up to the 12th month, and to 50% from the 13th to the 18th month. Surgeries, day hospitals and treatments for chronic disease (e.g., cancer) are exempted from these reductions.

Unlike other countries, such as the Netherlands, where workers can claim disability benefits after a long period on sick leave,[7] in Italy sick leave is not automatically connected to disability benefits.

Employees can be subject to monitoring to assess their health status on the days in which they are on sick leave. HVs verify whether the stated reason for the sick leave (based on a certificate issued by a GP) matches the current actual health conditions of the employee. This is truly an inspection aimed solely at verifying the congruity of the certificate presented by the worker. The doctor cannot prescribe any treatment or change the diagnosis and extend the sick leave duration set by the GP. Only the latter is allowed to present a new certificate for the worker. Moreover, a doctor-inspector cannot perform HVs in areas where she is also a GP to avoid potential conflicts of interest.[8] While INPS has been checking on sick leave of private sector workers since 2011, monitoring public sector employees is a new duty for the social security administration, which took up this task in 2017 from local health administrations.

Inspections are performed by a INPS doctor, checking whether the employee is sick and at home. The inspector goes directly to the residence of the employee to check her conditions, without any notice. There are two types of HVs: the so called "employer called" visits (ECV) ("Visite Datoriali") and the "INPS called" visits (ICV) ("Visite d'Ufficio"). Each type of visits accounts for about one half of the total HVs. The ECVs are performed after a request is issued by the public administration where the absent worker is employed. If

[7]Workers are eligible for WIA (Dutch abbreviation for the Work and Income Act) benefits if they have been ill for nearly 2 years (104 weeks) and, because of the illness or disability, only can earn 65% or less of their previous income.

[8]See: Decreto del Ministero del Lavoro e della Previdenza Sociale, October 12 2000; Circolare INPS n. 4 January 8 2001; Circolare INPS n.199 November 13 2001.

INPS has resources to carry out the ECV,[9] a doctor-inspector ("medico fiscale") is sent out to verify the sickness certificate.

ICVs follow the same procedure of the ECVs, but, in this case, it is the social security administration itself to decide whom should be subject to the inspection, and when it should take place. Certificates are selected for a visit on a daily basis, among the universe of ongoing sick leaves notified to the social security administration. Our experiment focuses only on this second type of visits that are fully under INPS control.

HVs can be carried out every day in the week and workers must be at home and available for an inspection between 9 and 12 am, and between 3 and 6 pm.[10] There are 3 possible outcomes for a HV, reported always to the public employer:

1. If a HV finds that the worker is at home but fit for duty, i.e. healthy, the worker has to be back to work the day after the visit.

2. If a HV finds that the public employee is not at home during the declared sickness period and the worker has no "force majeure" reason for this absence – such as hospitalization, need of life-saving therapies, or medical check-ups – then INPS reports the matter to the public employer.

3. If the employee is sick and at home, then INPS (based on the recommendation of the inspector) confirms the prognosis.

In the first two cases, the worker is considered to be irregularly on leave. From then on, the worker involved cannot claim new certificates for the same sickness event. There is no pre-determined sanction: possible disciplinary actions are left to the discretion of the public manager, who is allowed to fire the employee in extreme cases. However, this event is extremely rare. For instance, in 2019 only 117 workers (.002% of public employment) were laid off for unjustified absences from the workplace. Nevertheless, even without a statutory sanction, being found to be irregularly on sick leave may affect career

[9] In the private sector the procedure is similar, but the employer pays the monetary costs associated to the visit.

[10] Due to budgetary reasons ICVs were not performed on weekends until 2020 while ECVs were performed also on non-working days. This information was not revealed to workers.

prospects and wage dynamics: a manager could, for instance, reduce paid overtime hours or postpone promotions for employees who have been reported as irregularly on sick leave.

The selection of certificates for inspections in the private sector since 2011 is based on an algorithm denominated *Savio* ("Sistema Assegnazione Visite Ottimizzato") to better target HVs.

Savio operates according to the following steps:

1. A random sample of ongoing certificates is selected out of the universe of the certificates notified every day to INPS to reduce costs related to the identification of certificates not subject to HVs (see step 2).

2. Certificates exempted from HVs (because involving some chronic disease, such as cancer) are excluded from this randomly selected sample.

3. Among the non-exempt certificates, the algorithm generates a ranking via a machine learning procedure maximizing the probability of detecting irregularities.

4. A matching of visits to doctors is undertaken to minimize costs, given a target number of visits. This is implemented to optimize travel time of doctors (who are paid based on the distance between their office and the residence of the worker on sick leave), but also to avoid any arbitrariness in the choice of the workers to be subject to inspections.

Doctors receive every morning a list of HVs to be performed within that day and the sequence they should follow in undertaking HVs. Doctors are fully compliant with these rules, and they immediately report the outcome of the visit to the social security and to the public administration involved. No reporting is envisaged to the GP who issued the certificate.

2.3 The experiment

When INPS was given the task to perform ICVs on public sector employees, it was decided to replicate a setup similar to that operating in the private sector. The underlying

procedure, which should target visits to certificates more likely to be irregular, however, needs to build on a critical mass of visits to feed the machine learning algorithm. Thus, INPS decided to begin by randomly selecting the certificates potentially subject to ICVs in order to maximize the informational content of the data collected in the home visits.[11]

The decision was made at the INPS headquarters at the highest hierarchical levels and was not disclosed at lower levels.[12] Most importantly, doctors involved in the visits did not receive any information concerning the experiment and did not experience any change in their activity. This uniform application of well established procedures makes it unlikely that doctors behaved differently in the experimental setting ("*Hawtorne effect*").

The experiment took place in the 45 days between November 22, 2017 and January 5, 2018. It assigned the HV treatment through a "typical" stratified randomized experiment, following a four-step procedure. Visits were performed over 29 working days out of the 45 days of the experiment period.[13] During the period of the experiment, INPS performed about 4,200 HVs, that is, about 145 visits per day. The main steps, visually summarized in Figure A1 in the Appendix, are as follows:

1. On each working day from November 22, 2017 until January 5, 2018, INPS drew daily, for each local office, a random sample of certificates from the universe of absence certificates (around 400,000 certificates) to reduce administrative costs related to step 2.

2. Exempted certificates, i.e. those involving chronic or very serious diseases, a recent surgery, etc., were excluded.

3. Among the selected sample, a second random sample was drawn. Its size was determined in such a way that for each worker subject to HV there were an additional

[11] The machine learning procedure assigning visits to certificate was never fully in place due to restrictions imposed by a ruling of the Italian Privacy Authority. This ruling was largely unexpected and took place in Autumn 2018, almost one year after the experiment.

[12] The experiment design and its implementation was under the direct control of the authors as one of them, Tito Boeri, was at the time of the experiment President of INPS, while Edoardo di Porto and Paolo Naticchioni were managers at the Research Department which was responsible for the monitoring and the assessment of the experiment.

[13] Due to budgetary limitations in 2017 and 2020 INPS did not undertake home visits during week ends.

eleven workers in the control group. This sample, made up of about 60,000 certificates, was then used for the experiment. Selected certificates were randomly ranked, and the resulting order defined the HV priority.

4. Visits were assigned to doctors, optimizing travel time.

Thus, the experiment replicated the Savio procedure except for step 3.

Two caveats are worth mentioning at this stage.

First, as in the private sector, a certificate, selected in Step 1 and having a prognosis longer than one day, would re-enter the pool of certificates subject to random draw in each day of its duration. Similarly, a worker claiming multiple certificates in the period of the experiment could be present in the pool of eligible certificates multiple times. Hence, workers having long or multiple certificates in the period of the experiment are more likely to be assigned to a HV. This issue implies a positive correlation between the time spent on leave in the experiment and the probability of being treated. We overcome this issue by directly controlling for the total number of days spent by an individual on sick leave in the period of the experiment. Furthermore, to account for chronic illness, general health conditions, and propensity to use sick leave, we also control for the take-up of sick leave in the six months preceding the experiment. In particular, we include in our regression model the total number of days claimed, and of certificates, as well as their average duration in the six months before the experiment.

Second, certificates, rather than workers claiming sick leave, are randomly extracted replicating the procedures followed in the case of private sector employees. However, as behavioral responses to HVs occur at the level of the individual, we run the analysis at the individual (rather than certificate) level.

Some indications about the relevance of these two problems can be obtained from Figure A2. Panel (a) reports the number of days spent on sick leave by workers in the period of the experiment. The quite wide range of values in this distribution suggests that individuals have much different treatment probabilities. Controlling for time spent on

leave in the period of the experiment is therefore crucial in our setting. Panel (b) reports the number of certificates sent by individuals in the period of the experiment. In more than 75% of cases, individuals only have one active certificate, about 18% have two, 5% three, and a negligible minority more than three. Thus, collapsing the analysis at the individual level comes at little cost.

3 Data

3.1 Linking three datasets

Our analysis uses a so far unexploited dataset, spanning the period from 2016 to 2019. It offers a record linkage of three administrative sources released by INPS for the first time for this paper.

First, our work exploits a dataset on certificates sampled in the experiment period, which provides information on the sick leave and on the randomized ICV inspections. This dataset reports the start date of the disease, the beginning and the end dates of the period covered by the sick leave, the identifier of the INPS local office responsible for the inspection activity, the date of the visit, and its outcome. The diagnosis is not available due to privacy restrictions.

Second, we complement this information with data on certificates issued for individuals involved in the experiment from 2016 up to April 2019. We use this dataset to construct our dependent variables and to control for past use of sick leave. Also in this case we have information on the start date of the disease, start and end dates of each sick leave spell, and identifiers for the worker and for the INPS local office handling the certificate for the worker and managing the inspection activity.

Thirdly, we extend the analysis with a unique dataset on workers' careers in the public sector at monthly frequencies since 2016. The dataset includes information on wages, type of contract (part-time/full-time, permanent/temporary), occupation, field of activity within the public sector, and location of the worker at the municipality level.

We recover additional demographic information from Social Security Archives. The record linkage of these three sources generates a unique and, so far, unexploited dataset. We restrict the analysis to individuals aged 24 to 67 (the normal retirement age), and having an employment record in the public sector at the time of the experiment as well as at least one employment record in the public sector in the six months before the experiment. The focus is on incumbent workers, that is to say, employees who were already in the public sector at the time of the experiment. Thus, our results are not affected by new hires after the experiment.

Finally, we draw on a question of the 2015 European Working Conditions eliciting whether workers *deal directly with people who are not employees at your workplace such as customers, passengers, pupils, and patients* and on the extensive classification of occupations provided by the INPS archives to provide a classification of *street level bureaucrats*, that is, public employees operating at the front line and hence frequently subject to interactions with the general public (see Appendix B for details).

3.2 Descriptive statistics

Table 1 provides summary statistics on the workers who were involved in our experiment. About 72% of them are women and the average age is about 53. Some 95% of the workers have a permanent contract, and 6% a part-time job. On average, workers taking part in the experiment spent almost 45 days on sick leave in the 16 months after the experiment. The distribution is strongly skewed to the right, with the median (20 days) being less than half of the average.[14] The majority of sick leaves requested (about 49%) are of short duration (between 1 and 3 days). We also observe, however, a non-negligible share (about 16%) of certificates with durations exceeding 15 days (see Figure A3 in the Appendix for additional details). Almost 10% of the workers on leave and selected for the experiment were subject to inspection, and in about four cases out of five the prognosis was confirmed, i.e. the worker was found to be regularly on leave. In the remaining cases

[14]This distribution is obtained by avoiding double counting, that is a day cannot be covered by more than one sickness leave.

either the worker was found *fit for duty*, while the medical certificate issued by the GP was stating the opposite, or the worker was absent without any justification, that is, the doctor did not find the employee at home when she should have been according to the existing regulation. The valid reasons for these absences are only those strictly related to the medical treatment the worker is undergoing for the certified disease (or issues of proved 'force majeure', such as the need for life-saving treatment or hospitalization). We categorize these latter two cases as "irregular" absences.

The first column of Table 2 looks at the distribution of days of absence in the 6 months prior to the experiment. The number of days of sick leave is steeply increasing with age: workers aged more than 65 have, on average, about 17 days of leave more than employees under 35. Consistently with previous findings on the relationship between employment protection and absenteeism,[15] permanent contract holders have, on average, about 10 days more of absence than temporary workers. Absenteeism is also higher in central administrations than elsewhere in the public sector, and is declining in wages. Exposure to direct contacts with the public, as captured by our measure of street level bureaucrats, is not associated to a higher incidence of sick leave. The same applies to exposure to epidemiological risk (health sector workers).

The second column of Table 2 is confined to audited workers and looks at the characteristics of the workers found to be irregularly on leave. Four results stand out. First, women are about 4 percentage points less likely to be irregularly absent from work than men, while no clear pattern emerges in terms of age groups. Second, workers found to be irregularly on leave are more likely to live in metropolitan areas and work in southern regions. Third, irregular behaviors are more likely to be detected among short certificates than for certificates lasting 10 days or more. Certificates of duration between 1 and 4 days, for example, are 55 percentage points more likely to be found irregular than certificates lasting more than 10 days. Fourth, visits performed on Fridays are about 6 percentage points more likely to detect irregularities.

[15] See for instance Ichino and Riphahn (2005). Another interpretation is that workers on temporary contracts spend less time employed in the public sector, and hence have fewer opportunities to claim sick leave benefits than civil servants with open-ended contracts.

This asymmetric distribution of irregular sickness leaves over the working week is visually represented in Figure 2. Panel (a), in particular, displays the distribution of irregular outcomes by date in which the HV takes place. There is a spike in the share of irregular leaves detected on Friday, December 22, that is, the last day of work before the Xmas break. While, on average, about one visit out of five detects an irregular behavior, on December 22, 2017 more than 40% of the workers inspected were found to be irregularly on sick leave. Moreover, there is a concentration of irregular certificates on Fridays. This can be better appreciated by looking at panel (b), panel (c) and panel (d) of Figure 2. The concentration on Fridays of irregular sick leave is more evident in central and southern regions than in the northern part of Italy. It should be noted that the higher incidence of irregular certificates on Fridays is not related to a higher number of HVs on Fridays. Figure A4 in the Appendix shows that the average number of visits per day is rather stable over week days (at about 150 HVs per day). The only exception are Tuesdays, when more visits are performed.

4 Empirical strategy

4.1 Baseline specification

To estimate the causal effect of HVs we exploit the random allocation of visits to sickness leave spells in the period of the experiment. We estimate the following cross-sectional OLS regression at individual level:

$$Y_{ij} = \alpha + \beta HV_i + X_i\zeta + Z_i\delta + Sede_j + D_i + \eta_{ij} \tag{1}$$

where Y_{ij} is the dependent variable for individual i at INPS local office (Sede) j, HV_i is a dummy equal to one if the individual received an inspection in the period of the experiment, and X_i is a set of controls for demographic and job characteristics while η_{ij} is a random error. We also include measures of past sick leave which control for long-term

health conditions and the tendency to draw sickness benefits: the vector Z_i includes the number of days on leave, the number of sick leave certificates, and the average duration of the certificates in the 6 months before the experiment.[16] Since the experiment takes place separately in each INPS office, treated and controls have to be compared at the office level. Hence, we include INPS local office fixed effects ($Sede_j$).[17]

As discussed in Section 2.3, the treatment probability is mechanically linked to the number of days of requested leave during the experiment. To deal with this issue, we introduce in our specification a set of fixed effects D_i that captures the number of days of eligibility in the experiment for each worker. In doing so, we compare treated and controls with the same ex-ante probability to be subject to a HV. In other words, we compare treated and control individuals within the same bin in terms of days spent on leave in the period of the experiment. Furthermore, conditioning on D_i provides estimates within groups of individuals with similar treatment probabilities. We cluster standard errors at local office level. Under the assumption that HVs are conditionally uncorrelated with unobservable individual traits or shocks, the parameter β identifies the causal effect of HVs on the future use of sick leave benefits.

4.2 Balancing

In order to assess whether the randomization procedure was successful in identifying a similar treatment and control group, we assess the difference between treated and controls first by looking at the differences in average characteristics and then computing normalized differences (Imbens and Wooldridge, 2009) for demographic and job characteristics at the individual level. The results of this exercise are reported in Table 3. Treatment (workers subject to a HV) and control groups (workers not subject to a HV) display striking similarities. Small compositional differences are, for the most, no longer statistically

[16] We also include the duration of the first certificate that the worker has in the period of the experiment. This duration is computed based on the start and end date of the disease as reported in the certificate when it was issued, and it is hence unaffected by the HV. This control provides a more precise measure of the health status of the individual during the experiment. Note that this variable is only partially accounted for by the number of days spent by the worker on sickness leave during the experiment, since the sickness spell could have started before the experiment or could end after the period of the experiment.

[17] Note that each Italian province includes an INPS office, and more than one for the wider provinces.

significant when control is made for the number of days spent on sickness leave (Column 4). Moreover, normalized differences are always well below the threshold value of 0.25 defined by Imbens and Rubin (2015) indicating that they should not affect our estimates. Overall, normalized differences for variables not mechanically related to the treatment[18] confirm that the randomization was successful in selecting proper treatment and control groups.

A further test of the balancing between treatment and control groups is provided in the Appendix, where we regress the probability of receiving a HV against demographic and job characteristics at the individual level (Table A1). In this multivariate approach, only minimal differences in the probability of receiving a HV can be detected for women, older workers (between 60 and 66 years of age) and workers with longer sick leave, while no differences can be detected in terms of other age groups, sector, job characteristics and previous use of sick leave. This confirms that treated and control workers are remarkably similar.

4.3 Difference in differences specification

Despite the remarkable similarity between treated and control workers, it is still possible that treated workers are different in some unobserved dimension (i.e., health status) with respect to control workers. In addition, treated workers might have been on a declining trend in the use of sick leave and this might lead to overestimate the decline in sick leave use as a consequence of the experiment. To further control for time-invariant unobserved heterogeneity across individuals and assess possible trend in sick leave use, we estimated the impact of HVs also within a difference in differences framework. This specification also allows testing for the presence of a differential trend in the use of sickness leave between treated and control individuals. In particular, we estimated the following panel OLS regression:

[18] Normalized differences are above the 0.25 threshold only for measures of the use of sick leave in the period of the experiment and in the six months before (bottom 5 rows). This is mechanical, since the treatment was repeatedly administered

$$Y_{it} = \alpha + \sum_{j \in \{-12,-11,..,18\}/\{-1\}} \beta_j \mathbb{1}(t - t_e = j) \times HV_i + \sum_{j \in \{-12,-11,..,18\}/\{-1\}} \gamma_j \mathbb{1}(t - t_e = j) \times D_i +$$
$$+ \sum_{j \in \{-12,-11,..,18\}/\{-1\}} \gamma_j \mathbb{1}(t - t_e = j) \times Sede_i + \delta_i + \lambda_t + \varepsilon_{it} \quad (2)$$

Where the dependent variable is the number of days spent on sick leave in month t by individual i, $\mathbb{1}(t - t_e = j)$ is a dummy variable taking value one when the difference between month t and the start of the experiment t_e, set to November 2017, is equal to j, δ_i is an individual fixed effect and λ_t is a month fixed effect.[19] We also include dummies for the days spent on sick leave during the experiment (D_i) and dummies for the INPS local office at the time of the experiment ($Sede_i$) both interacted with time dummies. Thus, we compare, over time, individuals who had the same exposure to the experiment and were under the jurisdiction of the same local office. Standard errors are clustered at the INPS local office level ($Sede_i$).

4.4 HVs and exposure to the treatment

Another possible concern with our setting is that HVs may affect the exposure to the treatment.

In order to deal with this concern, we instrumented the actual number of days on sick leave in the experiment by the number of days on leave in the period of the experiment as implied by start and end dates of the first certificate when it was issued.[20] This strategy has two advantages: a) the first certificate is fully exogenous, since it is not affected in any way by the experiment; b) after accounting for the duration of the certificate, the variation induced by the instrument only depends on when the first certificate was issued and to what extent its duration overlaps with the experiment.

[19] Note that the period of the experiment encompasses November 2017, December 2017 and January 2018. So it consists of the periods from zero to two and we look at the impact of the treatment from month 3 to 18, that is in the 16 months after the experiment, consistently with the baseline model.

[20] This variable is the same as the duration of the first certificate in the period of the experiment if the first certificate of the worker is fully contained in the period of the experiment.

As a further robustness check, we restricted the sample to a group of individuals with only one certificate in the period of the experiment, with start and end dates fully within the time interval of the experiment. We then replaced the actual number of days spent on leave in the experiment with the potential duration of the leave (as defined by the certificate when it was issued), which is predetermined with respect to the treatment and, hence, exogenous.

5 Results

5.1 HVs and future sick leave claims

The main outcome of interest is the number of days of sick leave claimed in the 16 months following the experiment. This captures potential deterrence effects, that is, whether receiving a random home visit discourages future claims of sick leave.

Table 4 reports our main results in Column (1). We report here results for our preferred specification which includes local office fixed effects, fixed effects for the number of days spent on leave during the experiment and controls for demographic and jobs characteristics and past sick leave use.[21] We find that being treated leads to a decline in the number of days on sick leave in the 16 months following the experiment by about 3.6 days, which corresponds to about 8% of the baseline days on leave, i.e. time spent on leave for workers in the control group over the 16 months after the experiment. This effect is statistically significant at 5%. The observed decline in the number of days on sickness benefits by audited individuals conflates two different components: the change in the number of certificates and the change in the average duration of the certificates. Column (2) and Column (3) consider separately the two margins and show that the adjustment in the long run takes place mostly at the intensive margin (i.e. the length of certificates). The number of certificates declines by about 0.1 and this decline is not statistically different from zero (-1.4% relative to the control group), while the average duration declines by 0.7

[21] We report results by including sequentially these elements in Table A2 in the Appendix. These results show that past use of sick leave and days on leave in the experiment are the most crucial controls while demographic and job characteristics play a minor role as could be expected from the balancing checks.

days per certificate (-10.6% with respect to the control).[22] The fact that the reduction comes mainly from the intensive margin is consistent with the fact that short prognosis for some of the most frequent types of illness (e.g., flu, headache) are more difficult to be contested in the case of a HV. While the GP issuing the certificates is not informed by social security about the HV either before or after the visit, it is plausible that the worker being inspected reports ex-post the occurrence of the HV to the GP making him more reluctant to issue relatively long certificates.

Next, we look at the time span over which the behavioral change takes place. We compute the cumulative number of days of sick leave at monthly frequencies in the 16 months after the experiment, and then we run a separate regression at each time horizon. We perform the same analysis on the number of sick leaves claimed and their average duration at each time horizon. Regression coefficients for the effect of a HV and their 95% percent confidence intervals are plotted in Figure 3. Panel (a) reports the pattern for the cumulative number of days on sick leave, Panel (b) for the cumulative number of certificates, and Panel (c) for the average duration of those certificates. The decline in the number of days builds up over time and progressively increases over the 16 months horizon. Both the extensive (Panel b) and the intensive (Panel c) margins contribute to this pattern, but the cumulative decline in the number of certificates stabilizes after about 10 months and becomes slightly smaller after 13 months since the experiment and is no longer significant at 5% after the 11^{th} month. The long-lasting effect is therefore driven by a shorter duration of certificates sent by treated workers over the whole observation period.[23]

5.2 Do HVs induce presenteeism?

A concern with the reduction in the days of sick leave originated by HVs is that it may induce presenteeism, causing a deterioration of the health conditions of the worker who

[22] We impute a zero average duration when no certificate is sent after the experiment

[23] Our results are not due to workers leaving the public sector as a consequence of the HV: Figure A5 in the Appendix shows that attrition is concentrated among workers aged 55 and above and our results hold if we restrict the sample to workers being present in our data for the entire 16 months period, as shown in Table A3 in the Appendix.

was on sick leave and, in case of infectious diseases, contagion of the co-workers.

Unfortunately, we do not have access to data on coworkers, hence we must rely only on information on the reported health status of the individuals involved in the experiment. One way to evaluate the risk of presenteeism is to look at relapses for visited workers. Indeed, if workers go back to work, but they are not fully healed, they might suffer again from the same disease for which they claimed the previous certificate.[24] We can proxy relapses by using the date of the start of the disease (the exact diagnosis is not available due to privacy concerns): if multiple certificates have the same start date for the disease, but they prescribe the beginning of the sickness leave at different dates, then the certificates after the first have to be considered as extensions of the original disease.

To assess the impact of HVs on this margin, we disentangle certificates related to new diseases from certificates related to previous diseases, and we estimate the impact of HVs on the two types of certificates. Table 5 reports our estimates. We find that HVs do not lead to a higher number of relapses, while they induce a decline in the number of days on sick leave for both categories of certificates, although the effect is stronger for certificates denoting a new disease. Hence, HVs do not seem to have been detrimental for the inspected workers' health, and they are unlikely to have had negative effects on the health status of coworkers.

5.3 Difference in differences: Dynamic effects of HVs

Our results might stem from differential trends in the use of sick leave across different individuals. In order to account for these and potentially additional sources of unobserved heterogeneity across individuals[25] we exploit the longitudinal dimension of our data, and we construct a panel at monthly frequencies from November 2016 up to May 2019. This allows us to run a difference in differences regression in event study form as outlined in

[24] Marie and Vall Castelló (2022), for example, show that, as a consequence of lower benefit generosity, public sector workers in Spain decrease their sick leave time and increase their rate of relapse (+8%).

[25] The probability of being treated depends on the number of days on leave in the experiment and on general health conditions. While we control for the number of days on leave in the experiment and for several measures of past use of sick leave, we might still miss aspects of the health of the worker which are relevant for both the probability of treatment and the future use of sick leave.

Equation 2. The model includes individual, time, days spent on sick leave in the period of the experiment, and INPS Local office fixed effects. We also interact fixed effects for INPS local office and time spent on sick leave in the period of the experiment. Thus, our difference in differences model compares, within each month, individuals with the same time on sick leave in the experiment and individuals in the same INPS local office during the experiment. We use a Poisson regression model to account for the higher number of zeros in the data at monthly frequencies. We group months in bins of three to gain precision, and use as reference the period between 4 and 6 months before November 2017.

Results are reported in Figure 4. Treated and control individuals are on a parallel trend in the period before the treatment. Moreover, the coefficients for the period after the experiment show a persistent negative effect on the number of days on sick leave by month. The coefficient here has a semi elasticity interpretation, and it shows that treated individuals reduce the number of days on leave by about 10% in the months following the experiment, which is close to the 8% estimated in the baseline model.[26]

Overall, our difference in differences reassuringly provides corroborating evidence in support of our results and excludes that they are driven by differential trends between control and treatment groups.

5.4 HVs and careers of workers

Another important outcome to be considered is whether HVs affect the career of the workers involved. This may be due to a number of reasons. Workers found to be irregularly absent from work may be sanctioned by their superiors in terms either of total compensation (e.g., by reducing their involvement in overtime work) or of mobility across jobs in the public administrations (disciplinary layoffs are extremely rare in the public sector). Workers receiving an HV, even when the latter validates the certificate, may update their beliefs about monitoring in the public sector and adjust accordingly

[26]Since the days spent on leave in the experiment are directly related to the dependent variable in the period of the experiment, we also repeated the analysis by excluding this period from the data (November 2017-January 2018). This does not affect our results, as shown in Figure A6.

their level of effort. Perceptions about the fairness of the system may also play a role in conditioning the future performance, and the career of the workers involved.

To assess whether HVs induce changes in the career of workers, we relied on difference in differences estimates, since they allow to better control for unobserved heterogeneity and net out possible differences in levels across groups of workers.

Our measures of individual careers exploit the longitudinal features of the social security archive, to which data on HVs are linked. We looked at the evolution of total pay, at separations from the public sector (either to non-employment or to the private sector), and at transitions across jobs within the public sector. In order to characterize the nature of these transitions (e.g., whether they denote career improvement or deterioration) we combined information on occupational mobility with data on the associated wage variation relative to the average wage in that occupation.[27] We also looked at the probability that a worker becomes a manager, capturing a promotion for employees not having reached that status before the HV.

We report results in Figure 5. Interestingly, we do not see an impact of the HV on the career of the average worker across any of the dimensions considered. Reassuringly, also in this case, we do not see any deviation in job outcomes between treated and control workers in the periods before the experiment. In the 16 months after the experiment we observe a small decline in the wage (about 25 Euro per month) and a small increase in the probability of exiting the public sector (0.5 percentage points) for the treated workers, but in all cases the effects are very far from being statistically different from zero. No specific pattern in terms of occupational changes is detected.

5.5 HVs and program substitution

Another outcome of interest, notably when assessing the fiscal impact of HVs, is related to the possibility that the inspected workers substitute away time on sickness leave with other benefits, which could provide them with additional sources of income while reducing

[27] We do not have access to contractual wages for those occupations, hence we rely on observed wages which are still likely to create a reasonable ranking in terms of pay across occupations.

their time at the workplace.

We considered two main alternatives to sickness leave: retirement and take up of disability benefits. We report results in Figure A7 in the Appendix. In both cases, we do not see any evidence of workers taking up retirement or disability benefits after a HV with respect to the control group. This comes at little surprise since, as stated above, we did not observe significant variations in the employment probability, and there are strict rules for access to disability benefits in Italy.

5.6 Effects along the income distribution

As pointed out by Slemrod (Slemrod, 2019) we still know too little about the implications of compliance and enforcement policies along the income distribution. The few attempts to date to provide evidence on willingness to evade and reaction to audit by income level focus on tax compliance (DeBacker et al., 2020; Guyton et al., 2021; Johns and Slemrod, 2010). There is also very little empirical work quantifying the costs and benefits of tax enforcement along the income distribution, with the only exception of Boning et al. (2023). We don't have clear priors as of the sign of the correlation between income and audit induced behaviors in the sick benefit system. Individuals with lower earnings are surely associated with worse health conditions and a higher willingness to claim sick benefit, but their reaction to an audit depends crucially on their risk attitude and their expectations as to the sanctions applied in case of detection of opportunistic behavior.

Estimates for the effect of HV on the cumulative number of days on sick leave in the 16 months after the experiment are provided in Panel (a) of Figure 6. Regressions are estimated separately by quintile of earnings, and are specified as in Equation 1.[28] Interestingly, we find that the deterrence effect is rather stable along the earnings distribution. Potential differences in risk aversion seem to be compensated by offsetting differences in expected sanctions or, more simply, individuals do not differ along these two dimensions. From a cost-saving perspective it is preferable to target high wage earners since the replacement

[28] Earnings are computed as the average of positive monthly salaries between January and November 2017.

rate and the costs of HVs are constant over the distribution and wages at the top are up to 2.5 times larger than at the bottom of the distribution.

Summarizing, HVs appear to reduce future sick leave of the workers involved, mainly by acting on the duration of the leave. The effect on this intensive margin is also persistent over the observation period, unlike the effect on the number of sick leave claims (extensive margin). We do not see significant effects either on workers' careers and access to alternative programs involving compensation while not working. Finally, there is no evidence of significant differences in the response to HVs along the earning distribution.

5.7 Robustness checks

We implement several robustness checks of our results.

First, we use an instrumental variable approach which leverages on how much individuals are exposed to the experiment based on the timing of their first certificate. In practice, we instrument the time spent on sick leave in the period of the experiment with how much the worker would be on sick leave during the experiment as implied by the start and end dates reported on her first certificate when it was issued. As discussed in section 4.4, after controlling for the duration of the first certificate, this instrument exploits only the timing of the first certificate.[29] Results of this check are reported in Table A4. Column (1) reports baseline estimates from Column (1) of Table 4 while Column (2) shows that controlling linearly for the number of days on leave in the experiment has a negligible impact on the estimates.[30] Column (3), finally, reports the results of our IV strategy. The Cragg-Donald F-test shows that the instrument is strong while the coefficient is extremely close to the results of our main specification reported in Column (1). This confirms that issues related to time spent on leave in the period of the experiment should be minor.

Second, we consider a subset of workers who only have one certificate, fully included in the period of the experiment, and then replace the time spent on leave during the experiment

[29]The identifying assumption of the IV is that the exposure to the experiment as implied by the timing of the certificate only affects the outcome through the days spent on leave in the experiment, which appears reasonable.

[30]We use a linear control in this specification since it substantially simplifies the IV estimation.

with the duration implied by the start and end dates of the certificate, when it was issued. In this case, our covariate matches closely the exposure to the experiment of the worker, and it has the advantage that it cannot be affected by the treatment since it was set at the beginning of the first sick leave occurring during the experiment. Results for this subset of workers are reported in Table A5 in the Appendix. The impact of HVs is very similar to the one in the baseline estimates, but it is less precisely estimated. This is most likely due to the strong decline in the sample size that we experience once we restrict the sample to certificates fully included in the period of the experiment. The coefficients are, however, extremely similar to the ones of our baseline specification, which confirms that issues possibly related to days spent on leave in the experiment are second order.

Third, we account for the possible role of unobserved heterogeneity by employing the methodology developed by Oster (2019). In particular, we simulate the effects of different levels of unobserved heterogeneity up to a 25% increase in the adjusted R-squared and of the relationship between our treatment variable and the unobserved heterogeneity. The most relevant omitted factor is the health condition of the worker, which correlates positively with both the dependent variable and the HV probability. Thus, we mostly consider a positive correlation, but we try also with small negative values. Results, reported in Figure A8 in the Appendix, show that for small levels of correlations (delta), the coefficient remains reasonably close to the baseline estimates, even for large levels of R squared. However, as the measure of the relation between the unobserved factor and the treatment variable increases, the treatment effect becomes larger in magnitude. This suggests that, under reasonable assumptions, our estimates are at worst underestimating the deterrence impact of HVs on future sick leave claims.

Fourth, we perform a randomized inference exercise. To do so, we replace the treatment in our preferred specification with a random treatment, obtained by imposing a dummy variable equal to one to random observations in the sample. The number of observations is chosen to match the share of treated in the true treatment (9.8%). We replicate this process 1,000 times, collect the coefficients for the fake treatment, and compare them with the coefficient obtained with the true treatment. We plot the distribution of resulting

coefficients together with the original treatment effects in Figure A9 in the Appendix. Results show that the "fake" coefficient is always above our coefficient. This suggests that our coefficient captures a treatment effect, which would be almost impossible to replicate with random draws from the data.

Fifth, we treat more explicitly the stratification of our sample by time spent on sick leave during the period of the experiment by running our model separately for small subgroups defined by the number of days spent on leave in the experiment. We then take the treatment effect for each of these strata and plot them in Figure A10. The size of the marker is proportional to the sample covered by the subgroup. The treatment effect is consistently negative for all but one subgroup, and it is particularly large and negative for individuals who spent less time in the experiment. For individuals with longer time in the experiment, the treatment effect is smaller (or even positive in one case) and very imprecisely estimated. These results support our claim that the impact of HVs is largely negative. We also estimate our treatment effect separately for each number of days spent in the experiment and define an aggregate coefficient by using the number of treated individuals in each stratum as weights (Athey and Imbens, 2017). We find a treatment effect which is in line with our baseline estimates.

Finally, in order to verify that we are jointly identifying treatment effects and denying "false positive" results, we implement a multiple hypothesis testing procedure developed by List et al. (2021). This battery of tests, under weak assumptions, is asymptotically balanced and controls for family-wise error rates and for the probability of one false rejection. Table A6 in the Appendix, reports the results of our tests as well as the standard (albeit less rigorous) Bonferroni and Holm tests. The results confirm that we can jointly identify our treatment effects for the cumulative days on sick leave after the experiment, the number of certificates, and the average certificate duration. Table A6 also shows that our estimates obtained through linear regressions (Column 1) are robust to alternative specifications (e.g., the regression framework of Lin 2013), and, if anything, are underestimating the effects of the treatment.

6 Potential mechanisms

The most intuitive explanation for the documented effects of HVs on future sick leave claims is deterrence of abuse, as workers revise their perceptions about the risk associated to opportunistic behavior.

In our setting, workers taking sick leave face a sort of double lottery since there are not well-defined rules concerning the sanctions to be applied to those being found irregularly absent. Thus, the updating may involve either the audit risk or the perceived penalty associated with being irregularly absent, or both. While all workers being subject to HVs acquire additional information about audit risk, and may revise their priors in this respect, only workers being found irregularly on sick leave acquire additional information on the sanctions associated to detection of opportunistic behavior and therefore may revise their priors in this respect.[31] Thus, it is important to analyze the differential response based on the outcome of the visit.

6.1 Behavioral response by visit outcome

So far, we investigated the average impact of receiving HVs on workers' behavior irrespective of the outcomes of the visits. The problem is that the treatment *being audited and found irregularly absent* is not exogenous, and we have no information about the regular or irregular status of the leave taken by the workers who are not subject to a HV. Yet, comparing the behavioral response of the two types of audited workers – regular and irregular – is useful to understand the nature of the learning process associated with receiving a HV.

We investigate the differential impact of HVs on future sick leave claims depending on the outcome of the visit by estimating our difference in differences model, allowing for different coefficients for workers found regularly and irregularly on leave. This strategy has the advantage that it can exploit the longitudinal features of our data to get rid of

[31] In principle also their coworkers may learn more about the sanctions, but this is a dimension that our data do not allow to explore.

time-invariant unobserved heterogeneity among workers who were on leave either regularly and irregularly. For instance, the two types of workers may differ from the average worker in the the number of days in sick leave, and we can net out this difference by comparing workers over time in a difference in differences framework. Moreover, under the assumption that individual propensity to misuse sick benefit is time invariant, a difference in differences estimate rules out this possible confounding effect. Overall, this strategy allows getting causal estimates for the impact of the HV on workers, as far as differences across workers are time-invariant.

Results, reported in Figure 7, show the impact of HV by outcome of the visit on days on leave and on the number of certificates per month. The parallel trends in the periods before the experiment, which hold for both workers found regularly and irregularly on leave during the experiment, suggest that this strategy absorbs any possible level difference between the two groups recovering causal parameters for the two groups of workers.[32] Panel (a), which reports the impact on the number of days on sick leave, points to a stronger effect on workers found irregularly on leave: the decline in the number of monthly sick leave claims for workers found irregularly on leave notably in the months immediately following the experiment and the coefficient remains statistically significant at 5% until about 13 months after the experiment. We also observe a milder decline in the number of days on sick leave for workers found regularly on leave, which becomes larger over time and is statistically significant at 5% about 10 months after the experiment. In the longer run, we do not see differences in the magnitude of the effect on workers found regularly on irregularly on leave, although the impact of the latter group is more imprecisely estimated (workers found irregularly on leave represent only one fifth of all workers subject to HV). The temporary impact on irregular workers is in line with evidence of temporary effects of audits in other settings (see for example Bertoni et al., 2021 for the impact of monitoring during school tests). The larger reaction of individuals non-compliant with regulation is also consistent with evidence from a random tax audit in Norway by Hebous et al. (2020) where audited individuals reduced future use of tax deductions if found misreporting.

[32]Results hold even if the period of the experiment (from 0 to 2) is excluded from the sample.

Panel (b) of Figure 7 also shows a larger and more persistent impact on the number of certificates sent for workers irregularly on leave. In the aftermath of the experiment, the effects on workers irregularly on leave is similar to the one on workers regularly on leave, but the effect is more persistent for workers irregularly on leave. After 16 months, both effects are no longer statistically significant and the effect for workers regularly on leave is virtually zero.

As documented in Section 3, irregular outcomes are more frequent among short sick leaves. This may contribute to explain why only workers being found irregularly on leave adjust more along the extensive margin. This interpretation is corroborated by evidence on the response along the workweek. Figure A11 displays the impact of HVs for the number of certificates by the weekly start and end days of the certificates according to the outcome of the visit. In particular, coefficients are obtained by estimating separate regressions for the number of certificates starting or ending on specific days of the week in the 16 months after January 2018 and then plotting the coefficients of interest by visit outcome together with their 95% confidence intervals. We find no effects of HVs on workers found regularly on leave, while workers found irregularly on leave reduce the number of sick leaves claimed or ending on specific days. The largest decline is registered for certificates issued at the beginning of the week (Monday) and ending during the weekend (Saturday and Sunday).[33] It should be noted that for some sectors, such as the health sector, Saturdays and Sundays could be normal working days, hence with potentially irregular absences from work.

The result above is in line with a reduction in strategic behavior in sick leave claims. Figure A12 in the Appendix extends the analysis to the number of absences by day of the week in the 16 months after the experiment. Coefficients are generally uniform over the week but increasing in magnitude, with the largest effects observed on Fridays and Saturdays (limited to workers found irregularly on leave).

We also report coefficients for our baseline OLS regression in Table A7 in the Appendix for the effects at 16 months after the experiment on the number of days on sick leave, the

[33] Figure A13 in the Appendix shows that certificates start more frequently on Mondays and end more frequently on Fridays.

number of certificates and the average certificate duration. Estimates point once more to larger effects for workers irregularly on leave.

6.2 Implicit sanctions and irregular leaves

As documented in Section 5.4, HVs do not seem to have an effect on careers of the workers involved. Is this still true when we focus specifically on workers found to be irregularly on sick leave?

Figure 8 replicates our analysis of career effects of HVs, by allowing this time for the two types of visit outcomes. Panel (a) reports results for monthly earnings in the public sector, Panel (b) looks at the probability of employment in the public sector, Panel (c) at the probability of being a manager and Panel (d) at the impact on the occupation earnings. Workers found regularly on sick leave experience no change in their career after the visit while workers found irregularly on leave face penalties in terms of wage reductions (Panel a), higher non-employment probability (Panel b), and occupation earnings (Panel d), although the latter effect is not statistically significant. No impact can be detected on the probability of becoming a manager (Panel c). The wage penalty for workers found irregularly on leave widens over time, reaching a peak of about 100 Euros per month one year after the experiment, and then it goes down to about 50 Euros losing statistical significance at the end of the observation period. Similarly, the probability of non-employment in the public sector increases up to 2.5 percentage points one year after the experiment. Also in this case, the effect fades away at the end of the observation period. The cumulative wage loss for workers found irregularly on leave amounts to about 500 Euros or 1.4% of the average take home pay in the 16 months after the experiment (about 35,000 Euros).[34]

Overall, the effects of HVs on future sick claim are stronger among workers who acquire new information about sanctions and the profile of the decline of future sick leave claims is consistent with a reduction of opportunistic behavior. At the same time, the implicit

[34] We cumulate only the effects referred to coefficients significant at least at the 10% level.

sanctions that could be possibly ascertained among workers detected to be irregularly on leave are fairly small. As argued above, workers being irregularly on sick leave face a sort of double lottery: the first lottery concerns the assignment of audits; the second lottery is related to the level of the sanctions applied by their supervisors. In presence of risk aversion, this double lottery can magnify the deterrence effects of HVs.[35] It is also possible that workers are induced by HVs to overestimate the HV probability in case of further absences from work, in line with *scarecrow effects* (Bérgolo et al., 2019). Insofar as workers do not know the probability distribution of sanctions, uncertainty aversion may also play a role in this context. Moreover, workers could be facing additional sanctions in terms of reputation, hostility from colleagues who had to make up for their absence, and face a higher workload once back to work, all effects that we are not able to investigate with the current data.

7 Cost-effectiveness of home visits

Our results indicate that HVs reduce the use of sick leave among audited workers, notably among those irregularly on leave. This does not necessarily imply that HVs are desirable from a public finance perspective. Indeed, sending doctors to visit workers comes at a monetary cost which may well outweigh gains in terms of lower sickness benefit expenditure. Assessing the overall impact of this work-intensive monitoring system appears therefore of

[35] Intuitively, suppose that worker's utility is given by $u(c,a) = log(c) + a\Gamma$ where c is consumption, and a is absence from work that can take only two values: 1 if the worker is on leave (and healthy) and 0 otherwise. Γ measures the utility of the individual who can "get away with it", that is, is not detected in his opportunistic behavior by the imperfect HV technology. Consider for simplicity that the wage is the only source of income of the worker. A regular worker would therefore enjoy $u(w,0) = log(w)$ while the expected utility of a worker irregularly on leave will be $u(w,1) = (1-d)(log(w)+\Gamma)) + dlog(w^l)$ where d is the detection probability, and $\sigma == w - w^l < 0$ is the sanction necessary to deter opportunistic behavior. If the premium (penalty for irregular absentees) is non-stochastic, the worker will not take an irregular leave as long as

$$log(w) - log(w^l) \geq \frac{1-d}{d}\Gamma \tag{3}$$

In other words, the percentage wage increase granted to workers in order to deter misbehavior is decreasing in the detection probability and increasing in the utility of "getting away with it". In this context, ex-ante uncertainty as to the actual sanction in case of misbehavior is akin to uncertain detection, and the second derivative of the sanction (premium) with respect to the detection probability is increasing. This implies that a mean preserving spread of the distribution of potential sanctions will also deter misbehavior because $\frac{1}{2}(log(w^l+k) + log(w^l-k)) < log(w^l)$ for any $k > 0$. Put it another way, an expected sanction preventing opportunistic behavior can be lower than a non-stochastic sanction.

paramount importance.

We perform a simple back of the envelope cost-benefit analysis taking the standpoint of the social security administration. First, we collect information on the administrative costs of a HV: the cost of a single visit is generally contained, ranging between 25 and 50 Euros depending on the distance between the residence of the worker and the office of the doctor assigned to the visit. To be conservative, we use the upper bound of the cost per visit. We then consider the savings for each visit. By using our main coefficient reported in Column (1) of Table 4, which implies a reduction of 3.6 days over 16 months for each visit, an average salary of 86.3 Euro per day, and a 100% replacement rate (provided in the first 9 months of leave), this implies a reduction in expenditure for the Social Security of about 311 Euros per visit. Thus, net savings are in the order of 261 (311-50) Euros per visit or involve a 6.2 lower expenditure per Euro spent (5.2 net savings) for a total of about 1.1 million Euros net surplus for the social security from all HVs performed during the experiment.

Assuming that detection technology costs and probabilities, treatment effects, and number of visits per day are all constant throughout the year, we can compute the average gain for the public sector. Running the HVs has a total cost of about 2.6 million Euros. Our effect implies a decline in sick leave expenditure of about 16.4 million Euros. The net gain for social security is then about 13.8 million Euros per year.

These effects are based on a 20% detection probability for irregular behavior. The application of Savio algorithm in the private sector, described in Section 2, can provide some indication of what would happen by targeting individuals more likely to be irregularly on leave. Boscarino et al. (2018) document that visits assigned by the algorithm had a detection rate close to 40%, i.e., twice as large as the detection probability in the case of random assignments. The average impact of HV would then be about 4.4 days of sick leave and each Euro spent on visits would lead to 7.6 Euros saved.[36] Total net benefits

[36]This is obtained as the weighted average between the impact for workers regularly on leave and workers irregularly on leave: $S_{irr}\beta_{irr} + S_{reg}\beta_{reg} = 0.4 * 7.94 + 0.6 * 2.05 \approx 4.4$, where S_irr is the share of irregular outcomes under targeting, and β_{irr} and β_{reg} are the coefficients for the effect of HVs on future days on sick leave in the following 16 months.

would be about 17.5 million Euros.

Our simple cost-benefit test does not consider several factors. First, we did not include the gains in terms of productivity and services offered to citizens as a consequence of lower absence rates. Assuming that wages are in line with productivity, these additional benefits could amount to about 86 Euros per day, the average daily wage of workers in our sample. Second, a more comprehensive assessment of costs could consider the expenses related to the administrative personnel managing the assignment of certificates. This cost, however, is likely to be fairly modest: indeed, most of the administrative costs come from an infrastructure which would be in place anyway to manage the ECVs paid by the private employers. Third, we did not consider costs that other workers would face to reorganize their tasks to cope with the absence of their colleague as well as spillovers on coworkers, who might be deterred to claim sick leave after observing a colleague being inspected. These externalities cannot be estimated with our data.

Our back of the envelope calculation does not consider the costs of distortions induced by revenue collection. We provide a measure of the welfare effects of HVs by computing the *Marginal Value of Public Funds* (MVPF) (Hendren and Sprung-Keyser 2020; Finkelstein and Hendren 2020). This appears particularly attractive as it relates budgetary effects to utility gains of the individuals involved and provides a consistent framework for comparisons across policies. The computation of the MVPF requires two key items. On the one hand, an evaluation should be made about how much the beneficiaries would be willing to pay in order to finance the extra spending, that is their willingness to pay for the policy (numerator); on the other hand, the net cost of the policy for the government (denominator) should be ascertained. We start by considering the denominator of the MVPF, that is, the net cost of the government for one Euro expenditure on HVs. As discussed above, spending an additional Euro on HVs involves a net reduction in expenditure of 5.2 Euros. Next, we turn to the monetary value that inspected workers assign to public expenditure on HVs. We can assume that workers would be willing to pay the entire reduction in gross expenditure in sickness benefits associated to one Euro of HVs, that is, 6.2 Euros. Thus, the implied MVPF is -6.2/-5.2 or about 1.2. This is well

below many of the estimates for the MVPF of taxes in the US: as reported by Hendren and Sprung-Keyser (2020), they range between 1.16 and 44.23 and the average MVPF for the top income tax is about 3. Hence, HVs appear to involve fewer distortions than standard revenue raising policy tools.

Drawing on the results in Section 5.6, we compute MVPFs at different levels of the earning distribution. As shown in Panel (b) of Figure 6 and in greater detail in Table A8 in the Appendix, the MVPF decreases from around 1.3 in the first quintile to 1.1 in the last quintile. As long as welfare weights are not increasing in wages (they are actually declining in standard welfare functions), this suggests that it is optimal to target workers with high earnings.

8 Conclusions

In this paper, we analyze the results of a randomized control trial for home visits checking on sickness benefit claims in the public sector. Our experiment concerns the universe of public employees in Italy, a population of about 3.2 million workers. The national scale of the project allow us to take into account of potential spillovers of the treatment. The scalability of our results is supported by the fact that we evaluate the effects of procedures that can be readily replicated in different institutional contexts. We draw on unique administrative data on sick leave and work histories in the public sector. We find that receiving a HV decreases by about 3.6 days the duration of sick leave over the 16 months following the experiment. The effect is persistent and survives to a large number of robustness checks

We consistently find that the reductions in future leave claims are stronger among workers found to be irregularly on leave than among workers whose GP certificate is confirmed at the visit. Furthermore, reductions are stronger on Fridays and Saturdays, and involve mostly certificates sent on Mondays and ending during the weekend, possibly extending week-end holidays. This, together with the fact that reductions of sick leave claims among workers irregularly on leave act also on the number of certificates and not only on their

duration, is consistent with the idea that HVs operate as a deterrent to abuse of sick leave.

Interestingly, such a deterrence operates even if there are no strict rules concerning the sanctions to be applied to workers irregularly on leave and those implicit sanctions that we capture by looking at work histories after HVs are fairly small. This may be due to the fact that workers face a sort of double lottery and do not know the probability distribution of the sanctions even when they have rational expectations about the detection probability.

Our cost-benefit and welfare analysis of the effects of HVs suggest that audits can be welfare improving, and even more so if targeted to certificates most likely to be associated to social security abuse and to workers at the top of the earning distribution. As far as we know, only Boning et al. (2023) had so far estimated the cost-effectiveness of policies enforcing tax compliance along the earning distribution.

Our findings are particularly relevant at a time in which countries are extending paid sick leave based on a reconsideration of epidemiological risk after the Covid-19 pandemic. Evidence on policy tools to tackle moral hazard in the use of sick leave (and potentially other health related transfers such as disability benefits) could provide useful guidance in designing more balanced and effective systems, whose costs are far from negligible.

References

Adams-Prassl, A., Boneva, T., Golin, M., and Rauh, C. (2023). The value of sick pay. *European Economic Review*, 151:104314.

Advani, A., Elming, W., and Shaw, J. (2017). The Dynamic Effects of Tax Audits. *The Review of Economics and Statistics*, pages 1–45.

Allingham, M. G. and Sandmo, A. (1972). Income tax evasion: A theoretical analysis. *Journal of public economics*, 1(3-4):323–338.

Alpino, M., Hauge, K. E., Kotsadam, A., and Markussen, S. (2022). Effects of Dialogue Meetings on Sickness Absence — Evidence from a Large Field Experiment. *Journal of Health Economics*, 83:102615.

Athey, S. and Imbens, G. W. (2017). The econometrics of randomized experiments. In *Handbook of economic field experiments*, volume 1, pages 73–140. Elsevier.

Bergeron, A., Tourek, G., and Weigel, J. (2021). The State Capacity Ceiling on Tax Rates: Evidence from Randomized Tax Abatements in the DRC. *Job Market Paper*.

Bérgolo, M. L., Ceni, R., Cruces, G., Giaccobasso, M., and Perez-Truglia, R. (2019). Tax Audits as Scarecrows: Evidence from a Large-scale Field Experiment. *IZA DP No. 12335*.

Bertoni, M., Brunello, G., De Benedetto, M. A., and De Paola, M. (2021). Does Monitoring Deter Future Cheating? The Case of External Examiners in Italian Schools. *Economics Letters*, 201:109742.

Bertoni, M., Brunello, G., and Rocco, L. (2013). When the Cat is Near, the Mice won't Play: The Effect of External Examiners in Italian Schools. *Journal of Public Economics*, 104:65–77.

Böckerman, P., Kanninen, O., and Suoniemi, I. (2018). A Kink that Makes You Sick: The Effect of Sick Pay on Absence. *Journal of Applied Econometrics*, 33(4):568–579.

Boning, W. C., Hendren, N., Sprung-Keyser, B., and Stuart, E. (2023). A welfare analysis of tax audits across the income distribution. *NBER Working paper No. 31376*.

Boscarino, R., di Porto, E., and Naticchioni, P. (2018). SAVIO Shut Down: Effetti sulle Visite Mediche di Controllo. *INPS DCSR Studi e Analisi, Nota n. 2*.

Colonnelli, E., Prem, M., and Teso, E. (2020). Patronage and selection in public sector organizations. *American Economic Review*, 110(10):3071–99.

Correia, S. (2019). REGHDFE: Stata Module to Perform Linear or Instrumental-Variable Regression Absorbing any Number of High-Dimensional Fixed Effects. *Boston College Department of Economics*.

Cronin, C. J., Harris, M. C., and Ziebarth, N. R. (2022). The Anatomy of US Sick Leave Schemes: Evidence from Public School Teachers. *NBER Working paper No. 29956*.

De Neve, J.-E., Imbert, C., Spinnewijn, J., Tsankova, T., and Luts, M. (2021). How to Improve Tax Compliance? Evidence from Population-Wide Experiments in Belgium. *Journal of Political Economy*, forthcoming.

De Paola, M., Scoppa, V., and Pupo, V. (2014). Absenteeism in the Italian Public Sector: The Effects of Changes in Sick Leave Policy. *Journal of Labor Economics*, 32(2):337–360.

DeBacker, J., Heim, B., Tran, A., and Yuskavage, A. (2020). Tax noncompliance and measures of income inequality. *Tax Notes*.

DeBacker, J., Heim, B. T., Tran, A., and Yuskavage, A. (2018). Once bitten, twice shy? The lasting impact of enforcement on tax compliance. *The Journal of Law and Economics*, 61(1):1–35.

D'Amuri, F. (2017). Monitoring and Disincentives in Containing Paid Sick Leave. *Labour Economics*, 49:74–83.

Fenizia, A. (2022). Managers and productivity in the public sector. *Econometrica*, 90(3):1063–1084.

Ferber, R., Hirsch, W. Z., et al. (1982). Social experimentation and economic policy. *Cambridge Books*.

Ferman, B., Torsvik, G., and Vaage, K. (2021). Skipping the Doctor: Evidence from a Case with Extended Self-Certification of Paid Sick Leave. *Journal of Population Economics*, pages 1–37.

Finkelstein, A. and Hendren, N. (2020). Welfare Analysis Meets Causal Inference. *Journal of Economic Perspectives*, 34(4):146–67.

Gemmell, N. and Ratto, M. (2012). Behavioral Responses to Taxpayer Audits: Evidence from Random Taxpayer Inquiries. *National Tax Journal*, 65(1):33–57.

Gittleman, M. and Pierce, B. (2012). Compensation for state and local government workers. *Journal of Economic Perspectives*, 26(1):217–42.

Godøy, A. and Dale-Olsen, H. (2018). Spillovers from Gatekeeping–Peer Effects in Absenteeism. *Journal of Public Economics*, 167:190–204.

Gomes, P. (2018). Heterogeneity and the Public Sector Wage Policy. *International Economic Review*, 59(3):1469–1489.

Guyton, J., Langetieg, P., Reck, D., Risch, M., and Zucman, G. (2021). Tax Evasion at the Top of the Income Distribution: Theory and Evidence. *National Bureau of Economic Research, Working Paper N. 28542*.

Harrison, G. W. and List, J. A. (2004). Field experiments. *Journal of Economic literature*, 42(4):1009–1055.

Hartman, L., Hesselius, P., and Johansson, P. (2013). Effects of eligibility screening in the sickness insurance: Evidence from a field experiment. *Labour Economics*, 20:48–56.

Hebous, S., Jia, Z., Løyland, K., Thoresen, T. O., and Øvrum, A. (2020). Do Audits Improve Future Tax Compliance in the Absence of Penalties? Evidence from Random Audits in Norway. *CESifo Working Paper N. 8480*.

Hendren, N. and Sprung-Keyser, B. (2020). A Unified Welfare Analysis of Government Policies. *The Quarterly Journal of Economics*, 135(3):1209–1318.

Hernæs, Ø. (2018). Activation against Absenteeism–Evidence from a Sickness Insurance Reform in Norway. *Journal of Health Economics*, 62:60–68.

Hesselius, P., Johansson, P., and Nilsson, J. P. (2009). Sick of Your Colleagues' Absence? *Journal of the European Economic Association*, 7(2/3):583–594.

Hesselius, P., Johansson, P., and Vikström, J. (2013). Social Behaviour in Work Absence. *The Scandinavian Journal of Economics*, 115(4):995–1019.

Ichino, A. and Riphahn, R. T. (2005). The Effect of Employment Protection on Worker Effort: Absenteeism during and after Probation. *Journal of the European Economic Association*, 3(1):120–143.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Imbens, G. W. and Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, 47(1):5–86.

INPS (2023). Dtori di lavoro privati e pubblici: dinamiche, incentivi, crisi energetica. *XXII Rapporto annuale*.

Johns, A. and Slemrod, J. (2010). The distribution of income tax noncompliance. *National Tax Journal*, 63(3):397–418.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., and Saez, E. (2011). Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark. *Econometrica*, 79(3):651–692.

Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining freedman's critique. *The Annals of Applied Statistics*, 7(1):295–318.

Lipsky, M. (1980). Street-level bureaucracy: Dilemmas of the individual in public service. *New York: Russell Sage Foundation*, 71.

List, J. A. (2020). Non est Disputandum de Generalizability? A Glimpse into The External Validity Trial. *NBER Working Paper N. 27535*.

List, J. A., Shaikh, A. M., Vayalinkal, A., et al. (2021). Multiple testing with covariate adjustment in experimental economics. *mimeo, The Field Experiments Website*.

Maclean, J. C., Pichler, S., and Ziebarth, N. R. (2020). Mandated Sick Pay: Coverage, Utilization, and Welfare Effects. *NBER Working Paper N. 26832*.

Marie, O. and Vall Castelló, J. (2022). Sick Leave Cuts and (Unhealthy) Returns to Work. *Journal of Labor Economics*, forthcoming.

Markussen, S., Mykletun, A., and Røed, K. (2012). The Case for Presenteeism—Evidence from Norway's Sickness Insurance Program. *Journal of Public Economics*, 96(11-12):959–972.

Markussen, S., Røed, K., and Schreiner, R. C. (2018). Can Compulsory Dialogues Nudge Sick-Listed Workers back to Work? *The Economic Journal*, 128(610):1276–1303.

Nelson, K., Fredriksson, D., Korpi, T., Korpi, W., Palme, J., and Sjöberg, O. (2020). The social policy indicators (spin) database. *International Journal of Social Welfare*, 29(3):285–289.

Olken, B. A. (2007). Monitoring Corruption: Evidence from a Field Experiment in Indonesia. *Journal of Political Economy*, 115(2):200–249.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2):187–204.

Pichler, S., Wen, K., and Ziebarth, N. R. (2020). Positive Health Externalities of Mandating Paid Sick Leave. *IZA Discussion Paper Series N. 13530*.

Pichler, S. and Ziebarth, N. R. (2017). The Pros and Cons of Sick Pay Schemes: Testing for Contagious Presenteeism and Noncontagious Absenteeism Behavior. *Journal of Public Economics*, 156:14–33.

Pomeranz, D. (2015). No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax. *American Economic Review*, 105(8):2539–69.

Pomeranz, D., Marshall, C., and Castellon, P. (2014). Randomized Tax Enforcement Messages: a Policy Tool for Improving Audit Strategies. *Tax Administration Review*, 36(January 2014):1–21.

Scognamiglio, A. (2020). Paid Sick Leave and Employee Absences. *LABOUR*, 34(3):305–322.

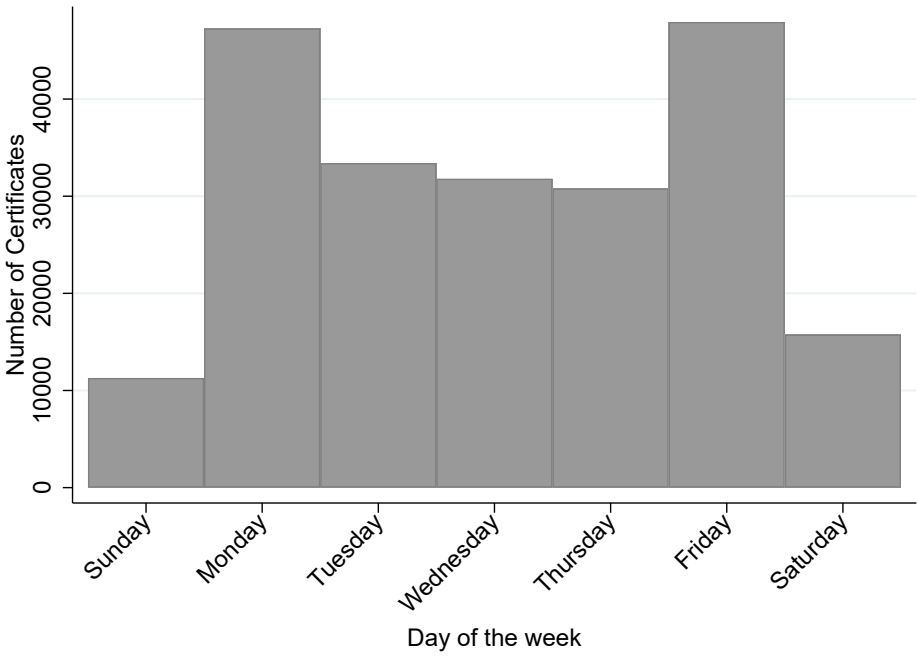
Slemrod, J. (2019). Tax compliance and enforcement. *Journal of Economic Literature*, 57(4):904–54.

Slemrod, J. and Gillitzer, C. (2013). *Tax Systems*. MIT Press.

Ziebarth, N. R. (2013). Long-term absenteeism and moral hazard—evidence from a natural experiment. *Labour Economics*, 24:277–292.

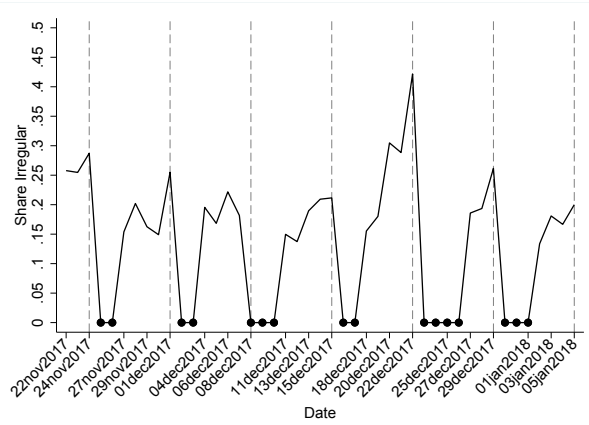
Figures

Figure 1: Certificates by start and end Day of the Week

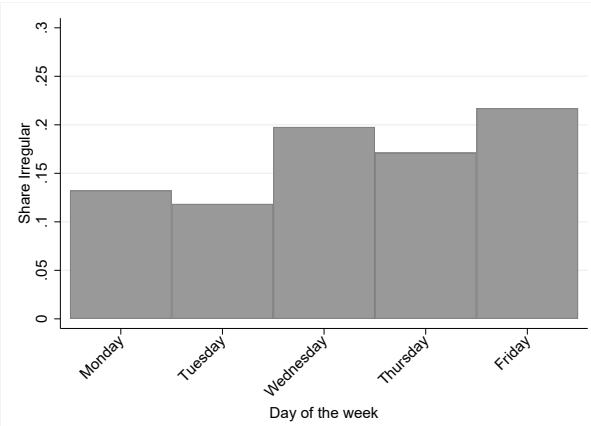


Note: Statistics are based on data for public sector workers involved in the experiment. We consider sick leave spells which started between May and October 2017. The figure reports the number of certificates by start and end day of the week. Each certificate is counted both for the start and end day of the week.

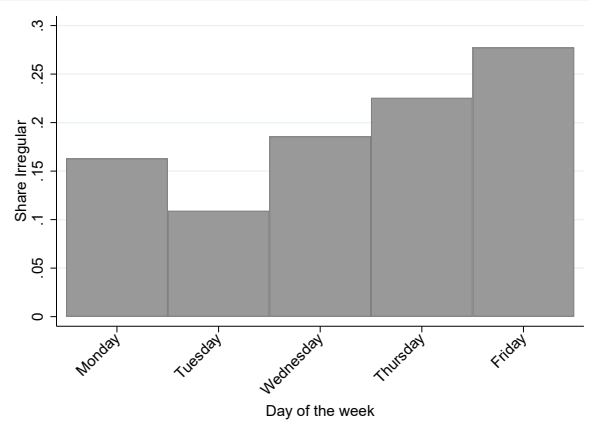
Figure 2: Share of visits with irregular outcome by day of the week, date, and geographic area



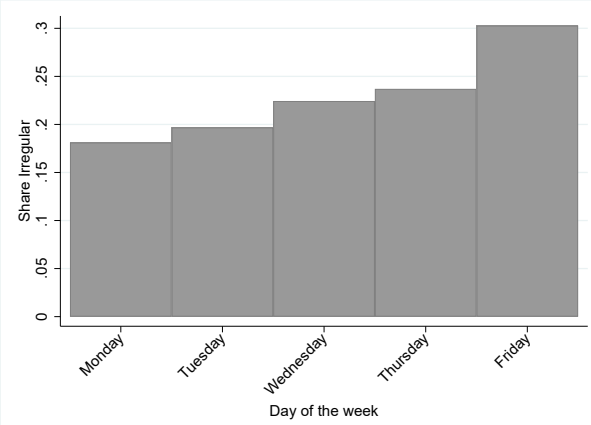
(a) Date



(b) Day of the Week: North



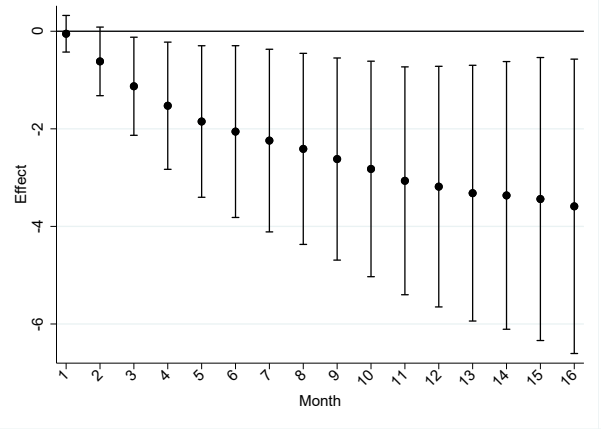
(c) Day of the Week: Centre



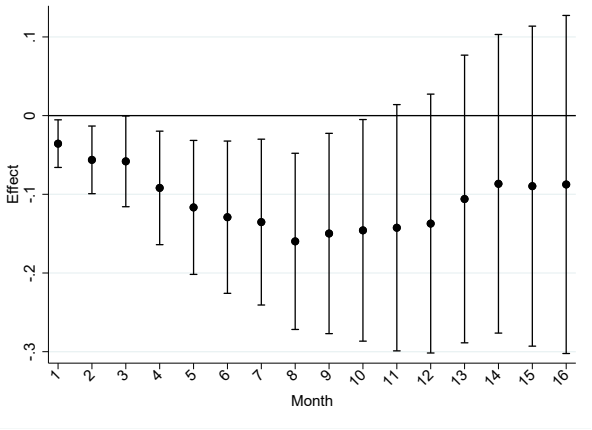
(d) Day of the Week: South

Note: The figures report the share of visits with Irregular outcome in the period of the experiment by date in Panel (a) and by day of the week and geographic area in panels from (b) to (d). During non-working days (mostly Saturday and Sunday), corresponding to the zeros in Panel (a), no INPS called HV was performed due to budgetary reasons.

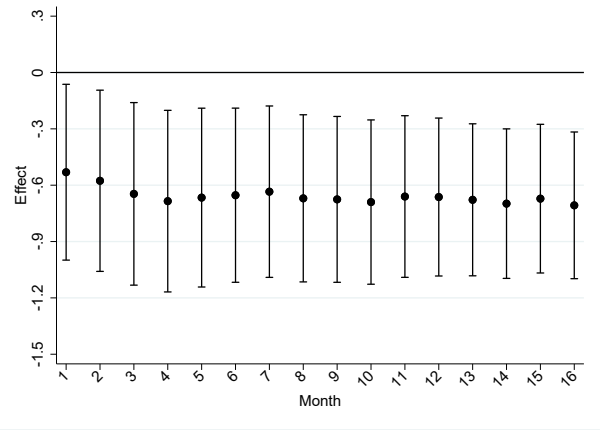
Figure 3: Effect of HV on sick leave over 16 months after the experiment



(a) Cumulative Days on Sick Leave



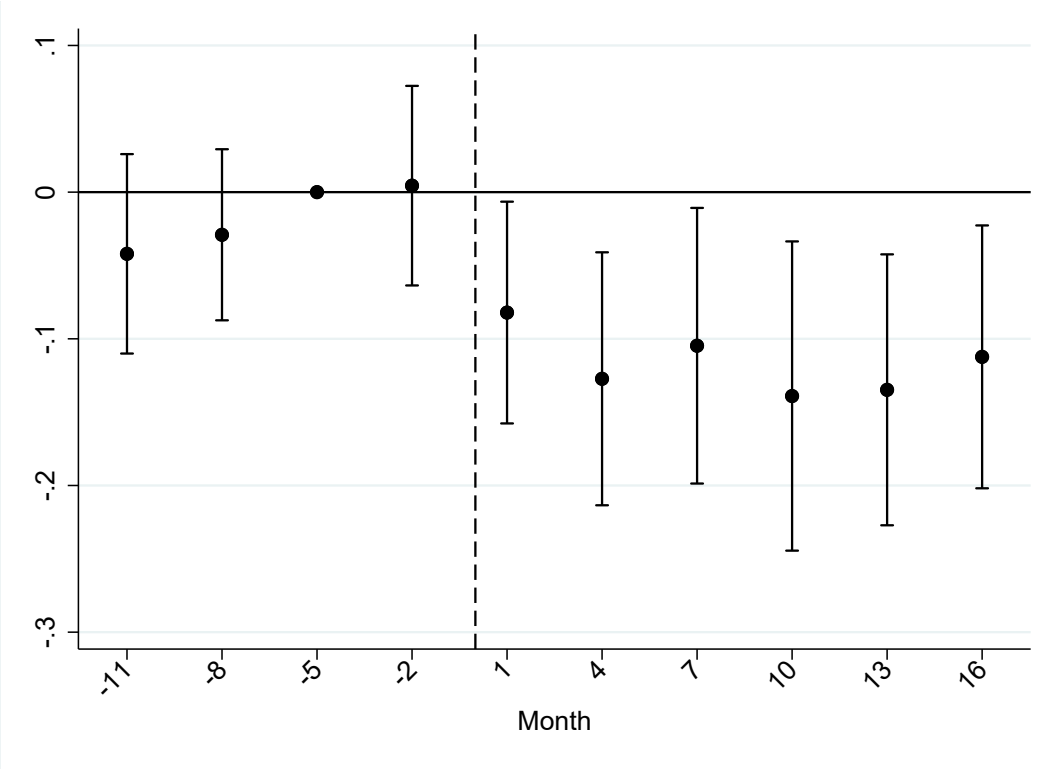
(b) Cumulative Number of Certificates



(c) Average Certificate Duration

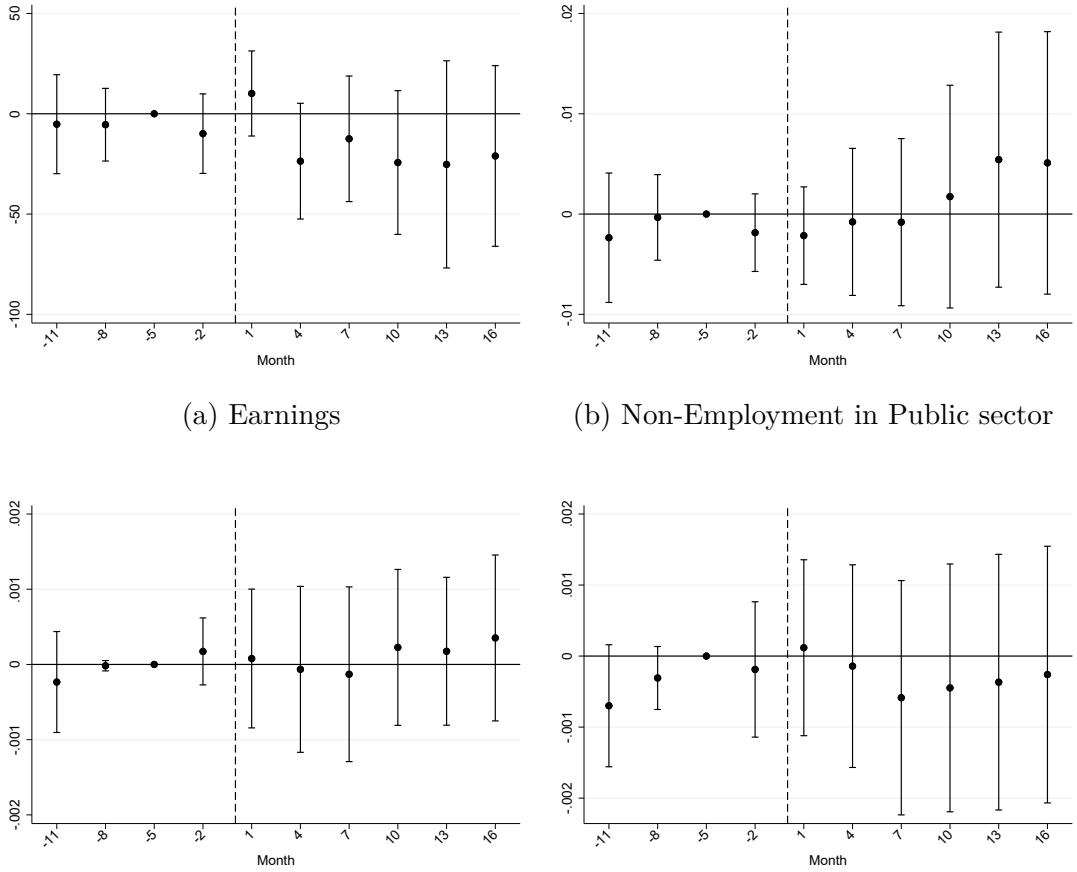
Note: The figure reports estimates for the effect of HV on cumulative number of days spent on sick leave (Panel a), cumulative number of certificates (Panel b), and average duration of certificates (Panel c) in the 16 months after the end of the experiment. The cumulative number of certificates corresponds to the cumulative number of sick leave spells, and the average certificate duration corresponds to the average duration of sick leave spells. Regression run separately for each month, and they include a dummy for receiving HV, demographic characteristics (gender, age category dummies), job characteristics (sector dummies, average salary in the 6 months before the experiment, part-time dummy, dummy for permanent contract), fixed effects at INPS local office level, number of days spent on sick leave in the period of the experiment (11/22/2017-01/05/2018), number of certificates, average duration, number of days spent on sick leave in the 6 months before the experiment, and duration first certificate. Standard errors clustered at local office level. Coefficient and 95% confidence interval of HV dummy reported.

Figure 4: Difference in differences effect on number of days on leave



Note: The figure reports effect of HV interacted with its leads and lags together with 95% confidence interval. The control group is workers not subject to HV during the period of the experiment. Regression includes periods from 12 months before the November 2017 up to 16 months after January 2018. Months grouped in three months periods for presentation purpose (reported at the mid month of the three-month period, e.g. the coefficient for 1 covers coefficients for months from 0 up to 2). Reference period is the period between -4 and -6 months before November 2017. The period zero corresponds to November 2017. Regression includes INPS local office fixed effects (at the time of the experiment), time spent on leave in the period of the experiment (dummies), both interacted with month groups fixed effects, as well as individual fixed effects. Regression estimated using Poisson Count regression. Standard errors clustered at INPS local office level.

Figure 5: Difference in differences estimates for workers' career



(a) Earnings

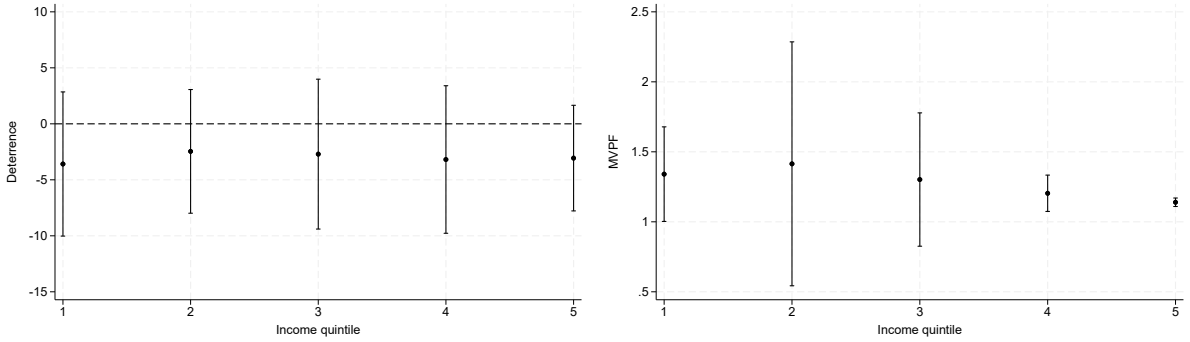
(b) Non-Employment in Public sector

(c) Manager

(d) Occupation Earnings

Note: The figure reports effects of HV and of its leads and lags on earnings in the public sector, non-employment in the public sector, probability of being a manager and average earnings in the current job with respect to the job at the time of the experiment together with 95% confidence interval. Regression includes periods from 12 months before the November 2017 up to 16 months after January 2018. Months grouped in three months periods for presentation purpose (reported at the mid-month of the three-month period, e.g. the coefficient for 1 covers coefficients for months from 0 up to 2). The period zero corresponds to November 2017. Reference period is the period between -4 and -6 months before November 2017. Regression includes INPS local office fixed effects (at the time of the experiment), time spent on leave in the period of the experiment (dummies), both interacted with month groups fixed effects, as well as individual fixed effects. Standard errors clustered at INPS local office level.

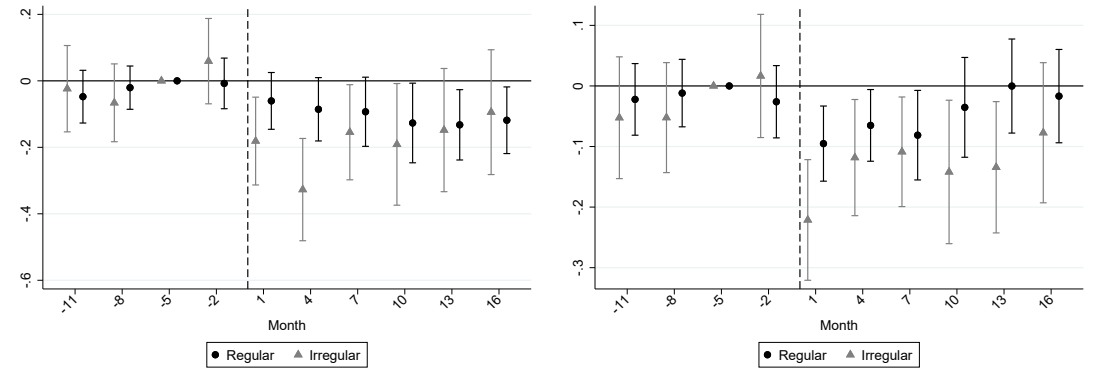
Figure 6: Effect of HV and MVPF by income quintile



(a) Impact HV: Quintile Income (b) MVPF: Quintile Income

Note: The figure reports estimates for the effect of HV on cumulative number of days on sick leave use in the 16 months following the experiment in Panel (a) and the associated marginal value of public funds in Panel (b). Regressions estimated separately for each category and reference income computed as the average monthly positive earnings between January and November 2017 (0 excluded). Regressions include a dummy for receiving HV and its interaction with relevant dummies, demographic characteristics (gender, age), job characteristics (sector dummies, average salary before the experiment, part-time dummy, dummy for permanent contract), fixed effects at INPS local office level, number of days spent on sick leave in the period of the experiment (11/22/2017-01/05/2018), number of certificates, average duration, and number of days spent on sick leave in the 6 months before the experiment. Metropolis are the two main administrative center in Italy: Rome (the capital of the country) and Milan (the main economic center). Standard errors clustered at local office level. Coefficients and 95% confidence intervals reported. Confidence intervals for the MVPF computed by using the delta method.

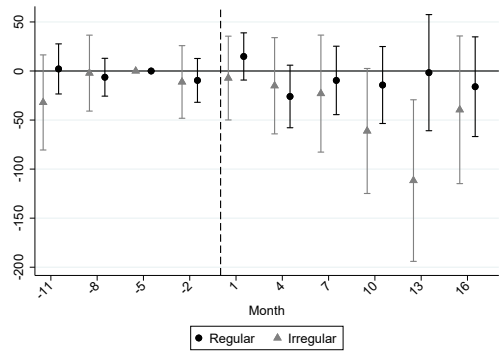
Figure 7: Difference in differences for number of days on sick leave and number of certificates: regular and irregular.



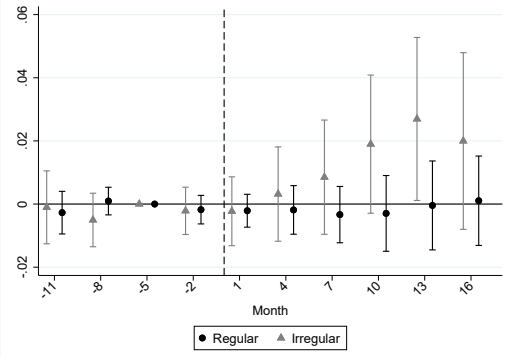
(a) Days on sick leave (b) Number of certificates

Note: The figure reports effect of HV for workers found regularly and irregularly on leave interacted with theirs leads and lags together with 95% confidence interval. Regressions estimated using Poisson Count regression and use, as dependent variables, the number of days on sick leave per month in Panel (a) and the number of certificates (sick leave spells) in Panel (b). Regressions include periods from 12 months before the November 2017 up to 16 months after January 2018. Months grouped in three months periods for presentation purpose (reported at the mid-month of the three-month period, e.g. the coefficient for 1 covers coefficients for months from 0 up to 2). The period zero corresponds to November 2017. Reference period is the period between -4 and -6 months before November 2017. Regression includes INPS local office fixed effects (at the time of the experiment), time spent on leave in the period of the experiment (dummies), both interacted with month groups fixed effects, as well as individual fixed effects. Standard errors clustered at INPS local office level.

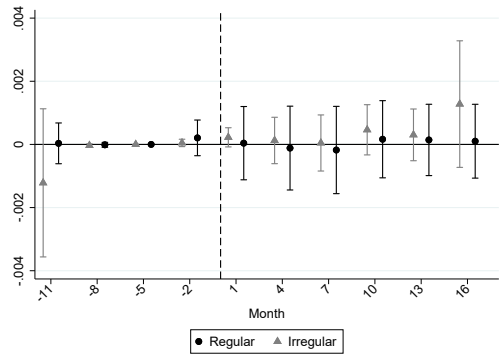
Figure 8: Difference in differences estimates for the impact of HV on workers' career in the publicsSector by visit outcome.



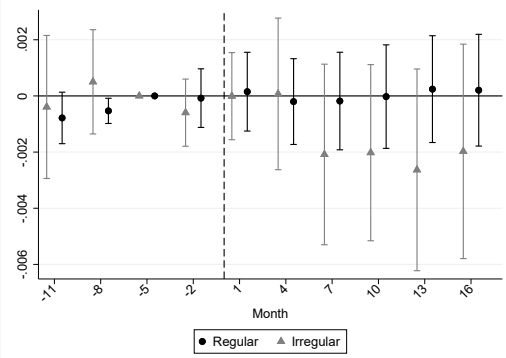
(a) Earnings



(b) Non-Employment in Public sector



(c) Manager



(d) Occupation Earnings

Note: The figure reports effects of HV and of its leads and lags on earnings in the public sector, non-employment in the public sector, probability of being a manager and average earnings in the current job with respect to the job at the time of the experiment together with 95% confidence interval. Regression includes periods from 12 months before the November 2017 up to 16 months after January 2018. Months grouped in three months periods for presentation purpose (reported at the mid month of the three month period, e.g. the coefficient for 1 covers coefficients for months from 0 up to 2). The period zero corresponds to November 2017. Reference period is the period between -4 and -6 months before November 2017. Regression includes INPS local office fixed effects (at the time of the experiment), time spent on leave in the period of the experiment (dummies), both interacted with month groups fixed effects, as well as individual fixed effects. Standard errors clustered at INPS local office level.

Tables

Table 1: Summary statistics at the individual level

Variables	Average	Se	Minimum	Median	Max
Female	0.725	0.446	0	1	1
Age	53.366	8.473	24	55	67
North	0.396	0.489	0	0	1
Center	0.177	0.382	0	0	1
South	0.427	0.495	0	0	1
Metropolis	0.119	0.323	0	0	1
School Sector	0.396	0.489	0	0	1
Central Administration	0.060	0.238	0	0	1
Local Administration	0.234	0.423	0	0	1
Health Sector	0.310	0.463	0	0	1
Permanent Contract	0.948	0.222	0	1	1
Part Time	0.060	0.238	0	0	1
Manager	0.028	0.166	0	0	1
(Log) Average Wage (six month bef. exp)	7.605	0.737	0	8	10
Days on sick leave in following 16 months	45.019	65.556	0	20	483
Certificates in following 16 months	6.083	7.435	0	4	190
Average Certificate duration in following 16 months	6.980	8.457	0	4	92
Days on Leave in Experiment	13.043	12.036	1	8	45
Total certificates (6 months bef.)	2.719	3.156	0	2	59
Total number of days (6 months bef.)	23.376	34.041	0	9	184
Average duration cert. (6 months bef.)	8.388	10.439	0	5	93
Duration first certificate	14.433	12.596	3	10	93
Home Visits and outcome: individual					
Home Visit	0.098	0.297	0	0	1
Home Visit Outcome: Regular	0.078	0.268	0	0	1
Home Visit Outcome: Irregular	0.020	0.140	0	0	1
Home Visits and outcome: certificate					
Certificates subject to Home Visit	0.080	0.271	0	0	1
Outcome Home Visit: Regular	0.064	0.245	0	0	1
Outcome Home Visit: Irregular	0.016	0.125	0	0	1
# Workers	42,707				

Note: Summary statistics at the individual level for public sector employees who had at least one ongoing sick leave certificate in the period of the experiment (11/22/2017-01/05/2018) and were randomly selected in the experiment. Metropolis identifies individuals living in large cities (Milan and Rome) and (log) Mean Monthly Earnings is the log of average earnings in the public sector for the worker from May to October 2017. Number of Certificates (bef. exp.), Number of Days (bef. exp.), and Mean Duration Certificate (bef. exp.) are the number of certificates, the total number of days on sick leave, and the average duration of certificates in the 6 months before the experiment.

Table 2: Descriptive for previous days on leave and irregularity

VARIABLES	(1) Past 6 months	(2) Irregular outcome
Female	-0.454	-0.045**
	(0.462)	(0.019)
Age: 36-40	4.137***	-0.032
	(0.917)	(0.052)
Age: 41-45	4.287***	-0.019
	(0.805)	(0.053)
Age: 46-50	6.114***	-0.053
	(0.826)	(0.053)
Age: 51-55	7.427***	-0.047
	(0.877)	(0.050)
Age: 56-60	8.266***	-0.031
	(0.918)	(0.047)
Age: 61-65	11.951***	-0.023
	(0.932)	(0.050)
Age: 66-67	17.341***	-0.019
	(1.554)	(0.061)
Street Level Bureaucrat	0.759	0.006
	(0.493)	(0.019)
Central Administration	6.963***	-0.036
	(1.369)	(0.037)
Local Administration	-0.838	0.004
	(0.663)	(0.023)
Health Sector	0.009	-0.022
	(0.479)	(0.014)
Permanent Contract	9.636***	-0.077*
	(0.687)	(0.045)
Part Time	1.101	-0.028
	(1.085)	(0.026)
Manager	0.962	-0.019
	(1.359)	(0.031)
(Log) Average Wage (six month bef. exp)	-1.552***	0.003
	(0.322)	(0.006)
Total certificates (6 months bef.)		0.001
		(0.004)
Total number of days (6 months bef.)		0.000
		(0.000)
Average duration cert. (6 months bef.)		-0.001**
		(0.001)
Metropolis	-2.312**	0.056**
	(1.083)	(0.028)
Center	-1.114	0.003
	(1.402)	(0.026)
South	2.365	0.066***
	(1.537)	(0.021)
Duration Certificate: 1-4		0.550***
		(0.064)
Duration Certificate: 5-7		0.234***
		(0.042)
Duration Certificate: 8-9		0.065*
		(0.038)
Friday		0.060***
		(0.017)
Observations	39,507	4,086
R-squared	0.021	0.062
Mean Dep		.199

Note: The table reports estimates of OLS regression with dependent variable the number of days in the six months before the experiment and linear probability model with dependent variable one if the worker is found irregularly on sick leave and zero otherwise. Column (1) uses data at indivudal level for both treated and control workers, while Column (2) uses data for visits to workers and their outcomes. Regressions are estimated by OLS with *reghdfe* stata command developed by (Correia, 2019). Metropolis is a dummy taking value one for Rome (the capital of the country) and Milan (the major economic centre of the country). (log) Mean Monthly Earnings is the log of average earnings in the public sector for the worker from May to October 2017. Number of Certificates (bef. exp.), Number of Days (bef. exp.), and Mean Duration Certificate (bef. exp.) are the number of certificates, the total number of days on sick leave, and the average duration of certificates in the 6 months before the experiment. Sede fixed effects are INPS local office fixed effects, while Date fixed effects are fixed effects for the day in which the HV was performed. Sample size excluding singletons reported. Standard errors clustered at local office level. Level of significance: 0.1 *, 0.05 **, 0.01 ***.

Table 3: Normalized differences for individual characteristics of treated and control workers

Variable	(1) Mean Treated	(2) Mean Control	(3) T-Stat	(4) T-Stat adjusted	(5) Normalized Difference
Female	0.740	0.724	2.147	3.574	0.037
Age	54.111	53.285	4.206	-0.198	0.098
North	0.380	0.398	-0.485	0.734	-0.036
Center	0.176	0.177	-0.055	0.152	-0.003
South	0.444	0.425	0.557	-0.834	0.038
Metropolis	0.103	0.120	-1.097	-0.173	-0.055
School Sector	0.447	0.391	4.852	1.542	0.114
Central Administration	0.066	0.059	1.293	1.120	0.029
Local Administration	0.196	0.238	-3.194	-1.291	-0.102
Health Sector	0.291	0.312	-1.974	-1.011	-0.046
Permanent Contract	0.968	0.946	5.941	2.036	0.108
Part Time	0.049	0.062	-2.031	-1.252	-0.054
Manager	0.037	0.028	2.924	1.078	0.053
(Log) Average Wage (six month bef. exp)	7.606	7.605	0.068	-0.093	0.001
Days on Leave in Experiment	25.132	11.730	33.724		1.108
Total certificates (6 months bef.)	3.643	2.618	13.633	6.293	0.326
Total number of days (6 months bef.)	42.663	21.281	22.244	6.832	0.564
Average duration cert. (6 months bef.)	15.015	7.669	26.979	7.798	0.640
Duration first certificate	25.302	13.252	30.222	8.914	0.934

Note: The table reports t-statics for the difference between the characteristics of treated and controls at baseline (Column 3) and after controlling for the number of days on leave in the experiment (Column 4) and normalized differences (Column 5). Normalized differences computed as $\Delta = \frac{\bar{X}_T - \bar{X}_C}{\left(\frac{(S_T^2 + S_C^2)}{2}\right)^{\frac{1}{2}}}$ and the critical threshold is 0.25. (log) Mean Monthly Earnings is the log of average earnings in the public sector for the worker from May to October 2017.

Table 4: Effect of HV on future use of sick leave

Variables	(1) #Days	(2) #Certificates	(3) Mean duration
HV	-3.589**	-0.088	-0.707***
	(1.539)	(0.110)	(0.199)
Observations	42,657	42,657	42,657
R-squared	0.162	0.410	0.195
Mean Dep	43.335	6.076	6.686
Controls	YES	YES	YES
Past Cert.	YES	YES	YES
Sede FE	YES	YES	YES
Days in the Experiment FE	YES	YES	YES

Note: The table reports estimates for the effect of HV on cumulative days on sick leave in the 16 months after the experiment (# Days in 16 Months), cumulative number of certificates (# Cert in 16 months), and average certificate duration (Mean days in 16 months). The cumulative number of certificates corresponds to the number of sick leave spells and the average certificate duration corresponds to the average duration of sick leave spells. Regressions are estimated with OLS with *reghdfe* stata command developed by (Correia, 2019). HV is a dummy equal to one if the worker was visited in the period of the experiment. Sede FE are local INPS office fixed effects. Past Cert. includes: the number of certificates, the total number of days on sick leave, and the average duration of certificates in the 6 months before the experiment. Controls include: demographic characteristics (gender, and age dummies), and job characteristics (sector dummies, average salary in the 6 months before the experiment, part time dummy, dummy for permanent contract). Mean dep is the average for the dependent variable for individuals who did not receive a HV (i.e. the control group). Sample size excluding singletons reported. Standard errors clustered at local office level. Level of significance: 0.1 *, 0.05 **, 0.01 ***.

Table 5: Impact of HV on relapse

VARIABLES	(1) #Cert: Renewal	(2) #Cert: No Renewal	(3) #Days: Renewal	(4) #Days: No Renewal
HV	-0.049	-0.038	-1.371**	-2.218*
	(0.032)	(0.101)	(0.655)	(1.217)
Observations	42,657	42,657	42,657	42,657
R-squared	0.109	0.401	0.076	0.148
Mean Dep	0.611	5.465	7.743	35.592
Controls	YES	YES	YES	YES
Past Cert.	YES	YES	YES	YES
Sede FE	YES	YES	YES	YES
Days in the Experiment FE	YES	YES	YES	YES

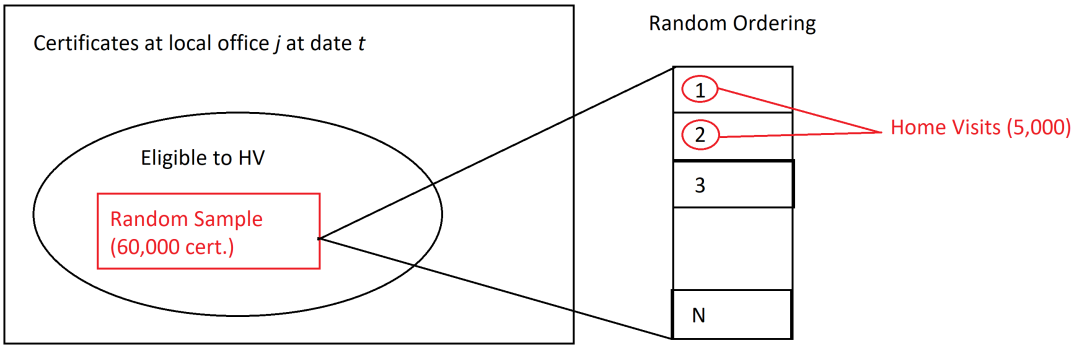
Note: The table reports estimates for the effect of HV on cumulative days on sick leave in the 16 months after the experiment (# Days). No renewals are days on leave for which the certificate is the first certificate with a that start date for the disease while renewals are days on sick leave for which the certificate is not the first one with that start date for the disease. Regressions are estimated with OLS with the *reghdfe* stata command developed by (Correia, 2019). HV is a dummy equal to one if the worker was visited in the period of the experiment. Sede FE are local INPS office fixed effects. Past Cert. includes: the number of certificates, the total number of days on sick leave, and the average duration of certificates in the 6 months before the experiment. Controls include: demographic characteristics (gender, and age dummies), and job characteristics (sector dummies, average salary in the 6 months before the experiment, part time dummy, dummy for permanent contract). Mean dep is the average for the dependent variable for individuals who did not receive a HV (i.e. the control group). Sample size excluding singletons reported. Standard errors clustered at local office level. Level of significance: 0.1 *, 0.05 **, 0.01 ***.

ONLINE APPENDIX

A Figures and Tables

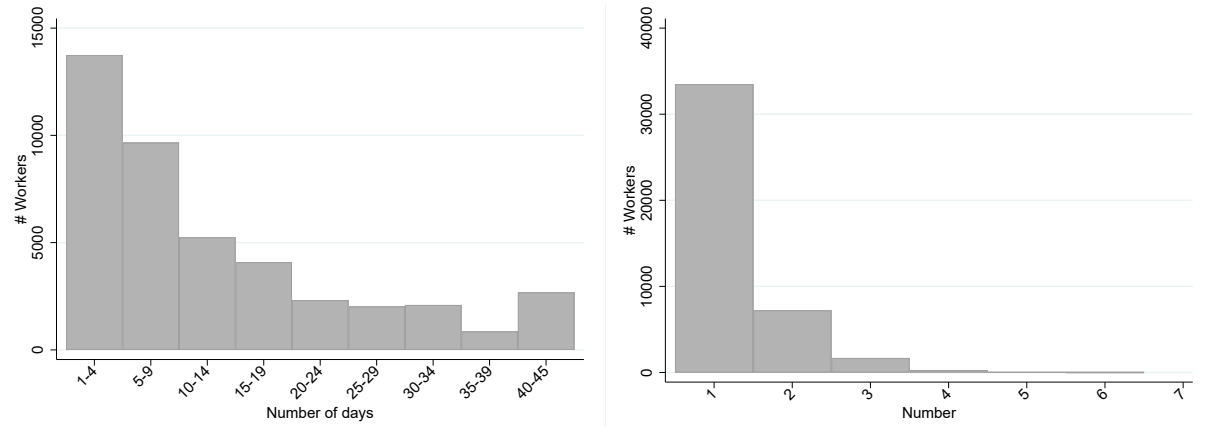
Figures

Figure A1: Structure of the experiment



Note: Description of the structure of the experiment. A random set of medical certificates is drawn at INPS local office (*sede*) j and date t level. Each certificate corresponds to a sick leave spell. Then, exempt certificates are removed (e.g. cancer), and a second random sample is drawn among certificates eligible to HV. Among these certificates, a random order, which determines which certificates will be inspected, is assigned for HV, and, finally, certificates are assigned to doctors to minimize mobility costs.

Figure A2: Number of workers by days spent on sick leave and number of certificates during theexperiment.

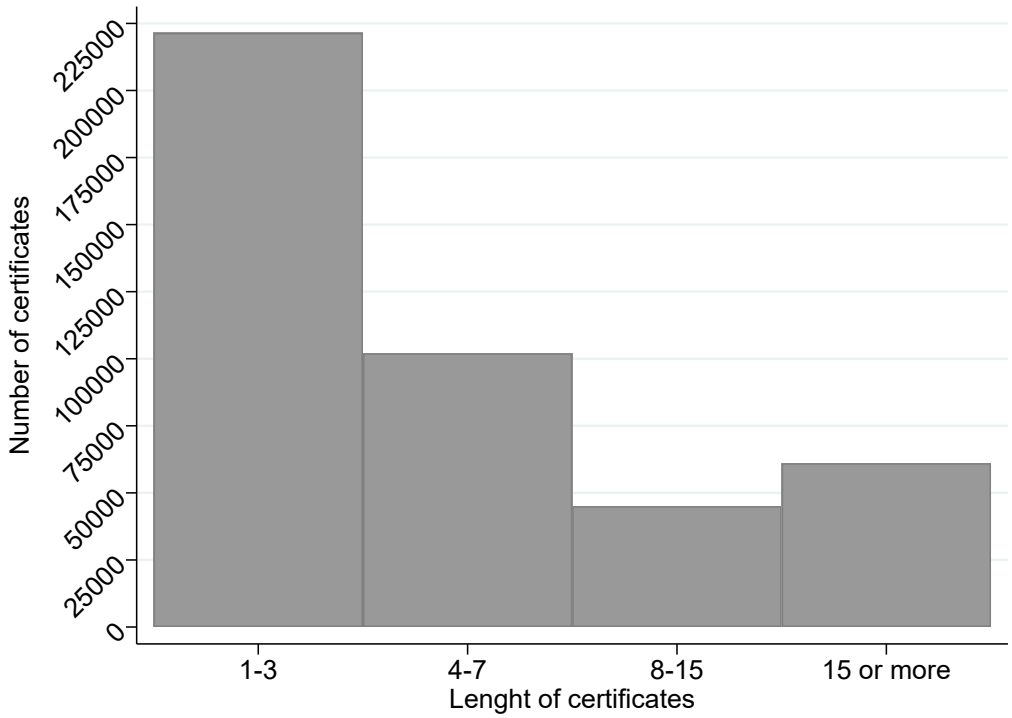


(a) Number of days

(b) Number of certificates

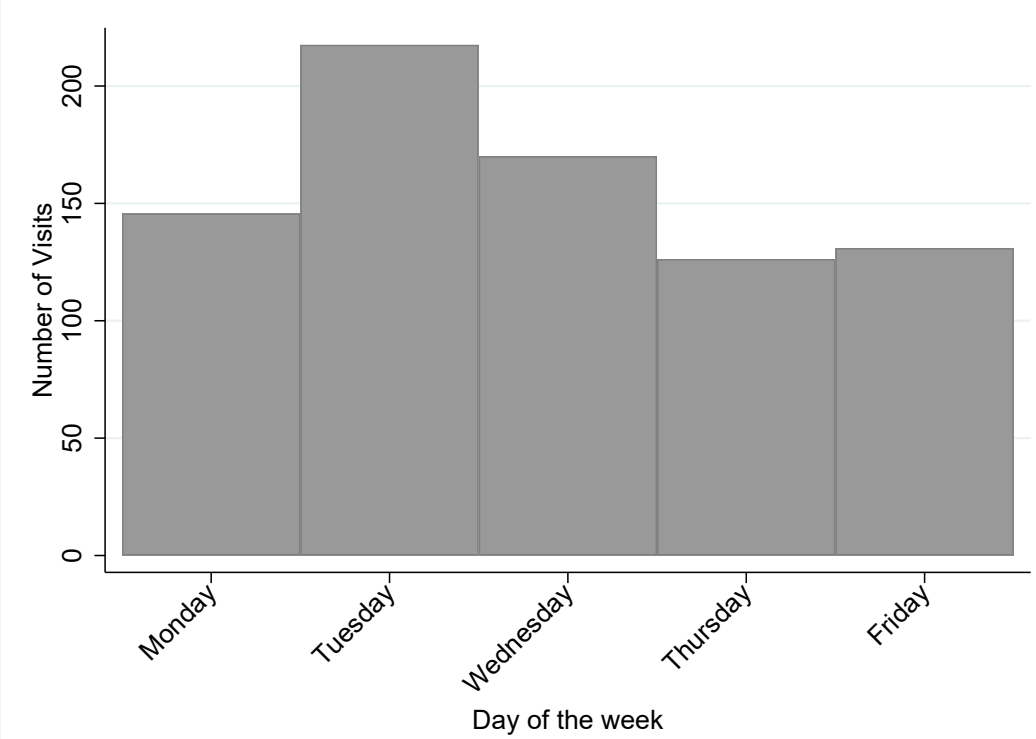
Note: The figure reports the number of individuals by number of days spent on sick leave in the period of the experiment (11/22/2017-1/5/2018) in Panel (a) and by number of certificates sent in the period of the experiment in Panel (b). The number of certificates corresponds to the number of sick leave spells.

Figure A3: Distribution of certificates length after the experiment.



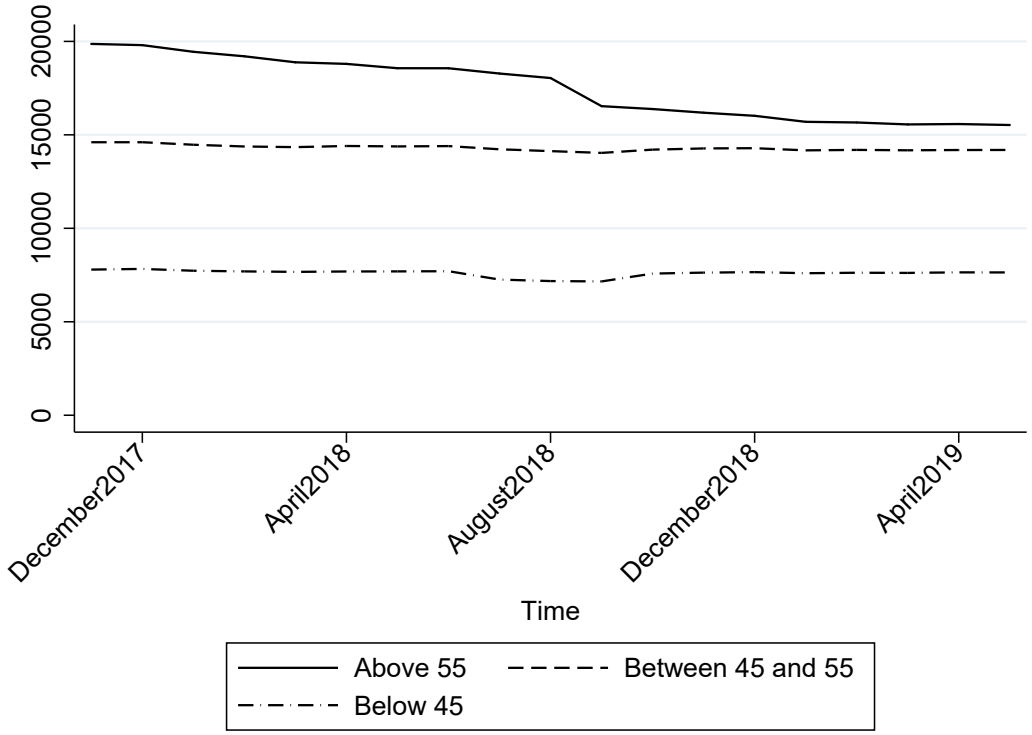
Note: The figure reports the number of certificates claimed by workers sampled in the experiment in the 16 months after the experiment (January 2018-April 2019).

Figure A4: Average number of visits by day of the week.



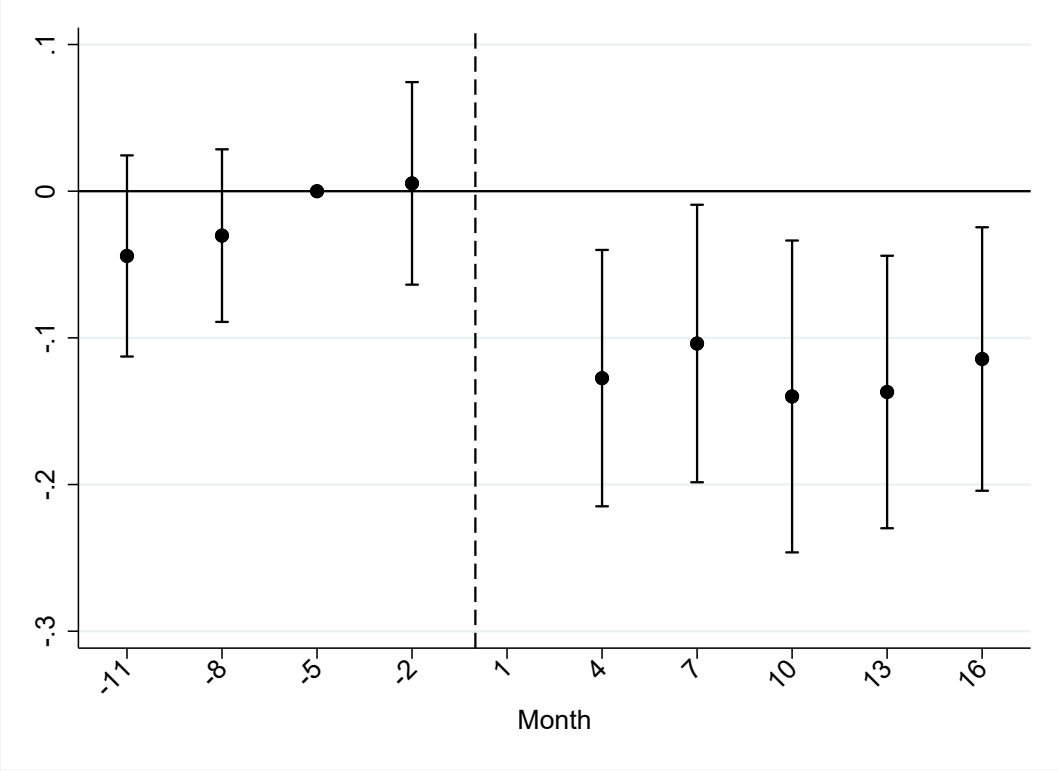
Note: The figure reports the average number of visits by day of the week in the period of the experiment. The figure includes only days in which visits were performed.

Figure A5: Attrition over panel.



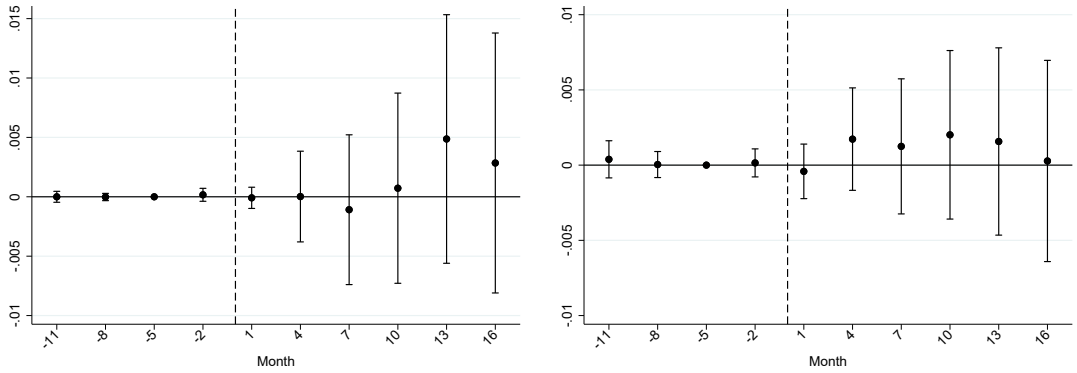
Note: Number of public sector workers still employed at different time horizons by age group.

Figure A6: Difference in differences effect on number of days on leave



Note: The figure reports effect of HV interacted with its leads and lags together with 95% confidence interval. The control group is workers not subject to HV during the period of the experiment. Regression includes periods from 12 months before the November 2017 up to 16 months after January 2018. Months grouped in three months periods for presentation purpose (reported at the mid-month of the three-month period, e.g. the coefficient for 1 covers coefficients for months from 0 up to 2). Reference period is the period between -4 and -6 months before November 2017. The period zero corresponds to November 2017. The sample excludes the months of the experiment between November 2017 and January 2018. Regression includes INPS local office fixed effects (at the time of the experiment), time spent on leave in the period of the experiment (dummies), both interacted with month groups fixed effects, as well as individual fixed effects. Regression estimated using Poisson Count regression. Standard errors clustered at INPS local office level.

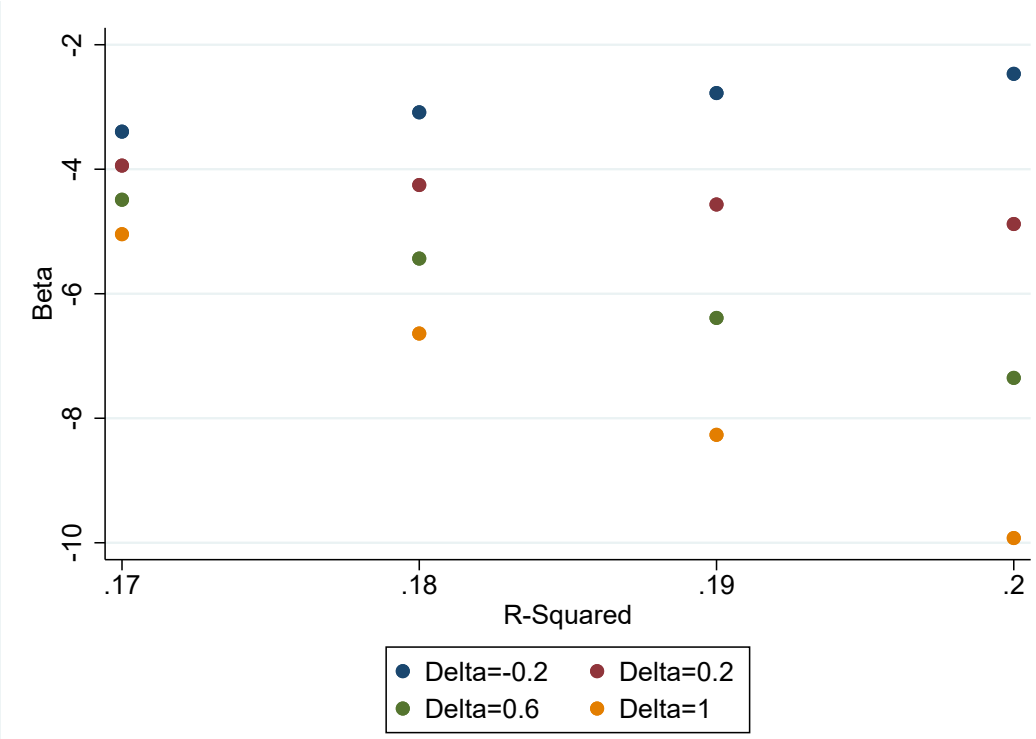
Figure A7: Difference in differences estimates for take up of other benefits other benefits.



(a) Retirement (b) Disability Benefit

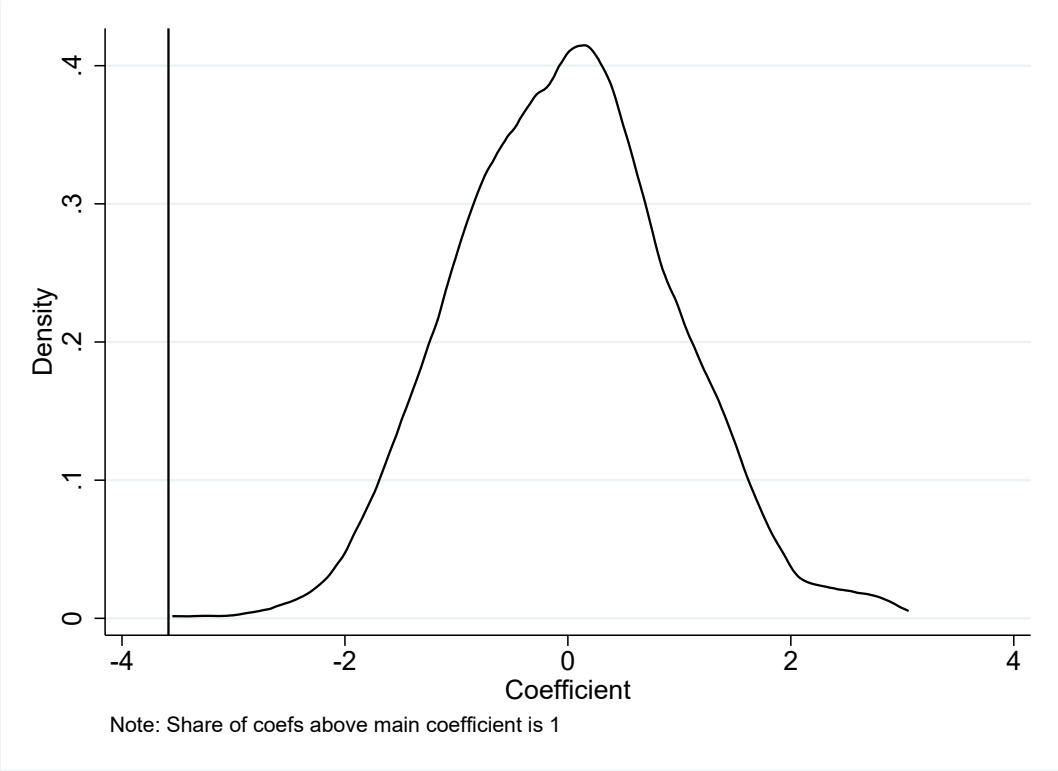
Note: The figure reports effects of HV and of its leads and lags on retirement, and take-up of disability benefits together with 95% confidence interval. Regression includes periods from 12 months before the November 2017 up to 16 months after January 2018. Months grouped in three months periods for presentation purpose (reported at the mid-month of the three-month period, e.g. the coefficient for 1 covers coefficients for months from 0 up to 2). The period zero corresponds to November 2017. Reference period is the period between -4 and -6 months before November 2017. Regression includes INPS local office fixed effects (at the time of the experiment), time spent on leave in the period of the experiment (dummies), both interacted with month groups fixed effects, as well as individual fixed effects. Standard errors clustered at INPS local office level.

Figure A8: Impact of unobserved heterogeneity on treatment effect



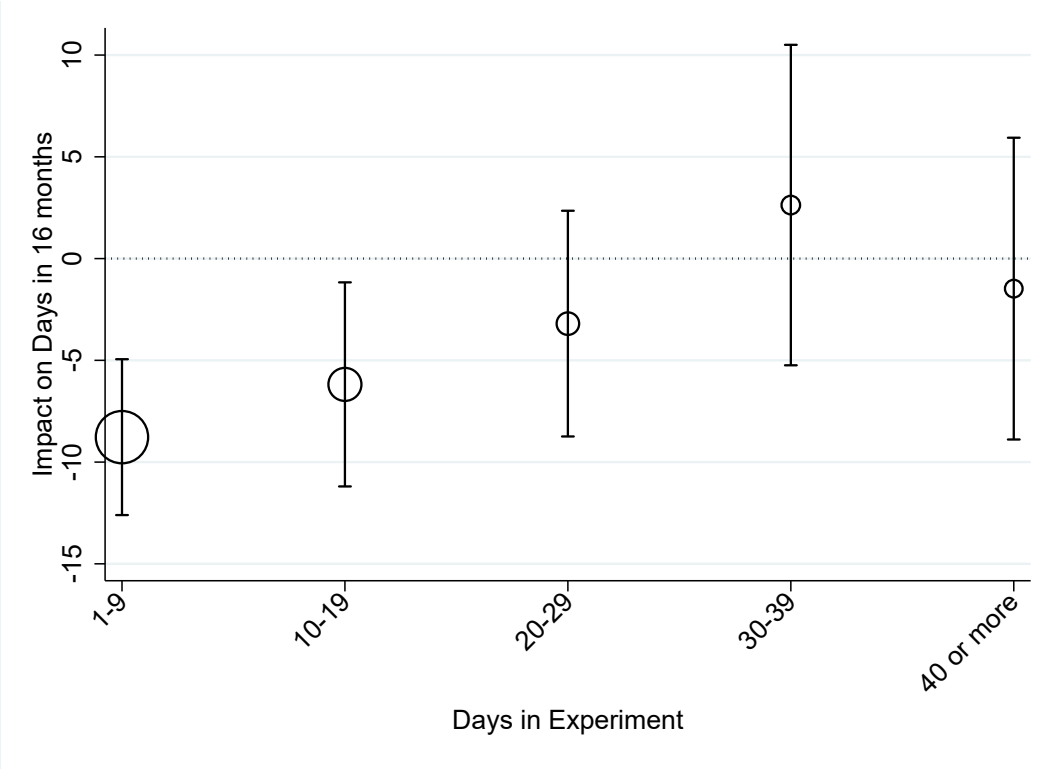
Note: The figure reports point estimates for effect of HV on the number of days on sick leave in the 16 months following the experiment for changes in the overall regression R-squared and correlation level between an unobserved component and HV according to the methodology of Oster (2019). The regression is run with OLS, and it includes a dummy for receiving HV outcome, demographic characteristics (gender, age), job characteristics (sector dummies, average salary in the 6 months before the experiment, part-time dummy, dummy for permanent contract), fixed effects at INPS local office level, number of days spent on sick leave in the period of the experiment (11/22/2017-01/05/2018), number of certificates, average duration, and number of days spent on sick leave in the 6 months before the experiment..

Figure A9: Comparison of our treatment effect on number of days on sick leave (16 months) with randomly assigned treatments.



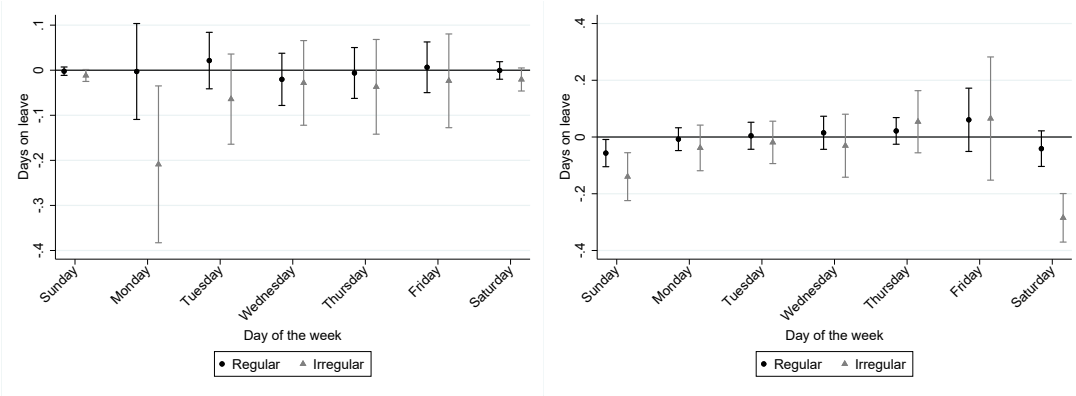
Note: The figure reports effect of HV on the number of days on sick leave in the 16 months following the experiment (vertical line) with the distribution of treatment effects obtained by randomly assigning a fake treatment to observations in the sample. We replace the treatment with a dummy equal to one for random observations in the sample and collect the relative treatment coefficients. The process is repeated 1,000 times, and the number of individuals assigned one is chosen to match the share of treated in the original sample. The vertical line corresponds to the treatment effect obtained with our preferred specification and real HV assignment. The note below the graph reports the share of fake treatments effects above our estimated treatment effect with the real HVs. Regressions include a dummy for receiving HV outcome, demographic characteristics (gender, age), job characteristics (sector dummies, average salary in the 6 months before the experiment, part-time dummy, dummy for permanent contract), fixed effects at INPS local office level, number of days spent on sick leave in the period of the experiment (11/22/2017-01/05/2018), number of certificates, average duration, and number of days spent on sick leave in the 6 months before the experiment. Standard errors clustered at local office level.

Figure A10: Impact of HV by number of days on sick leave in the experiment



Note: The figure reports effect of HV on the number of days on sick leave in the 16 months following the experiment by bins of number of days spent on sickness leave in the period of the experiment by workers. OLS regressions are run separately for each group of workers, and the resulting estimate for the impact of the HV and related standard errors are reported. Regressions include a dummy for receiving HV outcome, demographic characteristics (gender, age), job characteristics (sector dummies, average salary in the 6 months before the experiment, part-time dummy, dummy for permanent contract), fixed effects at INPS local office level, number of days spent on sick leave in the period of the experiment (11/22/2017-01/05/2018), number of certificates, average duration, and number of days spent on sick leave in the 6 months before the experiment. Standard errors clustered at local office level.

Figure A11: Effect of HV on number of certificates by start and end day of the week.

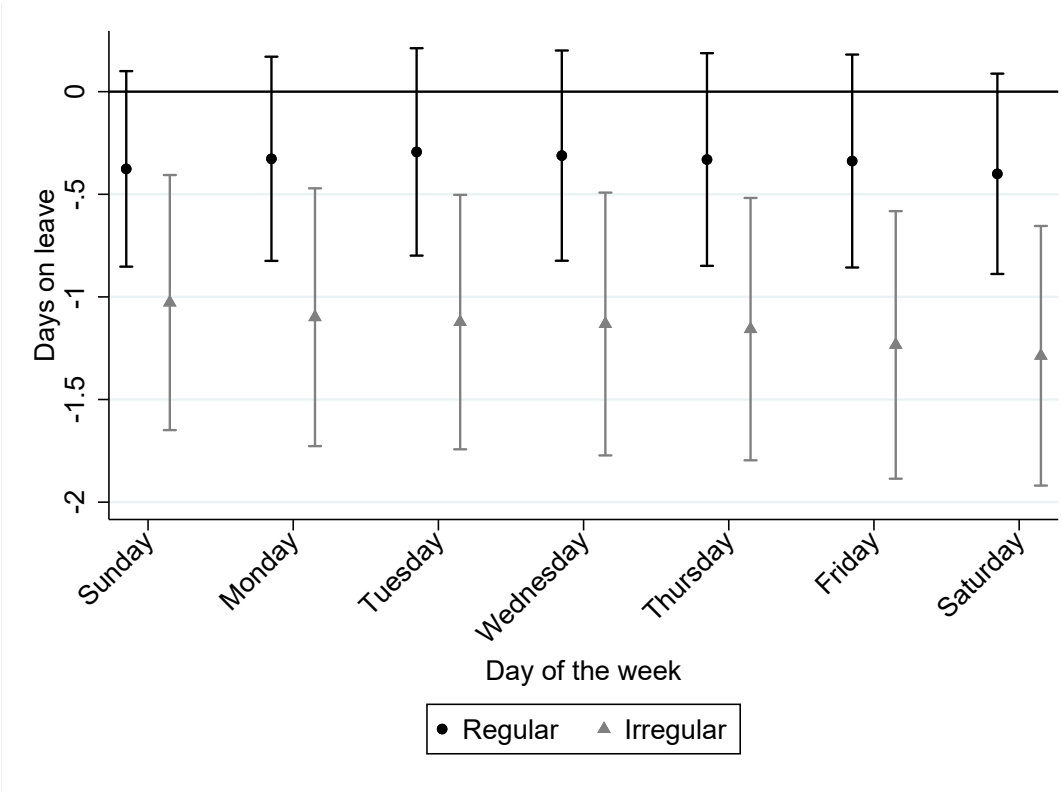


(a) # Certificate by start day of the week

(b) # Certificate by end day of the week

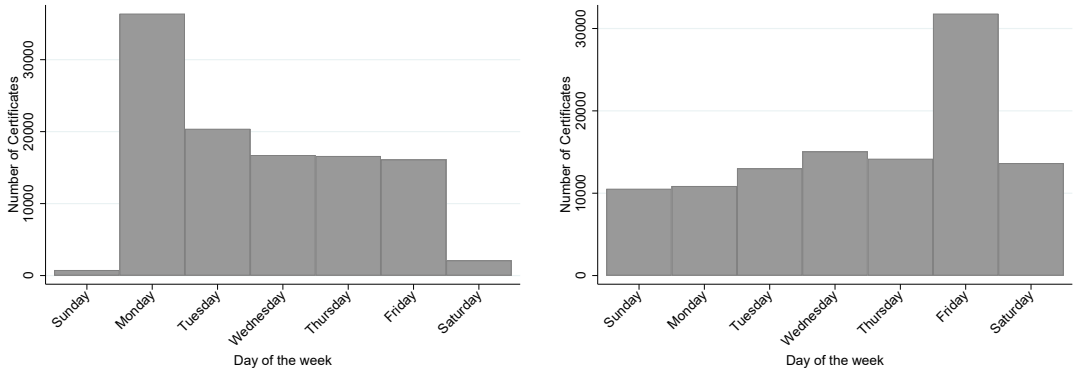
Note: The figure reports the effect of HV by outcome of the visit on the number of certificates by day of the week start and end. Coefficients obtained by separate regressions by day of the week. Coefficients obtained from OLS regressions including indicators for workers subject to HV with regular and irregular outcome, demographic characteristics (gender, age), job characteristics (sector dummies, average salary in the 6 months before the experiment, part time dummy, dummy for permanent contract), fixed effects at INPS local office level, number of days spent on sick leave in the period of the experiment (11/22/2017-01/05/2018), number of certificates, average duration, and number of days spent on sick leave in the 6 months before the experiment. Figures include confidence interval at 95% with clustered standard errors at INPS local office level.

Figure A12: Effect of HV on sick leave by day of the week over 16 months after the experiment by visit outcome.



Note: The figure reports coefficients for effect of HV on the number of absences by day of the week according to the outcome of the visit. Coefficients obtained by separate regressions by day of the week. Coefficients obtained from OLS regressions including indicators for workers subject to HV with regular and irregular outcome, demographic characteristics (gender, age), job characteristics (sector dummies, average salary in the 6 months before the experiment, part-time dummy, dummy for permanent contract), fixed effects at INPS local office level, number of days spent on sick leave in the period of the experiment (11/22/2017-01/05/2018), number of certificates, average duration, and number of days spent on sick leave in the 6 months before the experiment. Figure includes confidence interval at 95% with clustered standard errors at INPS local office level.

Figure A13: Start and end day of the week for sick leave spells.



(a) Day of the Week: Start (b) Day of the Week: End

Note: Statistics are based on data for public sector workers involved in the experiment. We consider sick leave spells which started between May and October 2017. Panel (a) reports the day of the week in which sick leave spells start while Panel (b) reports the day of the week in which sick leave spells end.

Tables

Table A1: Balancing regressions for the probability of receiving HV

Variables	(1) HV	(2) HV	(3) HV
Female	0.011***	0.010***	0.009***
	(0.003)	(0.003)	(0.003)
Age: 36-40	0.001	-0.001	-0.001
	(0.009)	(0.008)	(0.008)
Age: 41-45	0.002	0.000	0.001
	(0.008)	(0.007)	(0.007)
Age: 46-50	0.001	-0.001	-0.001
	(0.008)	(0.007)	(0.007)
Age: 51-55	-0.008	-0.008	-0.007
	(0.007)	(0.006)	(0.006)
Age: 56-60	-0.012	-0.012	-0.011
	(0.008)	(0.007)	(0.007)
Age: 61-65	-0.010	-0.006	-0.006
	(0.008)	(0.007)	(0.007)
Age: 66-67	0.044***	0.042***	0.043***
	(0.013)	(0.012)	(0.012)
Central Administration	0.007	-0.002	-0.001
	(0.007)	(0.007)	(0.007)
Local Administration	-0.004	0.000	0.001
	(0.004)	(0.004)	(0.004)
Health Sector	-0.007*	-0.005	-0.004
	(0.004)	(0.004)	(0.004)
Permanent Contract	0.008	0.003	0.003
	(0.006)	(0.005)	(0.005)
Part Time	-0.008	-0.007	-0.007
	(0.006)	(0.005)	(0.006)
Manager	0.013	0.011	0.012
	(0.010)	(0.009)	(0.009)
(Log) Average Wage (six month bef. exp)	-0.000	-0.000	-0.000
	(0.002)	(0.002)	(0.002)
Days on Leave in Experiment	0.006***	0.007***	
	(0.000)	(0.000)	
Total certificates (6 months bef.)	0.003***	0.001	0.001
	(0.001)	(0.001)	(0.001)
Total number of days (6 months bef.)	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Average duration cert. (6 months bef.)	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Duration first certificate	0.002***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)
Metropolis	-0.003	0.002	0.002
	(0.009)	(0.005)	(0.005)
Observations	42,660	42,657	42,657
R-squared	0.120	0.183	0.185
Joint P-value	0.000	0.000	0.000
Sede FE	NO	YES	YES
Days in the Experiment FE	NO	NO	YES

Note: The table reports results for a linear probability model at the individual level with dependent variable a dummy equal to one if the worker is subject to a HV in the period of the experiment (11/22/2017-01/05/2018) and zero otherwise. All regressions are estimated by OLS. with the *reghdfe* stata command developed by (Correia, 2019). From Column 2, the regression includes fixed effects for the INPS local office (*Sede*). Column 3 includes dummies for the number of days on sick leave in the period of the experiment (Duration sick leave in exp.). (log) Mean Monthly Earnings is the log of average earnings in the public sector for the worker from May to October 2017. Number of Certificates (bef. exp.), Number of Days (bef. exp.), and Mean Duration Certificate (bef. exp.) are the number of certificates, the total number of days on sick leave, and the average duration of certificates in the 6 months before the experiment. Sample size excluding singletons reported. P-value for the joint significance of coefficients but past use of sick leave reported at the bottom of the Table. Standard errors clustered at local office level. Level of significance: 0.1 *, 0.05 **, 0.01 ***.

Table A2: Effect of HV on cumulative days on sick leave in the 16 months after the experiment: adding controls

VARIABLES	(1) #Days	(2) #Days	(3) #Days	(4) #Days	(5) #Days
HV	17.187***	12.285***	-2.318	-3.852**	-3.589**
	(1.679)	(1.788)	(1.796)	(1.550)	(1.539)
Total certificates (6 months bef.)				4.981***	5.010***
				(0.172)	(0.170)
Total number of days (6 months bef.)				0.200***	0.197***
				(0.026)	(0.025)
Average duration cert. (6 months bef.)				-0.271***	-0.304***
				(0.054)	(0.053)
Duration first certificate				0.353***	0.343***
				(0.053)	(0.052)
Female					0.217
					(0.826)
Age: 36-40					2.199
					(1.444)
Age: 41-45					5.156***
					(1.385)
Age: 46-50					5.094***
					(1.418)
Age: 51-55					7.620***
					(1.469)
Age: 56-60					9.830***
					(1.420)
Age: 61-65					16.153***
					(1.562)
Age: 66-67					-3.051
					(2.506)
Central Administration					-13.560***
					(1.565)
Local Administration					-9.004***
					(1.038)
Health Sector					-5.108***
					(0.952)
Permanent Contract					6.234***
					(1.196)
Part Time					2.321*
					(1.248)
Manager					-10.628***
					(1.539)
(Log) Average Wage (six month bef. exp)					-1.822***
					(0.445)
Observations	42,707	42,704	42,704	42,704	42,657
R-squared	0.006	0.030	0.066	0.152	0.162
Mean Dep	43.335	43.335	43.335	43.335	43.335
Sede FE	NO	YES	YES	YES	YES
Days in the Experiment FE	NO	NO	YES	YES	YES

Note: The table reports estimates for the effect of HV on cumulative days on sick leave in the 16 months after the experiment (# Days). Regressions are estimated with OLS with the *reghdfe* stata command developed by (Correia, 2019). HV is a dummy equal to one if the worker was subject to HV in the period of the experiment. Days in the experiment is the number of days spent on sick leave in the period of the experiment (11/22/2017-05/01/2018). (log) Mean Monthly Earnings is the log of average earnings in the public sector for the worker from May to October 2017. Number of Certificates (bef. exp.), Number of Days (bef. exp.), and Mean Duration Certificate (bef. exp.) are the number of certificates, the total number of days on sick leave, and the average duration of certificates in the 6 months before the experiment. Duration first certificate is the duration (End date-Start date) of the first certificate with which a worker appear in the experiment. Mean Dep is the average for the dependent variable for individuals who did not receive a HV (i.e. the control group). Sede FE are local INPS office fixed effects. Sample size excluding singletons reported. Standard errors clustered at local office level. Level of significance: 0.1 *, 0.05 **, 0.01 ***.

Table A3: Extensive vs intensive margin for sick leave: Balanced panel

VARIABLES	(1) # Days	(2) # Certificates	(3) Mean Days
HV	-4.829***	-0.197	-0.836***
	(1.763)	(0.132)	(0.196)
Observations	33,430	33,430	33,430
R-squared	0.162	0.455	0.152
Mean Dep	41.532	6.248	6.171
Controls	YES	YES	YES
Past Cert.	YES	YES	YES
Sede FE	YES	YES	YES
Days in the Experiment FE	YES	YES	YES

Note: The table reports estimates for the effect of HVs on cumulative days on sick leave in the 16 months after the experiment (# Days in 16 Months), cumulative number of certificates (# Cert in 16 months), and average certificate duration (Mean days in 16 months) for workers present in all the 16 months of the experiment. The cumulative number of certificates corresponds to the number of sick leave spells and the average certificate duration corresponds to the average duration of sick leave spells. Regressions are estimated with OLS with *reghdfe* stata command developed by (Correia, 2019). HV is a dummy equal to one if the worker was visited in the period of the experiment. Sede FE are local INPS office fixed effects. Past Cert. includes: the number of certificates, the total number of days on sick leave, the average duration of certificates in the 16 months before the experiment, and the duration of the first certificate. Controls include: demographic characteristics (gender, and age dummies), and job characteristics (sector dummies, average salary in the 6 months before the experiment, part time dummy, dummy for permanent contract). Mean dep is the average for the dependent variable for individuals who did not receive a HV (i.e. the control group). Sample size excluding singletons reported. Standard errors clustered at local office level. Level of significance: 0.1 *, 0.05 **, 0.01 ***.

Table A4: Effect of HV on cumulative days on sick leave in the 16 months after the experiment: IV for days in the experiment

VARIABLES	(1) Baseline	(2) Linear	(3) IV
HV	-3.589**	-3.668**	-3.407**
	(1.539)	(1.540)	(1.645)
Observations	42,657	42,657	42,657
R-squared	0.162	0.161	0.138
Mean Dep	43.335	43.335	43.335
Controls	YES	YES	YES
Past Cert.	YES	YES	YES
Sede FE	YES	YES	YES
Days in the Experiment FE	YES	NO	NO
Cragg-Donald F-test			12375.196

Note: The table reports estimates for the effect of HV on cumulative days on sick leave in the 16 months after the experiment. Regressions are estimated with OLS with *reghdfe* stata command developed by (Correia, 2019). HV is a dummy equal to one if the worker was visited in the period of the experiment. Sede FE are local INPS office fixed effects. Past Cert. includes: the number of certificates, the total number of days on sick leave, and the average duration of certificates in the 6 months before the experiment. Controls include: demographic characteristics (gender, and age dummies), and job characteristics (sector dummies, average salary in the 6 months before the experiment, part time dummy, dummy for permanent contract). Mean dep is the average for the dependent variable for individuals who did not receive a HV (i.e. the control group). Effective duration is the number of days spent on sick leave in the period of the experiment (11/22/2017-05/01/2018). Column (1) reports effects of HV with fixed effects for time spent on sick leave in the experiment as in Table 4. Column (2) reports results by including time spent on leave in the period of the experiment linearly. Column (3) estimates the impact of HV after instrumenting the number of days spent on leave in the period of the experiment by the time implied by the start and end date of the first certificate when it was issued. Standard errors clustered at local office level. Level of significance: 0.1 *, 0.05 **, 0.01 ***.

Table A5: Effect of HV on cumulative days on sick leave in 16 months after the experiment: workers with only one certificate fully included in the experiment

VARIABLES	(1) #Days	(2) #Days	(3) #Days	(4) #Days	(5) #Days
HV	12.375***	3.762	-3.433	-4.400	-4.064
	(2.546)	(2.715)	(2.841)	(2.762)	(2.740)
Observations	18,727	18,725	18,722	18,722	18,697
R-squared	0.002	0.034	0.050	0.172	0.183
Mean Dep	34.687	34.687	34.687	34.687	34.687
Sede FE	NO	YES	YES	YES	YES
Duration first certificate FE	NO	NO	YES	YES	YES
Past Cert.	NO	NO	NO	YES	YES
Controls	NO	NO	NO	NO	YES

Note: The table reports estimates for the effect of HV on cumulative days on sick leave in the 16 months after the experiment (# Days). Regressions are estimated with OLS with the *reghdfe* stata command developed by (Correia, 2019). The sample is restricted to workers with only one ongoing certificate in the period of the experiment fully included in the experiment. HV is a dummy equal to one if the worker was visited in the period of the experiment. Sede FE are local INPS office fixed effects. Duration first certificate fixed effects are fixed effects for the length of the certificate they report during the experiment. This is computed as the end date minus the start date of the certificate and it can not be affected by the HV. Past Cert. includes: the number of certificates, the total number of days on sick leave, and the average duration of certificates in the 16 months before the experiment. Controls include: demographic characteristics (gender, and age dummies), and job characteristics (sector dummies, average salary in the 6 months before the experiment, part time dummy, dummy for permanent contract). Mean dep is the average for the dependent variable for individuals who did not receive a HV (i.e. the control group). Sample size excluding singletons reported. Standard errors clustered at local office level. Level of significance: 0.1 *, 0.05 **, 0.01 ***.

Table A6: Multiple hypothesis test for main outcomes

Outcome	(1) Baseline	(2) Coef	(3) Remark 3.2	(4) Thm 3.1	(5) Remark 3.8	(6) Bonf.	(7) Holm.
Days in 16 months	-3.589	-5.717	.008	.015	.015	.024	.016
Certificates in 16 months	-0.088	-.221	.337	.337	.337	1	.337
Average Days per Certificate	-0.707	-.836	.001	.003	.003	.004	.004

Note: The table reports estimates and p-values for multiple hypothesis test for the effect of HV on cumulative days on sick leave in the 16 months after the experiment (Days in 16 Months), cumulative number of certificates (Certificates in 16 months), and average certificate duration (Average Days per Certificate). The cumulative number of certificates corresponds to the number of sick leave spells and the average certificate duration corresponds to the average duration of sick leave spells. Estimation is based on the command MHTEXP2 which implements the methodology outlined in List et al. (2021). The first Column reports the coefficients reported in Table 4, computed with OLS regressions; the second Column reports coefficients computed with the regression framework of Lin (2013); following columns report p-values of tests reported in List et al. (2021). Column from (3) to (5) report p-values for the joint significance of the effects according to the framework of List et al. (2021) while Column (6) and Column (7) report usual provide the standard (albeit less rigorous) Bonferroni and Holm tests.

Table A7: Extensive vs intensive margin for sick leave use by visit outcome

VARIABLES	(1) #Days in 16 months	(2) #Cert. in 16 months	(3) Mean Days in 16 months
HV: Regular	-2.361	-0.004	-0.477**
	(1.781)	(0.130)	(0.229)
HV: Irregular	-8.094***	-0.394*	-1.551***
	(2.222)	(0.209)	(0.265)
Observations	42,657	42,657	42,657
R-squared	0.162	0.410	0.196
Mean Dep	43.335	6.076	6.686
Controls	YES	YES	YES
Past Cert.	YES	YES	YES
Sede FE	YES	YES	YES
Days in the Experiment FE	YES	YES	YES

Note: The table reports estimates for the effect of HV, by outcome of the inspection, on cumulative days on sick leave in the 16 months after the experiment (# Days in 16 Months), cumulative number of certificates (# Cert in 16 months), and average certificate duration (Mean days in 16 months). The cumulative number of certificates corresponds to the number of sick leave spells and the average certificate duration corresponds to the average duration of sick leave spells. Regressions are estimated with OLS with *reghdfe* stata command developed by (Correia, 2019). HV is a dummy equal to one if the worker was visited in the period of the experiment. Sede FE are local INPS office fixed effects. Past Cert. includes: the number of certificates, the total number of days on sick leave, and the average duration of certificates in the 16 months before the experiment. Controls include: demographic characteristics (gender, and age dummies), and job characteristics (sector dummies, average salary in the 6 months before the experiment, part time dummy, dummy for permanent contract). Mean dep is the average for the dependent variable for individuals who did not receive a HV (i.e. the control group). Sample size excluding singletons reported. Standard errors clustered at local office level. Level of significance: 0.1 *, 0.05 **, 0.01 ***.

Table A8: HV Impact and MVPF by earning quintile

Earning Quintile	(1) Deterrence	(2) Daily Wage	(3) Benefit Expenditure decline	(4) MVPF
1	-3.588	54.846	-196.788	1.341
2	-2.46	69.384	-170.686	1.414
3	-2.705	79.692	-215.568	1.302
4	-3.187	92.615	-295.165	1.203
5	-3.061	133.269	-407.937	1.14

Note: The table reports estimates for the effect of HV by quintile of earnings in the public sector. Column (1) reports the impact of HV by earning quintile as shown in Panel (a) of Figure 6, Column (2) reports the daily wage of workers by earning quintile, Column (3) reports the decline in expenditure on sick leave benefits as a consequence of the HV, and, finally, Column (4) reports the corresponding MVPF as reported in Panel (b) of Figure 6.

B Street level public sector workers

Table B1: Occupations by involvement with the public

Job	Mean	Median
Health professionals	2,07	1
Sales workers	2,14	1
Teaching professionals	2,44	1
Health associate professionals	2,55	1
Customer services clerks	2,57	1
Hospitality, retail and other services managers	2,59	2
Personal service workers	3,01	2
Personal care workers	3,13	2
Legal, social, cultural and related associate professionals	3,21	2
Business and administration associate professionals	3,47	3
Protective services workers	3,53	3
Chief executives, senior officials and legislators	3,60	4
Legal, social and cultural professionals	3,65	3
Street and related sales and service workers	3,75	3
Drivers and mobile plant operators	3,81	4
Other clerical support workers	3,83	4
Production and specialised services managers	3,90	4
Administrative and commercial managers	3,91	4
Business and administration professionals	4,15	5
Information and communications technicians	4,16	5
General and keyboard clerks	4,19	5
Food preparation assistants	4,46	5
Science and engineering professionals	4,59	5
Electrical and electronic trades workers	4,62	5
Building and related trades workers, excluding electricians	4,73	5
Numerical and material recording clerks	4,76	5
Information and communications technology professionals	4,83	5
Refuse workers and other elementary workers	4,90	6
Science and engineering associate professionals	5,06	6
Cleaners and helpers	5,10	6
Metal, machinery and related trades workers	5,15	6
Handicraft and printing workers	5,18	6
Commissioned armed forces officers	5,25	6
Food processing, wood working, garment and other craft and related trades workers	5,27	6
Market-oriented skilled agricultural workers	5,39	6
Non-commissioned armed forces officers	5,50	6
Market-oriented skilled forestry, fishery and hunting workers	5,58	6
Labourers in mining, construction, manufacturing and transport	5,64	7
Armed forces occupations, other ranks	5,73	6
Agricultural, forestry and fishery labourers	5,90	7
Assemblers	6,14	7
Subsistence farmers, fishers, hunters and gatherers	6,24	7
Stationary plant and machine operators	6,37	7

Source: EWCS (European Working Conditions Survey), 2015. Occupations, at the ISCO-08 2-digit level, are ordered according to the respondents' mean answer to the question "Does your main paid job involve dealing directly with people who are not employees at your workplace such as customers, passengers, pupils, patients, etc.?". Possible answers: (1) All of the time, (2) Almost all of the time, (3) Around 3/4 of the time, (4) Around half of the time, (5) Around 1/4 of the time, (6) Almost never, (7) Never.