Monitoring and Sickness Absence Behaviour: Who Goes Back to Work Rather Than Visit the Doctor?

Yaroslav Yakymovych
Institute for Housing and Urban Research, Uppsala University

August 2025

Introduction

Paid sick leave from work is a right in most developed countries

- Shields individuals from the income effects of health shocks, allows consumption smoothing

How to best design the system?

- Monitoring used to make sure that workers return to their jobs when healthy enough to do so
- Opportunity cost of medical professionals' time is high
- Can it be cost-efficient to reduce monitoring of individuals whose sickness absence it does not affect?
- Sickness insurance/absence salient issue during the Covid-19 pandemic as countries relaxed rules (OECD 2020)

This Paper

What characterises workers with different sensitivity to monitoring?

- Setting: unique randomised controlled experiment
- Estimate causal effects of monitoring on sickness absence

Heterogeneity analysis using GRF (Athey et al., 2019)

- Fully data-driven, tests for heterogeneity across large number of covariates and thresholds, avoids overfitting, nonparametric

Findings:

- Those who react strongly to monitoring have high past sickness absence, low socioeconomic status, are men, live in poor neighbourhoods, have peers with high sick leave uptake
- Possible to be more cost-efficient with a targeted monitoring policy - if desirable

Background

We know quite a lot about the patterns of sickness absence...

- Sickness absence higher among women (Paringer, 1983), public sector employees (Frick and Malo, 2008), low-paid workers (Barmby et al., 2002), non-key workers (Hensvik and Rosenqvist, 2019), large workplaces (Winkelmann, 1999), certain neighbourhoods (Lindbeck et al., 2016)...
- But is high sickness absence in a group connected to higher sensitivity to monitoring?
- Large cross-country differences in sickness absence
- Reducing/increasing income replacement rates decreases/increases absenteeism, absenteeism falls after first day of absence excluded from coverage (Henrekson and Persson, 2004; Johansson and Palme, 2005; Ziebarth and Karlsson, 2010; Böckerman et al., 2018)

Background

...but quite little about how workers react to monitoring

- Studies based on public sector workers in Italy (Boeri et al., 2021) and a Norwegian municipality (Ferman et al., 2023) arrive at opposite conclusions
- Earlier work on the Swedish experiment: Hartman et al. (2013); Hesselius et al. (2013); Johansson et al. (2019)

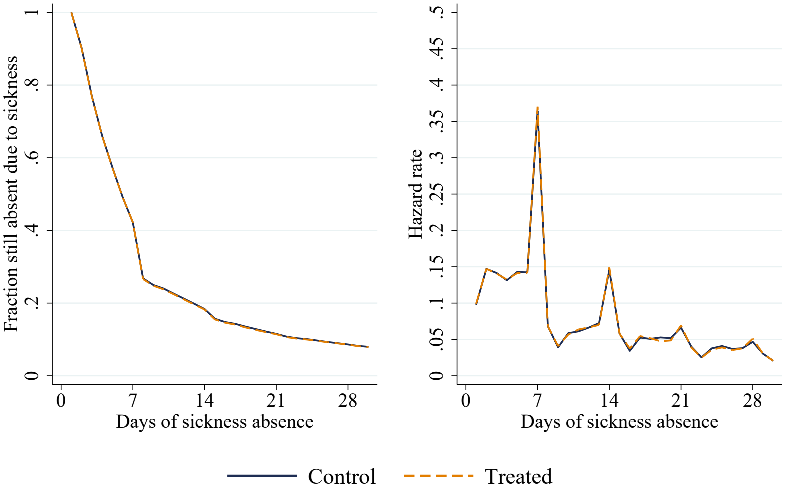
The Experiment

Randomised controlled trial in two Swedish regions (Gothenburg city and Jämtland region) in July-December 1988

- 270 000 insured individuals affected
- *Controls* (born on odd dates) could take out 7 days of sick leave before providing a doctor's certificate; *Treated* (born on even dates) could take out 14 days of sick leave before providing a doctor's certificate
- Seen as unsuccessful due to substantial increase in sickness absence among the treated

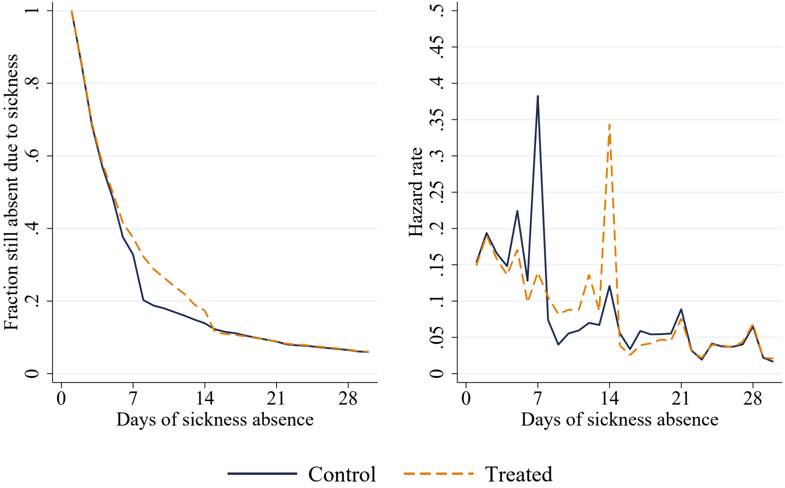
Map | Sickness insurance system during the experiment

Before the Experiment



Propensity score histogram

During the Experiment



Data

Individual-level register data

- All sickness absence spells before, during and after the experiment; allow for detailed description of workers' families, careers, socioeconomic status, neighbourhoods...
- Only employed individuals; workers in the central government sector not part of the experiment and are excluded

Outcome: duration of sickness absence spell in days

- Probability of sick leave spell ending on days 8-14 gives very similar results

Heterogeneity Analysis: GRF (Athey et al, 2019)

Traditional approach: Researcher splits sample based on a characteristic and threshold of interest

GRF: Splits the sample based on the variables and threshold values which maximise treatment effect heterogeneity

- Aggregates a large number of **causal trees** - each tree is based on a random subsample of individuals

Advantages of GRF:

- Can analyse many variables and thresholds, identifies relationships which hold consistently across subsamples (no overfitting), flexible with regard to functional form, tolerates missing values

When estimating: Split into **training set** (80 % of families) and **held-out test set** (remaining 20 %) Implementation details

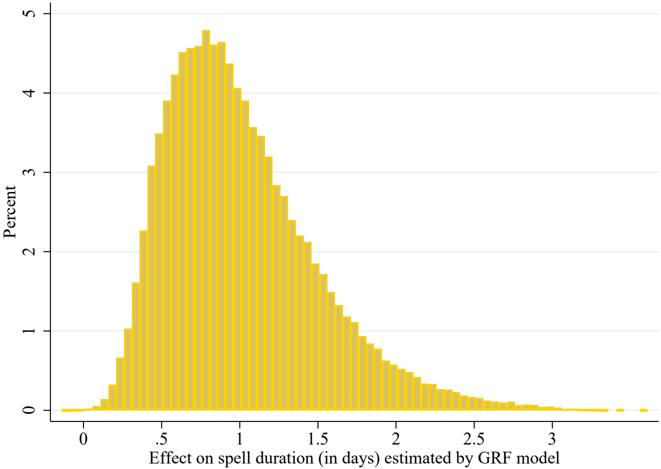
Potential Drivers of Sickness Absence

Aim to include as many factors that have been identified as connected to sickness absence by previous literature as possible

Literature

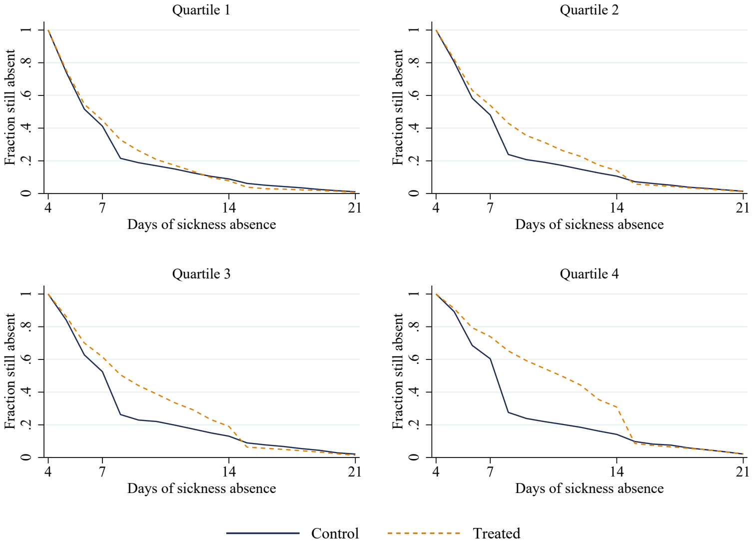
- **Demographics**: age, gender, immigrant background
- **Family**: marital status, share of family income, N children, days spent caring for sick children, partner's pre-period sick leave...
- **Career**: education, L income, share income from main job, from self-employment, from social payments, industry dummies, public sector...
- **Workplace**: N employees, income rank, tenure, average sick leave...
- **Neighbourhood**: socioeconomic characteristics, average sickness absence...
- **Health** (pre-experiment): Days sickness absence, N short absence spells...

Treatment Effect Estimates



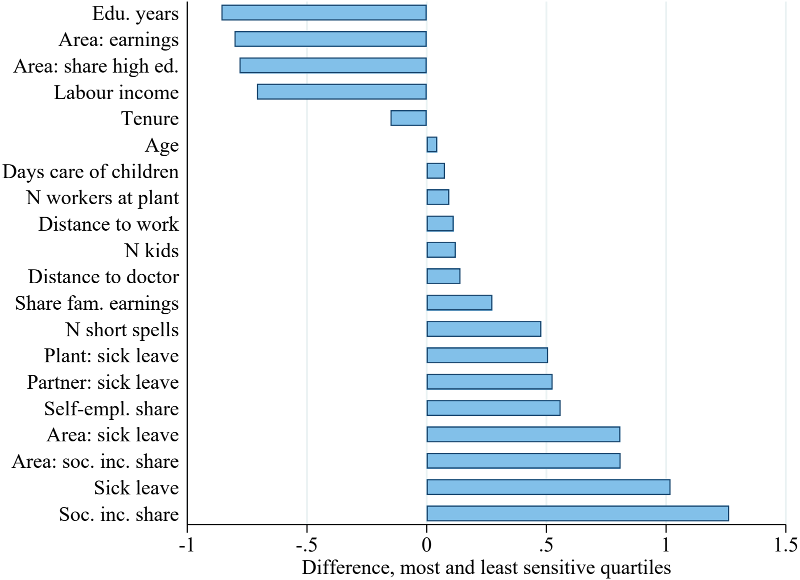
BLP test | Pr(Return to work in week 2) as outcome | Comparison to Lasso

Validation on the Test Set

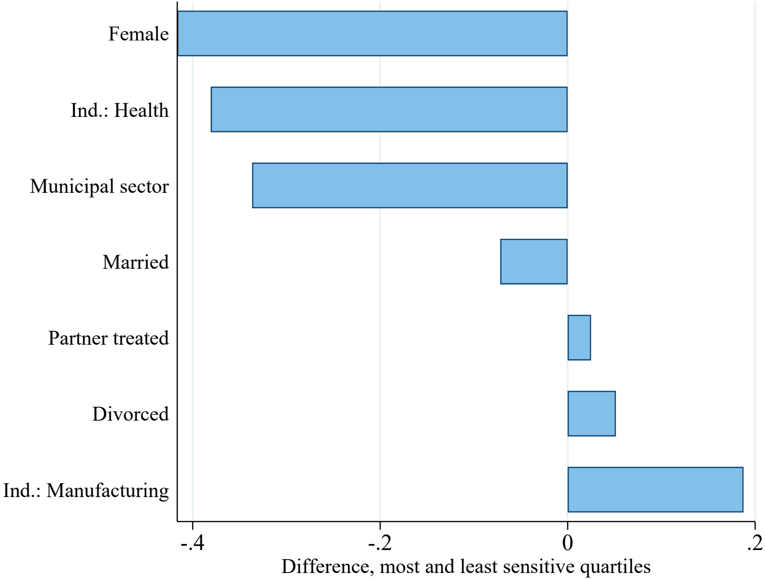


Hazard graphs

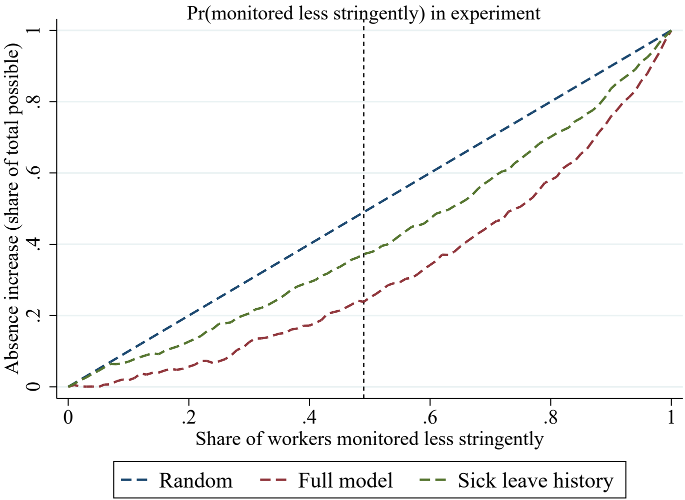
Who Reacts to Monitoring? (Selected Variables)



Who Reacts to Monitoring? (Selected Variables)

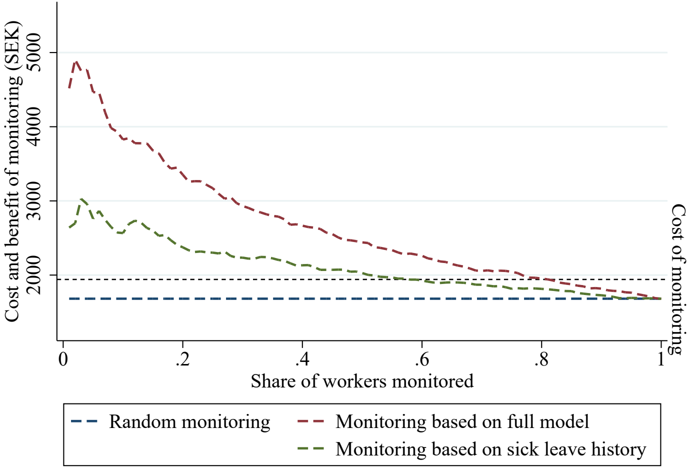


Targeted Monitoring Policy



Comparison to LASSO | Policy tree | QINI of different policies

Cost-Benefit Analysis



Assumptions: cost of visit=1941 SEK (≈ 190 EUR, from Västra Götaland region),

benefit=reduction in days of absence × "daily wage" including payroll tax

Concluding Remarks

Use GRF to study whether it is efficient to relax sickness absence monitoring in a targeted way

- Analyse a large number of characteristics at the same time in a data-driven way

Workers who are sensitive to monitoring:

- Have high sick leave uptake in previous periods, low socioeconomic status, are men, live in socioeconomically disadvantaged neighbourhoods, have peers with high sick leave uptake

If monitoring is targeted:

- Same reduction in share of workers monitored as in the experiment can take place at half the cost in terms of increased absence
- Estimated to be cost efficient, unlike non-targeted monitoring

References

Angelov, Nikolay; Johansson, Per; Lindahl, Erica; Lindström, Elly-Ann. Kvinnors och mäns sjukfrånvaro, IFAU Report 2011:2

Susan Athey. Julie Tibshirani. Stefan Wager. "Generalized random forests." Ann. Statist. 47 (2) 1148 - 1178, April 2019

Tim A. Barmby, Marco G. Ercolani, John G. Treble, Sickness Absence: An International Comparison, The Economic Journal, Volume 112, Issue 480, June 2002

Laura Hartman, Patrik Hesselius, Per Johansson, Effects of eligibility screening in the sickness insurance: Evidence from a field experiment, Labour Economics, Volume 20, 2013, Pages 48-56

References

Lena Hensvik and Olof Rosenqvist, Keeping the Production Line Running: Internal Substitution and Employee Absence, J. Human Resources January 1, 2019 54:200-224

Patrik Hesselius, J. Peter Nilsson, Per Johansson, Sick of Your Colleagues' Absence?, Journal of the European Economic Association, Volume 7, Issue 2-3, 1 May 2009

Hesselius, Patrik; Johansson, Per; Vikström, Johan, Monitoring and norms in sickness insurance: Empirical evidence from a natural experiment, IFAU Working paper 2008:8

Henrekson, Magnus, and Mats Persson. "The effects on sick leave of changes in the sickness insurance system." Journal of Labor economics 22.1 (2004): 87-113.

References

Johansson, Per and Karimi, Arizo and Nilsson, Peter, Gender Differences in Shirking: Monitoring or Social Preferences? Evidence from a Field Experiment. IZA Discussion Paper No. 8133

Johansson, Per, and Mårten Palme. "Assessing the effect of public policy on worker absenteeism." Journal of Human Resources (2002): 381-409.

Lindgren, Karl-Oskar, Workplace size and sickness absence transitions, IFAU Working paper 2012:26

Jos N. van Ommeren, Eva Gutiérrez-i-Puigarnau, Are workers with a long commute less productive? An empirical analysis of absenteeism, Regional Science and Urban Economics, Volume 41, Issue 1, 2011

Riksförsäkringsverket, Utvidgad egen sjukskrivning, 1989-08-14, AD 1989-954:01

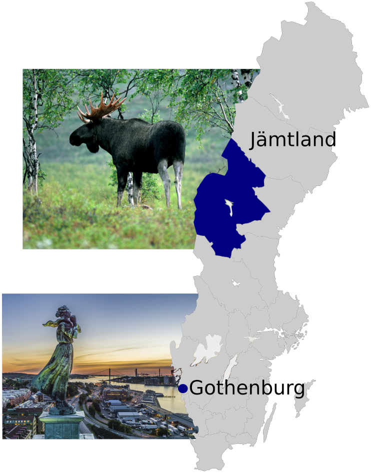
Rainer Winkelmann (1999) Wages, firm size and absenteeism, Applied Economics Letters, 6:6, 337-341

References

Ziebarth, Nicolas R., and Martin Karlsson. "The effects of expanding the generosity of the statutory sickness insurance system." Journal of Applied Econometrics 29.2 (2014): 208-230.

Böckerman, Petri, Ohto Kanninen, and Ilpo Suoniemi. "A kink that makes you sick: The effect of sick pay on absence." Journal of Applied Econometrics 33.4 (2018): 568-579.

Map



Back

Sickness Insurance System During the Experiment

Somewhat more generous than current system, but similar in many respects

- Most workers received 90% income replacement rates (80% today)
- All reimbursements from government-run insurance funds (employer responsible for the first 14 days today)
- Coverage from first day of absence (no "qualification day"/karensdag) - but focus on spells of durations where this is unlikely to be a factor
- Monitoring by doctors after seven days of absence - like today
- Leave for taking care of sick children (VAB) - similar monitoring and replacement rates, unaffected by experiment
- Similar provisions are in place in many European countries today (Palme and Persson, 2020)

Back

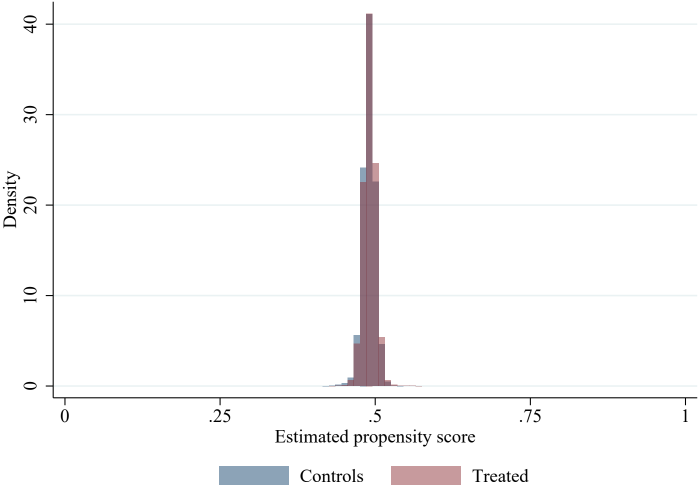
Determinants of Sickness Absence

A significant number of factors have been identified as potentially connected to sickness absence in the literature

- Gender and income (Hartman, Hesselius and Johansson, 2013)
- Previous health history (Avdic and Johansson 2013)
- Family situation (Barmby et al. 2002)
- Education (Johansson and Nilsson 2008)
- Work environment/peers (Hesselius, Johansson, Vikström, 2008, Johansson, Karimi, Nilsson 2014)
- Plant characteristics (Winkelmann 1999), key position at plant (Hensvik and Rosenqvist 2015), tenure (Bratberg and Monstad, 2015)
- Public or private sector, industry (Barmby et al., 2002, Frick and Malo, 2008)

Back

Propensity Scores Among Treated and Controls



Back

GRF setup

Clustering: on the family level

Training set: 80 percent sample of families

Held-out test set: 20 percent of families

Number of trees = 5000

Main specification: tuned parameters

- Minimum node size=1, maximum imbalance=0.034, imbalance penalty=1.4, N variables sampled=21 instead of 28, sample fraction=0.44, honesty fraction=0.71
- Default GRF parameters give similar results

Back

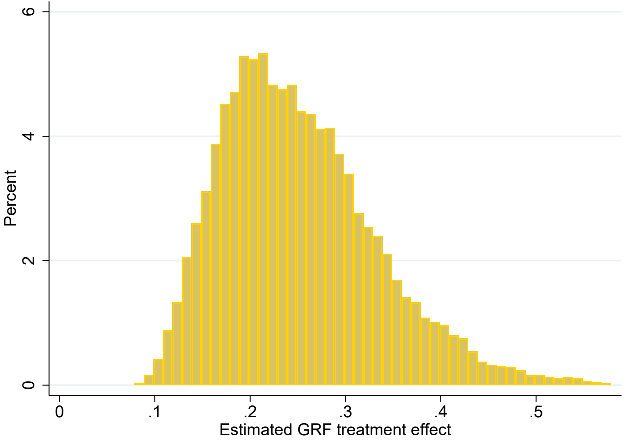
Best Linear Predictor Test

Tests how well the forest's estimates fit the actual duration of spells

	Estimate	Std. Error
α	1.01	0.002
β	1.28	0.018

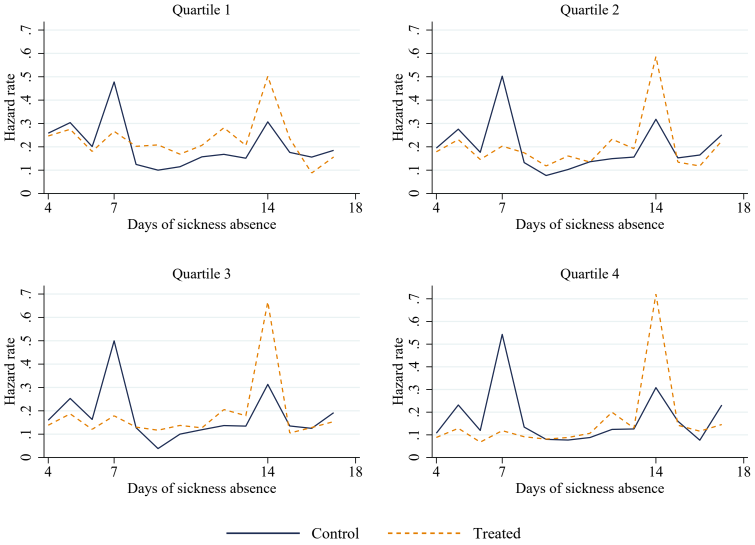
Back

Probability of Returning to Work in Week 2: Treatment Effect Estimates



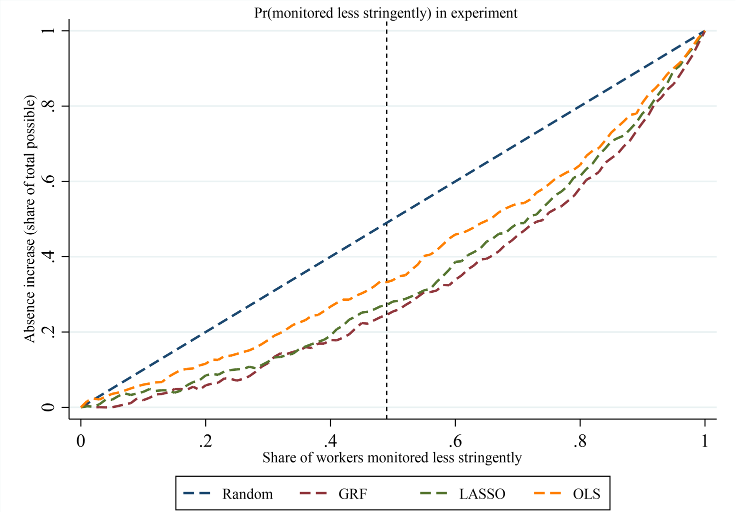
Correlation with duration estimates: 0.90 Back

Hazard Rates by Predicted Treatment Effects



Back

Targeted Monitoring Policy: Comparison to OLS and LASSO



Back

QINI Coefficients of Different Targeting Policies

	QINI	Std. Error
Full GRF	0.202	0.015
Sickness absence history	0.092	0.015
Social payment income share	0.093	0.014

Back