Received: 10 December 2018 | Revised: 3 December 2019 | Accepted: 11 January 2020

DOI: 10.1002/hec.4002

RESEARCH ARTICLE

Health Economics WILEY

Health effects of reduced workload for older employees

Espen Bratberg[1,2] | Tor Helge Holmås[1] | Karin Monstad[1]

[1]Social Science, NORCE Norwegian Research Centre, Bergen, Norway

[2]Department of Economics, University of Bergen, Bergen, Norway

Correspondence
Karin Monstad, Social Science, NORCE Norwegian Research Centre, Nygårdsgaten 112, 5008 Bergen, Norway.
Email: kamo@norceresearch.no

Funding information
Norwegian Research Council, Grant/Award Number: 237808

Abstract

To keep elder employees in the labour force, introducing age-dependent job conditions can be a policy measure. However, we know little about the effect of such initiatives. We investigate the effects of a particular programme in Norway that reduces the workload of teachers at age 55 but maintains the same wage. Evaluation of this programme is well suited to a difference-in-difference analysis, where the control group is teachers slightly too young to be eligible for the workload reduction. Using full population register data for the period 2006–2013, we analyse the effects of the programme on health as indicated by sickness absence and health care utilization. We find that whereas there is no effect among women, the workload reduction causes a decrease in sickness absence and an improvement in mental health among males. These results, which are robust to a placebo test, to extending the pretreatment period, and to dropping single birth cohorts, are driven by a subgroup of men whose prior health status is poor.

KEYWORDS

absenteeism, aging, health, sickness absence, working conditions, work intensity, workload

JEL CLASSIFICATION

I18; J14; J22

1 | INTRODUCTION

Achieving increased labour market activity of older individuals has been high on the political agenda for several years, with the Organization for Economic Co-operation and Development (OECD) noting that population aging is one of the most important challenges facing its members. An older society may place a rising economic burden on the working-age population, put public finances under heavy pressures, and reduce growth in living standards (OECD, 2006). This concern motivates policies for retaining employed workers in the labour force, including measures to improve working life conditions of older employees. The current analysis investigates the causal effect of such a preventive initiative, being the reduction in workload given to teachers in Norway at the age of 55, that is, 7 years before they have the option of early retirement. This age is of great interest when measuring the potential for active and healthy aging, as demonstrated by the Active Aging Index, developed by United Nations' Commission for Europe and the European Commission. We analyse the intention-to-treat effect of a negotiated workload reduction on sickness absence and health, exploiting the fact that employees only a year or less apart in age face different workload demands in the same period.

In Norway, teaching requirements are set in nationwide agreements between teachers' unions and their counterpart. The intention behind the initiative analysed, introduced in 2006, was to relieve older teachers of some workload by reducing the number of lessons per week by 5.8% (Norwegian Association of Local and Regional Authorities, 2006). The workload reduction affected teachers in primary, lower secondary, and upper secondary schools. All teachers of the relevant age have the same percentage reduction in workload, irrespective of union membership, the subject taught, grade, and school level, whereas contracted hours and labour income was not affected. Further details on the initiative are found in Section 2.4.

An implicit premise of this initiative is that health deteriorates with age, implying lower productivity. In health economics theory, the development of health over the life cycle is a key issue. Within the classical model of Grossman (1972), eternal life is possible, if the health deterioration that comes with age is outweighed by health production from market and nonmarket goods (e.g., medical care and exercise, respectively). Labour is then primarily a source of income, although it also gives less room for leisure that is an important health production input. Extending on this model, Muurinen and Le Grand (1985) argue that when generating earnings, workers with low human and/or financial capital are constrained in their choices and may therefore prefer to substitute health for wealth, which would imply that the health of poorer and less educated workers would deteriorate more rapidly. This is also the prediction from Case and Deaton's model, given that perfect health repair is impossible (Case & Deaton, 2005).

Empirically, it is difficult to distinguish between the health effect of workload (job intensity), other work conditions such as job injury risk, or hours worked. For instance, in a boom or following a reform of the standard work week, it is plausible that both hours worked and workload will change, and income is typically affected. The bulk of the existing literature does not aim at making such a distinction. A recent example is the work by Cygan-Rehm and Wunder (2018), who note that due to data limitations, their estimated effect of an extra hour's work time is a "total effect," that is, it "incorporates any potential channels through which increased working time might affect health such as, e.g., changes in worker's tasks and pressure at work."

In the literature that specifically investigates the relationship between working conditions and health, the working conditions studied are rather diverse but can be categorized as physical demands and psychosocial risk factors (for a recent overview, see Defebvre, 2018). Most related to our work are the analyses studying the impact of psychosocial factors. Using retrospective panel data, Defebvre (2018) analyses the effect of treatment based on self-declared exposure to a set of various psychosocial risk factors. He concludes that such exposure—separately or combined with physical demands—is associated with a higher number of self-declared chronic diseases (Defebvre, 2018). Cottini and Lucifora (2013) find that high work intensity (working at very high speed with tight deadlines at least half of the time) has adverse effects on mental health, after controlling for a number of other work conditions and long work hours. Their study uses repeated cross-sectional survey data (Cottini & Lucifora, 2013).

In the bulk of the literature on workload and health, health is proxied by self-reported health. An exception is the work by Ose (2005), who applies sickness absence as a proxy for health. There have been diverging views on sickness absence as a global indicator of health. Kivimäki et al. (2003) conclude that medically certified sickness absence can be used as a measure of ill health in working populations. However, this literature also recognizes that sickness absence can be related to factors other than health. Absence can also be used as an indicator of effort (Hesselius, Nilsson, & Johansson, 2009) and is reduced when job environments become less secure (Bratberg & Monstad, 2015;Arai & Thoursie, 2005; Ichino & Riphahn, 2005). In this analysis, we study use of health care (general practitioner [GP] visits) as well as the effect on certified sickness absence.

In general, analyses of the relationship between workload and health are faced by several endogeneity concerns. First, the choice of education and occupation, and the hiring of employees, is likely to be related to individual health. Next, health-related exits from the labour force may bias effect estimates downwards (the so-called "Healthy worker effect"). Our research design implicitly addresses these sources of selection bias. We study one occupation and a sample of employees who are still working, at approximately the same age. The assignment of treatment depends on month of birth within a narrow age window, that is, exogenous to individual health. Furthermore, the analysis benefits from panel data that enable us to apply a difference-in-difference (DD) approach and the fixed effects estimator to control for time-constant unobserved heterogeneity. Finally, using administrative registry data, we avoid problems of recall or justification biases, which is a concern in surveys of self-reported health.

Therefore, the main contribution of our analysis is that we address the endogeneity of workload with respect to individual health and isolate the effect of workload from the effect of income and working hours. Employees receive full compensation in terms of income with the change in workload, and the initiative/programme does not affect hours worked. We find some indications that the workload reduction causes a decrease in sickness absence and an improvement in mental health among males. A subgroup of men, whose prior health status is poor, experience a considerable health improvement in terms of certified sickness absence and GP visits for mental health problems. Among women, sickness absence is unaffected, but women of poor pretreatment health have fewer GP visits for cardiovascular problems.

This paper proceeds as follows. Section 2 provides some institutional background on the Norwegian labour market in general and the education sector in particular. Section 3 presents our empirical strategy and Section 4 the requisite data. Section 5 details our results and provides several robustness checks. Section 6 concludes.

2 | INSTITUTIONAL BACKGROUND

2.1 | Sickness insurance

In Norway, sickness insurance is mandatory and regulated by law, covering all employees who have been with the same employer for at least 2 weeks, with a 100% replacement rate from the first day. A medical certificate is required for spells of absence of more than 3 or 8 days, depending on whether the employer has signed the Tripartite Agreement on a more inclusive working life ("IA Agreement") or not. The first 16 days of absence are paid by the employer (the employer period), whereas the remaining period is paid by social insurance. The maximum period of benefits is 1 year, including the employer period. The level of sickness absence is high, particularly among women (Markussen, Røed, Røgeberg, & Gaure, 2011).

Unlike many countries where long-term unemployment precedes early retirement, in Norway the early retirement pathway is typically via health-related benefits, that is, long-term sickness absence followed by disability benefits (OECD, 2013). Therefore, a reduction in sickness absence is the subject of much political attention.

2.2 | Health care sector

Norway's health care system is characterized by mandatory health insurance and is largely financed by general taxes. Patients' out-of-pocket payments are in practice very low due to a tight annual expenditure cap. All individuals may list with a GP. Administrative registers for remuneration of GPs are an important source for research on health care utilization both at the individual and population level.

2.3 | Norwegian labour market

At the national level, the primary policy measures to retain older workers in the workforce are the following. First, a 2011 pension reform provides workers with incentives to stay at work longer. It primarily affects private sector workers and cohorts born after 1953, whereas public sector workers born 1954 and later will be gradually affected because their pension will be subject to new indexation rules and life-expectancy adjustments (OECD, 2013). Second, the IA Agreement, which was launched in 2001 and has been extended several times since, is intended to promote initiatives at the company level. Lastly, by law, workers aged 60 years and above are entitled to 1 week extra holiday, irrespective of sector or industry.

2.4 | The treatment

In the 2006 agreement between Union of Education Norway and The Norwegian Association of Local and Regional Authorities, it is made explicit that the initiative should result in a workload reduction for teachers who pass a certain age threshold. Although the local employer (school) might assign these teachers "other pedagogical tasks," for instance mentoring younger colleagues, the reduction in lessons should not be outweighed by an increase in other tasks and duties (The Norwegian Association of Local and Regional Authorities, 2006). The age threshold is such that, in a given school year, teachers whose 55th birthday is before December 31 will receive a 5.8% reduction in teaching requirements, whereas their slightly younger colleagues, who turn 55 during the school year, but after December 31, will not.

Teaching requirements differ by subject and school level (the required number of lessons per week ranges from 26 in primary school to 16.4 for a few subjects in upper secondary). Given that all teachers of the relevant age have the same *percentage* reduction in workload, it amounts to 1 to 1.5 fewer lessons per week. (Teachers have other duties besides giving lessons. It is stipulated that classroom teaching and tasks directly linked to it, such as preparation and correcting homework and tests, amounts to 67% of contracted hours (Union of Education Norway, 2012).

The education sector is dominated by public schools. Teachers' employers (municipalities for primary schools and lower secondary and counties for upper secondary) have essentially the same senior policy towards teachers as towards other employees, with the exception of workload regulations. The workload reduction at age 55 is quantifiable and standard across the public education sector. It is also easy to communicate, and there is no income loss; hence, there is every reason to consider that the uptake would be high. In principle, a teacher may turn down the option of reduced workload if he/she prefers teaching the same number of lessons as before rather than potentially being assigned additional "pedagogical tasks." However, the wage income is the same whether the option is accepted or not (for a given number of contracted hours).

The workload reduction can affect health and sickness absence in several ways. A less stressful work situation should in itself be health improving. In addition to this direct health effect, workers may interpret the workload reduction as an acknowledgement of greater job demands, and workers who consider that they are treated well (given a workload reduction) respond with having less sickness absence (Fehr & Gächter, 2000). Following the same logic, their colleagues who have to wait another year to receive the same workload reduction could respond by increasing their sickness absence, in which case the effect would be underestimated.

3 | EMPIRICAL STRATEGY

The challenges to estimating causal impacts of policy measures are well known. Unless assignment to a treatment, D, is independent of the potential outcomes, a plain comparison of treated and untreated units typically will be biased because it mixes up the treatment effect with selection to treatment. With repeated observations, the problem is alleviated: instead of assuming that potential outcomes are statistically independent of the treatment, it suffices that in absence of treatment, treated and untreated follow a common trend. Formally, with treatment in t,

$$E\left(Y_t^0 - Y_{t-1}^0 | D=1\right) = E\left(Y_t^0 - Y_{t-1}^0 | D=0\right), \tag{1}$$

where Y^0 denotes the untreated outcome. This is the identifying assumption for DD methods.

The design of the workload reduction agreement makes it well suited for evaluation with DD. The Norwegian school year starts in mid-August and ends in late June. In the beginning of the school year, teaching plans are made where teachers are allocated to classes and courses. The plan takes into account whether teachers are entitled to the age-contingent workload reduction during the school year. Notably, only teachers whose 55th birthday is before December 31 receive the reduction, whereas colleagues who turn 55 during the school year but after December 31, do not. Having been born in the last (treated) rather than the first (controls) part of the year therefore acts as a natural experiment available for a DD strategy. By design, the estimated effects of this experiment relates to the first (school) year of treatment because in later years, both the treatment and the control groups are eligible for the workload reduction. Because we do not observe the actual workload of each individual teacher, the estimated effects should be interpreted as intention-to-treat effect estimates.

In what follows, t denotes the school year that begins in t (i.e., t = 2007 refers to the school year 2007–2008). For school year t, the treatment group includes teachers who turn 55 before New Year's Eve (for t = 2007, those who are born July–December 1952). The control group consists of teachers whose 55th birthday is in school year t, but after December 31 (those who are born January–June 1953 for t = 2007).

We apply a standard regression DD where Y_{it} denotes the outcome of teacher i, who turns 55 in school year t. The dummy variable D indicates the treatment group. Another dummy, T, equals 1 in the treatment year and 0 in the year before the treatment year, that is, the school year when the teacher turns 54. In the estimated equation presented below, the coefficient β_1 captures the effect of getting 1 year older in the control group. Because we pool several school years, we also include year fixed effects, γ_t. Finally, individual fixed effects, γ_i, are included to pick up age effects within a birth cohort and other individual heterogeneity. The individual fixed effects also include the pretreatment difference between the treated and control group. Then, the DD estimator is δ in

$$Y_{it} = \beta_0 + \beta_1 T_t + \delta(D_i \times T_t) + \gamma_i + \gamma_t + u_{it}, \tag{2}$$

and the equation is estimated gender-wise by the within groups (fixed effect) estimator with standard errors clustered at the individual level.

By restricting the analysis to teachers born in July–December in t − 55 (treated) and January–June t − 54 (controls), we obtain a control group that is quite close to the treatment group—on average 6 months younger. Moreover, the treated and controls are homogenous in education and occupation. These facts are in favour of the common trend assumption, which is also checked by a placebo test and by extending the regression equation with more years before treatment.

As discussed in more detail below, the first *calendar* year with available data is 2006, and the last is 2014. Thus, the first *school* year of treatment (T = 1) is 2007/2008, with T = 0 in 2006/2007. The oldest treatment cohort is born in 1952, with controls born in 1953. The youngest treatment/controls are born in 1958/59 with treatment (school) year 2013/2014.

4 | DATA

4.1 | Sources

The key data source is the FD-Trygd database, which links administrative information from the Norwegian Labour and Welfare Administration and Statistics Norway. This database covers all Norwegians from 1992 onwards and provides information on gender, month, and year of birth, along with detailed information on certified sick leave covered by social insurance (start and stop dates), disability, work history (date of job entry and exit, sector, industry, occupation, and contracted hours), and the level and type of education, and so forth.

For the purpose of this project, data from FD-Trygd was merged with an administrative dataset containing individual level health information available from 2006 onwards. The Norway Control and Payment of Health Reimbursement register holds information on all invoices sent by GPs for remuneration, that is, for each patient contact (consultation), whether at the GP's regular office or at an emergency centre. At each consultation, the GP records a diagnosis according to the International Classification of Primary Care, Second edition classification. For each consultation, we also know whether a sickness certificate was issued.

4.2 | Outcomes

We consider several outcomes, related to a school year. The Norwegian Labour and Welfare Administration data contains information on sickness absence remunerated to employers from National Insurance. These are episodes that last at least 16 days, but the data provides information on the full duration of an episode, including the part paid by employers. We use this information to define the variable *sickdays* measuring mean absence days per month. From the Norway Control and Payment of Health Reimbursement data, we obtain *GP visits* (per year) and *sick notes*, indicating the number of sickness certificates issued per year. The sick notes variable tells whether a sickness certificate was issued, not the duration of absence. We also define three variables related to the diagnosis recorded at the visit; psychiatric (*P*), musculoskeletal (*L*), and cardiological (*K*), that is, diagnoses within International Classification of Primary Care, Second edition Chapters P, L, or K, respectively. The former two represent the most frequent illnesses identified in certified sickness absence, whereas the latter represents potentially serious impairments. Each variable measures the number of occurrences per year.

4.3 | Sample

For each school year starting 2007–2013, we identify teachers born in July–December in t − 55 (treatment group) or in January–June in (t + 1) − 55 (control group). The population of teachers is defined based on information about industry, sector, education, and (where possible) occupation. Within the relevant industries (school levels), we restrict the sample to individuals whose level of education is compatible with teaching at that level. For instance, to be qualified as a teacher at upper secondary, academic track, 1 year of study at a university or university college is required (Sjaastad, Carlsten, Wollscheid, Reiling, & Federici, 2016). As the workload reduction draws on specific employer–union agreements, the sample is restricted to teachers in public schools, that is, schools owned and operated by either municipalities or counties.

The sample consists of observations of teachers employed in August at the start of school years t and $(t - 1)$. We identified 15,985 individuals in the relevant cohorts and school years who met the criteria outlined above. After dropping 406 individuals who were not present in both years, the final sample included 15,579 teachers, comprising a treatment group of 2,449 men and 4,860 women born 1952–1958 and a control group of 2,650 men and 5,620 women born 1953–1959.

4.4 | Descriptive statistics

Table 1 provides summary statistics at the beginning of the school year before treatment. Except for month of birth, the treatment and control groups are similar with respect to marital status, nationality, education, school level, and labour supply. We note that at pretreatment, women have more sickdays and sick notes than men. There are also relatively more part-time workers among women. For men, sickness absence (days and notes) is higher in the treatment group than in the control group, and they have more GP consultations because of mental health problems than men in the control group. There are no such differences for women. Figures 1 and 2 show monthly data on sickdays and sick notes before aggregation into yearly averages. There is seasonal variation, but we also see that treated men have more absence than the untreated in $t - 1$, as reported in Table 1. Even though the treated sample is older at baseline, this difference is somewhat unexpected, as the treatment group is not selected on any other criteria than date of birth. Figures S1–S4 in

TABLE 1 Summary statistics at baseline

	Men				Women			
	Controls		Treated		Controls		Treated	
Description	**Mean**	***SD***	**Mean**	***SD***	**Mean**	***SD***	**Mean**	***SD***
Female								
Month of birth	3.575	1.660	9.468	1.744	3.568	1.682	9.416	1.711
Norwegian born	0.944		0.954		0.952		0.948	
Education level								
Upper secondary	0.025		0.032		0.004		0.003	
Bachelor	0.789		0.801		0.905		0.897	
Master or more	0.186		0.167		0.091		0.101	
School level								
Compulsory	0.494		0.500		0.785		0.770	
Upper secondary	0.506		0.500		0.215		0.230	
Labour supply								
Full time	0.844		0.838		0.751		0.750	
Part time	0.156		0.162		0.249		0.250	
Hours/week	34.820	6.637	34.777	6.856	33.016	7.223	33.172	7.143
Outcomes[a]								
Sickdays/month	1.272	4.257	1.683	5.217	2.084	5.478	2.101	5.537
Sick notes	0.606	1.689	0.718	1.914	1.041	2.201	1.032	2.177
GP visits	1.913	2.695	2.034	2.802	2.652	2.993	2.677	2.927
GP visits (*P*)[b]	0.414	1.875	0.579	2.671	0.658	2.293	0.693	2.375
GP visits, (*L*)[b]	0.798	2.116	0.795	2.265	1.172	2.677	1.214	2.653
GP visits, (*K*)[b]	0.592	1.906	0.683	2.268	0.436	1.514	0.409	1.442
N	*2,650*		2,449		5,620		4,860	

Note. Treatment cohorts are born July–December 1952–1958; control cohorts are born January–June 1953–1959.

Abbreviations: GP, general practitioner; *SD*, standard deviation.

[a]Except for sickdays, all outcomes relate to a school year (10-month period).

[b](*P*): psychiatric diagnosis, (*L*): musculoskeletal diagnosis, (*K*): cardiovascular diagnosis (chapters P, L, and K in the ICPC-2 codebook).

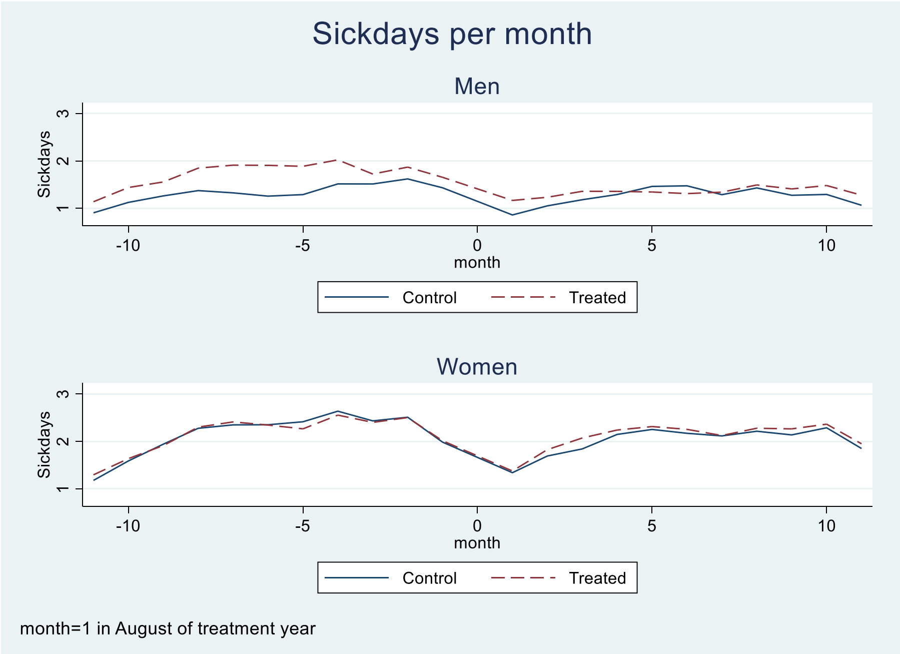

FIGURE 1 Mean number of sickdays per month, by treatment status and gender [Colour figure can be viewed at wileyonlinelibrary.com]

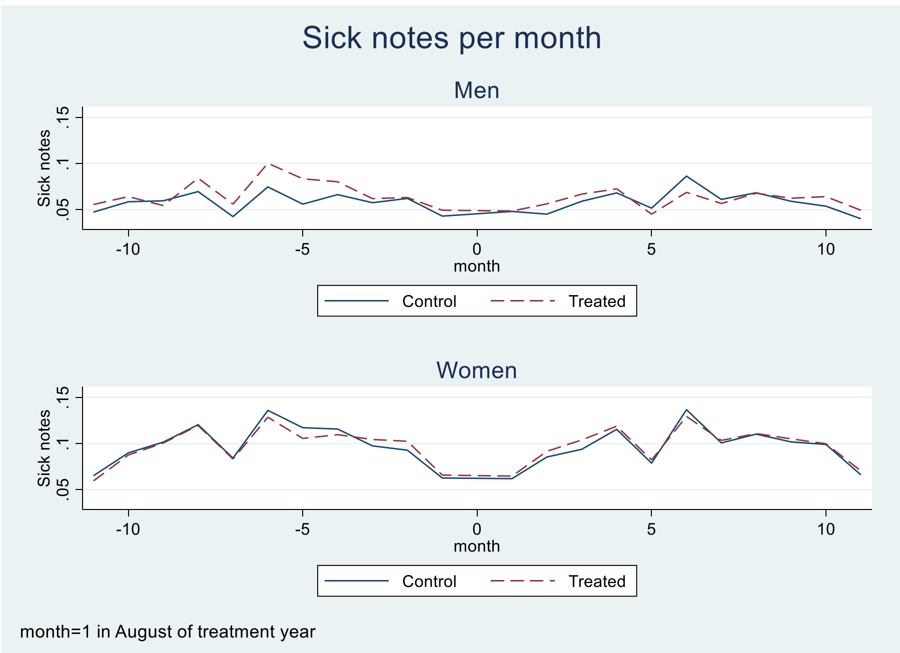

FIGURE 2 Mean number of sick notes per month, by treatment status and gender [Colour figure can be viewed at wileyonlinelibrary.com]

the online appendix show plots according to Figure 1 and 2 broken down by year, and it does not appear that this pattern is driven by specific years or birth cohorts. In the empirical analysis, where we estimate the outcomes separately for men and women, the DD design takes care of any pretreatment differences in outcome levels between the treatment and control groups. The outcome pattern of men shows that it is important to probe the common trend assumption. We also explore whether results are driven by outcome patterns in particular years.

5 | RESULTS

5.1 | Main results

Table 2 presents the main DD results. In addition to the treatment effect, we report estimates for the posttreatment effect (δ and β_1 in Equation [(2)], respectively). We also report estimates from a separate regression testing whether

TABLE 2 Effects of work load reduction, full sample. Difference-in-difference estimates with individual fixed effects

		Men		Women	
Outcomes:		**Estimates**	**Sample mean at baseline**	**Estimates**	**Sample mean at baseline**
Sickdays	Treatment effect	−0.311 (0.163)	1.470	0.075 (0.137)	2.092
	Posttreatment	0.016 (0.103)		−0.014 (0.092)	
	Pretreatment diff.	0.294 (0.277)		0.194 (0.214)	
Sick notes	Treatment effect	−0.092 (0.056)	0.660	0.031 (0.051)	1.037
	Posttreatment	0.023 (0.038)		−0.002 (0.035)	
	Pretreatment diff.	0.092 (0.109)		0.100 (0.088)	
GP visits	Treatment effect	0.009 (0.077)	1.971	−0.023 (0.058)	2.663
	Posttreatment	0.110* (0.053)		0.007 (0.040)	
	Pretreatment diff.	0.070 (0.180)		0.135 (0.122)	
GP visits, P diagnosis	Treatment effect	−0.104 (0.063)	0.493	0.028 (0.048)	0.674
	Posttreatment	0.043 (0.040)		0.013 (0.034)	
	Pretreatment diff.	0.190 (0.118)		0.058 (0.103)	
GP visits, L diagnoses	Treatment effect	−0.024 (0.064)	0.797	0.007 (0.057)	1.191
	Post treatment	0.030 (0.046)		0.057 (0.039)	
	Pretreatment diff.	0.127 (0.129)		0.079 (0.105)	
GP visits, K diagnoses	Treatment effect	0.084 (0.064)	0.636	−0.020 (0.030)	0.423
	Post treatment	0.070 (0.039)		0.038 (0.021)	
	Pretreatment diff.	−0.103 (0.135)		−0.003 (0.066)	
	Observations	10198		20960	

Note. Standard errors clustered at the individual level in parentheses. Treatment cohorts are born July–December 1952–1958; control cohorts are born January–June 1953–1959. For all outcomes, time and treatment effects are estimated with individual fixed effects and controls for school year. In separate regressions, we test whether pretreatment means differ statistically, when controls for school year and month of birth are included.

Abbreviations: diff., difference; GP, general practitioner

*$p < .05$. **$p < .01$. ***$p < .001$.

pretreatment differences are statistically significant. Overall, the estimates do not reach statistical significance at 5% level. For women, all estimates are far from statistical significance. However, there is indication of an effect for *men*: the estimated effect on monthly sickdays is −0.31, with a p value of.06. For sick notes per year, the estimate is −0.092 ($p = .10$). The effect on GP visits due to mental health problems is −0.105, at a similar precision level. The (imprecisely estimated) effect on sickdays is quite large: a 21% reduction from the average 1.47. The effect on the closely related outcome sick notes is also large, amounting to a 14% reduction, which makes sense, as a sickness certification may vary in length. The fact that the two measures of sickness absence come from different sources, but point in the same direction, adds credibility to the interpretation of a decrease in sickness absence among men. There are also (statistically weak) indications that lower workload causes a reduction in GP visits for psychological issues. In comparison, the estimate of β_1, the aging effect, is much smaller in magnitude and not statistically significant in any of these estimations.

Next, we explore heterogeneity along several dimensions in addition to gender: previous health status, school level, and hours worked.

In Table 3, the sample is split according to previous health status, proxied by the number of GP visits 2 years before treatment. The majority have no or one visit, and it is not obvious where to cut the sample. Therefore, we report results where the sample is split at the 8th, 9th, and 10th gender-wise decile of visits.

The results clearly indicate that the effects, where found, are driven by teachers with a history of previous poor health. In the 70/30 split, the effects on sick notes and sick days are statistically significant for the "poor health" group of men, and the finding on mental health visits remain. The effects are large compared with the sample means at baseline, suggesting an average reduction of about 40%. However, even in this subgroup, the majority has little absence, so most of the action is among those who have more than average absence. In the "healthy" group, there are no effects. For the other cuts, the point estimates for the poor health subgroup are similar but less precisely estimated. The larger standard errors are probably a mechanical effect of smaller sample sizes. Interestingly, we now also estimate a negative effect on visits due to cardiovascular problems for women in the "unhealthy" group (in the 70/30 and 80/20 splits).

Heterogeneity by school level and hours worked are reported in Appendix Tables S1 and S2. In short, the estimates are more imprecise than in the main analysis, and no clear inferences can be made, even though the results on sickdays for men go in the same direction.

TABLE 3 Effects of work load reduction, by gender and prior health status. Difference-in-difference estimates with individual fixed effects

	(I)				(II)				(III)			
	Men		Women		Men		Women		Men		Women	
Outcomes	70%	30%	70%	30%	80%	20%	80%	20%	90%	10%	90%	10%
Sickdays	−0.156	−1.169*	0.230	−0.124	−0.263	−1.117	0.146	0.049	−0.313	−1.657	0.133	0.068
	(0.211)	(0.500)	(0.169)	(0.383)	(0.204)	(0.641)	(0.166)	(0.467)	(0.202)	(1.017)	(0.164)	(0.679)
Baseline mean	1.012	2.798	1.490	3.215	1.122	3.155	1.619	3.380	1.292	3.633	1.798	3.521
Sick notes	−0.034	−0.362*	0.075	−0.017	−0.073	−0.304	0.061	−0.004	−0.076	−0.604	0.031	0.213
	(0.073)	(0.180)	(0.063)	(0.148)	(0.072)	(0.228)	(0.062)	(0.183)	(0.071)	(0.357)	(0.062)	(0.274)
Baseline mean	0.434	1.288	0.722	1.785	0.486	1.462	0.795	1.913	0.564	1.723	0.888	2.199
GP visits	0.154	−0.348	0.028	−0.082	0.113	−0.364	0.008	−0.042	0.053	−0.350	−0.049	0.433
	(0.101)	(0.236)	(0.072)	(0.172)	(0.097)	(0.317)	(0.070)	(0.217)	(0.096)	(0.494)	(0.069)	(0.343)
Baseline mean	1.420	3.568	1.994	4.514	1.550	4.005	2.139	4.938	1.725	4.898	2.359	5.778
GP visits, *P*	0.039	−0.493*	0.102	−0.081	−0.031	−0.416	0.076	−0.046	−0.068	−0.499	0.074	−0.163
diagnoses	(0.070)	(0.225)	(0.059)	(0.148)	(0.071)	(0.297)	(0.058)	(0.183)	(0.074)	(0.495)	(0.058)	(0.280)
Baseline mean	0.280	1.132	0.452	1.279	0.343	1.255	0.506	1.392	0.414	1.518	0.566	1.752
GP visits, *L*	−0.063	0.227	0.057	0.186	−0.064	0.413	0.038	0.311	−0.002	0.217	0.029	0.665
diagnoses	(0.088)	(0.199)	(0.071)	(0.171)	(0.084)	(0.259)	(0.069)	(0.215)	(0.083)	(0.416)	(0.068)	(0.343)
Baseline mean	0.623	1.510	0.879	2.063	0.665	1.744	0.948	2.257	0.746	2.092	1.056	2.599
GP visits, *K*	0.096	−0.222	0.043	−0.247**	0.073	−0.257	0.014	−0.243*	0.020	−0.104	−0.012	−0.258
diagnoses	(0.066)	(0.209)	(0.037)	(0.090)	(0.065)	(0.290)	(0.035)	(0.113)	(0.070)	(0.436)	(0.036)	(0.165)
Baseline mean	0.368	1.345	0.292	0.743	0.421	1.573	0.326	0.783	0.512	1.901	0.369	0.858
Observations	4696	1756	10324	3872	5258	1194	11444	2752	5884	568	12856	1340

Note. Standard errors clustered at the individual level in parentheses. Main sample is split by gender and individuals' number of GP visits two years before treatment. Cut-offs at 8th, 9th, and 10th gender-wise decile of visits (see groups of columns marked I, II, and III, respectively, where estimates for the poorest health group within each gender are reported in the right column). Treatment cohorts are born July–December 1953–1958; control cohorts are born January–June 1954–1959. For all outcomes, treatment effects are estimated with individual fixed effects and controls for school year and time effects (posttreatment indicator).
Abbreviation: GP, general practitioner.
*$p < .05$. **$p < .01$. ***$p < .001$.

5.2 | Placebo test

The results thus far indicate a favourable treatment effect on male teachers' sickness absence and psychological health, with negative estimates. However, this finding may be spurious if the assumption of a common trend is violated. As one check of the identifying assumption, we perform a placebo test. Here, "treatment" is 1 year earlier, that is, in the school year when "treated" teachers all have turned 54 before the autumn semester ends. Given that, in reality, there is no treatment at that age, we would not expect to find negative coefficient estimates of any statistical power on the placebo treatment. Table 4 details the placebo results and confirms our expectation. None of the coefficients for sick days, sick notes, or GP visits for mental health issues are statistically significant, and some of the point estimates have the "wrong sign," including a statistically significant positive effect on one diagnosis variable for men. Because the "treated" group is older than the "control" group, one possible explanation for positive estimates could be an aging effect that our design does not fully control for. If such an effect exists, it would lead to an underestimation of the true treatment effect. We conclude that the placebo tests provide no evidence calling into question the main results.

5.3 | Extending the pretreatment period

With more than one observation before treatment, it is possible to ease the common trend assumption by including group specific trends in the model. Table 5 reports results from the extended model for sickdays and sick notes, where we have added two extra lags. Because our data on sick notes is available from 2006 onwards, the extension is only possible for treatment/control cohorts 1954/55–1958/59. Our data for sickdays goes further back, so we can do this analysis for the same cohorts as in Table 2. The results are consistent with the main results, with a negative and larger effect on sickdays, whereas the effect on sick notes is still imprecisely estimated. In conclusion, this additional analysis supports the main results.

TABLE 4 Placebo treatment, full sample. Difference-in-difference estimates with individual fixed effects

Outcomes	Men	Women
Sickdays	0.306 (0.183)	−0.050 (0.147)
Sick notes	0.065 (0.064)	−0.044 (0.055)
GP visits	0.080 (0.086)	0.034 (0.063)
GP visits, *P* diagnoses	0.076 (0.065)	−0.034 (0.051)
GP visits, *L* diagnoses	−0.011 (0.074)	0.023 (0.062)
GP visits, *K* diagnoses	0.136* (0.064)	−0.011 (0.033)
N	8,202	17,572

Note. Standard errors clustered at the individual level in parentheses. Placebo treatment 1 year before actual treatment, that is, treatment cohorts are born July–December 1953–1958, control cohorts are born January–June 1954–1959. For all outcomes, treatment effects are estimated with individual fixed effects and controls for school year and time effects (posttreatment indicator).
Abbreviation: GP, general practitioner.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

TABLE 5 Effect on sickdays and sick notes. Difference-in-difference estimates with individual fixed effects, with three pretreatment years

	Sickdays		Sick notes	
	Men	Women	Men	Women
Trend	−0.078 (0.174)	−0.077 (0.172)	0.000 (0.027)	0.038 (0.024)
Trend×treated	0.161 (0.084)	0.032 (0.071)	0.043 (0.037)	−0.024 (0.032)
Treatment year	0.024 (0.140)	−0.015 (0.124)	0.043 (0.064)	−0.038 (0.058)
Treatment effect	−0.449* (0.212)	0.052 (0.178)	−0.161 (0.094)	0.066 (0.080)
Observations	20,224	41,629	12,793	28,201

Note. Standard errors clustered at the individual level in parentheses. Sickdays: treatment cohorts are born July–December 1952–1958, control cohorts are born January–June 1953–1959. Sick notes: treatment cohorts are born July–December 1954–1958, control cohorts are born January–June 1955–1959.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Even though Figures S1–S4 in the online appendix did not reveal any clear patterns in absence and sick notes across treatment years (confer Section 4.4), we investigated the sensitivity of our results to dropping particular years from the analysis. If the results are driven by individual outliers or macro effects that are not picked up by the year fixed effects, we would expect that to show up here. Figures S5–S8 (analysis with one pretreatment observation) and S9–S12 (three pretreatment observations) in the online Appendix reveal no systematic changes in the estimates. Comparing to the main results in Table 2, we see that the point estimate on sickdays for men (−0.31 in Table 2) is consistently below −0.2 but with large confidence intervals, and the sick notes estimate (−0.092 in Table 2) is always below −0.07.

6 | DISCUSSION AND CONCLUSION

The contributions of this analysis are twofold. First, we investigate a research question that is important but seldom analysed, namely the effect of policy initiatives aimed at improving job conditions for older workers, in this case, an age-dependent reduction in workload. Second, to estimate the causal impact of workload on the labour supply of older employees, we utilize a setting where workload varies by employee's month of birth, which is exogenous with respect to both the employer and employees. This initiative has not been evaluated before. The estimated effect is an intention-to-treat effect because we cannot observe whether the individual employee makes use of the option. The initiative evaluated is modest, but—if there is any impact at all—we expect it to have favourable effects on health and sickness absence.

Our main result is that the workload reduction causes a decrease in sickness absence among males, and there are indications that this is from less psychological strain. We support these results with a placebo test. The effects for men are of large magnitude; a reduction in sick days and sick notes of 21 and 14%, respectively, albeit imprecisely estimated. If there exists another treatment other than the workload reduction offered to employees at the age of 55, it would bias the results. However, we know of no such treatment. Although a relevant concern could be that the results for men to some extent hinge on the higher absence level in the treated relative to the control group before treatment, the sensitivity analysis does not indicate that this difference is due to idiosyncrasies related to particular years or birth cohorts.

Our analysis has detected large heterogeneity in the response to workload reduction among different subgroups. The results for men are driven by male teachers whose prior health status is poor, that is, their number of GP visits in the pretreatment year is in the highest 30%. We find this to be an interesting result, showing that for these men, even a modest workload reduction has an impact on sickness absence and health. Among women, there are no clear effects, except a reduction in number of GP visits for *K* diagnoses for women in the poorest health group.

Differences in job characteristics do not easily explain the difference in effect estimates between men and women because we compare employees within the same occupation. Our finding that male and female employees react differently to institutional changes is then in line with previous research. Not only are women's sickness absence levels higher than men's (see, for instance, Markussen et al., 2011), but more importantly, women also appear less responsive to workplace changes. Elsewhere, these include negative organizational shocks/change in job security (Bratberg & Monstad, 2015; Ichino & Riphahn, 2005); their absenteeism reacts differently to social security reform (Johansson & Palme, 2005), and their labour supply responses to a negative health shock are different from those of men (Trevisan & Zantomio, 2016). In general, this gender difference could arise from differences in options or preferences, possibly shaped by gender roles. Finally, we note that women are much more likely to work part-time than men at baseline, as Table 1 shows. This allocation might be health-related, indicating that women who are still employed at the age of about 53–54 to a greater extent than same-age men have adjusted to the work strain by reducing weekly hours worked. If so, effect estimates for women will be biased downwards.

Some caveats are in place. When six outcomes are considered, like in the present analysis, a possible concern is that testing several hypotheses increases the probability of Type I errors, that is, falsely rejecting a null hypothesis of no effect. A Bonferroni correction says that for rejection at a 5% significance level, the p value must be below.05/.06 = .008. That is very conservative and increases the probability of Type II error, but still the precision of our results as reported may be too optimistic.

On the other hand, the initiative may have favourable external effects that are not considered in this analysis. If teacher health improves and sickness absence is reduced, it would benefit pupils' learning and colleagues who do not need to step in as locums. Our results show a decrease in long-term sickness absence, there might be a decrease in sickness absence spells shorter than 16 days, as well. Also, we study employees at ages 53–55 and use a narrow time window to trace the causal effect, which relates to the first 10 months after the start of treatment. Because of the nature of the intervention, we cannot conclude with respect to later outcomes such as retirement age, the uptake of disability pension, or health status. However, we find it worth mentioning that in Norway, disability pension is typically preceded by long-term sickness absence. This adds relevance to our finding that the workload reduction particularly impacts on the sickness absence of men of poor health, who probably are most vulnerable to health-related labour market exit.

Despite the fact that the reduction in workload studied is modest and affects only part of the work duties of a teacher (those directly related to classroom teaching), we identify a strong effect among men. Nonetheless, it is noteworthy that men given less workload appear to have improved their health in the sense that consultations due to mental health problems are less frequent, with fewer days of certified sick leave and sick notes issued less often.

Overall, our results suggest that preventive measures taken to retain older workers can indeed have a favourable effect on their sickness absence and health. A cost–benefit analysis of this initiative is beyond the scope of this paper. Still, if we consider only two of the elements relevant for such an analysis, a back-of-the-envelope calculation would be that the value of less sickness absence among men (a reduction of 21% in the use of locums, for one third of the teacher sample) more than outweighs the extra wage costs that the workload reduction generates (an increase of 5.8% for the whole sample). However, there is a need for caution as preventive measures such as the workload reduction studied make older workers costlier and thus may yield them a competitive disadvantage in the labour market. Moreover, the large heterogeneity in the response to workload reduction at age 55 calls into question whether work life regulations should be targeted to those who benefit the most instead of towards all employees within the age group. However, such differentiation is difficult to implement.

ACKNOWLEDGEMENTS

We are grateful to the Editor and two anonymous referees for their valuable comments and suggestions. The usual disclaimer applies. We gratefully acknowledge financial support from the Norwegian Research Council (Grant No. 237808), as well as comments from participants at the European Conference on Population Economics 2017, Glasgow, and the National Conference on Social Security Research 2016. Many thanks to May–Britt Heimsæter, Union of Education Norway, and Kjersti Myklebust, the Norwegian Association of Local and Regional Authorities, for valuable information on workload regulations.

CONFLICT OF INTEREST

None.

ORCID

Espen Bratberg https://orcid.org/0000-0003-0973-1064
Karin Monstad https://orcid.org/0000-0002-9025-7031

REFERENCES

Arai, M., & Thoursie, P. S. (2005). Incentives and selection in cyclical absenteeism. *Labour Economics*, *12*(2), 269–280.

Bratberg, E., & Monstad, K. (2015). Worried sick? Worker responses to a financial shock. *Labour Economics*, *33*, 111–120.

Case, A., & Deaton, A. S. (2005). Broken down by work and sex: How our health declines. In *Analyses in the Economics of Aging* (pp. 185–212). University of Chicago Press.

Cottini, E., & Lucifora, C. (2013). Mental health and working conditions in Europe. *ILR Review*, *66*(4), 958–988.

Cygan-Rehm, K., & Wunder, C. (2018). Do working hours affect health? Evidence from statutory workweek regulations in Germany. *Labour Economics*, *53*, 162–171.

Defebvre, É. (2018). Harder, better, faster ... Yet stronger? Working conditions and self-declaration of chronic diseases. *Health economics*, *27*(3), e59–e76. https://doi.org/10.1002/hec.3619

Fehr, E., & Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, *14*(3), 159–181.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, *80*(2), 223–255.

Hesselius, P., Nilsson, J. P., & Johansson, P. (2009). Sick of your colleagues 'absence? *Journal of the European Economic Association*, *7*(2-3), 583–594.

Ichino, A., & 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.

Johansson, P., & Palme, M. (2005). Moral hazard and sickness insurance. *Journal of Public Economics*, *89*(9), 1879–1890.

Kivimäki, M., Head, J., Ferrie, J. E., Shipley, M. J., Vahtera, J., & Marmot, M. G. (2003). Sickness absence as a global measure of health: Evidence from mortality in the Whitehall II prospective cohort study. *BMJ*, *327*(7411), 364–369.

Markussen, S., Røed, K., Røgeberg, O. J., & Gaure, S. (2011). The anatomy of absenteeism. *Journal of Health Economics*, *30*(2), 277–292.

Muurinen, J. M., & Le Grand, J. (1985). The economic analysis of inequalities in health. *Social Science & Medicine*, *20*(10), 1029–1035. https://doi.org/10.1016/0277-9536(85)90259-x

Norwegian Association of Local and Regional Authorities (2006). "Tariffoppgjøret 2006—SFS 2213, Undervisningspersonalet i kommunal og fylkeskommunal grunnopplæring", B-rundskriv 05/2006. [circular B-05/2006. In Norwegian only].

OECD (2006). *Live longer, work longer*. Paris: OECD Publishing.

OECD (2013). *Ageing and employment policies: Norway 2013: Working better with age*. Paris: OECD Publishing. https://doi.org/10.1787/9789264201484-en

Ose, S. O. (2005). Working conditions, compensation and absenteeism. *Journal of Health Economics*, *24*(1), 161–188.

Sjaastad, J., Carlsten, T. C., Wollscheid, S., Reiling, R. B., & Federici, R. A. (2016). [Kartlegging av lærere uten formelle kvalifikasjoner. In Norwegian] "Unqualified teachers in Norwegian schools" (our translation). NIFU report no. 2016:15.

Trevisan, E., & Zantomio, F. (2016). The impact of acute health shocks on the labour supply of older workers: Evidence from sixteen European countries. *Labour Economics*, *43*, 171–185.

Union of Education Norway (2012). Lærerårsverket i grunnskolen [A teacher's man-labour year. In Norwegian only.]. https://www.utdanningsforbundet.no.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Bratberg E, Holmås TH, Monstad K. Health effects of reduced workload for older employees. *Health Economics*. 2020;29:554–566. https://doi.org/10.1002/hec.4002