WORKING PAPER 2018:11

Forward-looking moral hazard in social insurance:

Evidence from a natural experiment

Marcus Eliason
Per Johansson
Martin Nilsson

IFAU INSTITUTET FÖR ARBETSMARKNADS- OCH UTBILDNINGSPOLITISK UTVÄRDERING

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala.

IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

More information about IFAU and the institute's publications can be found on the website www.ifau.se

ISSN 1651-1166

Forward-looking moral hazard in social insurance: evidence from a natural experiment[a]

by

Marcus Eliason[b], Per Johansson[c] and Martin Nilsson[d]

5th September, 2018

Abstract

This study tests for forward-looking moral hazard in the social insurance system by exploiting a 1991 reform in Sweden. The replacement rate was reduced for short absences but not for long absences, which introduced a potential future cost of returning to work. Using this exogenous variation in the replacement rate and controlling for dynamic selection, we find that the potential future cost of returning to work decreased the outflow from absence by 10 percent. This finding suggests that long-term sickness absentees are forward-looking, and highlights the importance of taking forward-looking behavior into account when designing and evaluating social insurance programs.

Keywords: disability insurance, dynamic incentives, forward-looking behavior, moral hazard, natural experiment, sickness absence, sickness insurance
JEL-codes: H55, I12, I13, J22

[a]The authors thank Karl-Oskar Lindgren and seminar participants at Duke University, IFAU, and Uppsala University for helpful comments.
[b]IFAU – Institute for Evaluation of Labour Market and Education Policy; marcus.eliason@ifau.uu.se
[c]IFAU – Institute for Evaluation of Labour Market and Education Policy; Department of Statistics, Uppsala University; IZA – Institute of Labour Economics; per.johansson@statistik.uu.se
[d]IFAU – Institute for Evaluation of Labour Market and Education Policy; Department of Economic, Uppsala University

Table of contents

1 Introduction

Sickness and disability insurance increase the utility for the working population through risk-sharing, which allows for intertemporal consumption smoothing. However, as sickness and disability in many cases are difficult to verify, the policymaker faces the problem of balancing the benefits associated with a generous insurance program with the disincentive, or moral hazard, effects of the same program. In the literature, there is strong empirical evidence of rather substantial effects of economic incentives on unemployment and short-term sickness absence in the unemployment insurance (UI) and sickness insurance (SI) programs. However, studies on how economic incentives affect absence behavior among long-term sickness absentees and disabled individuals in the sickness insurance (SI), disability insurance (DI), and worker compensation insurance (WCI) programs are scarcer and less coherent. One potential and obvious explanation of the weaker evidence of behavioral responses to economic incentives among long-term sickness absentees and disabled individuals is that these individuals have a reduced work capacity that leaves little room for such adjustments. Another potential explanation is that there are dynamic incentives that differ across various programs. For example, costs associated with entering the insurance program – e.g., waiting periods, low short-term replacement rates, lengthy or complicated application/screening processes – will provide not only economic incentives to remain in work but also disincentives to return to work among those who have already entered the program.[1] A rational and forward-looking individual, who is not liquidity constrained, would take into consideration not only the direct gains from leaving the program but also these potential future costs of having to re-enter the same program. If these costs are high enough relative to the insurance level, a forward-looking individual might choose to stay in the program beyond what is actually necessary. Such "forward-looking moral hazard"[2] is more likely to be important in social insurance programs where the eligibility criteria is difficult to verify, and where the disincentives to return to work

[1] In the appendix, we provide a simple model that illustrates the absence behavior of a forward-looking individual in a sickness insurance program with such a cost of entering, or reentering, the program.

[2] The phrase "forward-looking moral hazard" was first coined in Autor et al. (2014), but originating from "forward-looking behavior in moral hazard" in Aron-Dine et al. (2012).

are not counteracted by a replacement rate that is either diminishing with duration or time limited. This situation often prevails in the SI, DI, and WCI programs (cf., Krueger & Meyer 2002), as opposed to the UI program where eligibility is quite clear, there are time limits for benefit receipt, and the replacement rate most often is decreasing with the duration spent in the program. Hence, dynamic incentives and forward-looking moral hazard might be more important aspects in the SI, DI, and WCI programs than in the UI program.

To test emprically for forward-looking moral hazard (i.e., that absence behavior respond to dynamic incentives) in social insurance, we employ a novel identification strategy that exploits a 1991 reform of the Swedish SI program. For long-term sickness absentees the reform introduced dynamic disincentives to (return to) work: The replacement rate for short- and mid-term absences was reduced from 90 percent of foregone earnings to 65 and 80 percent, respectively. Meanwhile, the replacement rate for longer absences remained unchanged at 90 percent of foregone earnings. Hence, the return to work was now associated with a potential future cost of re-entering the insurance program. That is, long-term absentees received 90 percent of foregone earnings while remaining absent from work, but after having returned to work they received only 65 percent in case of a relapse that required a new absence. Forward-looking moral hazard implies that long-term sickness absentees (on average) would respond to the potential future cost of returning to work by prolonging their current absence.

This study is related to two different fields of research. First, it is related to the literature on the responsiveness to economic incentives among long-term sickness absentees and disabled individuals in the SI, WCI, and DI programs. The studies on DI and WCI programs are mostly from the U.S. and Canada, and the findings are somewhat inconclusive. Several studies find that increased benefit levels are associated with reduced labour supply (e.g., Curington 1994, Meyer et al. 1995, Gruber 2000, Neuhauser & Raphael 2004), but there are also studies that have found no effects or only modest effects (e.g., Campolieti 2004, Chen & van der Klaauw 2008).[3]

The studies on SI programs are mostly from European countries, but have mainly

[3]Notable exceptions to the North American studies are two Norwegian studies that found significant return-to-work effects induced by financial incentives in the temporary disability insurance (TDI) and DI programs (Kostøl & Mogstad 2014, Fevang et al. 2017).

focused on short-term absence. These studies consistently show that short-term sickness absence respond to economic incentives as predicted by economic theory (e.g., Johansson & Palme 2002, 2005, Puhani & Sonderhof 2010, Ziebarth & Karlsson 2010, Pettersson-Lidbom & Thoursie 2013, De Paola et al. 2014, Ziebarth & Karlsson 2014, Aaviksoo & Kiivet 2016).

Research on the importance of economic incentives for long-term sickness absence is scarcer and less conclusive. Only two studies have investigated how long-term absence responds to changes in the replacement rate for long-term absence; these two studies report rather different findings. On the one hand, in Norway, Markussen et al. (2011) found a "dramatic" increase in return-to-work when approaching the one-year limit in the SI program, when absentees are transferred from a sickness benefit with a replacement rate of 100 percent to a rehabilitation benefit with a replacement rate of 66 percent. On the other hand, in Germany, Ziebarth (2013) found that reducing the replacement rate from 80 percent to 70 percent for long-term absence affected neither the incidence nor the duration of long-term sickness absence.

A few studies have investigated how long-term absence responds to a change in the waiting periods or the replacement rate for short-term absences. The findings from these studies are also mixed: Aaviksoo & Kiivet (2016) reported that in Estonia, the reduction of the replacement rate from 80 percent to 70 percent, together with an extension of the waiting period from one to three days, had negligible impact on longer-term sickness absence. Although the main focus of Ziebarth & Karlsson (2010, 2014) is on the effects of changes of statutory short-term sick pay on short-term sickness absence (less than 7 weeks), their supplementary analyses contain estimates on the incidence of long-term absence. They report that neither the reduction of the replacement rate for short-term sickness absence from 100 percent to 80 percent in 1996 nor the restoration to 100 percent in 1999 had any impact on incidence of long-term sickness absence. However, De Paola et al. (2014) found that in Italy, the reduction in sick pay (from 100 percent to 80–90 percent) for short-term absences, together with stricter monitoring, increased the duration of longer (more than 10 days) absences. Moreover, Pollak (2017) found that workers who were compensated by supplementary sick pay during the three-day waiting period in France had shorter absence periods, and Pettersson-Lidbom & Thoursie (2013) found that the abolishment of a waiting period of one day, and an increase in the benefit levels for sickness absences shorter than 14 days, increased the outflow from longer absences. Most closely related to the present study is Johansson & Palme (2005) that investigated the impact on both short- and long-term absence of the same 1991 reform.[4] They report that following the reduction in the replacement rate for shorter-term absences, long-term absence increased.

The observation that a change in the direct cost of short-term absence (e.g., a change in the replacement rate for short term absence) are associated with a change in long-term absence, reported in Johansson & Palme (2005), but also in Pettersson-Lidbom & Thoursie (2013), De Paola et al. (2014), and Pollak (2017), suggests that long-term sickness absentees are indeed forward-looking. However, changing the direct costs of short-term absence would also affect the composition of the population of long-term absentees through dynamic selection (i.e., the population of long-term absentees would be more selected on health), which would produce patterns of long-term sickness absence that are similar to those produced by forward-looking moral hazard. Thus, the identification of the forward-looking effect of a change in the potential future costs of returning to work is exceedingly difficult, because in most situations policies would simultaneously affect the direct cost of either short- or long-term absence. None of the previous studies cited above, including Johansson & Palme (2005), have separated the effect of a change in the potential future cost of returning to work from the effect of a change in the direct cost of absence and from the compositional effect through dynamic selection. The novelty in this study is that we exploit that for long-term sickness absentees the 1991 reform (i) introduced a potential future cost of returning to work without affecting the direct cost of (long-term) absence, and (ii) did not apply to ongoing absences. The former, i.e., (i), ensures that our estimate of the effect of the potential future cost of returning to work is not contaminated by an effect of a change in the direct cost of (long-term) absence, and the latter, i.e., (ii), allows us two avoid any compositional effects from dynamic selection. Hence, we can (arguably)

[4] Johansson & Palme (2005) was based on a sample containing only 1,396 blue-collar workers, while the data in the present study comprise the full population.

estimate the causal effect of the potential future cost of returning to work.

Second, the present study is also related to a growing literature on the empirical testing for forward-looking behavior. This field mainly concerns consumer demand in various contexts, such as the demand for college textbooks (Chevalier & Goolsbee 2009), cigarettes and alcohol (e.g., Gruber & Köszegi 2001, Tiezzi 2005, Pierani & Tiezzi 2011), and medical care or drugs (e.g., Long et al. 1998, Dalton et al. 2015, Abaluck et al. 2015, Einav et al. 2015, Aron-Dine et al. 2015, Alpert 2016, Kaplan & Zhang 2017). The studies on the consumption of alcohol and cigarettes generally suggest that consumers are forward-looking (e.g., Gruber & Köszegi 2001, Tiezzi 2005, Pierani & Tiezzi 2011) rather than myopic as hypothesized by some models of addiction.[5] Chevalier & Goolsbee (2009) also found that students are forward-looking in their demand for college textbooks. However, the studies on the utilization of medical care, mostly exploiting the dynamic pricing incentives in Medicare Part D, report conflicting results. Some studies have found evidence of forward-looking behavior in medical utilization (e.g., Einav et al. 2015, Aron-Dine et al. 2015, Alpert 2016, Kaplan & Zhang 2017), while others report results that are supportive of myopia (e.g., Long et al. 1998, Dalton et al. 2015, Abaluck et al. 2015). Closely related to this study is Autor et al. (2014) that investigated dynamic incentives in a private long-term DI program in the U.S. They report that workers seem to account for the expected duration of their disability in their decision on whether to seek benefits for impairments, suggesting that disabled workers are forward-looking. However, the results in Autor et al. (2014) might alternatively be explained by binding liquidity constraints.[6] Fortuitously, binding liquidity constraints are not relevant in this study.

To sum up: (i) there is strong causal evidence that unemployment and short-term sickness absence respond to (static) economic incentives; (ii) there is no conclusive evidence that absence behavior among long-term sickness absentees or disabled individuals in the SI, DI, and WCI programs respond to (static) economic incentives; (iii) there is

[5] Recent applications of the myopic model of addiction include Hidayat & Thabrany (2010), Miljkovic & Nganje (2008), and Yakovlev (2018).

[6] While Autor et al. (2014) state that their findings are "most consistent" with forward-looking moral hazard, they acknowledge that they cannot reject binding liquidity constraints as an alternative explanation. However, their "(imperfect) test" of binding liquidity constraints suggested that binding liquidity constraints were not crucial in explaining their findings.

no causal evidence that absence behavior responds to dynamic economic incentives in public social insurance; and (iv) there is no conclusive evidence on whether individuals are forward-looking or myopic, although a majority of studies support forward-looking behavior. Hence, our contribution is threefold: First, we provide the first causal evidence of forward-looking moral hazard in public social insurance. Second, we add to the sparse literature on how absence behavior among long-term sickness absentees or disabled individuals in social insurance programs respond to economic incentives in general. Third, we add to the literature on the empirical testing for forward-looking behavior. We find that the potential future cost of returning to work, which was introduced by the reform through the reduction of the replacement rate for shorter-term absence, causally decreased the transition back to work by 10 percent among long-term sickness absentees. This finding suggests that long-term sickness absentees not only respond to economic incentives but they are indeed forward-looking. Placebo and sensitivity analyses support our claim of a causal interpretation. Heterogeneity analyses also suggest that individuals who were likely to have a higher perceived risk of relapse were more likely to prolong their current absence.

The rest of the paper is structured as follows. In the next section, we describe the Swedish SI program, the 1991 reform, and the macroeconomic environment at the time. In Section 3, we explain our empirical strategy and our choice of reform cohort (i.e., study group) and comparison cohort (i.e., control group). In Section 4, we describe our administrative register data covering the entire Swedish working-age population. We present the results, including placebo, sensitivity, and subgroup analyses, in Section 5. Section 6 concludes.

2 Background: the sickness insurance program, the 1991 reform, and the macroeconomic environment

2.1 The Swedish sickness insurance program

All workers (employed and unemployed) are covered by the public SI program administrated by the Swedish Social Insurance Agency (SIA). During the first seven days of sickness absence, the individual is decisive of whether being sick and to what extent it warrants absence from work, and merely has to inform the employer (or the SIA if unemployed). As of the eighth day, a medical certificate issued by a general practitioner (GP) is required, which states the length and extent of sickness absence that is (expected to be) necessary. This certificate must be renewed on a monthly basis. However, although the GPs are supposed to function as gatekeepers, in practice this gatekeeping function has been difficult to enforce. Svärdsudd (2000) report that among consultations that included a consideration about sickness absence, a medical certificate was not issued in only 6 percent of cases, and in 87 percent of cases a certificate was issued even when the GP found sickness absence to be either not advisable or even harmful. Hence, statistics suggest that even after the first seven days, the individual was the one primary decisionmaker in the length of sickness absence

Both the replacement rate and the employer's responsibility for sickness benefits have changed on several occasions during the last few decades. During the time period studied here, there was neither a time limit for benefit recept, a qualifying period, nor a period of employer-provided sick pay. Both before and after the reform, the insurance replaced a part of foregone earnings up to the social security ceiling of 7.5 price-base amounts per year (see *Figure 1* in Section 2.2). In 1991, this price ceiling amounted to SEK 241,500 (appr. USD 26,200) per year and only about 8 percent of the labor force had labor earnings above that (Lidwall 2012).[7]

In addition to the compulsory sickness insurance program, a large part of the Swedish labor market was, and still is, covered by negotiated complementary sickness insurance programs regulated in agreements between labor unions and employers' confederations.[8] In general, these complementary insurances replaced about 10 percent of forgone earnings.

[7] Those above the price ceiling could, potentially, be used to construct the counterfactual case since they were unaffected by the reform. However, because of the social gradient in health, the fraction above the price ceiling is much lower among long-term sickness absentees and therefore too small to be useful in the empirical analyses.

[8] It is not possible to identify in the data who were, and who were not, covered by these complementary sickness insurance programs.

2.2 The reform

The reform, which took effect on March 1, 1991, was the most important budget cut proposed by the Swedish government in early 1991 to halt an accelerating budget deficit. The reform implied that the insurance scheme changed from a flat replacement rate of 90 percent of foregone earnings, to a scheme with a replacement rate that was stepwise increasing with the duration of absence. More precisely, the replacement rate was reduced to 65 percent for short-term absences (i.e., 1–3 days), and to 80 percent for medium-term absences (i.e., 4–90 days); meanwhile it remained unchanged at 90 percent for long-term absences (i.e., longer than 90 days).[9] The scheme applied to all new absences, but not to ongoing absences.

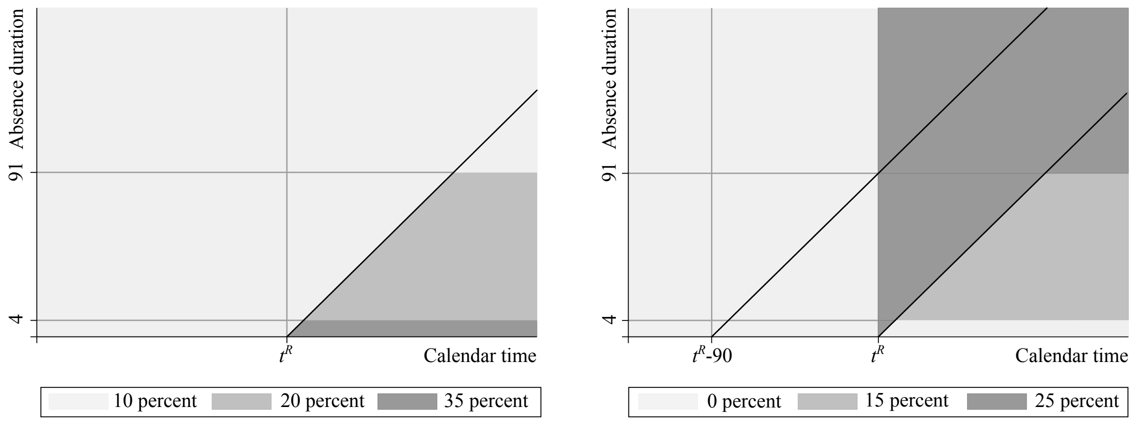
As noted in Section 2.1, part of the Swedish labor market is covered by negotiated complementary sickness insurance programs that in general replaced about 10 percent of forgone earnings. Since the reform aimed to affect all groups in the labour market equally, it included an additional reduction equal to the part of these complementary insurance programs exceeding 10 percent for short- and medium-term absences, and equal to the full amount for long-term absences. Hence, for those covered by complementary sickness insurance programs, the total replacement rate was in most cases reduced from a flat replacement rate of 100 percent of foregone earnings, to 75 percent for short-term absences, and to 90 percent for medium- and long-term absences.

The Lexis diagrams in *Figure 1* depict how the direct cost of absence (left) and potential future cost of returning to work (right) varied by calendar time (t^R denotes the date of the reform) on the x-axis and by absence duration on the y-axis. We define the direct cost of absence as the percentage share of foregone earnings not replaced by insurance, and the potential future cost of returning to work as the change in direct cost if a new absence period occurs after the individual returns to work.[10]

[9]This design of the new SI scheme, with a replacement rate increasing with the duration of absence, was motivated by a desire to cut public spending without adversely affecting a disadvantaged group.

[10]For ease of presentation, we ignore the possibility that the total potential future cost of returning to work cost also depends on the duration of the new absence.

Figure 1: Lexis diagram illustrating the direct cost of absence (left) and potential future cost of returning to work (right) by cohort, calendar time, and absence duration



Notes: The Lexis diagram on the left depicts the direct cost (i.e., the percentage share of earnings not replaced by insurance), and the Lexis diagram on the right depicts the potential future cost of returning to work (i.e., the change in the percentage share of earnings not replaced by insurance, if the individual returns to work, and then starts a new absence period), for cohorts whose absences started before or after t^R (i.e., the day of the reform) and by absence duration.

From the diagram on the left, it is clear that the reform changed the direct cost of absence from 10 percent to 35 percent for short-term absences, and to 20 percent for medium-term absences; meanwhile the direct cost of absence remained at 10 percent for long-term absences. From the diagram on the right, we observe that first, prior to t^R there was no potential future cost of returning to work, since the direct cost of absence was 10 percent regardless of the duration of absence. Second, a change in the direct cost of absence for short- and medium-term absences did not imply any potential future cost of returning to work for short-term absences,[11] but implied a potential future cost of returning to work of 15 percent for medium-term absences,[12] and of 25 percent for long-term absences.[13] Hence, for long-term sickness absentees, the reform did not affect the direct cost of absence (which remained at 10 percent), but introduced a potential future cost of returning to work (of 25 percent). Third, even though the reform did not apply to ongoing absences, it nevertheless introduced a potential future cost of returning to work

[11] The direct cost of absence would have remained at 35 percent if the individual starts a new absence after having returned to work.

[12] The direct cost of absence would have increased from 20 percent to 35 percent if the individual starts a new absence after having returned to work.

[13] The direct cost of absence would have increased from 10 percent to 35 percent if the individual starts a new absence after having returned to work.

of 25 percent for these ongoing absences. In the empirical analyses, this feature allowed us to avoid compositional effects from dynamic selection, which will be further explained in Section 3.[14]

2.3 The macroeconomic environment

The 1991 reform of the SI program occurred in a rather exceptional macroeconomic context, which warrants some discussion. *Figure 2* depicts the changing macroeconomic environment surrounding the time period under study (represented by the gray areas in the graphs). The last years of the 1980s were characterized by an exceptionally low unemployment rate (*Figure 2a*). Unemployment had been falling for several years, and in 1989, it fell to a low of 2.2 percent, while employment rose continually to a peak of 83.4 percent in 1990 (*Figure 2b*). By the end of the 1980s, these two measures indicated a more buoyant labour market than ever seen before. At the beginning of the 1990s, however, Sweden experienced a macroeconomic shock unparalleled since the Great Depression. By 1994, unemployment had risen to 10.6 percent, and total employment had fallen by 11 percentage since 1990.[15]

The unemployment and employment rates portray a decade of dramatically changing macroeconomic conditions, raising the question of how these conditions affected sickness and disability absence. In Sweden, sickness absence is strongly pro-cyclical, and in the late 1980s the total annual number of insured absence days had been increasing for several years, peaking at 107 million days, or 20 days per person, in 1988 (*Figure 2c*).[16] This is the context in which we should assess the reform; because the cost of the SI program increased substantially during the late 1980s, reforms were deemed necessary to halt an accelerating budget deficit. In the years that followed, the total annual number of insured absence days decreased gradually and then plummeted abruptly in 1992. The sharp fall

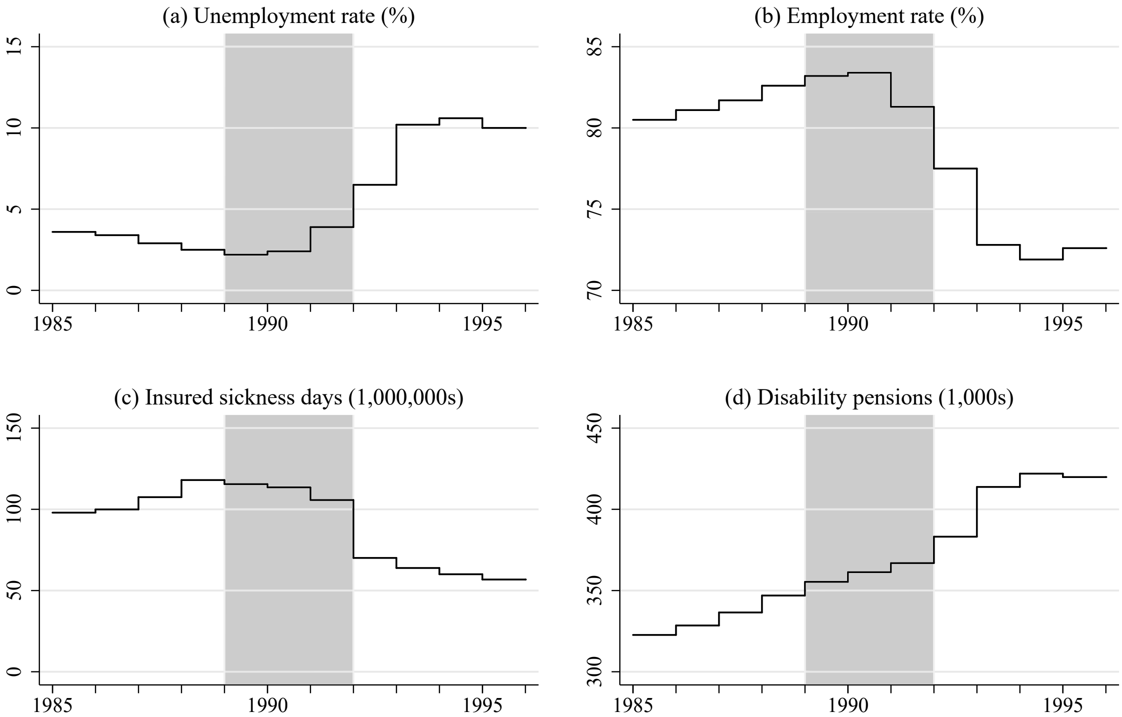
[14] Moreover, for all pre-reform cohorts the potential future cost of returning to work were also the same regardless of whether they were covered by complementary insurance, which implies that our analyses are not comprised even though we are unable to identify from the data who were, and who were not, covered by complementary insurance.

[15] The rising unemployment rate warranted cuts also of the replacement rate in the UI program. Because these reductions in the replacement rate were not implemented until July 1993, and the absences in our analyses either ended before, or were censored by, the end of 1991, they should not have affected any incentive effects caused by interactions between the UI and SI programs.

[16] The working age population was about 5.3 million.

did not, however, reflect a fall in actual absence. Instead, it signaled the introduction of employer-provided sick pay in January 1992; employers were now responsible for the compensation for the first 14 days (i.e., the first two weeks of absence were no longer registered as insured sickness absence).

Figure 2: The annual unemployment rate (a), the employment rate (b), the number of insured sickness absence days (c), and the number of disability pensions (d) during 1985–1995



Source: Statistics Sweden and the Social Insurance Agency.
Notes: The gray areas in the graphs mark the period under study.

For those with a reduced work capacity of at least 25 percent there was also a possibility to leave the labour force for early retirement through the DI program.[17] Until October 1991, those who were at least 60 years old could be granted DI for labor market reasons. After October 1991, a medical reason was required for all workers regardless of age. Between 1985 and 1992, the number of people receiving DI increased linearly from 323,000 to 383,000 (*Figure 2d*). In 1993, the trough of the recession, the number of granted disability

[17] The replacement rate in the DI program was lower than in the SI program. Hence, there were no economic incentives to be transferred from the SI program to the DI program. The disability benefit amounted to approximately 65 percent of previous earnings. Negotiated complementary disability insurance programs covering major parts of the labor market increased the total disability benefits to approximately 80 percent of previous earnings.

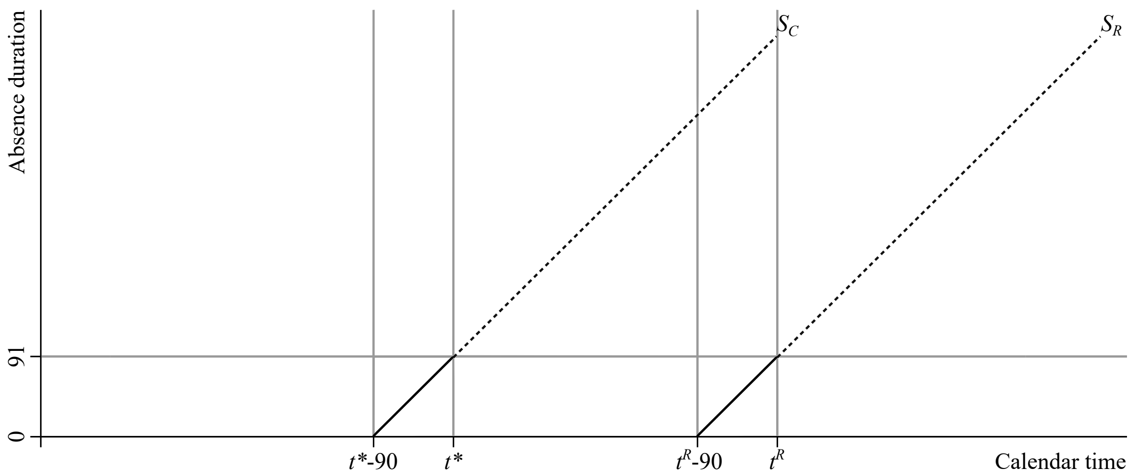
pensions increased by 30,000 to 413,000, and then further to 420,000–422,000 during 1994–1995.

Hence, we can conclude that the 1991 reform occurred during a period characterized by dramatically changing macroeconomic conditions. Importantly, most of the rise and fall of the four measures in *Figure 2* occurred either before or after the period under study. In Section 3, we will return to the issue of how and to what extent the macroeconomic context might have affected our analyses.

3 Empirical strategy

In this section, we will outline our empirical strategy for testing for forward-looking moral hazard among long-term sickness absentees. Our objective is to isolate the effect of the potential future cost of returning to work, which was introduced by the reform, on absence behavior. Using *Figure 3* as point of departure, we will explain how reform and comparison cohorts were chosen.

Figure 3: Illustration of potential reform and comparison cohorts



Notes: The figure illustrates a potential reform containing absences S_R beginning on day $t^R - 90$ and lasting at least until t^R (when the reform took effect) and similarly a potential comparison cohort containing absences S_C beginning on day $t^* - 90$ and lasting at least until t^*.

We recall from Section 2.2 that individuals whose absences began after t^R were exposed to both a change in the direct cost of absence and a potential future cost of returning to work even before becoming long-term absentees (i.e., reaching 91 days), which would create a dynamic selection problem, where the composition of long-term absentees under the new regime is expected to differ from that of long-term absentees under the old regime. Because the reform only applied to new absences, individuals who began their absence before t^R were not exposed to any change in the direct cost of absence. However, from t^R onward, these individuals were nonetheless exposed to the potential future cost of returning to work. Hence, the composition of long-term sickness absentees, who had not reached 91 days of absence at t^R, is also expected to have been affected by the reform. These compositional effects, due to dynamic selection, rule out all absences that began after $t^R - 90$ from the reform cohort.

Hence, let us for now define the reform cohort as those individuals whose absence began at $t^R - 90$ and who remained absent at least until $t^R - 1$ (i.e., such as absence S_R in *Figure 3*).[18] This cohort of long-term absentees was exposed to the potential future cost of returning to work from t^R onward; however, before t^R, they were not exposed to either a change in the direct cost of absence or any potential future cost of returning to work (i.e., there were no compositional effect of the reform). Similarly, let us define the comparison cohort as those individuals whose absence began at $t^* - 90$ and who remained absent at least until $t^* - 1$ (i.e., such as absence S_C in *Figure 3*), where t^* is an arbitrary date such that $t^* < t^R$. This comparison cohort represents the counterfactual case of not being exposed to any potential future costs of returning to work. A remaining issue, however, is how to chose t^*. The choice of t^* also determines the outcome window: with t^* close to t^R, it would be possible to estimate only short-run effects, since the absence periods in the comparison cohort must be censored before t^R. However, given this restriction on the outcome window any $t^* < t^R$ could be chosen, provided that the following two assumptions are valid: (i) there were no anticipation effects (i.e., the reform was not anticipated or at least people did not respond to it), and (ii) there were no calendar time effects (e.g., no seasonal variation in absence) that differently affected the outflow from absence within the reform and comparison cohorts or the selection into these two cohorts.

[18] Note that individuals who began their absence before $t^R - 90$ and who remained absent at least until $t^R - 1$ could potentially also be included in the reform cohort.

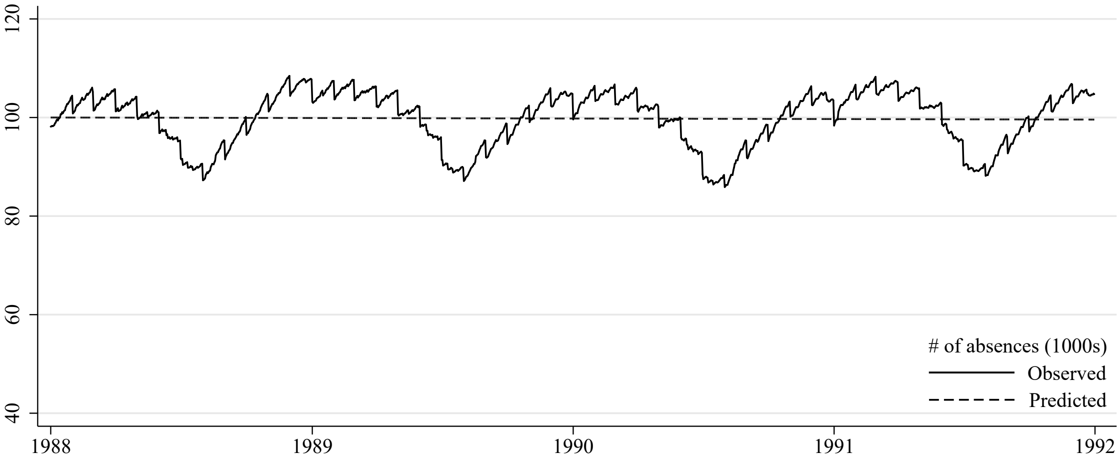
Because our focus is on forward-looking behavior, a highly relevant issue is whether there were any anticipatory responses to the reform. Although there were incentives to begin a new absence period before rather than after t^R, long-term sickness absentees (or anyone with an ongoing absence) had no incentives to respond in any way to the forthcoming reform. Hence, since we did not include any absences that began later than 90 days prior to the reform, these incentives should not have affected our sample or our analyses.

Although anticipation effects do not impose a threat to our identification strategy, a calendar time effect that might bias our analysis is a remaining concern. For at least two reasons one would expect such effects. First, as discussed in Section 2.3, the reform occured in the shift from a booming economy to a deep recession. A stylized fact is that (especially short-term) sickness absence is procyclical. The two dominating explanations to this procyclicality are that it is due to (i) a changing composition of the labor force over the business cycle, and (ii) a disciplining effect from the fear of job loss during recessions. Hence, given the worsening business cycle conditions during our study period (see *Figure 2*), procyclical sickness absence would imply that we underestimate the behavioral response to the potential future cost of returning to work that was introduced by the reform.

From *Figure 4* – that depicts the daily stock of long-term sickness absentees, with no more than 730 days of continuous absence,[19] and its corresponding linear trend – it is evident that despite the shifting macroeconomic environment there was no trend over time in long-term sickness absence. A second reason to expect calendar effects is seasonal variation in sickness absence. In *Figure 4*, a repeating monthly and yearly pattern in long-term sickness absence is clearly apparent: long-term sickness absence is most common during the winter and least common during the summer, and there are marked drops at the turn of the months. Taken together, these patterns suggest that we do not have to handle across, but only within, year variations in long-term sickness absence. The obvious (and perhaps only sensible) choice of t^* would, therefore, be the same day and month as t^R, but in the preceding year (i.e., March 1, 1990).

[19]The upper limit is a consequence of the lack of data prior to 1986, but it also corresponds well to the maximal observed duration of absence in the analyses.

Figure 4: The stock of long-term sickness absentees (in thousands), with no more than two years of continuous absence, during 1988–1992



Notes: The figure depicts the observed and predicted number of long-term sickness absentees, with no more than two years of continuous absence. The predicted values are obtained from a linear regression of the number of long-term sickness absentees on calendar time.

To gain statistical precision in our analyses, we expanded the time window defining our reform and comparison cohorts, from containing only those who reached exactly 90 days of absence at $t^* - 1$ or $t^R - 1$ (i.e., February 28, 1990 and 1991, respectively), to include all earlier cohorts that year, provided that the absences were still ongoing at $t^R - 1$ or $t^* - 1$, respectively. In the Lexis diagram in *Figure 5*, we illustrate the time windows defining both the reform and comparison cohorts used in the analyses, and the outcome periods when absence duration was measured (depicted by the gray areas). The reform cohort comprises all long-term sickness absentees who began their absence between January 1 and December 1, 1990. Similarly, the comparison cohort comprises all long-term sickness absentees who began their absence between January 1 and December 1, 1989.[20] [21] All absences were censored by the end of the year (i.e., by December 30, 1991 for the reform cohort and by December 30, 1990 for the comparison cohort).

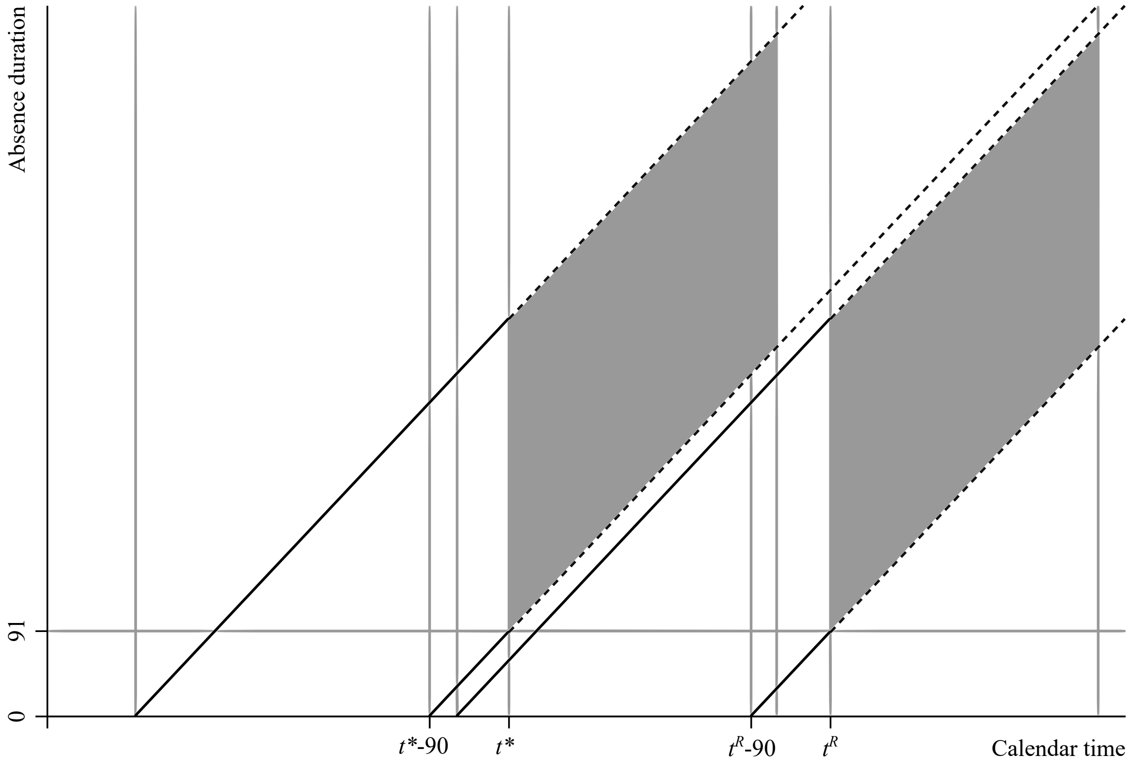
Ideally, these two cohorts would be identical in terms of all characteristics affecting the timing of the return to work. Then, the effect of the potential future cost of returning to work introduced by the reform could be estimated simply by taking the difference between the post-February 28 survival rates for the reform and comparison cohorts. A

[20] An alternative, but equivalent, way to express this is that the reform (comparison) cohort contains all individuals whose elapsed absence duration was 90–425 days on $t^R - 1$ ($t^* - 1$).

[21] For a placebo analysis presented in Section 5.3 we similarly sampled an additional cohort (1988).

positive difference would suggest that there exists forward-looking moral hazard in social insurance. In Section 4.2, we assess the extent to which the "assignment" to the two groups is as good as random (in terms of observable characteristics).

Figure 5: Illustration of the reform and comparison cohorts and the associated outcome periods



Notes: The Lexis diagram illustrates how we have defined the reform and comparison cohorts and their respective outcome periods (depicted by the gray areas). t^R denotes the day of the reform, i.e., March 1, 1991, and t^* denotes March 1, 1990. The solid diagonal lines depict the absence periods with the earliest and latest possible starting date to be included in the respective cohort.

4 Data

4.1 The data sources

The data that we used originate from Swedish administrative registers with universal coverage. Linking across the registers is possible because of the 10-digit personal identity number that is unique to each Swedish resident. Specifically, three registers/databases were used to create the data set: First, to identify all periods of long-term sickness absence we used the Sickness Benefit Register. This register is administered by the Swedish

Social Insurance Agency (SIA) and contains information on sickness insurance payments for each individual. Most importantly (for this study), from 1986 it contains both the start and end dates for every insured absence period.[22] Second, background characteristics were drawn from Statistics Sweden's longitudinal database LOUISE. This database contains comprehensive annual information from 1990, drawn from a number of administrative registers, for the nationally registered population aged 16–64 years. The aim of LOUISE is to enhance the conditions for research on sickness insurance and labor market issues requiring longitudinal individual data. However, the database does not cover any years before 1990, while our comparison cohort is sampled in 1989 (and the additional cohort used in the placebo analysis is sampled in 1988). Therefore, we had to draw information also from the more limited Employment Register that has data from 1985. The background information contained in both the Employment Register and LOUISE is limited to age, sex, immigration status, attained education level, county of residence, and annual earnings. From the Sickness Benefit Register we could obtain information also on sickness absence during the three preceding years.

4.2 Descriptive statistics

In *Table 1*, we present descriptive statistics for the reform and comparison cohorts. The two cohorts are quite similar: in both cohorts, the average age was 46 years, 57 percent were women, and 17 percent were foreign born. However, during the preceding three years there were also some minor, but nonetheless statistically significant, differences: average annual earnings were SEK 130,500 (appr. USD 14,100) in the reform cohort and SEK 128,000 (appr. USD 13,900) in the comparison cohort; average annual number of sickness absence days were 61.4 and 59.7, respectively; and average annual number of absence periods were 3.3 and 3.4, respectively.

[22] Until the end of 1991 the Sickness Benefit Register covered all absence periods. From 1992 onwards, only absence periods longer than 14 days were recorded because of the introduction of a two-week period of employer-provided sick pay at the beginning of each absence.

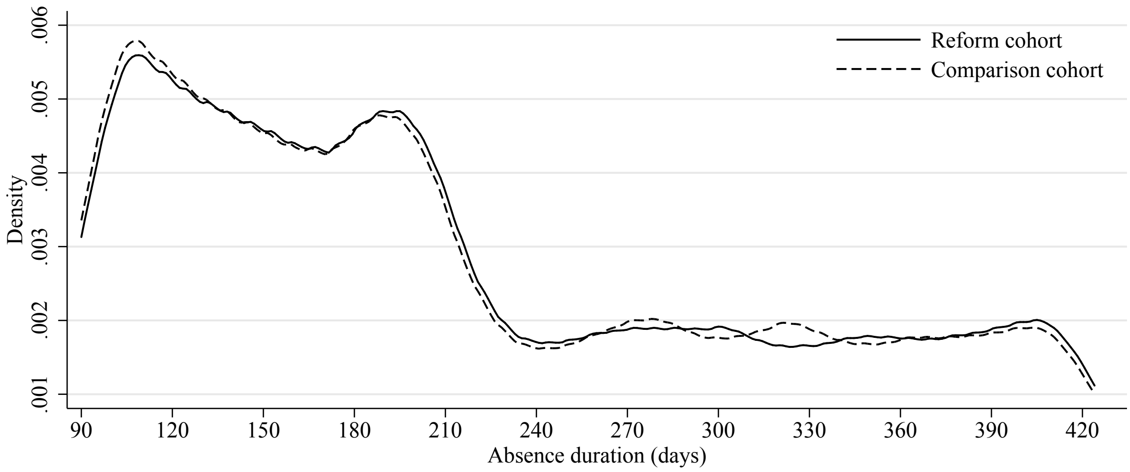
Table 1: Descriptive statistics

Covariates	Reform cohort Mean	(Std.dev.)	Comparison cohort Mean	(Std.dev.)	p-value[a]
Demographics					
Age	46.403	(12.177)	46.345	(12.310)	0.365
Female	0.571	(0.495)	0.573	(0.495)	0.435
Foreign born	0.165	(0.371)	0.167	(0.373)	0.233
Attained education					
Compulsory schooling	0.409	(0.492)	0.434	(0.496)	0.000
Upper secondary schooling	0.406	(0.491)	0.391	(0.488)	0.000
College/university	0.117	(0.322)	0.106	(0.308)	0.000
Unknown	0.069	(0.253)	0.069	(0.254)	0.591
Previous earnings/sickness[b]					
Earnings[c]	130.496	(83.621)	128.044	(81.727)	0.000
No. of absence days	61.405	(63.563)	59.733	(62.004)	0.000
No. of absence periods	3.254	(2.535)	3.441	(2.350)	0.000
No. of long absence periods	0.174	(0.268)	0.168	(0.263)	0.000
County of residence					
Stockholm	0.191	(0.393)	0.185	(0.389)	0.005
Uppsala	0.032	(0.176)	0.029	(0.167)	0.001
Södermanland	0.034	(0.182)	0.032	(0.177)	0.031
Östergötland	0.045	(0.208)	0.045	(0.208)	0.982
Jönköping	0.033	(0.178)	0.032	(0.175)	0.164
Kronoberg	0.018	(0.132)	0.018	(0.133)	0.786
Kalmar	0.026	(0.160)	0.026	(0.158)	0.384
Gotland	0.006	(0.079)	0.006	(0.077)	0.403
Blekinge	0.018	(0.133)	0.019	(0.138)	0.085
Kristianstad	0.029	(0.169)	0.031	(0.174)	0.018
Malmöhus	0.085	(0.279)	0.091	(0.288)	0.000
Halland	0.025	(0.156)	0.025	(0.156)	0.901
Göteborg	0.102	(0.303)	0.101	(0.302)	0.759
Älvsborg	0.044	(0.204)	0.047	(0.211)	0.009
Skaraborg	0.025	(0.156)	0.026	(0.160)	0.113
Värmland	0.034	(0.181)	0.033	(0.178)	0.250
Örebro	0.030	(0.170)	0.028	(0.165)	0.044
Västmanland	0.028	(0.166)	0.026	(0.159)	0.007
Dalarna	0.032	(0.176)	0.032	(0.177)	0.653
Gävleborg	0.041	(0.199)	0.042	(0.200)	0.482
Västernorrland	0.031	(0.173)	0.032	(0.176)	0.171
Jämtland	0.019	(0.135)	0.018	(0.134)	0.749
Västerbotten	0.039	(0.193)	0.037	(0.189)	0.108
Norrbotten	0.033	(0.178)	0.037	(0.189)	0.000
No. of observations		74 364		73 259	

[a] p-value from a t-test of equal means across the two cohorts.
[b] Annual averages over the three calendar years preceding the ongoing absence period.
[c] Reported in thousands of CPI-adjusted (2015 average values) Swedish kronor (SEK).

The elapsed duration of the ongoing absence period has been suppressed in *Table 1*. Because of its importance, we have instead plotted its full distribution in *Figure 6*. The densities for the two cohorts are very similar and the majority of the absences were between 90 and 210 days on February 28. All in all, we conclude that the two cohorts are similar, but not identical, with respect to observable characteristics. These differences might need to be adjusted for and we return to this issue in Section 5.1.

Figure 6: Kernel density estimates of the elapsed absence durations on February 28 for the reform and comparison cohorts, respectively



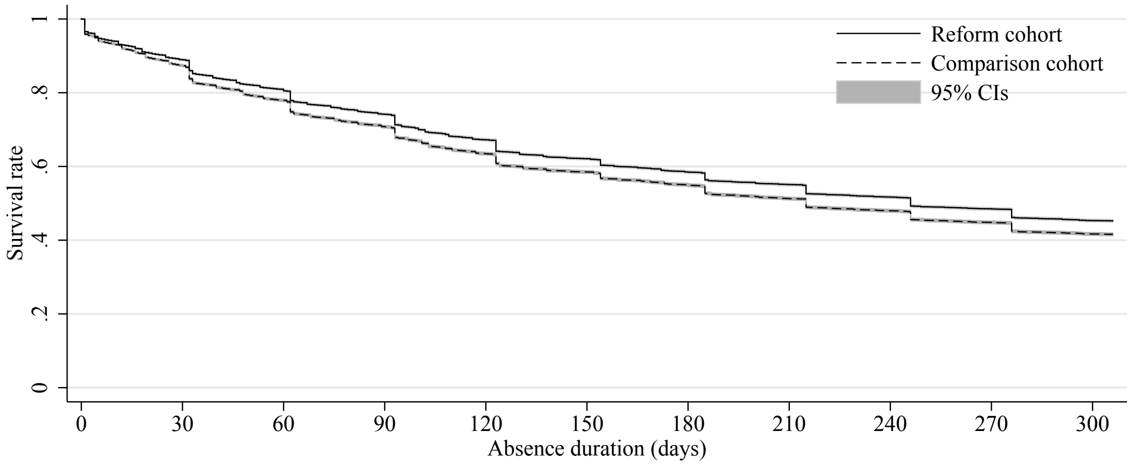
5 Results

In this section, we first present our main estimates of the behavioral response to the potential future cost of returning to work, which was introduced by the reform, among long-term sickness absentees. We then present the results from repeating the analysis for various subgroups (i.e., by sex and sickness absence history). In the following subsection, we present three additional analyses to assess the extent to which our previously presented estimates indeed have a causal interpretation.

5.1 Main results

If long-term sickness absentees are forward-looking, the potential future cost of returning to work, which was introduced by the reform, is expected to have prolonged the post-February 28 absence duration for the reform cohort relative to the comparison cohort. In *Figure 7*, such a prolonged absence duration is evident from the observed gap between the plotted survival functions of the two cohorts.[23] The gap increases to 3.7 percentage points during the first 105 days following February 28 and then remains at approximately this level. Hence, it seems that the increased potential future cost of returning to work made long-term sickness absentees more reluctant to return to work ("too" early).

Figure 7: Main results: Post-February 28 survival functions, with 95 percent confidence intervals (95% CIs), for the reform and comparison cohorts



Notes: The figure depicts the post-February 28 Kaplan-Meier survival curves, with 95% CIs, for the reform and comparison cohorts of long-term sickness absentees. The difference between the two curves can be interpreted as the behavioral response to the potential future cost of returning to work that was introduced by the reform.

To ascertain that the differences between the two cohorts (in terms of some observable characteristics, as observed in *Table 1*) have not affected our results, we also estimated a Cox Proportional Hazards Model (Cox PHM) including all characteristics in *Table 1* and *Figure 6*:

$$h(ReformCohort_i, \mathbf{x}_i, t) = h_0(t)\exp(\delta \cdot ReformCohort_i + \varphi'\mathbf{x}_i), \tag{1}$$

where $h(\cdot,t)$ is the hazard function at time t, $h_0(t)$ is the baseline hazard function at time t, δ is the parameter of interest, associated with the indicator of belonging to the reform cohort (i.e., *ReformCohort*), and φ is a vector of parameters associated with the

[23]The step-wise shape of the survival functions is due to a greater number of absence periods ending on the last day of each calendar month.

vector $\mathbf{x}_i$ of observed characteristics of individual i.[24] In *Table 2*, we report the estimates of δ expressed as hazard ratios (HRs) with 95 percent confidence intervals (95% CIs). Both the unadjusted and adjusted estimates are practically identical, and suggest that the potential future cost of returning to work that was introduced by the reform decreased the transition back to work by 10 percent.

Table 2: Main results: Unadjusted and adjusted estimates of the effect of the potential future cost of returning to work, expressed as hazard ratios (HRs) with 95 percent confidence intervals (95% CIs)

	Unadjusted		Adjusted		
	HR	95% CI	HR	95% CI	*N*
Reform cohort	0.898	(0.886–0.910)	0.897	(0.885–0.910)	147 623

Notes: The estimated hazard ratios are obtained from unadjusted and adjusted Cox PHMs of post-February 28 absence duration using the sample with the reform and comparison cohorts of long-term sickness absentees. The adjusted model includes age, female, foreign born, attained education, previous earnings, county of residence, the elapsed pre-Feb 28 duration of the ongoing absence period, and annual averages of absence days, absence periods, and long absence periods during the three preceding calendar years (net of the ongoing absence).

5.2 Subgroup analyses

In this subsection, we present the results from two subgroup analyses. For brevity, we only report the (unadjusted and adjusted) hazard ratio estimates from the Cox PHM. These estimates can be compared to those for the full sample reported in *Table 2*.

First, we have repeated the analysis, separately, for men and women, as absence behavior is known to differ by gender. On the one hand, women are more likely to be absent from work for health reasons than men (e.g., Paringer 1983, Mastekaasa & Olsen 1998, Broström et al. 2004, Angelov et al. 2013). On the other hand, men have been found to react more strongly to reductions in replacement rates (e.g., Johansson & Palme 1996, Henrekson & Persson 2004, Ziebarth & Karlsson 2014). The results from this analysis are reported in *Table 3*. The sample of women is considerably larger than the sample of men, which is in line with the fact that women, on average, are more likely to be absent from work for health reasons than men. However, the estimates actually suggest that the behavioral response to the potential future cost of returning to work was somewhat larger

[24] We have estimated both (covariate) adjusted models (i.e., including X_i) and unadjusted models (i.e., excluding X_i).

among women than among men, contrary to what could be expected based on previous studies showing that men react more strongly to changes in replacement rates.[25]

Table 3: A sub-group analysis by sex and previous sickness: Unadjusted and adjusted estimates of the effect of the potential future cost of returning to work, expressed as hazard ratios (HRs) with 95 percent confidence intervals (95% CIs)

	Unadjusted		Adjusted		
Sample	HR	95% CI	HR	95% CI	*N*
Sex					
Men	0.915	(0.896–0.934)	0.913	(0.894–0.932)	63,137
Women	0.886	(0.870–0.902)	0.888	(0.873–0.904)	84,486
Absence days[a]					
0–42/3	0.935	(0.910–0.960)	0.946	(0.920–0.972)	37,358
43/3–116/3	0.878	(0.854–0.902)	0.880	(0.856–0.904)	36,273
117/3–258/3	0.895	(0.871–0.920)	0.892	(0.868–0.917)	37,040
259/3–	0.887	(0.863–0.911)	0.890	(0.866–0.915)	36,952
Absence periods[a]					
0–4/3	0.926	(0.903–0.950)	0.922	(0.898–0.946)	44,271
5/3–8/3	0.915	(0.887–0.944)	0.901	(0.874–0.930)	28,952
9/3–13/3	0.899	(0.877–0.922)	0.898	(0.875–0.921)	41,914
14/3–	0.861	(0.837–0.886)	0.863	(0.839–0.888)	32,486
Long absence periods[a]					
0	0.907	(0.891–0.922)	0.904	(0.889–0.920)	94,798
1/3	0.867	(0.843–0.892)	0.866	(0.842–0.891)	34,946
2/3–	0.919	(0.883–0.956)	0.927	(0.891–0.965)	17,879

Notes: The estimated hazard ratios are obtained from unadjusted and adjusted Cox PHMs of post-February 28 absence duration, estimated separately for each subgroup, using the sample with the reform and comparison cohorts of long-term sickness absentees. The adjusted model includes age, female, foreign born, attained education, previous earnings, county of residence, the elapsed pre-Feb 28 duration of the ongoing absence period, and annual averages of absence days, absence periods, and long absence periods during the three preceding calendar years (net of the ongoing absence).
[a] Annual averages over the three preceding calendar years (net of the ongoing absence).

Second, we have also repeated the analysis for subgroups based on their sickness absence histories. Because the potential future cost of returning to work was realized only if an individual was starting a new absence period, the long-term sickness absentees' history of sickness absence may have affected the response to the reform. As the history of sickness absence has been found to predict future sickness absence (e.g., Koopmans et al. 2008, Roelen et al. 2011, Laaksonen et al. 2013), individuals with longer or more frequent

[25] To investigate whether this difference between men and women is statistically significant we have also estimated a pooled model with both sexes and an interaction between the indicator for the reform cohort and the indicator for being a woman. In the unadjusted and adjusted models, this interaction is statistically significant at the 5 and 10 percent level, respectively.

absences in the past are likely to have a more negative assessment of the risk of relapse. Hence, we should expect to find a larger effect among long-term sickness absentees with a history of sickness absence.

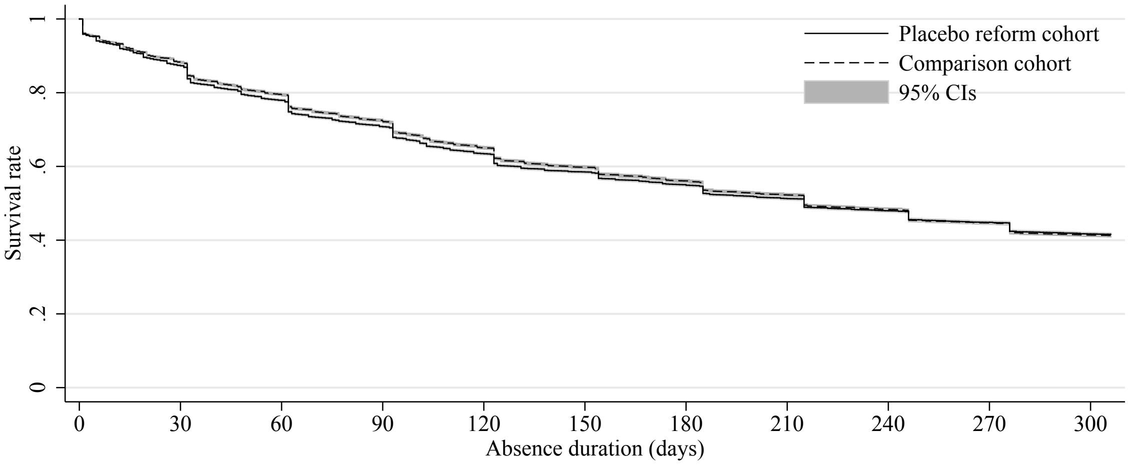
The estimates from this analysis, where the sample has been divided in three or four groups based on previous number of absence days, absence periods, and long absence periods, respectively, are also presented in *Table 3*. The general conclusion from these results is that the estimated effect of the potential future cost of returning to work that was introduced by the reform is, as expected, larger among individuals with a history of sickness. Both the number of previous days of sickness absence and the number of previous absence periods (short and long) seem to have consistently affected the behavioral response. However, the number of previous long absence periods do not appear to have affected the behavioral response. For individuals with the fewest previous absence periods (i.e., at most four periods during the three-year period), the hazard ratio is 0.93, while for individuals with the most previous absence periods (i.e., at least 14 periods during the three-year period), the hazard ratio is 0.86. Similarly, for those with the fewest previous absence days (i.e., at most 42 days during the three-year period) the hazard ratio is 0.94, while for those with the most previous absence days (i.e., at least 259 days during the three-year period) the hazard ratio is 0.89.[26]

5.3 Placebo and sensitivity analyses

As a test of our identification strategy, we have performed a placebo analysis, where the reform date was artificially changed to March 1, 1990 (i.e., one year before the actual reform). The cohort that hitherto had been used as the comparison cohort in the analyses then became the (placebo) reform cohort, and the preceding cohort became the new comparison cohort. Using these two cohorts, we repeated the analyses of Section 5.1. A statistically significant placebo effect would cast serious doubts on whether the previously reported estimates represent causal effects of the potential future cost of returning to work that was introduced by the reform.

[26] Based on pooled models with interactions between the indicator for the reform cohort and the sickness absence categories, these differences are statistically significant at the 5 percent level.

Figure 8: Placebo analysis: Post-February 28 survival functions, with 95 percent confidence intervals (95% CIs), for the placebo reform and comparison cohorts



Notes: The figure depicts the post-February 28 survival curves, with 95% CIs, for the placebo reform cohort (1989) and the new comparison cohort (1988) of long-term sickness absentees.

The survival curves for the two cohorts are quite close (*Figure 8*), and both the unadjusted and adjusted Cox PHM yield a precisely estimated null effect (*Table 4*). These results support our claim that the estimation strategy provides causal estimates of the effects of the potential future cost of returning to work.

Table 4: Placebo analysis: Unadjusted and adjusted placebo estimates of the effect of the potential future cost of returning to work, expressed as hazard ratios (HRs) with 95 precent confidence intervals (95% CIs)

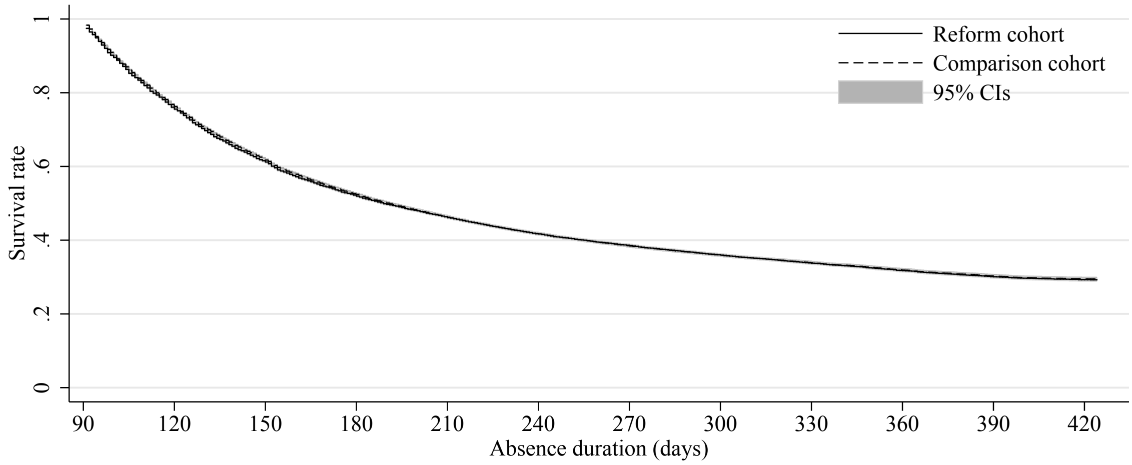
	Unadjusted		Adjusted		
	HR	95% CI	HR	95% CI	*N*
Placebo reform cohort	1.009	(0.996–1.023)	1.004	(0.991–1.018)	148 452

Notes: The estimated hazard ratios are obtained from unadjusted and adjusted Cox PHMs of post-February 28 absence duration using a placebo reform cohort (1989) and a comparison cohort (1988) of long-term sickness absentees. The adjusted model includes age, female, foreign born, attained education, previous earnings, county of residence, the elapsed pre-Feb 28 duration of the ongoing absence period, and annual averages of absence days, absence periods, and long absence periods during the three preceding calendar years (net of the ongoing absence).

In Section 5.1, we showed that controlling for the differences in observable characteristics did not affect the estimates. Although this finding is encouraging, it does not exclude the possibility that there might be unobserved factors that affect our estimates. However, if there are such unobserved factors, there is little reason to believe that they would affect absence duration after, but not before, t^R and t^*. Hence, we can compare the pre-

February 28 survival rates for the full reform and comparison cohorts (i.e., not only those with absences that lasted until February 28) of long-term sickness absentees. The more alike the pre-February 28 survival rates are, the less likely that there are differences in unobservable factors that may have affected absence duration.

Figure 9: Pre-February 28 analysis: Survival functions, with 95 percent confidence intervals (95% CIs), for the full reform and comparison cohorts



Notes: The figure depicts the pre-March 1 survival curves, with 95% CIs, for the full reform and comparison cohorts of long-term sickness absentees.

The survival curves for the two cohorts are not distinguishable by the eye (*Figure 9*). While the small unadjusted estimate from the Cox PHM is in fact statistically significant, the adjusted estimate is once again a precisely estimated null effect (*Table 5*). Hence, this second analysis also supports our claim that our estimation strategy provides causal estimates of the potential future cost of returning to work.

Table 5: Pre-February 28 analysis: Unadjusted and adjusted hazard ratios (HRs) with 95 percent confidence intervals (95% CIs)

	Unadjusted		Adjusted		
	HR	95% CI	HR	95% CI	*N*
Reform cohort	1.012	(1.002–1.022)	1.004	(0.994–1.014)	307,793

Notes: The estimated hazard ratios are obtained from unadjusted and adjusted Cox PHMs of pre-February 28 absence duration using the full reform and comparison cohorts of long-term sickness absentees. The adjusted model includes age, female, foreign born, attained education, previous earnings, county of residence, and annual averages of absence days, absence periods, and long absence periods during the three preceding calendar years (net of the ongoing absence).

In Section 3, we discussed how to define the reform and comparison cohorts. We started out with the extreme case of defining the two cohorts as only those who reached exactly 90 days of absence by February 28 (i.e., the long-term sickness absentees who began their absence either at $t^R - 90$ or at $t^* - 90$). To gain statistical precision, however, we expanded the sampling window to include all long-term absentees whose absence started earlier that same year and were still ongoing by February 28. To investigate the robustness of our results to the choice of the width of the sampling window, we reestimated the Cox PH models and stepwise increased the window widths by a (calendar) month at the time (see *Table 6*): starting with December (which includes only the absences that began on December 1), then including also November, and so forth until reaching the full window used in all previous analyses.

Table 6: Sensitivity analysis: Unadjusted and adjusted placebo estimates of the effect of the potential future cost of returning to work, expressed as hazard ratios (HRs) with 95 precent confidence intervals (95% CIs), using sampling windows of increasing width

	Unadjusted		Adjusted		
Sampling window	HR	95% CI	HR	95% CI	*N*
December	0.812	(0.694–0.951)	0.861	(0.731–1.013)	1,001
November–December	0.874	(0.849–0.899)	0.888	(0.862–0.914)	26,921
October–December	0.881	(0.862–0.901)	0.893	(0.874–0.913)	50,020
September–December	0.888	(0.871–0.905)	0.895	(0.878–0.912)	68,274
August–December	0.887	(0.872–0.902)	0.891	(0.876–0.906)	89,553
July–December	0.885	(0.871–0.900)	0.890	(0.876–0.905)	98,859
June–December	0.891	(0.877–0.905)	0.892	(0.878–0.906)	106,688
May–December	0.893	(0.879–0.906)	0.894	(0.881–0.908)	115,503
April–December	0.896	(0.883–0.909)	0.893	(0.880–0.906)	123,378
March–December	0.897	(0.885–0.910)	0.894	(0.882–0.907)	131,241
February–December	0.900	(0.887–0.912)	0.897	(0.884–0.909)	138,625
January–December	0.898	(0.886–0.910)	0.897	(0.885–0.910)	147,623

Notes: The estimated hazard ratios are obtained from unadjusted and adjusted Cox PHMs of the effect of the potential future cost of returning to work among long-term sickness absentees, using the reform and comparison cohorts of long-term sickness absentees and sampling windows of increasing width. The adjusted model includes age, female, foreign born, attained education, previous earnings, county of residence, the elapsed pre-Feb 28 duration of the ongoing absence period, and annual averages of absence days, absence periods, and long absence periods during the three preceding calendar years (net of the ongoing absence).

Our estimates do not seem to be sensitive to the choice of window width (if anything, the estimated effect of the potential future cost of returning to work is decreasing with the width of the window). All (adjusted) hazard ratios, except when applying the most narrow window including only December 1, are found within 0.988–0.897. For the one-day window of December 1, both the estimated unadjusted and adjusted effects of the potential future cost of returning to work are actually somewhat larger (HR: 0.812 and HR: 0.861, respectively), but because of the much smaller sample size, only the former is statistically significant.

6 Conclusions

In this study we tested for the presence of forward-looking moral hazard, i.e., a behavioral response to dynamic incentives, in social insurance. Costs associated with entering a social insurance program – e.g., waiting periods, lower short-term replacement rates, and lengthy or complicated application/screening processes – do not only provide (static) economic incentives to (remain in) work but also dynamic disincentives to (return to) work among those who have already entered the program. A rational and forward-looking individual, who is not liquidity constrained, would take into consideration not only the direct gains from leaving the program but also the potential future costs of returning to work. If these costs are high enough relative to the benefit level, a forward-looking individual might respond by remaining in the program for longer than necessary. Such forward-looking moral hazard is more likely to be important in social insurance programs where the eligibility criteria is difficult to verify, and where the disincentives to return to work are not counteracted by incentives to return to work such as a replacement rate that diminish with the duration in the program and/or a time limit for benefit receipt. Hence, different dynamic incentives across social insurance programs might potentially explain the less conclusive evidence on long-term sickness absentees and disabled individuals' behavioral responses to economic incentives in the SI, DI, and WCI programs, compared to the strong evidence in support of unemployment and short-term sickness responses to economic incentives in the UI and SI programs. However, the extent to which long-term sickness absentees and disabled individuals are forward-looking and respond to dynamic incentives in the social insurance programs is an open empirical question.

To empirically test this hypothesis, we exploited a 1991 reform in the Swedish SI program as a natural experiment. Before the reform there was a flat replacement rate, corresponding to 90 percent of foregone earnings. By reducing the replacement rate to 65 percent for short-term absences, and to 80 percent for medium-term absences, while leaving it unchanged at 90 percent for longer absences, the reform introduced a potential future cost of returning to work for long-term sickness absentees. That is, individuals who returned to work but subsequently suffered a relapse requiring a new absence period would now receive only 65 percent of foregone earnings. Forward-looking moral hazard would in this context imply that long-term sickness absentees (on average) would respond to the potential future cost of returning to work by prolonging their current absence (until their perceived risk of relapse was low enough). An empirical challenge, however, is that increased direct costs (i.e., reduced replacement rate) of short-term absence is likely to also affect the composition of the population of long-term absentees through dynamic selection (i.e., the population of long-term absentees would be more negatively selected in terms of health). Such dynamic selection would produce patterns of long-term sickness absence that are similar to those produced by forward-looking moral hazard. In this study, we exploited the fact that for long-term sickness absentees, the reform introduced a potential future cost of returning to work without affecting the direct cost of absence, and did not apply to ongoing absences. Hence, we could separate the impact of the potential future cost of returning to work from the impact of the direct cost of absence, and avoid compositional effects from dynamic selection.

We show that the potential future cost of returning to work, introduced by the reform through the reduced replacement rate for short- and medium-term absences, causally decreased the transition rate back to work by 10 percent among long-term sickness absentees. Placebo and sensitivity analyses support our claim of a causal interpretation. The results suggest that (i) long-term sickness absentees indeed take dynamic incentives into account, and that (ii) there is forward-looking moral hazard in social insurance.

Moreover, we do not find that men reacted to the potential future cost of returning to work more strongly than did women. This finding is not necessarily at odds with the stylized fact that men are more responsive to economic incentives than women. In contrast to previous studies on static economic incentives, the dynamic incentives investigated in this study involve both the discounting of the future and the assessment of the risk of relapse. If men discount the future more heavily, or make a more positive risk assessment, that would diminish the economic disincentives to return to work. Moreover, hitherto, we have not discussed whether the reform prolonged sickness absence through increased “shirking” or decreased “sickness presenteeism.” Because “shirkers” would have a lower risk of relapse relative to those who are still sick yet present at work, it is likely that most of the estimated forward-looking moral hazard effect is operating through decreased “sickness presenteeism.” Women might also be more willing to respond to economic incentives that imply preventive actions, such as reduced sickness presenteeism, rather than shirking.

Furthermore, we also found that those with more absence periods in the past seem to have responded more strongly to the reform. Since a potential future cost of returning to work is realized only if an individual starts a new absence period after having returned to work, this finding should be expected, given forward-looking behavior and that the perceived risk of relapse increases with the number and duration of past absence periods.

To conclude, we show not only that long-term sickness absentees respond to economic incentives but also that they do so in a forward-looking manner. We believe that these findings contribute to two different research fields. First, our (arguably) convincing causal estimates contribute to the sparse literature on the importance of economic incentives for long-term absence among sick and disabled individuals. Second, we contribute to a growing literature on the empirical testing for forward-looking behavior (in various contexts). From a policy perspective, the design and evaluation of social insurance programs require an understanding of whether people who are absent from work due to long-term sickness or disability take potential future costs into account, and how they might adjust their absence behavior. Finally, our findings suggest that in a social insurance program with imperfect screening, a cost of entering the program might create “locking-in” effects unless counteracted by a replacement rate that is either diminishing with duration or time limited.

References

Aaviksoo, E. & Kiivet, R. A. (2016), 'Influence of the sickness benefit reform on sickness absence', *Health Policy* **120**(9), 1070–1078.

Abaluck, J., Gruber, J. & Swanson, A. (2015), Prescription drug use under Medicare Part D: a linear model of nonlinear budget sets, Working Paper No. 20976, National Bureau of Economic Research.

Alpert, A. (2016), 'The anticipatory effects of Medicare Part D on drug utilization', *Journal of Health Economics* **49**, 28 – 45.

Angelov, N., Johansson, P. & Lindahl, E. (2013), Gender differences in sickness absence and the gender division of family responsibilities, Working Paper 2013:9, IFAU - Institute for Evaluation of Labour Market and Education Policy, Uppsala.

Aron-Dine, A., Einav, L., Finkelstein, A. & Cullen, M. (2015), 'Moral hazard in health insurance: do dynamic incentives matter?', *Review of Economics and Statistics* **97**(4), 725–741.

Aron-Dine, A., Einav, L., Finkelstein, A. & Cullen, M. R. (2012), Moral hazard in health insurance: How important is forward looking behavior?, Working Paper No. 17802, National Bureau of Economic Research.

Autor, D., Duggan, M. & Gruber, J. (2014), 'Moral hazard and claims deterrence in private disability insurance', *American Economic Journal: Applied Economics* **6**(4), 110–141.

Broström, G., Johansson, P. & Palme, M. (2004), 'Economic incentives and gender differences in work absence behavior', *Swedish Economic Policy Review* **11**, 33–63.

Campolieti, M. (2004), 'Disability insurance benefits and labor supply: Some additional evidence', *Journal of Labor Economics* **22**(4), 863–889.

Chen, S. & van der Klaauw, W. (2008), 'The work disincentive effects of the disability insurance program in the 1990s', *Journal of Econometrics* **142**(2), 757–784.

Chevalier, J. & Goolsbee, A. (2009), 'Are durable goods consumers forward-looking? evidence from college textbooks', *The Quarterly Journal of Economics* **124**(4), 1853–1884.

Curington, W. (1994), 'Compensation for permanent impairment and the duration of work absence: Evidence from four natural experiments', *Journal of Human Resources* **29**(3), 888–910.

Dalton, C. M., Gowrisankaran, G. & Town, R. (2015), Salience, myopia, and complex dynamic incentives: Evidence from Medicare Part D, Working Paper No. 21104, National Bureau of Economic Research.

De Paola, M., Scoppa, V. & 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.

Einav, L., Finkelstein, A. & Schrimpf, P. (2015), 'The response of drug expenditure to nonlinear contract design: evidence from Medicare Part D', *The Quarterly Journal of Economics* **130**(2), 841–899.

Fevang, E., Hardoy, I. & Røed, K. (2017), 'Temporary disability and economic incentives', *The Economic Journal* **126**(603), 1410–1432.

Gruber, J. (2000), 'Disability insurance benefits and labor supply', *Journal of Political Economy* **108**(6), 1162–1183.

Gruber, J. & Köszegi, B. (2001), 'Is addiction "rational"? theory and evidence', *The Quarterly Journal of Economics* **116**(4), 1261–1303.

Henrekson, M. & Persson, M. (2004), 'The effects on sick leave of changes in the sickness insurance system', *Journal of Labor Economics* **22**(1), 87–113.

Hidayat, B. & Thabrany, H. (2010), 'Cigarette smoking in indonesia: examination of a myopic model of addictive behaviour', *International Journal of Environmental Research and Public Health* **7**(6), 2473–2485.

Johansson, P. & Palme, M. (1996), ‘Do economic incentives affect work absence? empirical evidence using swedish micro data’, *Journal of Public Economics* **59**(2), 195–218.

Johansson, P. & Palme, M. (2002), ‘Assessing the effect of public policy on worker absenteeism’, *Journal of Human Resources* **37**(2), 381–409.

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

Kaplan, C. M. & Zhang, Y. (2017), ‘Anticipatory behavior in response to Medicare Part D’s coverage gap’, *Health Economics* **26**(3), 338–351.

Koopmans, P. C., Roelen, C. A. M. & Groothoff, J. W. (2008), ‘Risk of future sickness absence in frequent and long-term absentees’, *Occupational Medicine* **58**(4), 268–274.

Kostøl, A. R. & Mogstad, M. (2014), ‘How financial incentives induce disability insurance recipients to return to work’, *The American Economic Review* **104**(2), 624–655.

Krueger, A. B. & Meyer, B. D. (2002), ‘Labor supply effects of social insurance’, *Handbook of Public Economics* **4**, 2327–2392.

Laaksonen, M., He, L. & Pitkäniemi, J. (2013), ‘The durations of past sickness absences predict future absence episodes’, *Journal of Occupational and Environmental Medicine* **55**(1), 87–92.

Lidwall, U. (2012), Insured income in the social insurance 1991–2010, Social insurance report 2012:6, Social Insurance Agency, Stockholm.

Long, S. H., Marquis, M. & Rodgers, J. (1998), ‘Do people shift their use of health services over time to take advantage of insurance?’, *Journal of Health Economics* **17**(1), 105–115.

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

Mastekaasa, A. & Olsen, K. M. (1998), 'Gender, absenteeism, and job characteristics a fixed effects approach', *Work and Occupations* **25**(2), 195–228.

Meyer, B. D., Viscusi, W. K. & Durbin, D. L. (1995), 'Workers' compensation and injury duration: evidence from a natural experiment', *The American Economic Review* **85**(3), 322–340.

Miljkovic, D. & Nganje, W. (2008), 'Regional obesity determinants in the united states: A model of myopic addictive behavior in food consumption', *Agricultural Economics* **38**(3), 375–384.

Neuhauser, F. & Raphael, S. (2004), 'The effect of an increase in worker's compensation benefits on the duration and frequency of benefit receipt', *The Review of Economics and Statistics* **86**(1), 288–302.

Paringer, L. (1983), 'Women and absenteeism: health or economics?', *The American Economic Review* **73**(2), 123–127.

Pettersson-Lidbom, P. & Thoursie, P. S. (2013), 'Temporary disability insurance and labor supply: evidence from a natural experiment', *The Scandinavian Journal of Economics* **115**(2), 485–507.

Pierani, P. & Tiezzi, S. (2011), 'Infrequency of purchase, individual heterogeneity and rational addiction in single households' estimates of alcohol consumption', *Giornale degli Economisti e Annali di Economia* **70**(2), 93–116.

Pollak, C. (2017), 'The impact of a sick pay waiting period on sick leave patterns', *The European Journal of Health Economics* **18**(1), 13–31.

Puhani, P. A. & Sonderhof, K. (2010), 'The effects of a sick pay reform on absence and on health-related outcomes', *Journal of Health Economics* **29**(2), 285–302.

Roelen, C., Koopmans, P., Schreuder, J., Anema, J. & van der Beek, A. (2011), 'The history of registered sickness absence predicts future sickness absence.', *Occupational Medicine* **61**(2), 96–101.

Svärdsudd, Lars Englund, K. (2000), ‘Sick-listing habits among general practitioners in a swedish county’, *Scandinavian Journal of Primary Health Care* **18**(2), 81–86.

Tiezzi, S. (2005), ‘An empirical analysis of tobacco addiction in italy’, *The European Journal of Health Economics* **6**(3), 233–243.

Yakovlev, E. (2018), ‘Demand for alcohol consumption in russia and its implication for mortality’, *American Economic Journal: Applied Economics* **10**(1), 106–49.

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

Ziebarth, N. R. & Karlsson, M. (2010), ‘A natural experiment on sick pay cuts, sickness absence, and labor costs’, *Journal of Public Economics* **94**(11), 1108–1122.

Ziebarth, N. R. & Karlsson, M. (2014), ‘The effects of expanding the generosity of the statutory sickness insurance system’, *Journal of Applied Econometrics* **29**(2), 208–230.

Appendix: A simple model

In this section, we present a simple model to illustrate the absence behavior of a forward-looking – in comparison to a myopic – individual in a sickness insurance program where there is a cost of entering (and re-entering) the program. In the empirical analysis this cost is reflected by a lower replacement rate for short-term absence. However, the cost could also comprise waiting periods, lengthy or complicated application/screening processes, or the risk of having the application rejected. In the following, we will model this more general case, but treat it as a monetary cost.[27]

Assume that the individual's instantaneous utility (u_t) is a weighted sum of consumption (c_t) and leisure (l_t):

$$u_t = (1-\sigma_t)c_t + \sigma_t l_t \tag{2}$$

where σ_t denotes the sickness level (related to his/her work capacity), which for simplicity is assumed to be uniformly distributed over [0,1]. The larger σ_t, the sicker is the individual, and the greater the weight given to leisure (or recuperation time) and the less the weight given to consumption. A working individual receives a (fixed) net wage income w from h hours of work, which is used to fund consumption, and enjoys $T-h$ hours of leisure, where T is total time available. Hence, the instantaneous utility is $u_t = (1-\sigma_t)w + \sigma_t(T-h)$. An individual who is absent due to health problems instead receives sickness benefits equal to a fraction r of the net wage income w and enjoys T hours of leisure (or recuperation time), but also bears a fixed cost ϕ of entering the program. Hence, the instantaneous utility is $u_t = (1-\sigma_t)(rw-\phi) + \sigma_t T$. As the cost ϕ is only associated with entering the program, the instantaneous utility in a subsequent absence period is $u_t = (1-\sigma_t)rw + \sigma_t T$.

In the following, we derive testable predictions from a two-period model. Because our focus is on long-term sickness, we assume that the individual is absent prior to period 1 and then decides whether to remain absent or to return to work.[28] Using this model, we

[27] Our model follows Ziebarth (2013), but differs in that it includes a cost of entering the benefit state instead of a replacement rate that varies over periods.

[28] Hence, it is a three-period model, but in this setting the first period is not associated with any choices and therefore does not need to be modelled.

can determine the indifference sickness level σ_1^{a*} at which the individual is indifferent between returning to work and remaining absent. That is, the level of sickness that equalizes (i) the utility from returning to work in period 1 plus the discounted expected utility in period 2 (conditional on working in the previous period, denoted by superscript e), and (ii) the utility from remaining absent in period 1 plus the discounted expected utility in period 2 (conditional on being absent in the previous period, denoted by superscript a). More formally:

$$(1-\sigma_1^{a*})w+\sigma_1^{a*}(T-h)+\tfrac{1}{1+\rho}E[u_2^e]=(1-\sigma_1^{a*})rw+\sigma_1^{a*}T+\tfrac{1}{1+\rho}E[u_2^a], \tag{3}$$

where ρ denotes the individual's discount rate reflecting his or her time preferences. An individual who is absent in period 1 will remain absent in period 2 if $\sigma_2 \geq \sigma_2^{a*}$, where σ_2^{a*} is the indifference sickness level in period 2 conditional on being absent in period 1. An individual who instead is working in period 1 will become absent in period 2 if $\sigma_2 > \sigma_2^{e*}$, where σ_2^{e*} is the indifference sickness level in period 2 conditional on working in the previous period. Since σ_2 is uniformly distributed over $[0,1]$, we can define $E[u_2^e]$ and $E[u_2^a]$ as:[29]

$$E[u_2^e]=\sigma_2^{e*}\left[(1-\tfrac{\sigma_2^{w*}}{2})w+\tfrac{\sigma_2^{e*}}{2}(T-h)\right]+(1-\sigma_2^{e*})\left[(1-\tfrac{1+\sigma_2^{e*}}{2})(rw-\phi)+\tfrac{1+\sigma_2^{e*}}{2}T\right] \tag{4}$$

and

$$E[u_2^a]=\sigma_2^{a*}\left[(1-\tfrac{\sigma_2^{a*}}{2})w+\tfrac{\sigma_2^{a*}}{2}(T-h)\right]+(1-\sigma_2^{a*})\left[(1-\tfrac{1+\sigma_2^{a*}}{2})rw+\tfrac{1+\sigma_2^{a*}}{2}T\right] \tag{5}$$

The indifference sickness levels in period 2, can easily be derived as:

$$\sigma_2^{e*}=\frac{w-rw-\phi}{w-rw-\phi+h} \tag{6}$$

$$\sigma_2^{a*}=\frac{w-rw}{w-rw+h} \tag{7}$$

[29] Specifically, we use that $\Pr[\sigma_2 \leq \sigma_2^{v*}]=\sigma_2^{v*}$, $E[\sigma_2|\sigma_2 \leq \sigma_2^{v*}]=\sigma_2^{v*}/2$, and $E[\sigma_2|\sigma_2 > \sigma_2^{v*}]=(1+\sigma_2^{v*})/2$, where $v=a,e$.

By substituting equations (4)–(7) into (3) and solving for σ_1^{a*}, we obtain:

$$\sigma_1^{a*} = \sigma_2^{a*} + \tfrac{1}{(1+\rho)}\kappa, \tag{8}$$

where

$$\kappa = \frac{-\phi h^2}{2(w-rw-\phi+h)(w-rw+h)^2}. \tag{9}$$

Because σ_1 is uniformly distributed over $[0,1]$ the transition rate back to work in period 1, i.e., $E[\sigma_1 < \sigma_1^{a*}]$, is directly obtained from equation 8. The discounted term in equation 8 can be viewed as the impact, on the transition rate, of the potential future cost of reentering the program in period 2. To determine how a change in the cost of entering the program affects the transition rate for long-term sickness absentees, we take the partial derivative of equation 8 with respect to ϕ:

$$\frac{\partial \sigma_1^{a*}}{\partial \phi} = \frac{\partial \sigma_2^{a*}}{\partial \phi} + \frac{1}{1+\rho}\frac{\partial \kappa}{\partial \phi} = \frac{1}{1+\rho}\frac{-h^2}{2(w-rw-\phi+h)^2(w-rw+h)} < 0 \tag{10}$$

Equation 10 shows that – given that individuals are forward-looking – increased costs of entering the sickness insurance will decrease the transition rate back to work. For the myopic individual, σ_1^{a*} instead equals σ_2^{a*} and a change in ϕ will not affect the transition rate at all.

In our empirical analysis, we can then test whether long-term sickness absentees behave in a forward-looking or myopic manner. The cost ϕ then represents the difference between the benefit level for long- and short-term absences (i.e., $\phi = [r - r_{short}]w$). To make the theoretical model tractable, we have extracted from the more realistic situation where sickness in later periods are correlated with sickness in earlier periods, and where sickness can be affected by whether the individual is absent or working. These correlations would introduce dynamic selection into the model, and any direct empirical comparison of the populations in the respective regime would be biased. However, not only did the reform provide exogenous variation in ϕ, its unique features also allowed us to circumvent the bias associated with dynamic selection.