Out for Good: Labor Market Effects of Transitory and Persistent Health Shocks*

Mattis Beckmannshagen[†] Johannes König[‡]

August 7, 2023

Abstract

We introduce a novel method of deriving two health shock indicators in a general population survey based on a data-driven classification of sickness absences and hospitalizations. In light of different magnitudes and post-shock dynamics observed in event studies of employment, work hours, and income, the shock indicators are best described as transitory and persistent. Both types of health shocks are widespread with an annual incidence of about 1.7%, which rises steeply with age. Persistent health shocks reduce employment by 7.5 percentage points and imply a substantially reduced intention to return to work among younger workers.

JEL Code: I10 · J22 · H51
Keywords: labor supply · health shocks · event study · k-means clustering

---

* Mattis Beckmannshagen thanks the Hans-Böckler-Foundation for financial support. Johannes König thanks the German Research Foundation (DFG) for supporting his research through the project "Life-Cycle Inequality Dynamics" (LINDY); project number: 430271414. We thank Markus M. Grabka, Daniel Graeber, Jana Hamdan, Jonas Jessen, Robin Jessen, Martina Kaplanová, Maximilian Longmuir, Shushanik Margaryan, Lorenz Meister, Thomas Rieger, Carsten Schröder, Johannes Seebauer, Barbara Stacherl, and Yogam Tchokni for their comments and insights. We also thank the members of the LINDY project, and participants of the ESEM 2021, Applied Economics Conference 2021, SEHO 2022, EALE 2022, SOLE 2023, IAAE 2023, the Economics Seminar at Umeå University for helpful comments and discussions. We thank Annika Sperling for outstanding research assistance. We thank Deborah Anne Bowen for copy editing.

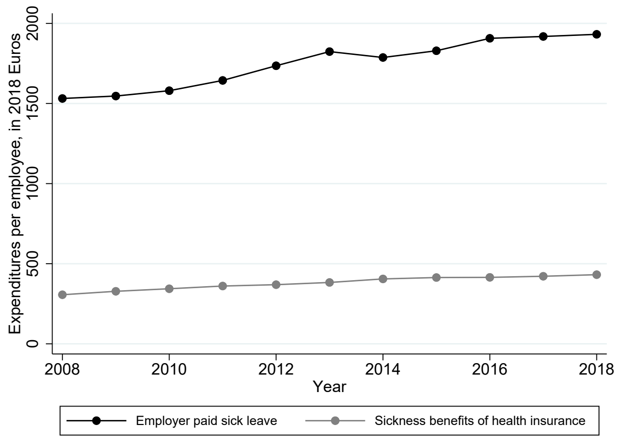
† SOEP at German Institute for Economic Research (DIW Berlin), Mohrenstr. 58, 10117 Berlin, Germany. E-mail: mbeckmannshagen@diw.de

‡ SOEP at German Institute for Economic Research (DIW Berlin) and Berlin School of Economics (BSE); Mohrenstr. 58, 10117 Berlin, Germany. E-mail: jkoenig@diw.de

1. Introduction

Health shocks are defined as sudden deteriorations of health caused by disease or accidents. They can affect individual well-being in multiple ways, not only reducing the quality of life but also limiting peoples capacities to act and participate in society. One of the capacities most severely impaired by health shocks is the capacity to work, that is, the ability to participate in the labor market. A reduced capacity to work can reduce lifetime earnings, since income losses are generally not fully insured. But reduced working capacity also has a possible indirect effect on earnings by preventing the accumulation of human capital. Accordingly, health shocks may not only have direct and immediate but also longer-lasting, indirect effects on individuals' labor market outcomes.

Health shocks affect a large percentage of the population. Almost one in three adults in developed countries suffers from two or more chronic illnesses during their lifetime, including hypertension, cancer, and diabetes (OECD, 2019). For adults between the ages of 50 and 59 who suffer from at least two such illnesses, the probability of being employed is more than 20 percentage points lower than for their healthy counterparts (OECD/European Union, 2016). Moreover, illness also has downstream effects beyond the direct effects on the labor market. When workers drop out of the labor market, their households experience a downward shift in the budget constraint. This loss can only be compensated by the employer, the state, or the family. The costs of compensation are high: In Germany, expenditures for sickness-related absences per employee, that is, wage continuation payments, have increased steadily over the past decades, as Figure 1 shows. From 2008 to 2018, real expenditures per employee rose from €1,800 to €2,400, an increase of 33%. In total, this amounts to employers spending more than €60 billion per year on wage continuation payments, while the public health insurance system spends €14 billion on sickness benefits alone (Federal Statistical Office, 2021a). Combined, the yearly costs for employers and the government are more than twice the yearly spending on unemployment insurance (roughly €80 billion vs. €35 billion[1] (Federal Statistical Office, 2021b)). While the immediate costs are high, the long-term costs are even higher as some workers leave the labor market for good and enter early retirement. As discussed in Buslei et al. (2019) and Engels et al. (2017), workers leaving the labor market put a significant strain on the welfare system, especially in a pay-as-you-go public pension system like Germany's.



Sources: Health expenditure accounts (Federal Statistical Office, 2021a) and Employment Statistics of the Federal Agency for Employment. Both publicly available at the Federal Statistical Office.

Figure 1. Real Annual Expenditures for Health-Related Absences per Employee: Employer and State

To assess the relevance of health for issues such as fiscal stress on the pension system, it is essential to consider the full range of health shocks in the population at risk. In this study, we quantify the effects of health shocks on employees in a representative survey of the German population (Socio-Economic Panel; Goebel et al. (2019); Schröder et al. (2020)). Within such a large population, health shocks naturally differ in severity across individuals. This inherent heterogeneity is crucial to address when estimating the effects of diverse health shocks.

[1] The estimated costs of sickness-related absences do not include expenditures for reduced earnings capacity pensions (*Erwerbsminderungsrente*).

Our three main contributions relate to the measurement of health shocks and the estimation of their causal effects on labor market outcomes in a representative, general population survey: First, we capture diminished working capacity based on sick days and nights hospitalized (or hospital nights in our shorthand).[2] Hence, our novel health shock indicators are directly connected to individuals' behavior and capture the heterogeneity in an individual's reaction to an adverse health event. We validate this by describing the indicators' relationship to previously employed measures of health such as health satisfaction and disease diagnoses, which we find to be strongly correlated (see Appendix A). Here, our choice of the SOEP data is vital, as these conventional health measures as well as labor market variables are all measured in the SOEP. A major benefit of our approach is that we can apply it to the entire population of interest, that is, individuals in dependent employment in Germany, which enables us to comprehensively assess the impact of ill health on employment and other labor market outcomes.[3] Second, we pursue a data-driven classification of shock heterogeneity by using k-means clustering with respect to the joint distribution of sick days and hospitalizations, focusing on the simplest and most parsimonious categorization into two shock indicators. This is necessary as we conduct our analysis on a large sample of the German workforce. We thoroughly stress-test our approach in a dedicated appendix and find that it stands up to scrutiny (see Appendix A).[4] While the clustering is based on the intensity of the shock, that is, the number of sick days and hospital nights in the year of the shock, the two shock indicators are best described as transitory and persistent as they exhibit vastly different effect dynamics in event studies, which align with the dynamics of shocks in the

[2] We do not consider information on doctor visits because these do not necessarily represent a significant absence from work. A doctor visit may be for a routine check-up and may therefore not even correspond to a health event.

[3] Other approaches such as those based on hospitalization records analyze a limited and selected sample.

[4] In this appendix, we vary our definition of when a shock occurs, examine the performance of the clustering, evaluate the relationship of the shock indicators to other health measures, and perform several robustness tests of our main results.

life-cycle income dynamics literature (Blundell et al., 2008a, 2015). Transitory shocks cause temporary, minor adjustments in the employment rate of at most 2 percentage points (pp), while persistent shocks cause large and long-lasting negative employment effects of at most 7.5 pp. Persistent shocks also negatively impact gross labor income and household net income. Distinguishing between these shocks is necessary, especially in a large, general population survey, to avoid combining shocks with different impacts on outcomes, which would result in less precise and less informative estimates. Further, persistent shocks have negative impacts on life-time utility as they can only partially be insured (Blundell et al., 2008b). By showing the effects of persistent shocks on both gross and net income, we can quantify the partial insurance through taxes and transfers as well as the family. The effects of the two shocks are also relevant for policy makers, as rehabilitation and re-entry to the labor market differ vastly between them. Third, we find meaningful differences in effect sizes across different groups of affected individuals. Persistent health shocks have particularly large employment effects on older and less educated workers. We also find more minor differences in employment between white-collar and blue-collar workers and between men and women. Most worryingly, we find that younger workers who drop out of the labor force following a persistent shock show a reduced intention to return to work (20 pp) compared to other young unemployed people.

Our pursuit of novel health shock indicators is motivated by a fundamental critique of existing approaches: Individual heterogeneity may obscure the severity of a given health event regardless of the use of subjective health measures, such as health satisfaction, or objective health measures, such as disease diagnoses (Britton and French, 2020; Blundell et al., 2021). For example, a cancer diagnosis may or may not be unexpected[5] and may or may not be severe and long-lasting. In some cases, a cancer diagnosis may entail chemotherapy with long-term impacts on working capacity, while in

[5]Someone with either a long family history of cancer or a long-term smoking habit is likely to consider cancer a significant health risk.

other cases, surgery alone may suffice and recovery may be fairly quick. Similar arguments apply to subjective health measures, as the perception of diseases and health states varies from individual to individual, as does adaptation to current health (Bound, 1991; Lindeboom and Kerkhofs, 2009; Hosseini et al., 2021b).

Britton and French (2020) and Blundell et al. (2021) model individual heterogeneity as measurement error. In this framework, subjective and objective health measures reflect the true underlying health status plus an error term. Our work arises from a different perspective: Rather than taking a stance on whether there is single, fundamental index of health, we look at direct measures of diminished working capacity due to bad health (sick days and hospitalizations). We then determine whether a sudden relevant change in these measures has taken place to investigate how individuals' labor market outcomes respond to such a change.[6] A major advantage of our approach is the focus on observable behavior, that is, foregoing work or receiving treatment to recover from illnesses. Hence, individual heterogeneity in the way individuals react to bad health is directly accounted for by our measures.

There are substantial differences in the intensity and longevity of health shocks. We therefore label these indicators persistent and transitory as they mirror the dynamics of income shocks in the life-cycle income dynamics literature. Even this simple classification into two shock indicators captures much of the heterogeneity in diminished working capacity. When we run the common diagnostic of calculating the residual sum of squares (RSS) along the number of indicators the k-means algorithm is allowed to discover, we find that introducing two brings about the most significant reduction in the residual sum of squares in both absolute and relative terms. With two indicators, the RSS drops by more than 50% compared to the initial RSS. To benchmark our indicators to more conventional measures

[6]For example, Britton and French (2020) and Blundell et al. (2021) apply, for example, instrumental variables estimation to address the issue of measurement error. We pursue a complementary approach by using measures of working capacity directly.

of bad health, we correlate our shock indicators to both objective and subjective measures of health. After a persistent shock, individuals are more likely to have been diagnosed with a severe chronic disease such as cancer, heart disease, or stroke. They also rate their health less favorably than individuals experiencing a transitory shock. However, after a persistent shock, the self-rated health measure recovers more quickly than labor market outcomes like employment. This underscores that an individual's stated health may be an imperfect proxy for their working capacity.

Since we use a representative survey, we can quantify the likelihood of experiencing a certain shock for the entire population at risk (employees). The probability of experiencing a transitory shock in any given year is about 1.68%, while it is 1.72% in the case of a persistent shock. However, the probability of shocks—especially persistent shocks—increases steeply with age. While the probability of both shocks is about equal between the ages of 36 and 49, for those over 50 the probability is 1.8% for transitory shocks and 2.3% for persistent shocks. A back of the envelope calculation reveals, that the likelihood of experiencing any health shock is about 80%.[7]

We conduct event study analyses to estimate the causal effects of experiencing a transitory or a persistent shock on employment, yearly working hours, gross labor income, the partner's gross labor income, and equivalent household net income. Our findings are:

1. There are large and persistent effects on the extensive employment margin for those experiencing a persistent health shock. One year after the shock, employment is about 5 pp lower among those affected. Three years after the shock, employment is 7.5 pp lower. This is especially severe for older individuals (>50 years of age), whose employment share drops by 13 pp three years after experiencing the shock, while the effect is around 5 pp for individuals up to age 50. Transitory shocks have no significant long-term effects on employ-

[7] This probability is calculated as $1-(1-0.0172-0.0168)^(65-18)$ and refers to the average likelihood of experiencing shocks. It also assumes no correlation between the shocks.

ment.

2. Those hit by a persistent shock reduce their labor supply by slightly more than 400 hours p.a. on average in the period of the shock, amounting to more than two months of full-time work. Those hit by a transitory shock reduce work by 180 hours. After both types of shocks, hours only partially recover to pre-shock levels.

3. Persistent shocks entail a substantial and long-lasting decline in gross labor income, which is reduced by around €3,000 p.a. even three years after the shock.

4. For the partner's gross labor income, we find no evidence of a reaction to either type of shock.

5. Persistent health shocks reduce household net income. However, the effect size is only 50% of the effect on gross labor income, indicating partial insurance by the family and the tax and transfer system. We do not find a significant effect of transitory shocks on household net income.

For individuals out of employment after a health shock, we also explore a subjective measure of their willingness to participate in the labor market in the future: the intention to return to work. This measure allows us to assess how long-lasting the employment effects of health shocks are. We split the analysis by age and find that after a transitory health shock, younger individuals indicate a stronger intention to reenter the workforce compared to other young individuals not in employment (94% vs. 81%). After a persistent shock, in contrast, we find that younger individuals are almost 20 pp less likely to indicate a positive intention to return to work. Among older workers, there is no statistically significant effect, as they indicate a very low baseline intention to return to work. The strongly diminished intention among younger individuals to seek employment after a persistent health shock is particularly worrisome against the backdrop of a pension system under demographic stress.

Our paper is related to several strands of literature in the fields of health and labor economics. The canonical model of health capital (Grossman, 1972) conceptualizes conceptualizes health as a durable capital stock that depreciates with age, that can be increased by investment, and that produces an output of healthy time. Individuals can allocate this healthy time freely between labor and leisure.Many empirical studies (e.g., Blundell et al., 2021; Hosseini et al., 2021a,b; Capatina, 2015; Kemptner, 2019) operationalize this concept of health capital or a single index of health either by using survey variables on self-assessed health or building indices from objective measures like disability classifications or disease diagnoses, or by combining both objective and subjective measures. Many of these studies then use the derived health index in structural life-cycle modeling.

A second strand of literature models health shocks rather than health capital and estimates the shocks' (immediate) causal impact. For example, García-Gómez et al. (2013) use acute hospitalization records and tax register data from the Netherlands to estimate the effect of such shocks on employment and income. Similarly, Schurer (2017) uses the SOEP and information on hospitalizations to examine labor supply responses after a health shock. For the United States, Dobkin et al. (2018) use hospital admissions to investigate the impact on labor market outcomes and beyond, as they also examine medical expenses, credit, and bankruptcy.

Our study ties in with this second strand of literature, since we also develop health shock indicators and investigate their impact on the labor market. However, we differ from previous studies in several key respects. First, we develop health shock indicators for the entire population at risk, that is, dependent employees in a general population survey. Second, we combine sick days and hospitalizations using clear, data-driven criteria (k-means clustering) to derive two distinct health shock indicators, which enables us to give a more detailed account of the effects of adverse health events. Hospitalizations generally indicate severe health shocks, leaving out more common and less severe health events. However, some illnesses do not require hospitalization but still reduce the capacity to work. We are able to capture the whole range of such adverse health events by drawing information from the joint distribution of sick days and hospitalizations. Third, we explore the impact of these health shocks on a comprehensive set of labor market outcomes and are further able to measure the impact on household-level variables, such as net income, which is more closely related to individual welfare. Our use of German data enables us to focus on the labor supply incentives and the imperfect insurance against income risk. In Germany, in contrast to the United States, medical expenses are generally fully covered by insurance, but income losses are only partially covered. More details on the institutional setting in Germany are given in Appendix B. Fourth, we are able to examine the dynamic trajectory of all of these outcomes over long time horizons, that is, at least three and, in robustness exercises, five post-shock years. Thus, we offer credible information on the long-term dynamics of labor market outcomes after health shocks. Fifth, we introduce a method to recover transitory and persistent shocks. Persistent shocks have important implications for individual welfare, but also for policy-making. Their long-lasting effects on individuals' labor market outcomes must be taken into consideration by policy makers when designing appropriate mitigation measures, such as the duration and extent of sickness benefits as well as rehabilitation and retraining programs. Connected with this, we document that persistent shocks have a negative effect on younger individuals' intention to return to work. It is crucial to consider how to entice these workers back to work because losing their productive capacity over the whole range of their potential labor market career is harmful to the fiscal integrity of the German welfare state.

The paper is organized as follows: In Section 2, we describe the data and derive our health shock classification. In Section 3, we lay out the empirical strategy. We present our results in Section 4 and discuss and compare them to the existing literature in Section 5. Section 6 concludes.

2. Data

Our study is based on data from the German Socio-Economic Panel (SOEP), a longitudinal representative household survey comprising around 30,000 respondents annually (Goebel et al., 2019). The SOEP contains a comprehensive list of socio-economic indicators, detailed labor market information, as well as subjective and objective health measures. For our analyses, we use 27 SOEP waves from the year 1994 to 2020.[8]

We restrict the sample to the working population aged 18 to 65. The sample offers labor market information for the years 1994 to 2019, with some of our variables being surveyed retrospectively. An overview of the number of observations is provided in Table 1.

To define our working sample, we exclude spells of motherhood, that is, absences from work due to the birth of a child. These spells are characterized by simultaneous changes in health and labor market status and are not informative for our causal analysis. We exclude the self-employed from our analysis because their access to health care generally differs from that of the rest of the population in Germany, and their income losses are not insured through the state, which affects their incentives to return to work.

In the event study design, which is our main analytical tool, we restrict the sample to observations that we can follow for seven consecutive years, that is, three relative periods before and after the health shock. As a result, we analyze the effects of health shocks that occur between 1997 and 2016.

As one might expect, the probability of experiencing a health shock is not independent of socio-demographic characteristics, which could potentially undermine our identification strategy. Table 2 shows descriptive statistics (means and standard deviations) for the two treatment and control groups. As described in Section 3, we use a combination of exact and Mahalanobis distance matching on a set of covariates[9] one period prior to the event

[8]We concentrate on this observation period due to data restrictions. Sick days were not surveyed in the SOEP in earlier years, interrupting the time series for one of our essential variables.

[9]We match 1:1, with the exact matching variables being age, gender, survey year,

Table 1. Observations in the Dataset

	SOEP total	Working age pop.	Working sample: Treat (trans)	Control (trans)	Treat (pers)	Control (pers)
N	95,685	84,553	2,049	2,049	1,889	1,889
Obs.	635,468	517,716	14,343	14,343	13,223	13,223

Note: N refers to unique individuals in the respective dataset; Obs. refers to person-year observations. Working-age population comprises individuals between 18 and 65. The working sample comprises individuals in the transitory shock group, the persistent shock group, and the respective control groups after matching. *Source*: SOEP v37.

taking place and match separately for the transitory and the persistent group, producing two distinct control groups. In Table 2, we present the basic descriptives for these two control groups, which underscore that the respective treatment and control groups are generally very similar in their socio-economic characteristics.

Table 2. Descriptive statistics of treatment and control groups

		Age	German	East	Fem.	Marr.	Educ.	Exp.
Treat (trans)	Mean	42.49	0.94	0.29	0.47	0.73	0.24	16.86
	SD	9.49	0.24	0.45	0.50	0.45	0.43	10.58
Control (trans)	Mean	42.49	0.96	0.23	0.47	0.75	0.27	16.53
	SD	9.49	0.20	0.42	0.50	0.43	0.44	10.59
Treat (pers)	Mean	44.11	0.93	0.30	0.48	0.74	0.22	18.42
	SD	9.54	0.26	0.46	0.50	0.44	0.42	11.07
Control (pers)	Mean	44.11	0.95	0.26	0.48	0.77	0.25	18.29
	SD	9.54	0.22	0.44	0.50	0.42	0.43	10.91

Note: Displayed are means and standard deviations of the treatment and control groups. Fem. refers to the female share, Marr. refers to the share married, Educ refers to the share tertiary education, while Exp refers to years of experience in full-time employment. *Source*: SOEP v37.

and employment. For the distance matching, we use a dummy for having German citizenship, a dummy for having children under the age of six in the household, a marriage dummy, three educational categories (primary, secondary, tertiary), years of full-time and part-time work, and dummies for blue-collar work and white-collar work.

2.1. Focal Variables

Our outcome variables fall into two groups: individual-level outcomes, which concern the labor market, and household-level outcomes, which concern the partner and the welfare of all members of the household.

Individual-Level Outcomes We consider three outcomes:

1. labor market participation, that is, being in regular employment[10];
2. yearly working hours adjusted for sick leave;
3. yearly personal labor income adjusted for sick leave.

The construction of the latter two measures requires us to adjust the existing measure of yearly hours provided in the equivalence file of the SOEP (Grabka, 2020). The existing measure of yearly hours is constructed by combining the SOEP's information on months spent in employment and the regular weekly working hours, but no attempt is made to correct for time spent away from work due to sickness. We use sick days to construct a corrected measure of yearly hours. We calculate

$$h_{i,t} = \tilde{h}_{i,t} - sickdays_{i,t} \times hpd_{i,t}\,, \tag{1}$$

where $\tilde{h}_{i,t}$ is the existing hours variable from the equivalence file, $sickdays_{i,t}$ is the number of sick days away from work, and $hpd_{i,t}$ is the average number of hours the individual works per day.[11] In Figure D.2 in the appendix, we show the distributions of sick-leave-adjusted and unadjusted hours, which are fairly similar, yet the adjusted distribution is uniformly shifted to the left (lower hours).

[10] Regular employment is defined as dependent employment, regardless of the number of hours worked. Individuals who are not in regular employment are apprentices, interns, and on-the-job trainees. We consider individuals as being in regular employment in a given year if they meet the above conditions at any point of the year.

[11] We construct $hpd_{i,t}$ from the recorded hours of work per week. Our assumption is that the individual works five days per week.

To adjust personal labor income, we use the information on sick days, aggregating to months, and then use microsimulation to calculate the replacement income.[12] We then reduce the unadjusted yearly income by the difference between replacement and employment income for the duration of the sickness spell.

Household-Level Outcomes We consider two outcomes:

1. partner labor income;
2. household net income.

Unlike the other outcomes, partner labor income or household net income do not require adjustment. Partner labor income is reported directly by the individual's partner, preventing the problem of misreporting a partner's income. Household net income is compiled by adding up all income sources, including non-labor incomes, and subtracting taxes and social security contributions calculated by microsimulation, as detailed in Grabka (2020). We needs-adjust this household net income with the modified OECD scale. All income variables are in 2018 euros according to the consumer price index.

Health Variables The two main health variables are *sick days*[13], that is, the number of days an employee is not working due to sickness absences registered with the employer,[14] and *hospitalizations*, that is, the number of overnight stays in a hospital. The advantage of using these variables for our analysis is that both capture health behavior related to the labor market. Both variables imply an incapacity to work, while hospitalizations

[12] We adjust monthly labor income because in the SOEP, it is recorded once and then spread across months of employment. We calculate replacement rates for every year according to the sickness benefits framework of the German health insurance system.

[13] Note that by definition, sick days are not recorded for the unemployed. Within the scope of our study, this is not a relevant limitation.

[14] In most cases, a sick employee has to notify their employer and submit a doctor's note verifying their medical status and the duration of the absence. Information on sick days is therefore very salient to the sick person, making it highly likely that it is reported accurately in the survey.

also indicate the need for inpatient treatment and thus signal more serious health issues.

We show the time series of sick days and hospital nights, restricted to observations with positive values, in Figure D.1. Both variables exhibit distributions with very long tails. Sick days exhibit a slight downward trend from 2002 on. Hospitalizations have gone down over time, especially since 2002, the year in which a strict reform on the maximum billable days in the hospital was introduced (see Appendix B).

Comparing administrative statistics compiled by the Institute for Employment Research (IAB), average sick days in 1993 were at 12.3, decreased to 8.1 in 2007, and then rose again to 10.6 in 2017 (Wanger et al., 2019). Figure D.3 in the Appendix shows that the SOEP data track these trends in administrative data well.

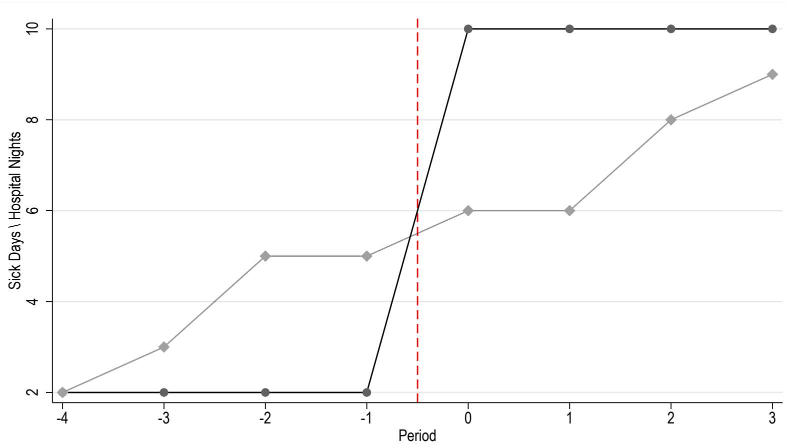
2.2. Health Shocks: Distinguishing Transitory and Persistent Treatment Groups

To derive health shock indicators, we first separate individuals with a sudden deterioration in health from the rest of the sample and then apply the clustering procedure.

Individual Deviation Condition Health (slowly) declines over the life cycle and these health trajectories are bound to be subject to individual heterogeneity. Thus, an adverse health event in isolation—a long absence from the job or a long hospitalization—does not necessarily represent a health shock because these events might be part of a declining trajectory or individual heterogeneity. Shocks are, by definition, sudden deviations from the current trajectory. We therefore exploit the panel dimension of our data and require that the health events that we classify as shocks are major deviations from an individual's health trajectory.

Our implementation of this requirement is as follows: We calculate an individual's median and standard deviation of sick days and nights hospitalized for a rolling window of five years. Thus, in each year, we construct a median and a standard deviation for both measures based on the current and four previous years. Only if a health event—that is, a change in either hospital nights or sick days from work—exceeds the individual's median by more than two standard deviations do we consider this person *shocked*. In the case that an individual satisfies the individual deviation condition more than once in their lifetime, we only consider the first instance.

In Figure 2, we illustrate two health trajectories: one person satisfies the condition and is classified as having experienced a health shock (black) and the other does not satisfy the condition (gray). As a robustness check, we lower the standard deviation condition to one standard deviation and show alternative results for employment in Figure A.1 in the appendix. The results do not substantially differ from those shown in the main analysis.



Note: This figure illustrates two stylized individual health trajectories. For the black line, the individual deviation condition is satisfied (median: 2; standard deviation: 3.6) but not for the gray line (median: 5; standard deviation: 1.6). The black trajectory would result in the person being considered treated.

Figure 2. Individual Deviation Condition (stylized)

Clustering Procedure Previous literature on health shocks distinguished primarily between healthy individuals and those experiencing *any* kind of health shock. This was true regardless of whether the shock measure was based on self-assessed health (**?**García-Gómez, 2011), disease diagnoses (Fadlon and Nielsen, 2021), or data on hospitalizations (García-Gómez et al., 2013; Schurer, 2017; Dobkin et al., 2018). However, as pointed out by Britton and French (2020), both health conditions and health shocks exhibit substantial between-individual heterogeneity. For example, a hospitalization due to a broken leg or a cancer diagnosis may lead to completely different paths for health, recovery, and labor market outcomes. As we are analyzing a sample of the entire German workforce, the heterogeneity in health conditions is much greater, compared to smaller samples from hospitalization records. To address the issue of shock heterogeneity in this large population, we need to distinguish between different intensities of shocks to assess the respective effects on labor market outcomes. Conducting our analysis without accounting for shock heterogeneity would result in a less informative and more imprecise combination of the two estimates presented here. To account for the heterogeneity, we introduce a data-driven method to distinguish between health shocks of higher or lower intensity. As the event study analysis will show, these shocks also differ in their impact on the dynamics of labor market outcomes. For this reason, we label them *transitory* and *persistent* health shocks.

We classify the two types of shocks using k-means clustering applied to the joint distribution of sick days and hospital nights (Friedman et al., 2017). The k-means algorithm sorts observations into k groups by minimizing the sum of squared deviations from the group-specific mean. Thus, the algorithm allows us to distinguish between shocks according to a clear criterion. Formally, it seeks to find the sets $\mathbf{S} = S_1, ..., S_k$ that minimize

$$\sum_{i=1}^{k} \sum_{\mathbf{x}_j \in S_i} ||\mathbf{x}_j - \mu_i||^2, \tag{2}$$

where $\mathbf{x}_j \in \mathbf{X}$ and $\mathbf{X}$ is the set of all observations and μ_i is the vector giving the means specific to group i. In our case the number of dimensions is two and we set $k = 2$. To evaluate whether $k = 2$ is a reasonable choice, we calculate and compare Eq. 2 for different values of k.[15] In Appendix A,

[15]The k-means algorithm is not guaranteed to deliver a unique, global optimum independently of the starting values. Thus, we use 100 different random starting value

we document that $k = 2$ is a reasonable compromise between group size, and therefore the reliability of the subsequent analysis, and the amount of heterogeneity that is captured.

Figure 3 illustrates the results of the clustering procedure for the first and last year of our observation period. In both years there is a border between transitory and persistent shocks, which shows the asymmetric trade-off between sick days and hospital nights: At low levels of hospital nights, only a very high number of sick days defines a persistent shock. Conversely, at high levels of hospital nights, a few sick days suffice to classify a shock as persistent. For example, in 1997, individuals are classified as experiencing a persistent shock at over seven hospital nights in conjunction with at least seven sick days. By contrast, at two hospital nights, an observation will be classified as a persistent shock only at a minimum of 82 sick days. The trade-off between sick days and hospital nights, and thus, the thresholds for transitory and persistent shocks, are not rigid across the years. For example, in 2016, at two hospital nights, an observation will already be classified as a persistent shock at a minimum of 52 sick days. Accordingly, the procedure implicitly weights the two dimensions differently, which is intuitive as hospital nights generally reflect more severe and long-lasting health conditions. In Appendix A, we show how relying solely on hospital nights would change the clustering results. The take-away from this test is that many observations with very low numbers of sick days and at least three hospital nights would move a shock from transitory to persistent, which is unappealing as such a short-term hospitalizations may be due to temporary interruptions such as a bone fracture or appendectomy. We conclude from the test that both variables contain information that is valuable for the classification.

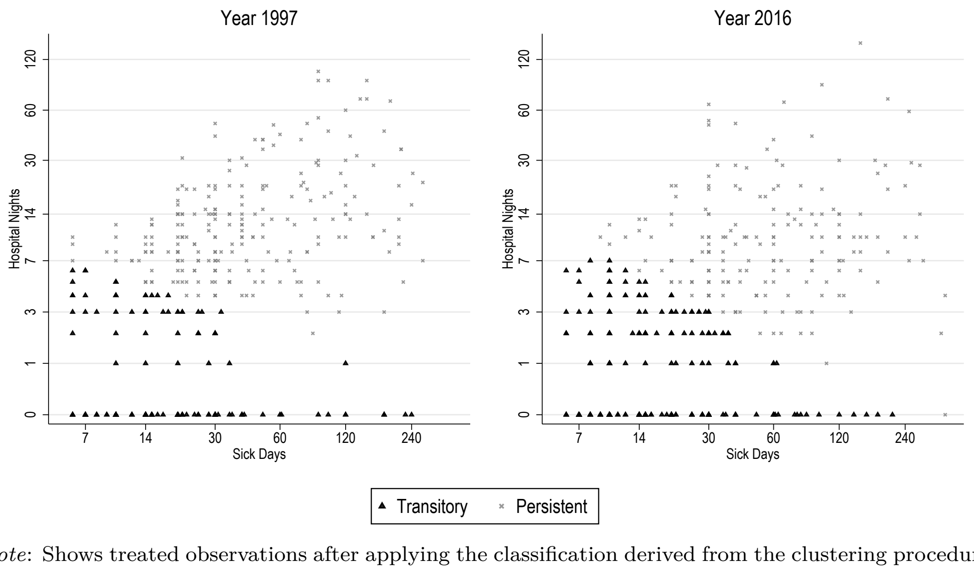
By allowing for flexible thresholds over time, we take into account that policy makers reduced the profitable duration of hospitalization in Ger-

combinations to ensure a stable, optimal solution. For a recent framework leveraging k-means clustering techniques to discover unobserved heterogeneity see the application by Bonhomme et al. (2019) in the field of firm- and worker fixed effects.

many over our observation period (see Figure D.1). Hence, similar health conditions imply fewer hospital nights in 2016 than in 1997, which is accounted for—as seen in Figure 3—by the clustering procedure.

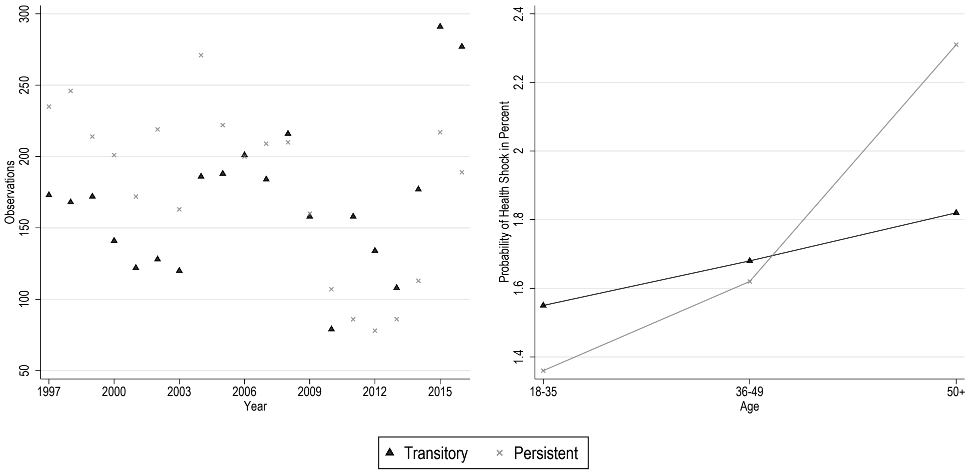
The left-hand panel of Figure 4 displays the absolute number of transitory and persistent shocks per year. The number of transitory and persistent shocks mostly varies between 100 and 300 per year without a clear, systematic, long-term upward or downward trend. Instead, the number of observed shocks correlates with the overall sample size, which changes over time due to panel attrition and the SOEP survey adding refresher samples.

The right-hand panel shows the probability of a given shock occurring in different age bands. The average probability of shock occurring for the at-risk population, that is, employed people, is is about 1.68% for a transitory and 1.72% for a persistent shock. As the figure shows, this probability varies widely with age, especially for persistent shocks. At the beginning of working life, the probability of experiencing a persistent shock is about 1.4%, while it is about 2.3% close to the end of working life.



Note: Shows treated observations after applying the classification derived from the clustering procedure in 1997 and 2016, i.e. the first and last years used as event periods for the event study. Many observations occupy the same points, which we do not represent graphically, i.e., the figure is not weighted. *Source*: SOEP v37.

Figure 3. Groups After Clustering - 1997 and 2016



Note: The left-hand panel shows observations per year with transitory or persistent shocks from 1997 to 2016. The right-hand panel shows the probability of experiencing a transitory or a persistent shock in a given year for different age bands. The population at risk are those employed in the period prior to the shock as in our event study framework. *Source*: SOEP v37.

Figure 4. Health Shocks Incidence

In Appendix A, we show several benchmarks of our health shock indicators against other objective and subjective measures of bad health. All indicate a strong and positive correlation with these alternative measures.

3. Empirical Strategy

Our empirical strategy relies on the comparison of individuals who experience a transitory or persistent health shock (treatment groups) with individuals who have not and will not experience a shock in the period of observation (never-treated/control group). We follow a two-step process:

Matching We match control units to treatment units one relative period prior to the shock based on a set of socio-demographic characteristics and other variables possibly affecting trend evolution. We combine two matching procedures: Mahalanobis distance matching (Mahalanobis, 1936) and exact matching.[16]

We match 1:1 with the exact matching variables being age, gender, survey year, and employment. For the distance matching, we use a dummy for having German citizenship, a dummy for having children under the age of six in the household, a marriage dummy, three educational categories (primary, secondary, tertiary), years of full-time and part-time work, and dummies for blue-collar work and white-collar work.

Finally, to align treatment and control observations on health trajectories pre-shock, we include the number of sick days in the three pre-shock periods. This is crucial for our identification strategy, which rests on the common trends of treatment and control observations. Intuitively, when the pre-shock health trajectories of treatment and control observations align and the only appreciable difference between the two is the strong deviation in health measures in period 0, we may take this as convincing evidence that the common trend assumption is satisfied.

For the estimation of effects on partner incomes, we restrict to individuals with a partner and additionally include a dummy indicating whether the partner has German citizenship, a quadratic for the partner's age, a dummy indicating whether the partner is employed, and the partners income in the pre-shock period as further covariates for the matching procedure. This allows us to obtain two control groups, one for the transitory and one for the persistent treatment group for each outcome variable.

To estimate the effects on household net income, we do not change the distance matching variables but simply add a partner dummy to the set of exact matching variables.

Estimation framework Based on these matched groups, we perform event study analyses by running OLS regressions with individual and year fixed

[16] The Mahalanobis distance is computed as follows: Let x and y be two vectors with observations on several variables. Then, $D_M(x, y) = \sqrt{(x-y)'CV^{-1}(x-y)}$ is the Mahalanobis distance of the two vectors, where CV is the covariance matrix associated with the variables in (x, y). Matching is implemented using the `kmatch` command in Stata (Jann, 2017).

effects as well as a set of dummies for the pre- and post-shock relative periods and their interactions with the treatment dummy. The coefficients of the post-shock interaction dummies measure the average treatment effect of a health shock on the treated (ATT). The regressions take the form

$$Y_{it} = \sum_{k=-3,k\neq-1}^{3} \gamma_k P_{it}^k + \sum_{k=-3,k\neq-1}^{3} \delta_k P_{it}^k \times T_i + \nu_i + \tau_t + \epsilon_{it}\,, \tag{3}$$

where Y_{it} is the outcome of interest for person i in year t, for example, employment or yearly hours, $\left\{P_{it}^k\right\}_{k=-3}^{3}$ is a set of relative period-dummies running from -3 to 3, but excluding $k = -1$, with a shock occurring in period k=0 if the person is in a treatment group, T_i is the respective treatment group dummy, ν_i is an individual fixed effect, τ_t is the year dummy, and ϵ_{it} is an idiosyncratic error. The coefficients of interest are the δ_k.

Again, our central identifying assumption is the common trend assumption, which states that in the absence of the shock, the evolution of the outcome for the treatment group would have been the same as for the control group.[17] If this assumption holds, we can interpret the differences in outcomes between control and treatment after the shock as causal. Beyond the fact that we align treatment and control group on socio-economic characteristics using our matching procedure, we are able to inspect pre-trends, that is, the outcome differences three and two periods prior to the treatment. In Appendix A, we add two additional pre-periods to get a better sense of the evolution of the outcomes across even longer time horizons before the shock. While this restricts our estimation sample considerably, we find no relevant differences compared to the estimated effects in our main analysis.

The event study design addresses the concern of reverse causality—the possibility that a labor market shock (e.g., dismissal, demotion, plant closure) can cause health problems (Haan and Myck, 2009; Marcus, 2013; Britton and French, 2020). Further, it rules out the possibility that other

[17]See, for example, Sun and Abraham (2021) and Goodman-Bacon (2021) for recent expositions on the topic.

contemporaneous confounders affect the outcome. Additionally, the event study enables us to understand the long-run dynamic effects of the health shocks.

4. Results

We show the results of our analysis by plotting the coefficients δ_k for employment, working hours, gross labor income, partner income, and equivalent household net income in Figure 5. For each outcome, we show two plots: one for the transitory shock group, and an analogous one for the persistent shock group. Each coefficient can be interpreted directly as the average period-specific treatment effect, that is, the average difference between the respective treatment and control group in a given period.

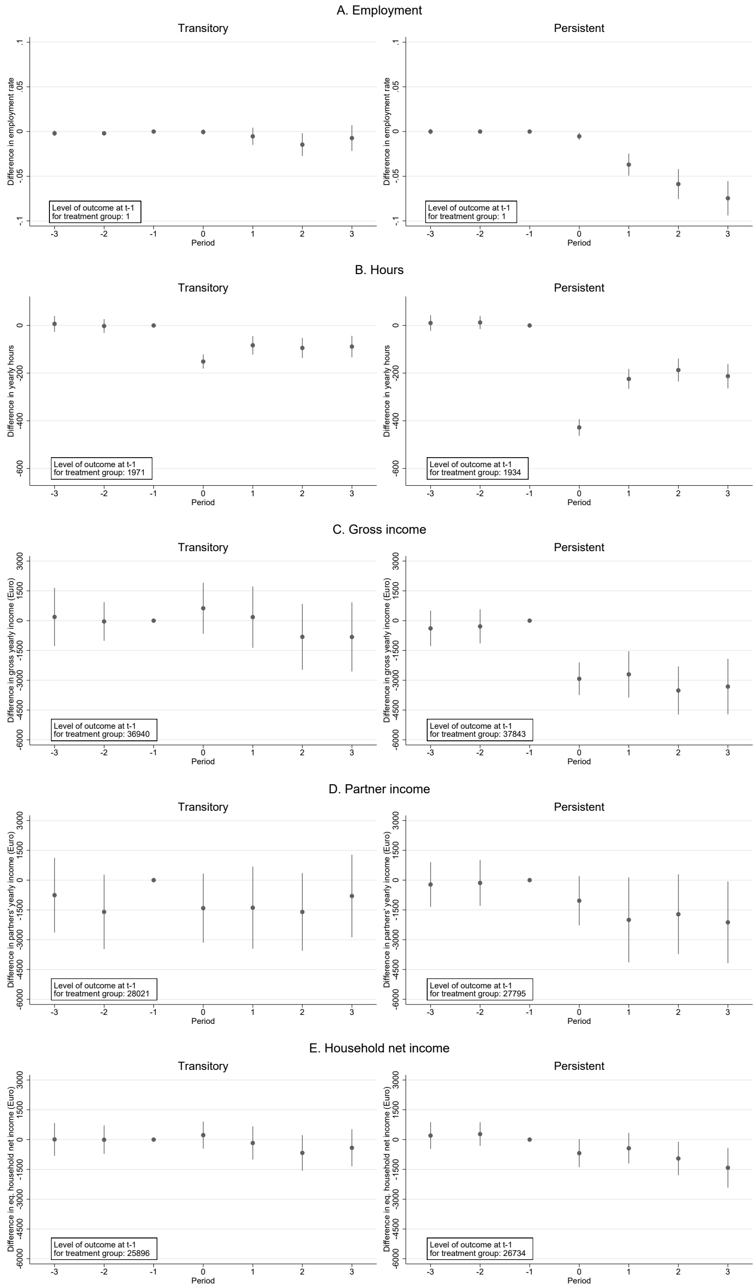
4.1. Main Analyses

Employment Panel A of Figure 5 displays the effects of either type of shock on employment. Transitory health shocks have small, negative, and mostly statistically insignificant employment effects as the employment rate is reduced by 1 to 2 pp for the treated two periods after the shock. In line with our shock terminology, three periods after the shock, the employment effect is basically indistinguishable from zero again. That is, after transitory health shocks, the affected individuals return to employment.

The employment effects of persistent health shocks, in contrast, are much larger and have a different trajectory. One period after the shock, employment drops by about 4 pp, continuing to drop consistently to about 7.5 pp in period 3.

Yearly Hours Panel B shows the effect of experiencing a transitory or a persistent shock on yearly hours.[18] Those who are not employed remain in the sample and work zero hours. Transitory shocks reduce hours in

[18]We adjust yearly hours to be consistent with the number of days registered sick. See Section 2.



Note: Shows period-specific coefficients according to Eq.(3). Bars give robust 95% confidence intervals of the respective coefficients. Number of observations: Transitory shock: 4,098; persistent shock: 3,778. *Source*: SOEP v37.

Figure 5. Main Results: ATTs of Transitory and Persistent Health Shocks

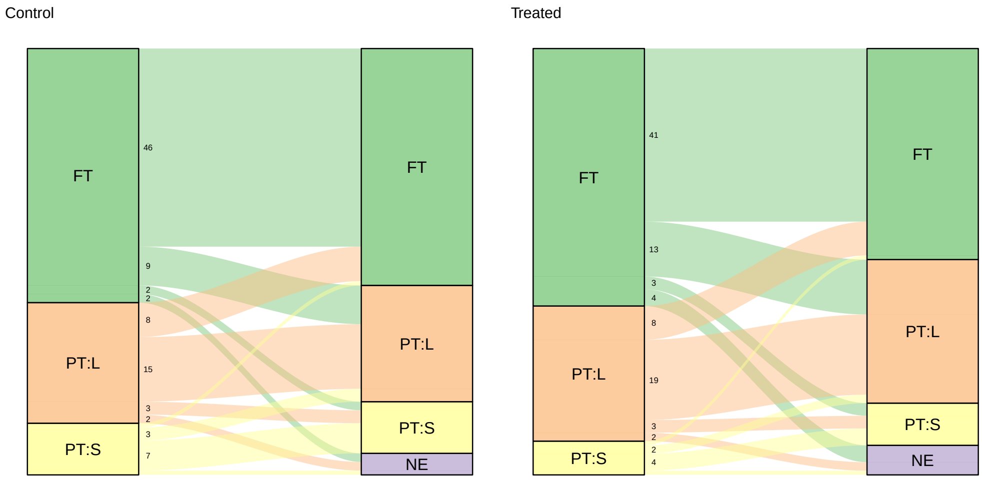
the period of the shock by around 180, which amounts to roughly one month of full-time work. The drop is followed by a quick but only partial recovery: Three periods after the shock, the yearly hours of those affected are still around 100 hours below the hours of the control group. Following persistent shocks, hours drop by slightly more than 400 in the period of the shock, which amounts to more than two months of full-time work. In the following periods, hours recover partially, but three periods after the shock, the average treatment effect still indicates a reduction of about 200 hours, which translates to about 1.2 months of full-time work.

Part of the sizable immediate effect in the period of the shock is mechanical, as the hours variable is adjusted to reflect the time spent in sickness absence. This mechanical effect operates on the whole sample of affected individuals, while in subsequent periods, if people manage to go back to work, the effect partially dissipates. The non-mechanical effect in periods 1 to 3 is a composite of both extensive and intensive margin adjustments, which we decompose in Figure 6.

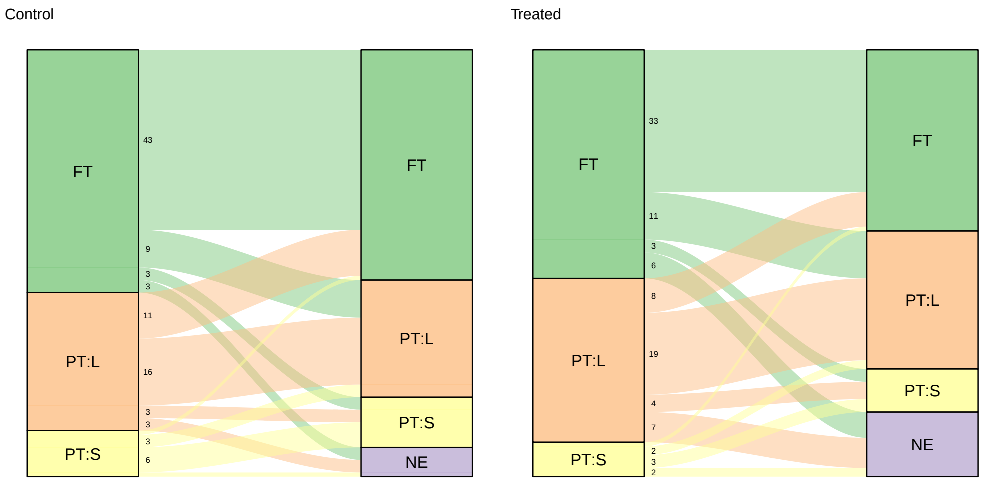
We show transitions from period -1 to period 3 between the following employment categories: 1) full-time (FT), defined as equal to or more than 2000 hours, 2) long part-time (PT:L), defined as from 1000 to 2000 hours, 3) short part-time (PT:S), defined as fewer than 1000 hours, and 4) not employed (NE), that is, zero hours.

Starting with transitory shocks in Panel A, we see that pre-shock, full-time workers in the treatment group have a 44% higher probability to move into long part-time employment (13% vs 9%). Further, about 4% of the treated full-time workers move out of employment, doubling the probability compared to the control group. For the other hours categories (PT:L,PT:S), transitions do not differ substantially.

Moving to persistent shocks in Panel B, we find that not only full-time workers but also workers in all hours groups are affected substantially. For all hours categories, the probability of moving out of employment at least doubles: For full-time workers, it increases from 3 to 6%, for long part-time workers, it increases from 3 to 7%, and for short part-time workers,



A.Transitory



B.Persistent

Note: Shows transition percentages for yearly hours categories for treatment and control group moving from period -1 to 3. *NE* means zero hours, *PT:S* means less than 1000 hours per year, *PT:L* means between 1000 and 2000 hours, *FT* is 2000 or more hours. Number labels for percentages smaller than 2 percent are not shown due to space constraints. *Source*: SOEP v37.

Figure 6. Transitions Between Hours Categories from Period -1 to 3

it increases from 1 to 2%. By contrast, the intensive margin adjustments, that is, switching from FT to PT:L or PT:S, are smaller compared to the transitory shock case. This suggests that persistent shocks fully incapacitate some affected workers, making them unable to participate in the labor market. This result highlights the relevance of our approach: When workers leave the labor market due to a health shock, finding new employment becomes more difficult than adjusting working hours in their current job. As a result, persistent shocks may entail knock-on effects, not just due to the limitation of working capacity, but also due to the scarring effect of leaving the current job.

Individual Labor Income Panel C of Figure 5 shows the effect of transitory or persistent shocks on yearly gross labor income. The sample remains constant, meaning those who are not employed earn €0 in our data. While the point estimates turn negative in the final two periods, we do not find any statistically significant effects of transitory health shocks on gross labor income. This is in stark contrast to persistent shocks, which have both statistically significant and economically meaningful effects on gross income: In all post-shock periods, gross income of those who experienced a persistent health shock is reduced by about €3,000, that is, an 8% drop from the baseline. The period-specific estimates indicate no trend, which also means no indication of recovery.[19]

Partner Labor Income Panel D in Figure 5 shows the effect of both shock types on partner labor income, allowing us to investigate potential added-worker effects. These occur when one partner increases their labor supply and, thus, their income to compensate for the income loss due the other partner's health shock. We consider all partners whose relationship status remained unchanged during all seven periods, and we do not condition on their employment status at the time of the shock. Unlike our results for the individual-level outcomes, we find no significant effects for either of the two shock types. The estimated confidence intervals are too large to strongly assert a null effect. In line with our findings, the literature does

[19]In Figure A.5 in Appendix A, we show effects for two additional post-shock periods. These indicate that the effect remains the same even five years after the shock.

not generally find added worker effects to be relevant (Dobkin et al., 2018; De Nardi et al., 2021).

Household Net Income Panel E of Figure 5 shows the effect of experiencing transitory or persistent shocks on yearly household net income. Estimating the effects on household net income allows us to calculate a pass-through coefficient of the shocks; a measure of insurance provided by the household and the state.

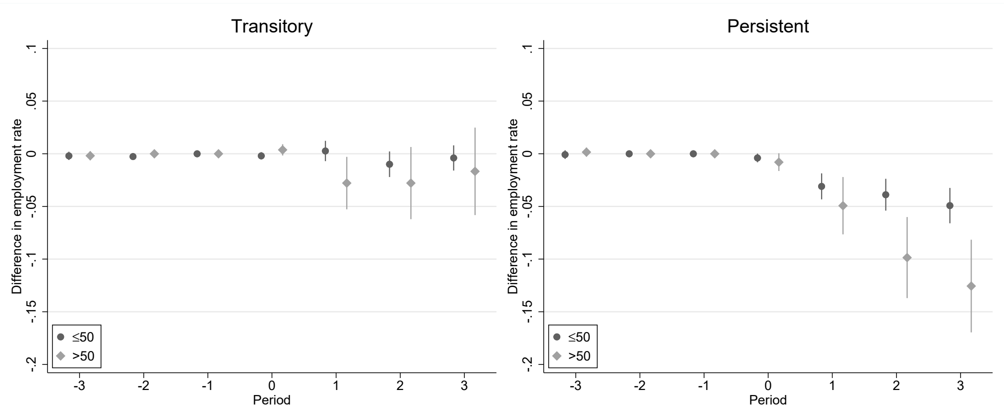
Transitory shocks have no significant effect on household net income, mirroring the results for employment and gross income. In contrast, persistent health shocks reduce household net income by about €750 in the period of the shock. This effect increases to €1,500 three periods after the shock. Since we find no significant effect of transitory shocks, we will not calculate a pass-through coefficient for transitory shocks. For persistent shocks, the effect on gross labor income was a decrease of about €3,000 in the third period, while the decrease in net household income amounted to roughly €1,500. Thus, in the final period, about 50% of the gross shock passes through to net household income, whereas in the periods immediately after the shock, the pass-through is smaller, pointing to more insurance through the tax and transfer system.

Finally and considering all outcomes, we do not find significant pre-shock trend deviations in any treatment group. In Appendix A, we provide analyses including two additional pre-shock periods, for which the same holds. In addition to our main results, this further strengthens our confidence that our identifying assumptions are satisfied.

4.2. Effect Heterogeneity

Age Age is the most natural margin of heterogeneity. Thus, we repeat our main analysis of employment effects after splitting the sample into individuals 50 years of age or younger (younger group), and individuals over

50 years of age (older group).[20] Figure 7 shows the effects of transitory and persistent health shocks on employment by age. After transitory shocks, point estimates are larger for the older group, although the confidence intervals often contain the point estimates for the younger group. Importantly, both age groups exhibit similar dynamics, with a small dip and recovery in employment. Persistent health shocks have similar effects on both age groups in periods 0 and 1, but diverge thereafter. Both age groups exhibit a reduction of 4 to 5 pp in employment in the first period. However, the older group detaches in the second and third period and shows effect sizes of 10 and 13 pp, respectively. For the younger group, in contrast, the effect sizes are smaller, ranging between 4 and 5 pp in periods 2 and 3. These age-specific dynamics are in line with the notion that many of the older workers are close to retirement and that a sudden deterioration in health will likely be a factor motivating individuals to drop out of the labor force and enter retirement.

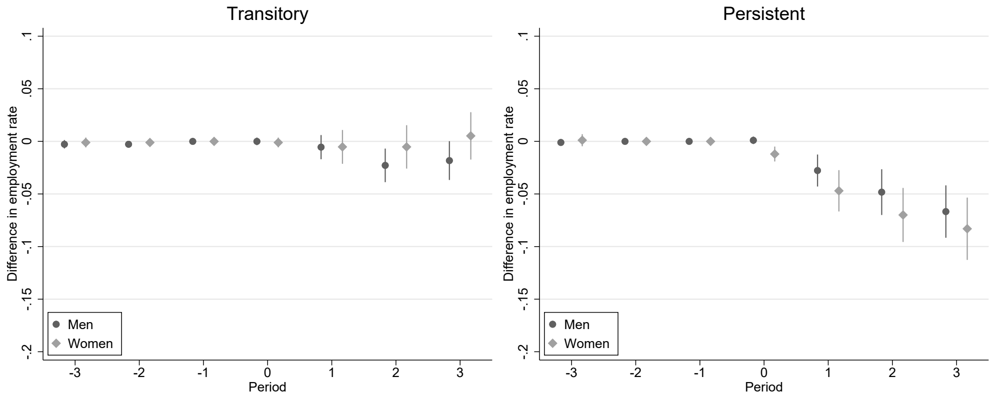


Note: Shows period-specific coefficients according to Eq.(3) for the treated split by age ($\leq$ 50 vs. $>$ 50). Bars give robust 95% confidence intervals of the respective coefficients. Number of observations: Transitory shock: 3,020 $\leq$ 50, 1,078 $>$ 50; persistent shock: 2,520 $\leq$50, 1,258 $>$ 50. *Source*: SOEP v37.

Figure 7. Effect Heterogeneity: Age

Gender Figure 8 displays effects of transitory and persistent health shocks separately for men and women. While for transitory shocks both dynamics

[20] We fix this age heterogeneity in period -1.

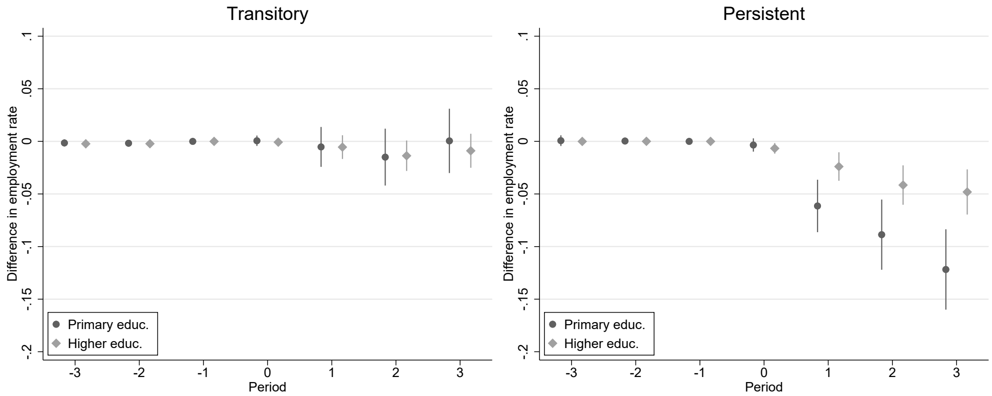


Note: Shows period-specific coefficients according to Eq. (3) for the treated split by gender (male vs. female). Bars give robust 95% confidence intervals of the respective coefficients. Number of observations: Transitory shock: 2,190 men, 1,908 women; persistent shock: 1,948 men, 1,830 women. *Source*: SOEP v37.

Figure 8. Effect Heterogeneity: Gender

and magnitudes appear very similar, persistent shocks tend to affect women more than men. Although for most post-shock periods confidence intervals for women contain the point estimates for men, point estimates for women are consistently lower than for men. This pattern is an expected outcome as women are generally less attached to the labor market (Killingsworth and Heckman, 1986; LaLumia, 2008; Bick and Fuchs-Schündeln, 2017; Kleven et al., 2019).

Education Several papers have documented differing health behaviors by education, both before and after adverse health events (Blundell et al., 2021; Britton and French, 2020). To explore this heterogeneity margin we split our sample into a group of individuals with only primary education and a group of individuals with secondary education or above. As Figure 9 shows, for transitory shocks, effects are essentially equal for both education groups for the entire post-shock period. By contrast, persistent health shocks induce entirely different employment trajectories for low and highly educated individuals. Those with only primary education exhibit an initial decline in employment of about 6 pp which increases to about 12 pp in period 3, while the employment rate of the highly educated drops by about



Note: Shows period-specific coefficients according to Eq. (3) for the treated, split by education (primary vs. higher). Bars give robust 95% confidence intervals of the respective coefficients. Number of observations: Transitory shock: 1,186 with primary education, 2,912 with higher education; persistent shock: 1,229 with primary education, 2,549 with higher education. *Source*: SOEP v37.

Figure 9. Effect Heterogeneity: Education

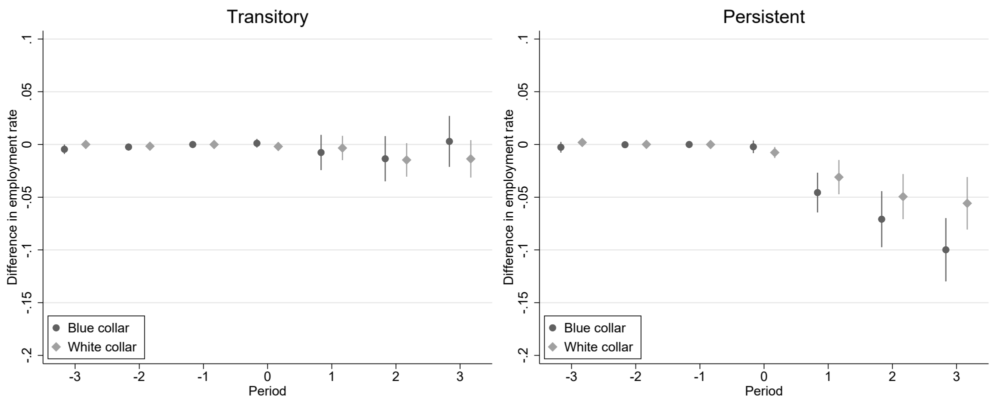
3 pp in period 1 and declines further to about 5 pp. This in line with recent findings from the US and the UK (Blundell et al., 2021).

White-collar and blue-collar collar workers Different occupational groups may react differently to similar health shocks. For example, there is a clear difference in the physicality of work between white-collar workers, who generally work in clerical occupations, and blue-collar workers, who generally perform manual tasks, predominantly in manufacturing.[21] Blue-collar workers might therefore be particularly affected by health shocks. If a health shock diminishes the physical capabilities of blue-collar workers, their working capacity might potentially be reduced to a larger extent than the working capacity of a white-collar worker, whose tasks are generally not physical.

Figure 10 shows that transitory health shocks have no significant effects on employment for either white-collar or blue-collar workers. After persistent health shocks, in contrast, the employment trajectories of blue-collar

[21] Our definition of white-collar workers contains all managerial employees (see Figure 11) as well as all other types of salaried employees. The group of blue-collar workers comprises all other types of workers, based on the SOEP variable `pgstib`.

and white-collar workers differ, with blue-collar workers showing a sharper decline in employment rates than white-collar workers. The difference is small one period after a health shock (roughly 5 pp vs. 3 pp), but the employment rate of blue-collar workers is reduced by 10 pp three periods after a health shock. This is about twice the size of the effect on white-collar workers.

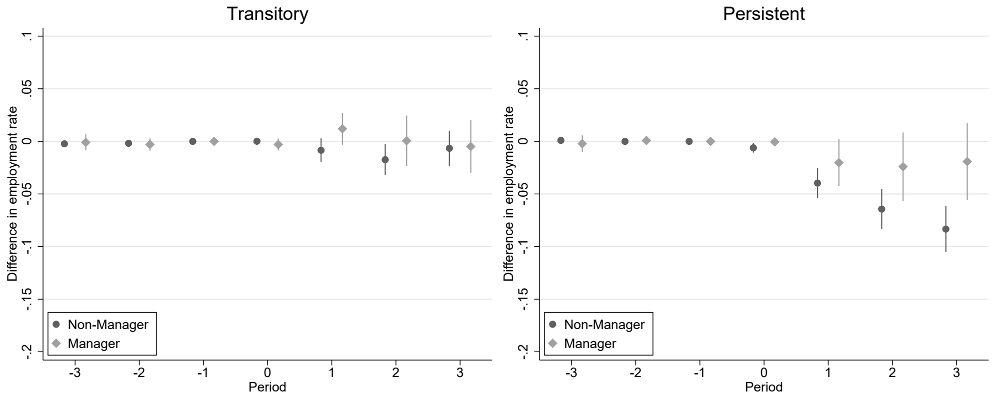


Note: Shows period-specific coefficients according to Eq. (3) for treated split by being a white-collar worker or not being a white-collar worker. Bars give robust 95% confidence intervals of the respective coefficients. Number of observations: Transitory shock: 1,709 blue-collar workers, 2,389 white-collar workers; persistent shock: 1,605 blue-collar workers, 2,173 white-collar workers. *Source*: SOEP v37.

Figure 10. Effect Heterogeneity: White-collar and blue-collar workers

For blue-collar workers, working capacity seems to diminish more strongly after a persistent health shock than for white-collar workers. However, we find a significant negative effect on employment for white-collar workers as well. Only for the very exceptional group of managers does the negative effect of persistent health shocks on employment vanish.

Managers Managers can be seen as a special case among white-collar workers. They have an exceptional position in organizational hierarchies and primarily complete tasks that are non-routine and that are cognitively but seldom physically demanding. Thus, three effects may be at play leading to potential effect heterogeneity for this group: First, managers have stronger incentive to work than other workers because of their higher wages.



Note: Shows period-specific coefficients according to Eq. (3) for the treated, split between managers and non-managers. Bars give robust 95% confidence intervals of the respective coefficients. Number of observations: Transitory shock: 3,341 non-managers, 757 managers; persistent shock: 3,144 non-managers, 634 managers. *Source*: SOEP v37.

Figure 11. Effect Heterogeneity: Managers

Second, managers are essential to a firm and are difficult to replace, creating strong incentives for firms to have managers return after a health shock. Third, since managers generally perform tasks that are less physically demanding, their recovery and return to work may be easier to accomplish.

Figure 11 shows that after transitory shocks, there are no significant differences between managers and non-managers. After persistent shocks, however, effects are more heterogeneous: In period 1, employment drops for both groups by around 2 to 3 pp, with confidence intervals largely overlapping. In the two periods thereafter, employment trajectories diverge. In period 3, managers still show a 2 pp drop but non-managers are affected more severely, with an 8 pp drop in employment. From these results, we find some support for the hypotheses stated above, as managers appear to exhibit stronger labor market attachment even after persistent health shocks.

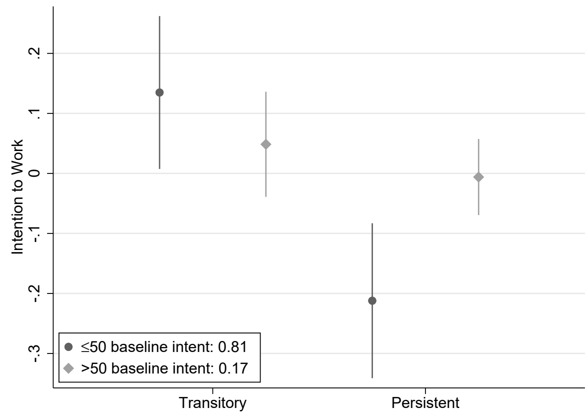
Intention to Return to Work In our main analysis, we find the largest adjustments to health shocks on the extensive margin (employment). To understand the labor market dynamics for individuals after a health shock, we investigate whether they intend but fail to reenter the labor market or whether they refrain from reentering in the first place. If the former were the case, there would be a strong public policy case for fostering these individuals' reentry to the labor market by helping to increase their probability of finding a job that is accommodating with respect to both their qualifications and their physical capacity.[22]

To determine whether individuals who dropped out of the labor market after a health shock are actively looking for and cannot find a job or are simply not looking for one, we use a question in the SOEP on the intention to work. For ease of interpretation, we code a dummy equal to one if individuals indicate a positive intention to return to work.[23] We only consider individuals who are out of employment after period 0, but pool all observations post-shock because the number of individuals who are out of work in our treatment groups is fairly low. As in the main analysis, we distinguish between people out of employment in the control group, who have not experienced a shock, and people out of employment who experienced either a transitory or a persistent shock. Further, we distinguish between those over the age of 50 and those 50 years of age or younger. We report the predictions of the intention-to-work indicator after running OLS regressions for the control and treatment groups by age groups. Figure 12 shows the results.

Individuals up to age 50 in the control group generally intend to return to work, with the mean of the intention-to-work indicator at about 0.81. For those in the control group over the age of 50, the mean is much lower, at about 0.17. Accordingly, when individuals over 50 end up out of employment—regardless of whether this was caused by a health shock or

[22] The German system has made some progress in this direction: When employees return to their old job after an illness, they can file to reenter at reduced capacity and then progressively increase their workload up to full capacity ("Wiedereingliederung", as per Sozialgesetzbuch IX).

[23] The SOEP asks those not in employment whether they are likely to obtain or resume employment in the future, with the answers falling into four categories: 1) "No, definitely not", 2) "Probably not", 3) "Probably", and 4) "Yes, definitely". We recode these categories into a dichotomous variable, in which categories 1 and 2 are coded as a zero and categories 3 and 4 are coded as a 1. This is the intention-to-work indicator that we use in subsequent analyses.



Note: Shows post-treatment coefficient differences aggregated for the three post-shock periods for treated and control units either being older than or up to 50 years of age. Bars give robust 95% confidence intervals of the respective coefficient differences. Number of observations: Transitory shock: 171 $\leq$ 50, 314 $>$ 50; persistent shock: 196 $\leq$50, 549 $>$ 50. *Source*: SOEP v37.

Figure 12. Intention-to-Work Coefficients After a Shock

not—there is a very low tendency to seek reemployment overall.

Compared to the control group, younger individuals in the treatment group show an increased intention to return to work after a transitory health shock by about 13 pp. Accordingly, after a transitory shock young individuals seek to return to work almost with certainty. After a persistent health shock, the pattern reverses, with the younger group out of employment indicating a significantly lower intention to reenter employment: The coefficient is at -0.20, that is, a 20 pp drop in the intention to reenter work compared to other young individuals out of employment. For the older age group—irrespective of whether non-employment is due to either type of health shock or any other cause—the intention to work is at a fixed and low level.

5. Discussion

The main results from our analyses can be summarized as follows: First, there are large extensive margin effects after persistent health shocks, and there exists pronounced heterogeneity with respect to age, education, and occupational class. The employment effects are milder for younger individuals, the more-educated, and managers. When we examine the intention to rejoin the labor force after a health-related spell of non-activity, young individuals who experienced transitory health shocks show a higher intention to return to work, whereas young individuals who experienced persistent health shocks show a much lower intention to return to the labor market.

Second, there are intensive margin adjustments after transitory shocks and more pronounced extensive margin effects across all hours categories after persistent shocks, while switches to lower hours categories hardly change.

Third, both on the individual level and on the household level, incomes decrease after a persistent shock, while transitory shocks have no statistically significant effect on household net income. The pass-through of persistent shocks from individual gross labor income to net household income is about 0.5, indicating significant insurance by state and family.

Benchmarking to Related Literature We can benchmark our results to other empirical studies on labor market effects of health shocks. Similar to our approach, studies such as Fadlon and Nielsen (2021), Dobkin et al. (2018) or García-Gómez et al. (2013) estimate reduced forms, each with different measures of adverse health events, finding differing severity of effects on labor market outcomes.[24]

Fadlon and Nielsen (2021) analyze the effects of fatal and non-fatal health shocks such as strokes and heart attacks on household labor supply and income in Denmark. For fatal health shocks, they find that widows increase their labor supply and obtain higher individual earnings after the death of their husbands. For non-fatal health shocks, they find that the labor force participation of sick individuals drops by 12 pp and annual earnings de-

[24] Further studies estimating reduced-form models of adverse health events are Meyer and Mok (2019) and Smith (2004). Also, Jolly and Theodoropoulos (2023) specifically examine spousal labor supply after health shocks. Similar to our result, they find only minimal changes.

crease by around €4,700, that is, an 18% drop from the baseline.[25] Fadlon and Nielsen (2021) do not distinguish between transitory and persistent health shocks, and yet their results for non-fatal shocks can be compared to our shock indicators: The 18% drop in annual earnings is larger than the 8% drop we report for persistent health shocks. However, they focus on a very specific set of adverse health events that is likely to have more severe implications for employment and income on average. Like us, they find no adjustment of partner labor supply or income in the case of non-fatal shocks. Shock pass-through to household net income is 50%, owing in part to the insurance mechanisms and the safety net provided by the Danish tax and transfer system, which is very similar to the German one.[26]

Dobkin et al. (2018) study the economic consequences of hospital admissions by adults aged 50 to 59 in the United States.[27] Three years after hospital admissions, they find a drop in employment of 11 pp and reduced labor earnings of around €9,300, a 24% drop from the baseline.[28] The effect on employment is comparable to our effect size for persistent shocks (7.5 pp overall, 13 pp for the older group), but the effect size for gross labor income is much larger. Dobkin et al. (2018) do not find significant effects on spousal earnings. Further, household net income does not significantly change after the shock, as estimates are imprecise. However, the point estimates for net household income indicate a drop of about €6,900 per year.

[25] Fadlon and Nielsen (2021) estimate a drop of 35,467 Danish crowns, which in September 2021 translates to around €4,700.

[26] The Danish health insurance system is fairly comparable to its German counterpart, although slightly less generous. Health insurance is funded through municipal income taxation at a flat rate, which, to the contributor to the system, is like paying into the German public health insurance system. Employers pay wage continuation for 30 days, just as they do in Germany, and employees with prolonged absences receive sickness benefits for up to 22 weeks thereafter within a given calendar year. Sickness benefits are slightly less generous than in Germany. For further details, see Online Appendix E of Fadlon and Nielsen (2021).

[27] The authors also report results for other age groups, such as individuals aged 60 to 64. However, we choose this age group as it compares well with our definition of the older group.

[28] Dobkin et al. (2018) report reduced earnings of $11,071, which in September 2021 was equivalent to around €9,300.

About 10% of the raw impact of the shock on earnings is compensated for by social security disability insurance payments (€745).[29]

García-Gómez et al. (2013) analyze the effect of acute hospitalizations on employment and labor income in the Netherlands. They find that employment drops by 7 pp two years after the shock, while personal post-tax and transfer income is reduced by €1,000.[30] Similar to us, they also find slightly stronger effects on employment for individuals older than 50 (1 pp more than the average effect). Effects on household net income—a drop of about €1,500—are larger than those on individual labor income because the spouse's probability to stay employed is reduced by 1.5 pp three years after the shock.

The estimated negative employment effects differ in magnitude across the aforementioned studies but cluster in a fairly narrow range (7 to 12 pp). This is remarkable, since both the health shock concepts as well as the institutional contexts differ between all of these studies. Our estimated effect sizes for persistent shocks are very close to those for other European states. Our estimated effects on labor income are in the middle of the range of the estimated effects in the aforementioned studies.

Evaluating the Results Overall, the magnitude of the effects of health shocks depends on country-specific particularities of the social security system. Our findings of large negative effects of health shocks on employment in Germany are worrying, especially against the backdrop of Germany's aging society and its comprehensive social security system. Old-age pensions and other social security benefits are financed through contributions from the actively working population. With large demographic groups such as the baby-boomers nearing or past retirement age, the public pension sys-

[29] Dobkin et al. (2018) report reduced household net income of $8,161, which in September 2021 was equal to around €6,900. The authors find social security disability insurance payments of $881, which amounts to €745. The implied pass-through of the gross income shock to net is 0.73, and thus much larger than in the German or Danish context.

[30] As the average effective tax and contribution rate in the Netherlands is around 38%, we can make a ballpark estimate that the effect on gross income is around €1,600.

tem is facing substantial financial challenges (Feld et al., 2020). To ensure the sustainability of the system, more contributions—whether through a larger workforce, a more productive workforce, or a workforce that retires later—would be crucial (Buslei et al., 2019).

In this respect, our results on employment effects should raise public concern, as people, even in the highly productive age range of 18-50, tend to drop out of the labor force after experiencing a persistent health shock. Additionally, a substantial fraction of these younger individuals do not intend to return to work. These individuals will not participate in the labor market or contribute to the social security system during the most productive phase of their lives. While we document individuals' low intentions to reenter the workforce, the ultimate reasons *why* they do not look for a subsequent job remain unknown. The most obvious reason is a diminished capacity to work. Further, one might suspect that individuals judge their prospects of finding an appropriate and well-paid job to be low. In both cases, there is room for a public policy response. While the former case calls for improved rehabilitation measures, the latter points towards a need for retraining and more efficient matching of individuals recovering from health shocks with jobs that suit their capacity to work (Mehnert et al., 2013; Rick et al., 2012). Our results also suggest substantial potential for improved reintegration into the labor market after recovery from illness. Compared to the large effects on the extensive employment margin, relatively few affected individuals move into part-time arrangements. Hence, at least for many, labor supply seems to be a binary decision: either full-time work or none at all. In contrast, the capacity to work ranges from a complete inability to carry out tasks in the workplace to only minor impairments that slightly limit working time. Thus, the rigidity of the labor market with respect to working schedules may lead to unused productive capacity. Pencavel (2016) reviews the reasons for desired and actual hours mismatch among workers, stating that these mismatches may stem from employers' hours mandates, which in turn reflect firms' price and production environments. One possible reason why firms demand full-time hours is that part-time work implies more start-up or quasi-fixed costs (more office space, transaction costs when sharing tasks), while another issue may be that employers require the joint presence of several inputs (e.g., two skill types of labor) and therefore restrict workers' hours choices (Deardorff and Stafford, 1976). Hence, the binary employment decision we observe for sick individuals may be deeply rooted in the production environment of the firm, which makes the flexibilization of work challenging.

Activating the unused working capacity of the formerly sick could help to mitigate the burden of a skilled-labor shortage and ensure the sustainability of the social security system. However, it may hinge not just on the intention to work on the supply side, but also on the incentives of the demand side to offer working arrangements in line with employees working capacity. Further, an intensified information asymmetry problem between job applicant and potential employer may exist, since it may be difficult for employers to assess applicants who are reentering the labor market after sick leave. It may therefore be prudent for policy makers to consider how they can influence these incentives on the demand side and signaling problems. Exploring whether reactivation policies should focus on the demand or the supply side is a promising avenue for future research.

Our results imply long-lasting earnings penalties for individuals who experience persistent health shocks. Encouragingly, these penalties do not affect household net income one-to-one. The partners labor supply appears not to provide insurance against these penalties, however, as we find no statistically significant effect on partner income. Shock pass-through is about 50% of the raw shock, leaving net income—a prime determinant of household welfare—much less affected than gross income.

6. Conclusion

This paper investigates the impact of transitory and persistent health shocks on labor market outcomes in Germany. To define health shocks, we follow a novel approach, relying on direct measures of health-related restrictions in working capacity: sick days and hospitalizations. We use these measures to derive two novel health shock indicators. We cross-validate this classification with other objective and subjective health measures, finding it to be strongly predictive of bad health, regardless of the alternative measure.

Using this novel classification, we applied an event study analysis to SOEP data covering the time period from 1994 to 2019. Our central finding is that health shocks imply significant negative employment effects: For those affected by transitory shocks, the employment rate temporarily drops by at most 2 pp with respect to the baseline period, while for those affected by a persistent shock, the employment rate drops by about 7.5 pp. Both older and less educated individuals show larger effect sizes. These extensive margin effects of persistent shocks occur across all hours categories, whereas intensive margin transitions are hardly affected. For transitory shocks, we find sizable downward adjustments to less intensive hours categories. Individual labor income decreases by about 8% after persistent health shocks, with no sign of recovering after three years. However, households are partially insured against the resulting income losses: Only half of the income loss passes through to household net income.

In trying to understand the long-lasting employment effects of health shocks, we compared the intention to work of non-employed individuals after transitory and persistent health shocks to that of other non-employed individuals. Remarkably, young individuals who experience a persistent health shock are 20 pp less likely to intend to return to work. This finding points to potential for improved public policies that would support these workers in their return to productive employment

References

Bick, A. and N. Fuchs-Schündeln (2017). Quantifying the disincentive effects of joint taxation on married women's labor supply. *American Economic Review 107*(5), 100–104.

Blundell, R., J. Britton, M. C. Dias, and E. French (2021). The Impact of Health on Labor Supply Near Retirement. *Journal of Human Resources*, 1217–9240.

Blundell, R., M. Graber, and M. Mogstad (2015). Labor income dynamics and the insurance from taxes, transfers, and the family. *Journal of Public Economics 127*, 58–73.

Blundell, R., L. Pistaferri, and I. Preston (2008a). Consumption Inequality and Partial Insurance. *American Economic Review 98*(5), 1887–1921.

Blundell, R., L. Pistaferri, and I. Preston (2008b, 12). Consumption inequality and partial insurance. *American Economic Review 98*(5), 1887–1891.

Bonhomme, S., T. Lamadon, and E. Manresa (2019). A distributional framework for matched employer employee data. *Econometrica 87*(3), 699–739.

Bound, J. (1991). Self-Reported Versus Objective Measures of Health in Retirement Models. *Journal of Human Resources 26*(1), 106–138.

Britton, J. and E. French (2020). Health and Employment amongst Older Workers. *Fiscal Studies 41*(1), 221–250.

Buslei, H., P. Gallego-Granados, J. Geyer, and P. Haan (2019). Rente mit 67: Der arbeitsmarkt für ältere wird entscheidend sein. *DIW Wochenbericht 86*(16/17), 275–283.

Capatina, E. (2015). Life-cycle effects of health risk. *Journal of Monetary Economics 74*, 67–88.

De Nardi, M., G. Fella, M. Knoef, G. Paz-Pardo, and R. Van Ooijen (2021). Family and government insurance: Wage, earnings, and income risks in the netherlands and the us. *Journal of Public Economics 193*, 104327.

Deardorff, A. V. and F. P. Stafford (1976). Compensation of Cooperating Factors. *Econometrica 44*(4), 671–684.

Dobkin, C., A. Finkelstein, R. Kluender, and M. J. Notowidigdo (2018). The Economic Consequences of Hospital Admissions. *American Economic Review 108*(2), 308–352.

DRV (2021). Erwerbsminderungsrenten im Zeitablauf 2021. Technical report, DRV Bund.

Engels, B., J. Geyer, and P. Haan (2017). Pension incentives and early retirement. *Labour Economics 47*, 216–231.

Fadlon, I. and T. H. Nielsen (2021). Family Labor Supply Responses to Severe Health Shocks: Evidence from Danish Administrative Records. *American Economic Journal: Applied Economics 13*(3), 1–30.

Federal Statistical Office (2021a). Genesis-online, 23611-0005: Gesundheitsausgaben: Deutschland, Jahre, Ausgabenträger, Erweiterter Leistungsbereich des Gesundheitswesens.

Federal Statistical Office (2021b). Genesis-online, 81000-0133: VGR des Bundes - Einnahmen und Ausgaben der Sozialversicherung: Deutschland, Jahre, Sozialversicherungszweige, Einnahme- und Ausgabearten.

Feld, L. P., V. Grimm, M. Schnitzer, A. Truger, and V. Wieland (2020). Corona-Krise gemeinsam bewältigen, Resilienz und Wachstum stärken. SVR Jahresgutachten 2020/21.

Friedman, J., T. Hastie, R. Tibshirani, et al. (2017). *The Elements of Statistical Learning*, Volume 1. Springer series in statistics, New York.

García-Gómez, P. (2011). Institutions, health shocks and labour market outcomes across Europe. *Journal of Health Economics 30*(1), 200–213.

García-Gómez, P., H. van Kippersluis, O. O'Donnell, and E. van Doorslaer (2013). Long- term and spillover effects of health shocks on employment and income. *Journal of Human Resources 48*(4), 873–909.

Goebel, J., M. M. Grabka, S. Liebig, M. Kroh, D. Richter, C. Schröder, and J. Schupp (2019). The German Socio-Economic Panel (SOEP). *Jahrbücher für Nationalökonomie und Statistik 239*(2), 345–360.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics 225*(2), 254–277.

Grabka, M. M. (2020). Soep-core v35-codebook for the pequiv file 1984-2018: Cnef variables with extended income information for the soep. Technical report, SOEP Survey Papers.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy 80*(2), 223–255.

Haan, P. and M. Myck (2009, 12). Dynamics of health and labor market risks. *Journal of Health Economics 28*(6), 1116–1125.

Hosseini, R., K. Kopecky, and K. Zhao (2021a). How Important Is Health Inequality for Lifetime Earnings Inequality?

Hosseini, R., K. Kopecky, and K. Zhao (2021b). The evolution of health over the life cycle. *Review of Economic Dynamics*.

Jann, B. (2017, May). KMATCH: Stata module module for multivariate-distance and propensity-score matching, including entropy balancing, inverse probability weighting, (coarsened) exact matching, and regression adjustment.

Jolly, N. A. and N. Theodoropoulos (2023). Health shocks and spousal labor supply: an international perspective. *Journal of Population Economics 36*(2), 973–1004.

Kemptner, D. (2019). Health-related life cycle risks and public insurance. *Journal of Health Economics 65*, 227–245.

Killingsworth, M. R. and J. J. Heckman (1986). Chapter 2 female labor supply: A survey. Volume 1 of *Handbook of Labor Economics*, pp. 103–204. Elsevier.

Kleven, H., C. Landais, and J. E. Søgaard (2019). Children and gender inequality: Evidence from Denmark. *American Economic Journal: Applied Economics 11*(4), 181–209.

LaLumia, S. (2008). The effects of joint taxation of married couples on labor supply and non-wage income. *Journal of Public Economics 92*(7), 1698–1719.

Lindeboom, M. and M. Kerkhofs (2009). Health and Work of the Elderly: Subjective Health Measures, Reporting Errors and Endogeneity in the Relationship between Health and Work. *Journal of Applied Econometrics 24*, 1024–1046.

Mahalanobis, P. C. (1936). On the generalised distance in statistics. In *Proceedings of the National Institute of Science of India*, Volume 12, pp. 49–55.

Marcus, J. (2013). The effect of unemployment on the mental health of spouses - Evidence from plant closures in Germany. *Journal of Health Economics 32*(3), 546–558.

Mehnert, A., A. d. Boer, and M. Feuerstein (2013, 6). Employment challenges for cancer survivors. *Cancer 119*, 2151–2159.

Meyer, B. D. and W. K. Mok (2019). Disability, earnings, income and consumption. *Journal of Public Economics 171*, 51–69.

OECD (2019). *Health at a Glance 2019*. Paris: OECD Publishing.

OECD/European Union (2016). The labour market impacts of ill-health. In *Health at a Glance: Europe 2016: State of Health in the EU Cycle*. Paris: OECD Publishing.

Pencavel, J. (2016). Whose preferences are revealed in hours of work? *Economic Inquiry 54*(1), 9–24.

Rick, O., E. M. Kalusche, T. Dauelsberg, V. König, C. Korsukéwitz, and U. Seifart (2012). Reintegrating Cancer Patients Into the Workplace. *Deutsches Arzteblatt 109*(42), 702–708.

Sarfati, H. (2020). Organisation for Economic Co-operation and Development. The future of work: OECD Employment Outlook. Paris, 2019.

Schröder, C., J. König, A. Fedorets, J. Goebel, M. M. Grabka, H. Lüthen, M. Metzing, F. Schikora, and S. Liebig (2020). The economic research potentials of the german socio-economic panel study. *German Economic Review 21*(3), 335–371.

Schurer, S. (2017). Bouncing back from health shocks: Locus of control and labor supply. *Journal of Economic Behavior and Organization 133*, 1–20.

Smith, J. P. (2004). Unraveling the SES: health connection. *Population and Development Review 30*, 108–132.

Sun, L. and S. Abraham (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics 225*(2), 175–199.

Wanger, S., T. Hartl, and F. Zimmert (2019). IAB-Arbeitszeitrechnung. Technical report, IAB.

Appendix

A. Robustness Exercises

Since we introduce a novel measure of health shocks, we provide a series of robustness checks to demonstrate the advantages and reliability of our approach. These robustness checks are structured in the same order as we proceed in section 2.2. First, we investigate our modeling choices with respect to the *individual deviation condition*, next with respect to the *clustering procedure*, then with respect to the *regression framework*. Finally, we show how our shock classifications relate to objective and subjective measures of health since they are foundational to many analyses of the labor supply effects of health changes.

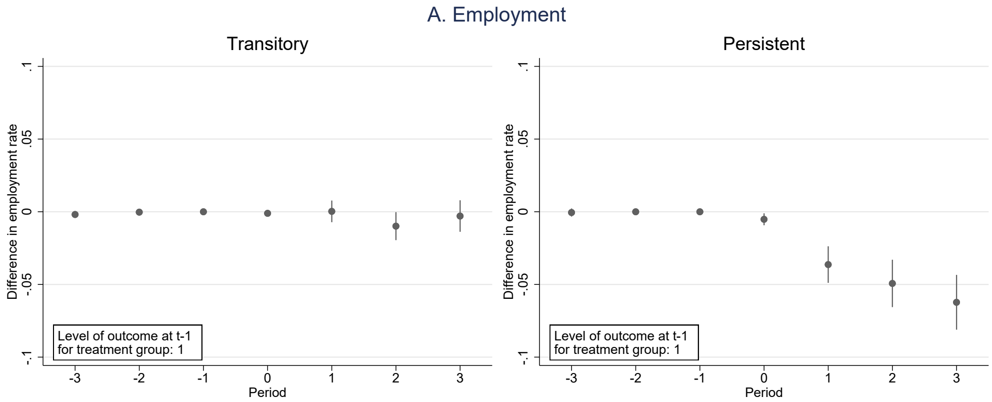
A.1. Robustness: Individual deviation condition

To check whether we obtain significantly different results with respect to the individual deviation condition, we derive the sample of shocked individuals based on the criterion that a deviation of 1 standard deviation from the individual median, instead of 2, has to occur in the period of the shock for either hospital nights or sick days. We show the alternative set of employment results in Figure A.1.

The indication from the figure is that there are no appreciable differences between the two shock criteria; both when considering effect magnitudes from an economic and from a statistical point of view. If anything, effect sizes marginally decrease, as is to be expected when defining health shocks as a smaller deviation from individuals' original health trajectory.

A.2. Robustness: Clustering procedure

The k-means clustering according to sick days and hospital nights might appear sensible because it a) is a data driven procedure that discovers heterogeneity in shock intensity based on the joint distribution of both variables and b) because it uses information on both sick days and hospital



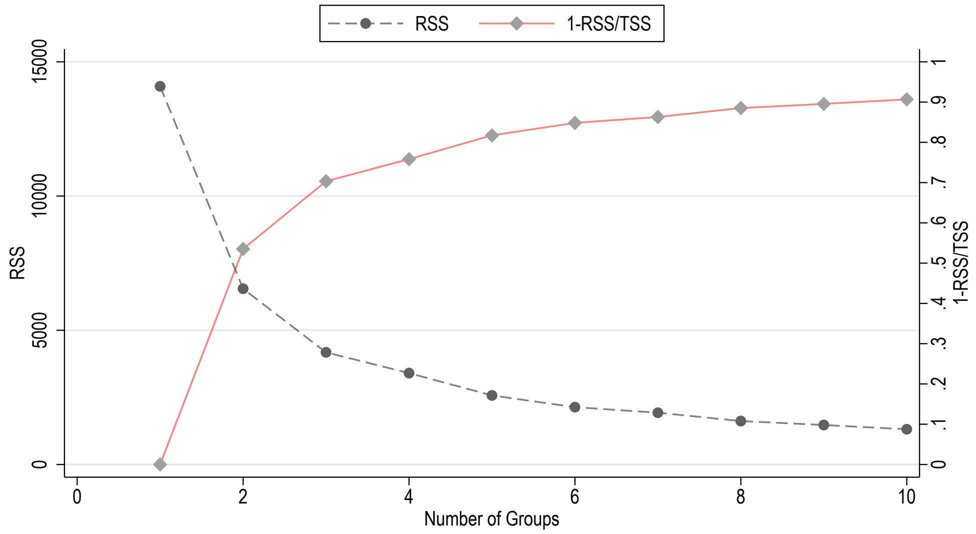
Note: Shows period-treatment-specific coefficients according to Eq.(3) if sick days or hospital nights have to be only one standard deviation above the individual median. Bars give robust 95% confidence intervals of the respective coefficients. Number of observations: Transitory shock: 7,472; persistent shock: 3,778. *Source*: SOEP v37.

Figure A.1. Employment After Health Shocks - 1SD Definition

nights, making the resulting classification based on the more fine-grained information from both distributions.

A very important issue is our choice of the number of groups we define before the clustering takes place. One may argue that although having two groups with the labels persistent and transitory is appealing when viewed from the perspective of the income dynamics literature, it is unappealing because of the large amount of heterogeneity that remains within each of these groups. A higher number of groups may be desirable to better represent the heterogeneity of the joint distribution of sick days and hospital nights. Whether two groups are sufficient can be determined with an elbow plot. The elbow plot shows the number of groups on the x-axis and the residual number of squares (RSS), in this case the average within-cluster sum of squares, on the y-axis. We show an elbow plot in Figure A.2.

The figure shows both the RSS and the ratio of RSS to the total number of squares, which in this case is simply the number of squares when there is only one group (one mean). The figure indicates that the largest absolute and relative decrease in the RSS occurs after introducing the second group. The RSS drops from close to 15,000 to slightly over 6,000, a more than 50% reduction, as can be seen on the right-hand side of the graph. While the

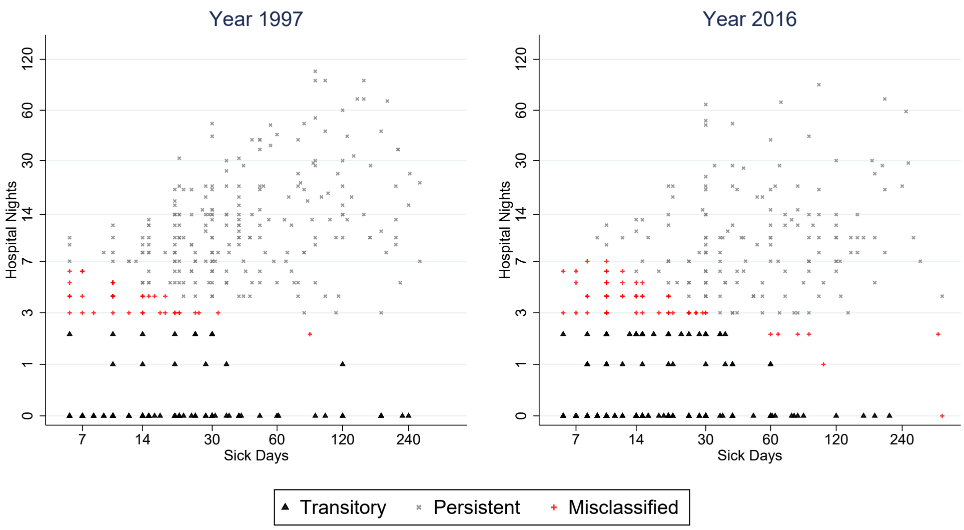


Note: The figure shows the residual sum of squares (RSS) summed for both sick days and hospital nights depending on the number of groups. It also shows an analogue to the R^2, that is, one minus the ratio of residual to total sum of squares (TSS). *Source:* SOEP v37.

Figure A.2. Elbow Plot

introduction of three groups brings about another sizable reduction in the RSS and increase in $1 - \frac{RSS}{TSS}$ ratio, they are not nearly as large as the first change. Past three groups, the subsequent reductions in RSS are fairly minor. Overall, much of the heterogeneity in sick days and hospital nights is well-captured by just two groups.

Another potential criticism of our approach is that the information from one of the dimensions, in particular hospital nights, may already suffice to capture all the relevant information to distinguish persistent from transitory shocks. For example, when health shocks in other papers are defined based on hospitalizations only (e.g., García-Gómez et al., 2013; Dobkin et al., 2018), this is in some sense equivalent to a setting where we cluster solely on hospital nights. To illustrate the implications of an alternative clustering that only considers hospital nights, we rerun our k-means clustering only based on hospital nights and show the observations that are classified differently than in our classification in the main body of the paper. We show an analogous graph to Figure 3 with the misclassification according to our original specification highlighted in Figure A.3.



Note: Shows treated observations after applying the classification only based on hospital nights and the misclassification with respect to the original procedure in 1997 and 2016. Many observations occupy the same points, which we do not represent graphically, that is, the figure is not weighted. *Source*: SOEP v37.

Figure A.3. Groups After Clustering Only on Hospital Nights – 1997 and 2016

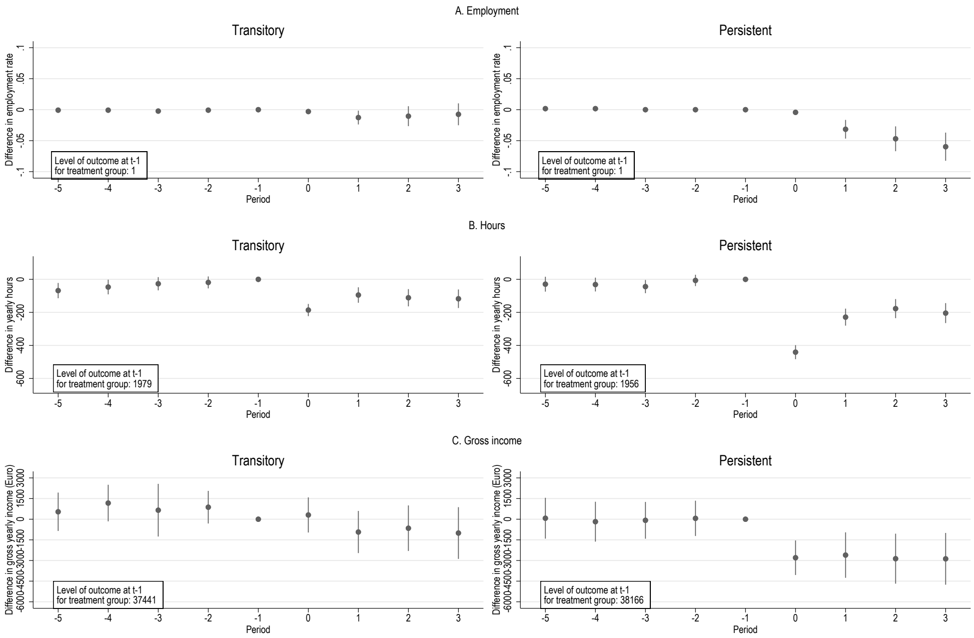
The figure shows that the bulk of observations for both groups still fall into the same clusters but that observations with few sick days and few hospital nights are misclassified. Therefore, many individuals with only minor health-related interruptions, such as smaller surgical interventions with a fast recovery, would be classified as persistently shocked. Additionally, some observations with high numbers of sick days but low numbers of hospital nights would end up classified as transitorily shocked, which is similarly unappealing, as these might indicate severe health conditions.

This exercise illustrates that both distributions contain valuable information about working capacity. By treating both variables as complementary to each other and classifying health shocks based on both distributions, we utilize the maximum amount of available information regarding individuals' working capacity.

A.3. Robustness: Regression framework

The regression framework has to strike an important balance: While it is important to have a sufficiently long pre- and post-shock period to observe the dynamic effects of health shocks, it is also important to consider the size of the sample that can be fit into the framework. We require for observations, both treatment and control, to be uncensored for seven continuous periods in the main analysis, which already limits the number of cases we can feed into the analysis. Extending the pre- or post-shock period leads to additional restrictions of the sample.

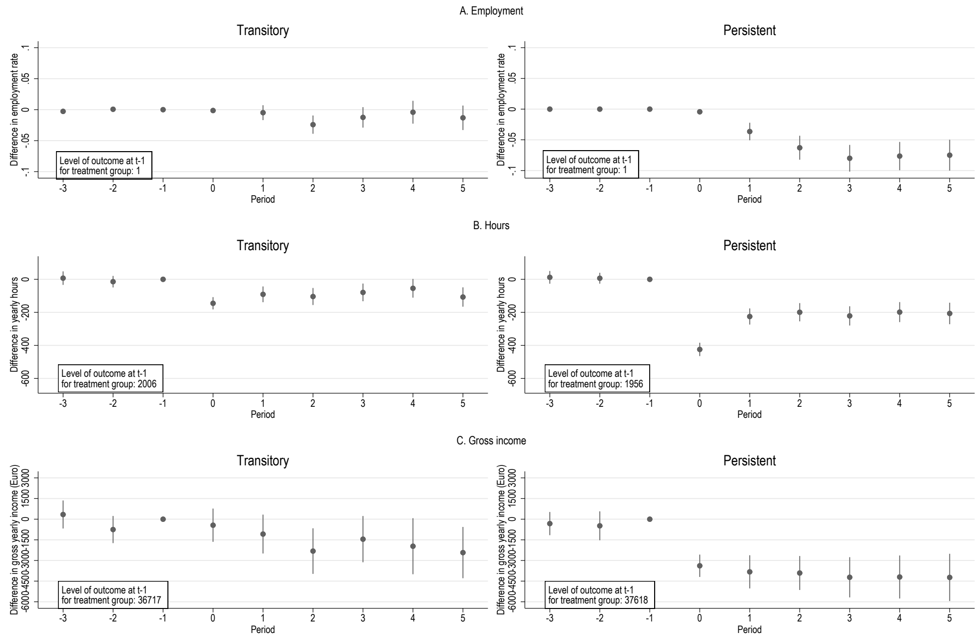
While the main analysis strikes a good balance between observation period and size of the sample, we consider robustness checks in which we a) extend the pre-shock period of the analysis and b) extend the post-shock period of the analysis. We show the event study results in Figures A.4 and A.5.



Note: The figure shows period-specific coefficients according to Eq.(3) for the treated. Bars give robust 95% confidence intervals of the respective coefficients. Number of observations: Transitory shock: 2,672; persistent shock: 2342. *Source*: SOEP v37.

Figure A.4. Main Outcomes: Longer Pre-Shock Period

Figure A.4, in which we extend the pre-shock period by two years, shows very similar results to our main results for both transitory and persistent shocks. Importantly, even though the number of observations per group has dropped by more than 1,000 observations, both the trajectories and effect magnitudes for both types of shocks are very comparable to the main analysis. The pre-trends, except for period -5 of the transitory hours graph, always encompass zero, lending further credence to the parallel trends assumption.



Note: The figure shows period-specific coefficients according to Eq.(3) for the treated. Bars give robust 95% confidence intervals of the respective coefficients. Number of observations: Transitory shock: 2,894; persistent shock: 2,800. *Source*: SOEP v37.

Figure A.5. Main Outcomes: Longer Post-Shock Period

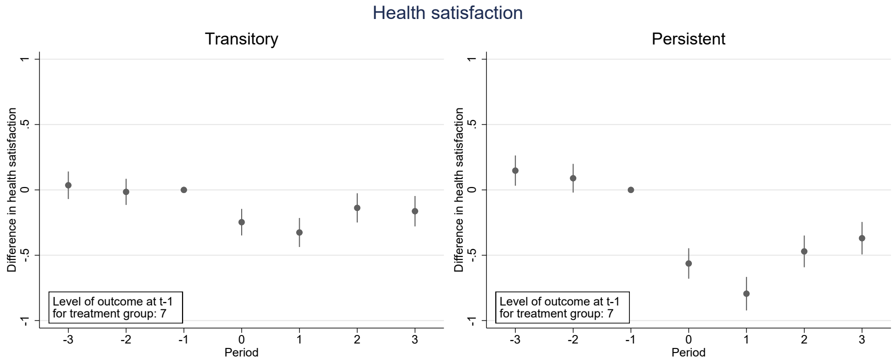
Figure A.5 shows the results with two additional post-shock periods. Similar to the longer pre-trend results, although the number of observations drops by more than 1,000 for both groups, results strongly resemble those in our main analysis. For the transitory shock group, both employment and hours return to pre-shock levels by at least period 4. The results on gross income after a transitory shock may indicate some scarring after the health change, but the results are very noisy. After a persistent shock, the outcomes are very stable: 7 to 8% drop out of employment after a persistent shock and this does not change even after five periods have elapsed. The same holds for the effects on hours and gross income.

A.4. Robustness: Relationship to other measures of health

In many empirical analyses of the labor market effects of health changes, researchers use objective, subjective or a combination of these measures to determine how changes in health drive labor supply, income, and other outcomes (Britton and French, 2020). We, and many other authors, have detailed the strengths and weaknesses of these approaches. While we have explicitly taken a different route in the quantification of the effects of health shocks, our approach still should have a strong relationship to these measures. We therefore offer some analyses of the correlation between our novel health shock measures and these commonly employed ones.

A straightforward way to relate our shock classification to subjective health measures is to use subjective health satisfaction as the outcome in our event study. Figure A.6 shows the effects of both types of shocks on the health satisfaction measure in our data. Health satisfaction drops sharply after either type of shock, but more so after a persistent health shock. After transitory health shocks, health satisfaction almost fully recovers and hardly any difference remains three periods after the shock. Following persistent shocks, we only observe a partial recovery as health satisfaction remains depressed three periods after the shock.

For objective health measures, the SOEP offers disease diagnoses. These disease diagnoses are only surveyed biennially starting in 2011. Thus, we can conduct our event study only on years from 2011 on, which severely restricts our sample. As the outcome variable, we construct disease prevalence dummies from the raw diagnoses, that is, we code dummies that are 1 if the person was diagnosed with an illness in the current period or in any previous period. Put simply, the dummy never switches back from 1 to 0 if it was ever 1. Further, to fit the biennial survey of the diagnoses to our



Note: The figure shows period-specific coefficients according to Eq.(3) for the treated. Bars give robust 95% confidence intervals of the respective coefficients. Number of observations: Transitory shock: 4,098; persistent shock: 3,778. *Source*: SOEP v37.

Figure A.6. Validation of Shock Measures: Health Satisfaction

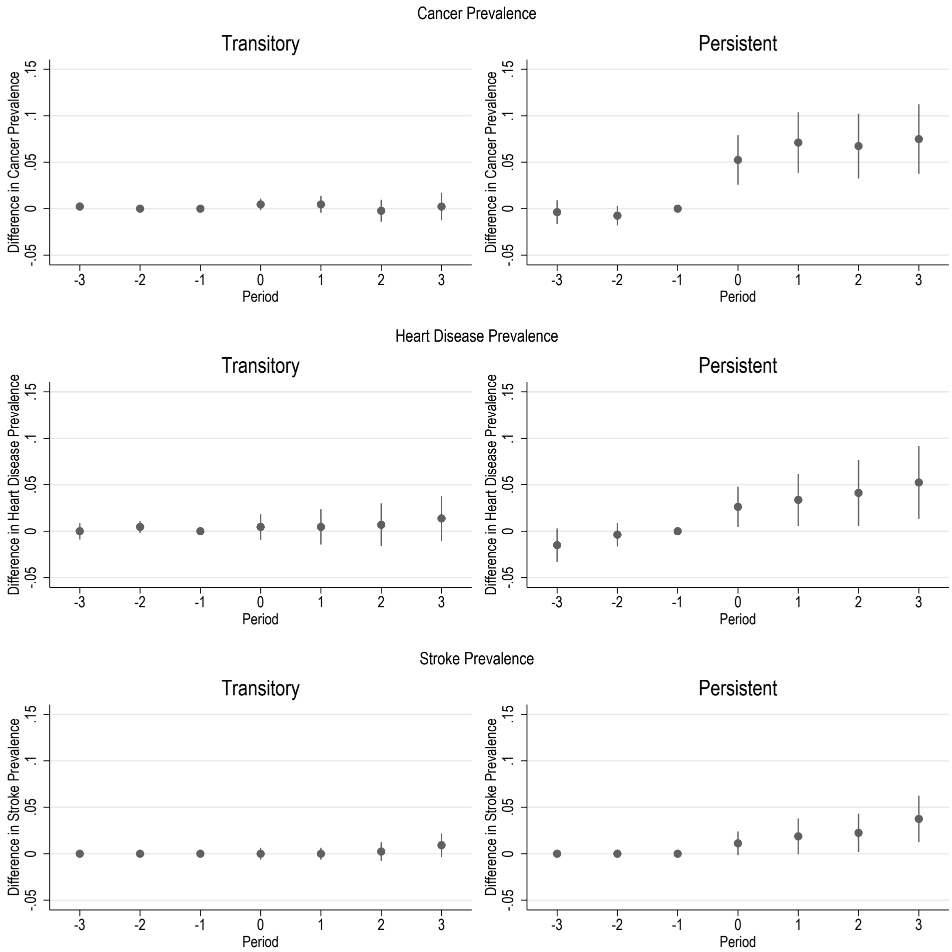
yearly event study, we fill values from the previous year for years in which the outcome was not part of the questionnaire.

Despite all these limitations, we show in Figure A.7 that our shock definitions align very well with chronic disease prevalence. For cancer, heart disease, and stroke prevalence, we find no significant increase after a transitory shock. By contrast, after persistent shocks, the prevalence of all these diseases significantly increases.

A.5. Accounting for Heterogeneous Dynamic Treatment Effects

The dynamic treatment effect literature has shown the importance of accounting for heterogeneous treatment effects along cohorts in event study designs (Sun and Abraham, 2021). Cohorts are groups defined by the calendar year that corresponds to the first relative time period with respect to treatment. Average treatment effects on the treated in a simple two-way fixed effects design are influenced by the proportion of each cohort in the dataset and, thus, may not give the true overall ATT. Rather, the ATT from two-way fixed effects may correspond to some other linear combination of the cohort-specific ATTs.

Sun and Abraham (2021) propose an estimator to address this problem, which we use to check the robustness of our results. The `eventstudyinteract`



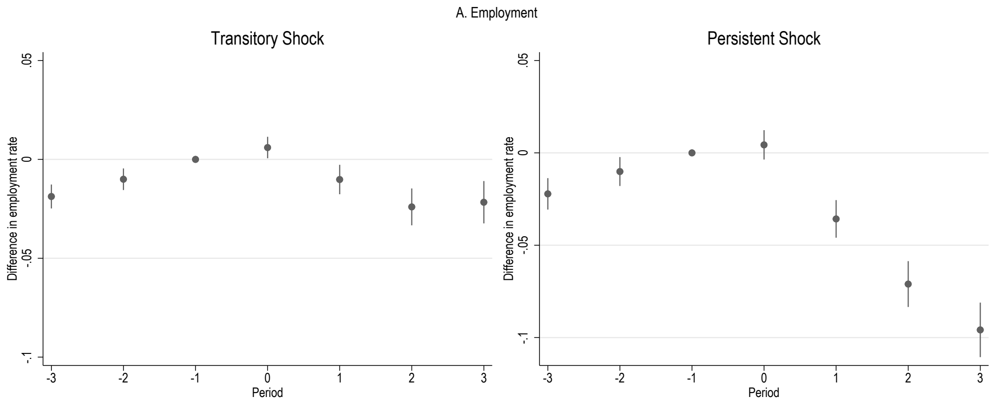
Note: The figure shows period-specific coefficients according to Eq.(3) for the treated. Bars give robust 95% confidence intervals of the respective coefficients. Number of observations: Transitory shock: 872; persistent shock: 534. *Source*: SOEP v37.

Figure A.7. Chronic Disease Prevalence

package for Stata provided by Sun enables us to implement the analysis.[31] The package produces estimates of the differences in outcomes along relative time periods compared to never-treated control units. We show these treatment effects with respect to employment in Figure A.8.

Both figures show qualitatively roughly equivalent trends to those shown in Figure 7, although they indicate a more pronounced pre-trend compared

[31] The package is available at https://economics.mit.edu/grad/lsun20/stata.



Note: Shows period-specific treatment effects estimated using the `eventstudyinteract`-package. Bars give robust 95% confidence intervals. Number of observations: Transitory shock: 4,098; persistent shock: 3,778. *Source*: SOEP v37.

Figure A.8. Employment Treatment Effects—Sun-Abraham Estimator

to our main results. In this specification, we are not able to fully implement our weighting based on the Mahalanobis distances computed in our matching procedure as the Sun-Abraham estimator implements its own weighting routine. Accordingly, it is to be expected that pre-trends do not fully align. However, the pre-trends follow a positive trend, indicating, if anything, a growing attachment to the labor market pre-shock. Effect sizes after the shocks are also of similar magnitude—for persistent shocks even somewhat stronger than in the main specification. We conclude that our main estimates do not suffer from substantial bias due to heterogeneity in dynamic treatment effects.

B. Institutional Background

In Germany, a system of institutions and regulations alleviates the negative effects of health problems. First, German employees enjoy broad employment protection stemming from the Unfair Dismissal Act (*Kündigungsschutzgesetz*). While this law does not guarantee employees full protection from termination due to illness, it does stipulate a number of conditions that make it difficult for employers to fire employees for health reasons. Further, under German law, employees have advocates in their place of employment in the form of work councils, which receive notice of all planned terminations and review these decisions. In practice, these measures lead to strong protection. The OECD reports that German employment protection ranked seventh on the employment protection index in 2019 among member countries (Sarfati, 2020).

Second, employees are entitled to employer-paid sick leave, which covers 100 percent of an employee's salary for up to six weeks.[32] Past this period of full replacement, employees become eligible for sickness benefits—paid by the public health insurance system[33]—for the duration of up to 78 weeks including the previous six weeks.[34] Sickness benefits do not provide employees with 100% of their regular salary. Generally, they cover 70% of regular gross income but not more than 90% of net income.[35] Once an employee's illness exceeds the 78-week threshold, their main option to receive further benefits is to apply for a partial or full reduction of earnings capacity with the public pension insurance. If individuals are not able to work in any job for at least three hours per day, they are granted a full-rate reduced earnings capacity pension. If individuals are able to work at least three but less than six hours per day, they are granted the half-rate reduced earnings capacity pension, which also permits them to work in a part-time job while receiving the pension. The approval of these pensions is based on assessments by physicians. The amount of the reduced earnings capacity pension depends on how much the individual has paid into the system so far and the pension value of the German public pension system ("Rentenwert"). In 2019, the average reduced earnings capacity pension was €835 before taxes (DRV, 2021). Especially for younger individuals who have not yet contributed substantially to the public pension system, the reduced earnings capacity pension will be very low, in some cases even lower than

[32] Between 1996 and 1999, this regulation changed, and sick pay was reduced to 80% of regular salary for those employees who were not protected by a collective bargaining agreement.

[33] Private health insurance providers pay similar amounts, but these contracts are opt-in.

[34] These 78 weeks are counted cumulatively within a period of three years.

[35] For high-income earners, the benefit is capped at 70% of the income ceiling for health insurance contributions, which was €4,537.50 in 2019.

the subsistence minimum defined by social assistance, which amounted to roughly €424 plus rent and heating assistance in 2019 for a single person. In this case, individuals are eligible for additional transfers until they reach the subsistence minimum.

Third, in relation to medical expenses, as German employees have been required since 2009 to have health insurance and, before that, were generally insured by a public health insurance provider, Germans usually do not have to pay out-of-pocket medical expenses to an extent comparable to the United States (Dobkin et al., 2018). In Germany, out-of-pocket medical expenses only occur under special circumstances when patients receive special treatment, such as single-patient rooms or treatment by the chief physician, and additional health services, such as orthodontic treatments and optometry. Finally, health care prices in Germany are slightly below the OECD average, while the United States ranks eighth among member states (OECD, 2019).

Over the duration of our sample, several reforms of the German health care system were introduced. Generally, these reforms were intended to reduce the expenditures of the system. In 1993, the "Gesundheitsstrukturgesetz" was introduced to allow Germans to freely choose between their health insurance providers. Further, copays for pharmaceuticals and hospitalizations were increased. In 1996, the "Beitragsentlastungsgesetz" additionally raised copays for pharmaceuticals and cut coverage of some health-related products like eyeglass frames. The "GKV-Neuordnungsgesetz" from 1997 lowered the replacement rate of sickness benefits from 80% of gross but not more than 100% of net earnings to 70% and 90%, respectively. In 2002, the "Beitragssatzsicherungsgesetz" lowered the flat rates for doctors, clinicians, and hospitalizations, leading to earlier discharge after hospitalization. In 2007, the "GKV-Wettbewerbsstärkungsgesetz" introduced compulsory health insurance for all Germans and established basic insurance contracts that have to be offered regardless of preexisting conditions. In 2011, the "Gesetz zur Neuordnung des Arzneimittelmarktes" slightly increased the contribution rates for public health insurance providers, and the health insurance providers were given more power in bargaining for lower pharmaceutical prices. Overall, these reforms have gradually reduced the generosity of the German health care and insurance system.

In summary, the German public health insurance system covers medical expenses almost completely in stark contrast to the United States. However, job and earnings losses are only partially insured by the employer and the government, making labor market outcomes the relevant variables to study in the German context.

C. SOEP Questions

150. What about hospital stays in the last year - were you admitted to a hospital for at least one night in 2017?

Yes ☐ No ☐ ➡ *Question **152!***

151. How many nights total did you spend in the hospital last year, that is, in 2017?

☐ nights

And how often did you have to go to the hospital in the year 2017?

☐ times

152. Were you on sick leave from work for more than 6 weeks at one time last year?

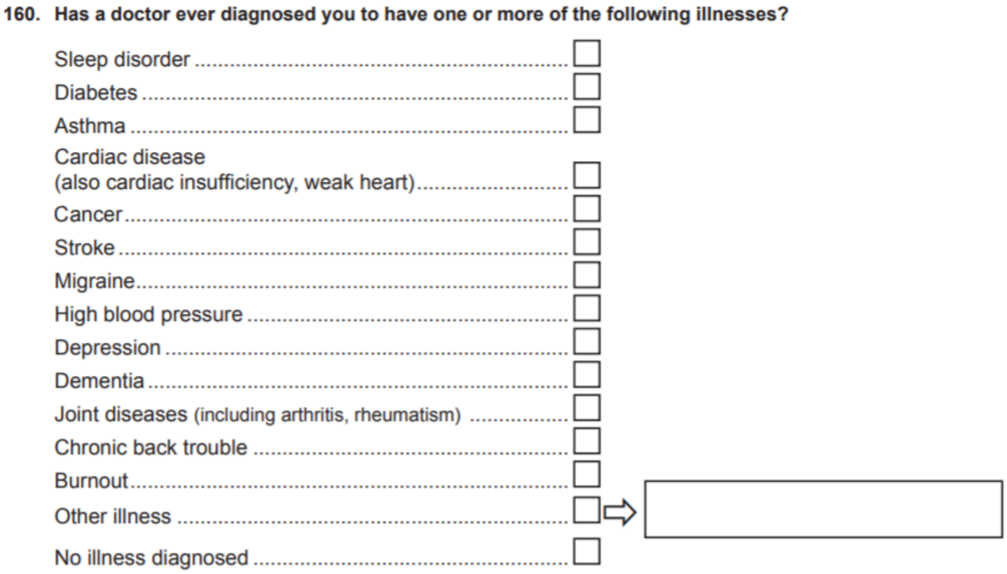
Yes, once ☐
Yes, several times ☐
No ☐
Does not apply, I was not employed in 2017 ☐ ➡ *Question **155!***

153. How many days were you unable to work in 2017 due to illness?

☞ *Please state the total number of days, not just the number of days for which you had an official note from your doctor.*

None ☐
A total of ☐ days

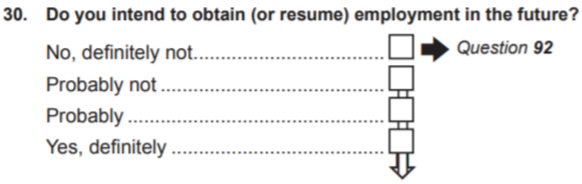
Figure C.9. Health Questions in the SOEP



160. Has a doctor ever diagnosed you to have one or more of the following illnesses?

- Sleep disorder ☐
- Diabetes ☐
- Asthma ☐
- Cardiac disease (also cardiac insufficiency, weak heart) ☐
- Cancer ☐
- Stroke ☐
- Migraine ☐
- High blood pressure ☐
- Depression ☐
- Dementia ☐
- Joint diseases (including arthritis, rheumatism) ☐
- Chronic back trouble ☐
- Burnout ☐
- Other illness ☐ ⇨
- No illness diagnosed ☐

Figure C.10. Questions on Chronic Illnesses in the SOEP

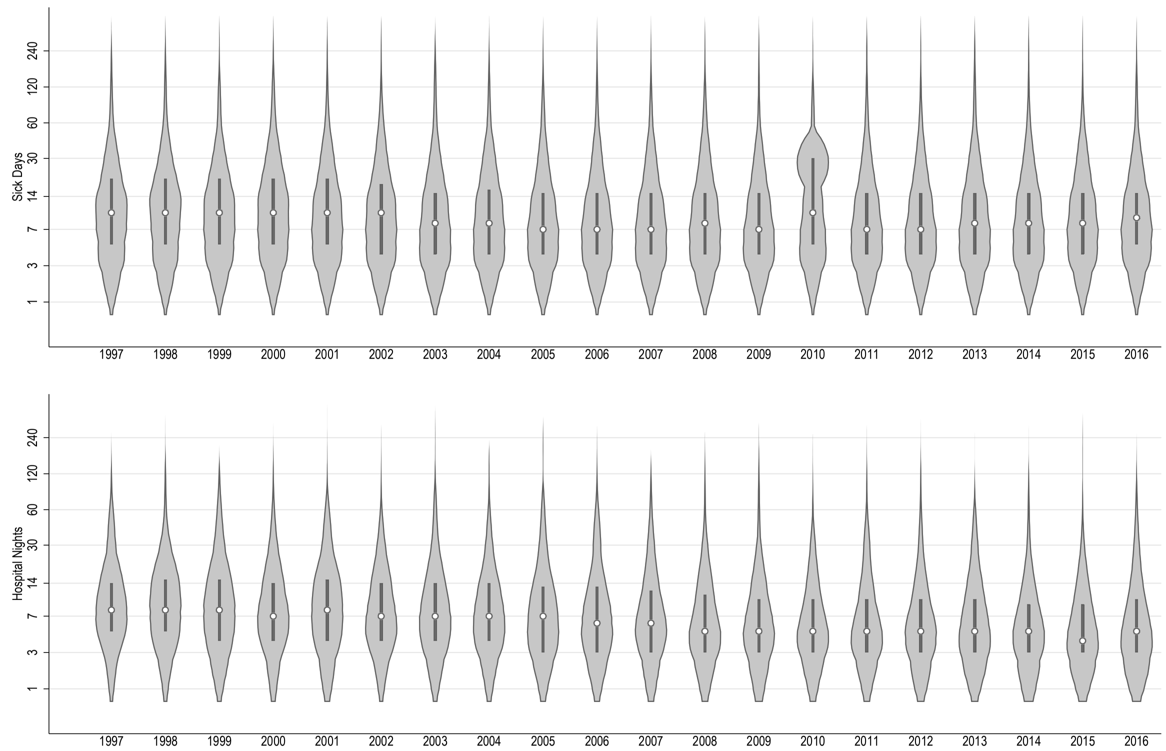


30. Do you intend to obtain (or resume) employment in the future?

- No, definitely not ☐ ➡ *Question* ***92***
- Probably not ☐
- Probably ☐
- Yes, definitely ☐

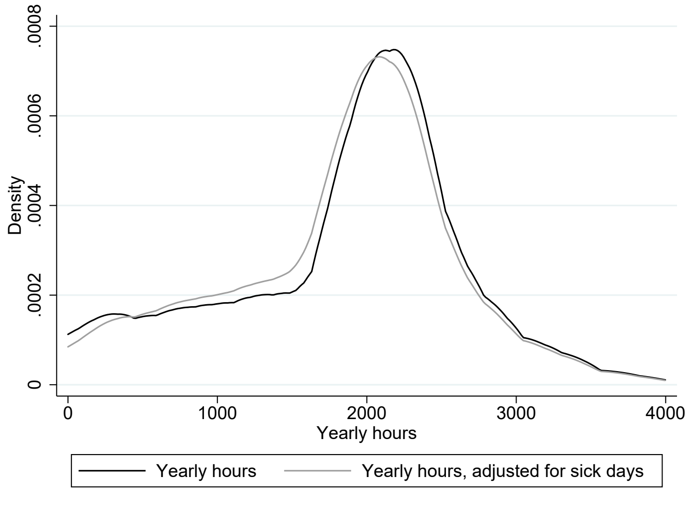
Figure C.11. Willingness to Work in the SOEP

D. Additional Tables and Figures



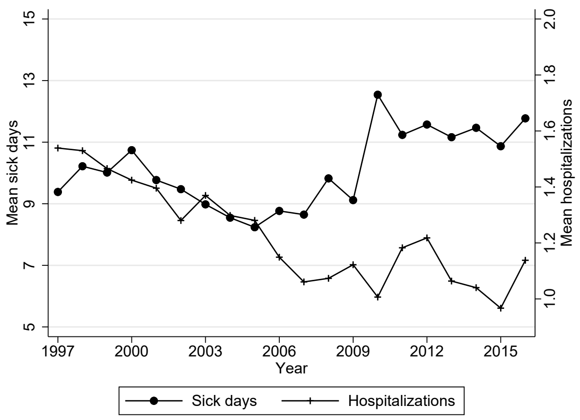
Note: Displayed are yearly violin plots of sick days and hospitalizations on a log scale. The point in the middle gives the median, while the bar shows the inter-quartile range. The shaded area indicates a density estimate. We restrict to the population with positive sick days or hospital nights. Observations are weighted using cross-sectional weights. *Source:* SOEP v37.

Figure D.1. Violin Plots of Sick Days and Hospitalizations



Note: Own calculations based on SOEP v37. The figure shows kernel densities for yearly hours for the working population.

Figure D.2. Pooled Distribution of Yearly Hours: Adjusted vs. Unadjusted



Note: Authors' calculations using SOEP v37. Based on working age population (18-65).

Figure D.3. Means of Sick Days and Hospitalizations