aging·health·labor
CHRISTIAN DOPPLER LABORATORY

How Physicians Affect Patients' Employment Outcomes Through Deciding on Sick Leave Durations

by

Alexander AHAMMER

Working Paper No. 1604

June 2016

Corresponding author: alexander.ahammer@jku.at

Christian Doppler Laboratory
Aging, Health and the Labor Market
cdecon.jku.at

Johannes Kepler University
Department of Economics
Altenberger Strasse 69
4040 Linz, Austria

Physicians Affect Patients' Employment Outcomes Through Deciding on Sick Leave Durations[*]

ALEXANDER AHAMMER

Department of Economics, Johannes Kepler University Linz
Christian Doppler Laboratory Aging, Health, and the Labor Market

December 7, 2016

Abstract

I analyze how general practitioners (GPs) indirectly affect their patients' employment outcomes by deciding on the length of sickness absences. I use an instrumental variables framework where spell durations are identified through supply-side certification measures. I find that a day of sick leave which is only certified because a worker's GP has a high propensity to certify sick leaves decreases employment probabilities persistently by 0.45–0.69 percentage points. Conversely, the risk of becoming unemployed increases by 0.28–0.44 percentage points due to the additional day of sick leave. These effects are mostly driven by men with comparably low job tenure and migratory background. Several robustness checks show that identification is not impaired by endogenous matching between patients and GPs. My results bear important implications for doctors: Whenever medically justifiable, it may be beneficial to certify shorter sick leaves in order to protect employment status of the patient.

JEL Classification: I10, J21, J60
Keywords: Sick leave duration, employment, general practitioners, supply-variation.

[*]*Correspondence:* Alexander Ahammer, Department of Economics, Johannes Kepler University Linz, Altenberger Straße 69, 4040 Linz, ph. +43/7312/2468-7370, email: alexander.ahammer@jku.at. I thank René Böheim, Peter Egger, Martin Halla, Michael Lechner, Gerald Pruckner, Nicole Schneeweis, Rudolf Winter-Ebmer, Anna Wurm, seminar participants in Innsbruck, Linz, at the 2016 Labor Seminar in St. Anton, and the WUWAETRIX-IV in Vienna, as well as conference participants at the 2016 EEA/ESEM in Geneva, the 2016 EuHEA in Hamburg, the 2016 ESPE in Berlin, and the NOeG-SEA 2016 in Bratislava for numerous helpful discussions and valuable comments. Furthermore, I am indebted to Tom Schober for providing parts of the data. Financial support from the Christian Doppler Laboratory Aging, Health, and the Labor Market is gratefully acknowledged. The usual disclaimer applies.

1 Introduction

Today, sick leaves are implemented in most industrialized countries around the globe as an institution that allows workers to recover from medical conditions without losing pay while being off work. Instead of protecting employment, however, a higher sick leave take-up rate has in fact been found to induce both unemployment and wage reductions (Andersen, 2010; Hansen, 2000; Markussen, 2012). Two mechanisms might explain this phenomenon: Either workers are penalized by their employers for being off work, or the sickness absence itself prevents the worker from engaging in regular activity, thereby entailing negative health effects which lead to lower productivity and employability later. The latter point has been raised by Markussen (2012) citing recent findings from the medical literature. While the association between sick leave take-up rate within a given time horizon and labor market outcomes seems to be well-understood in the empirical literature, effects of variations in the length of individual sick leaves have received little attention so far. The latter is indeed the more obvious and immediate decision variable for general practitioners (GPs) in day-to-day medical care.

Establishing a causal link between sick leave duration and labor market outcomes is complicated due to the existence of omitted variables: Effort and (job) motivation are important determinants of both spell durations and employment outcomes, neglecting such variables may cause serious bias in empirical analyses. Furthermore, patients may convince doctors to grant longer sick leaves if they expect to be laid-off soon, thus causality may simultaneously run in both directions. In order to account for these issues, I use the prescription behavior of Upper Austrian GPs as an instrumental variable for the duration of sick leave spells they certify. Doctors, even when holding health status of the patient fixed, differ substantially with regard to their prescription behavior, both across and within geographic regions (Aakvik *et al.*, 2010; Grytten and Sørensen, 2003; Phelps *et al.*, 1994). The reason is that physicians simply differ in their beliefs about the necessity and efficiency of different treatments, and both medical as well as legal leeway allows them to adjust their prescription behavior accordingly.

Using this supply-side variation as an instrumental variable ultimately yields a local average treatment effect (LATE; Angrist and Imbens, 1995; Imbens and Angrist, 1994) which captures precisely the effect of a *marginal* day of sick leave on labor market outcomes. Here the term marginal is used to describe a day of sick leave which is only certified because a worker's GP has an above-average certification propensity. Essentially, I therefore compare two identical workers who are equally sick but consult different doctors. While one doctor grants, for example, four days of sick leave, the other grants five days – in this case, the LATE captures exactly the effect of that one single day of sick leave on the worker's subsequent labor market status.

Consequently, my research question abstracts fundamentally from the vast literature on *absenteeism*, which terms work absence behavior despite being healthy. Absenteeism may in large part stem from moral hazard problems, thus having important welfare implications for policy makers on its own.[1] In the present paper, however, the decision whether to stay home for an additional day is *not* taken by the worker, but rather by the doctor who grants the sick leave. Isolating this particular channel is a consequence of the inherent mechanics my empirical strategy builds on: Embedded into the LATE framework, the effect of sick leave duration on employment is identified through workers whose spell length is only extended because they consult a doctor who has a high certification propensity, but not in the counterfactual scenario in which they consult a doctor with a low certification propensity.

Principal-agent problems – in the sense that patients and GPs may bargain over the length of a sick leave (see e.g., Nilsen *et al.*, 2011) do not pose problems for my empirical analysis. Estimated from a multilevel fixed-effects model, my indicator of prescription behavior is orthogonal to both time-variant and time-invariant patient characteristics. In the 2SLS framework, the LATE captures then the effect of a day of sick leave which is caused only be the variation in GPs' prescription behaviors, while patient-side bargained days of sick leave are out of the equation. Furthermore, estimated effects do not just reflect different characteristics of the patient population (for example along demographic or socio-economic lines), because these are controlled for and not captured by the LATE anyway.

Generally, the effect of spell duration on employment is *a priori* undetermined. Two mecha-

[1] See Brown and Sessions (1996) for a comprehensive survey on both the theoretical as well as the empirical literature on absenteeism. Recent contributions using Austrian data include Böheim and Leoni (2014) and Halla *et al.* (2015). Halla *et al.* analyze how the distribution of worker, firm, and government shares of sick leave remunerations affect absence behavior in Upper Austria. Their findings are somehow mixed: While increasing workers' and firms' cost shares seems to reduce subsequent health cost by a substantial amount, a higher government share contributes surprisingly little to health outcomes. Böheim and Leoni (2014) exploit a particular discontinuity in the social security system which determined whether firms had to pay a deductible for sick leaves of blue-collar workers. Moral hazard induced by this deductible is found to have no effect on sickness absences of blue-collar workers both on the extensive and intensive margin.

nisms could be triggered: (1) the marginal day of sick leave allows workers to regenerate better and longer (e.g., by reducing the stress level or engaging in rehabilitation activities) which increases their employability, or (2) the employer uses sick leaves as a screening device and perceives the longer absence either as a signal of absenteeism or as a persistent loss in productivity and penalizes the worker. The effect on productivity itself is *a priori* undetermined as well: While it may increase due to the longer regeneration period, workers are off the job at the same time and possibly loose touch with their colleagues or miss out on new developments related to work tasks.

I contribute to the literature in three important ways: First, my instrumental variables framework allows me to isolate the impact physicians have on workers' employment outcomes through their decisions on the length of sick leave spells. To my knowledge, this particular channel has not been explored thus far. Second, I deviate from the existing literature by using durations of individual sick leave spells as my main explanatory variable, rather than analyzing aggregate sickness absence measures. As argued above, this is clearly the more relevant decision variable for GPs and thereby entails more practicable policy recommendations. Third, I suggest a new robustness check that utilizes random patient-GP matches to show that my analysis is not impaired by endogenous sorting of patients to GPs. By (1) conditioning on a large set of covariates and incorporating worker-level fixed-effects in my regressions, and (2) providing an array of sensitivity analyses, I argue that remaining biases shrouding causal effects should be negligible. On a side note, most of the existing evidence linking sick leaves to labor market outcomes stems from Scandinavian countries. Although Austria has a similar social security system and economic structure in general, it is still important to consider different countries as well in order to gain a more comprehensive picture.

Using social security data and health records from Upper Austria, I find that a marginal day of sick leave decreases employment probabilities persistently during the first 18 months after the end of the sick leave by 0.45–0.69 percentage points. The risk of becoming unemployed due to the additional absence day is between 0.28 percentage points and 0.44 percentage points, but this effect approaches zero comparably quicker with significant effects being found almost exclusively during the first six months after the sick leave. Stratifying the population into different subsamples, I find that these effects are largely driven by men with low job tenure and migratory background.

These results are valid only if mobility between patients and GPs is conditionally exogenous. Within the course of my sensitivity analyses, I estimate employment probabilities for subsamples of the population where either mobility is restricted *a priori*, moves to new GPs can be assumed to be exogenous, or where patient-GP sorting is random by nature. On weekends, for instance, GPs rotate to provide emergency care, assignment between patients and doctors is then more or less random and depends merely on the rotation schedule. Additional robustness checks include restricting the sample to areas with low competition amongst GPs, to smaller towns with less than 18,705 inhabitants, to patient-GP matches where the geographical distance is less than 10 kilometers, to moves to new GPs where the zip code of the patient changes as well, and finally to patients who never change GPs during the observation period. Effects are robust to all of these sample restrictions.

My findings have important implications for policy makers, and more importantly, for doctors. In line with the existing literature, I show that each additional day of sick leave is in fact detrimental in terms of patients' employment outcomes, and a large part of this negative effect can be explained by high certification propensity doctors granting longer sick leaves. In case of doubt, doctors should therefore certify shorter sick leaves whenever possible in order to protect employment status of their patients.

1.1 Review of the Literature

The association between sick leave take-up and labor market outcomes has increasingly been gaining attention from both labor and health economists in recent years, originating mainly in Scandinavia. Using Norwegian administrative data, Markussen (2012), for instance, finds that a one percentage point increase in a worker's take-up rate is associated with a 0.5 percentage point reduction in the probability of being employed, and a 1.2% reduction in earnings two years later. Similar to this paper, identification of the take-up rate is based on propensity-to-prescribe measures estimated from a competing risks survival model. In order to address the problem of endogenous sorting between patients and doctors, Markussen estimates his model on subsamples of the population where mobility is either restricted or as good as random. In particular, he considers (1) only patients that did not change their GP during the observation period, and (2) patients who move to a new GP because their old one retired, arguing that allocation to the new physician is then more or less random. Effects are robust to both sample restrictions.

Studies focusing on wage outcomes include Hansen (2000) and Andersen (2010). Swedish workers are covered by a national health insurance reimbursing their earnings while being sick. Hansen (2000) the effects of a reform in 1991 which led to a substantial reduction in the replacement rate. Using pre- and post-reform indicators as instrumental variables, Hansen finds negative wage effects due to an increase in the sick leave take-up rate, but only for women. Andersen (2010) exploits a similar policy reform in Denmark which changed the reimbursement scheme for sick leaves, placing an additional financial burden on municipalities which, as argued by Andersen, should provide incentives for them to speed up case work for workers on sick leave. Andersen finds that a one-month increase in aggregated sick leaves reduces wages up to two years later by 4.4%–5.5%, which is a rather small yet statistically significant effect.

To my knowledge, there is only one study which takes the length of individual sick leave spells explicitly into account: Hesselius (2007) splits sick leave durations into short (1–7 days), medium (8–28 days), and long (more than 28 days) spells and analyzes how the number of sick leaves taken in each of these categories affects unemployment risk. Using Cox proportional hazards models, he finds that unemployment risk increases monotonically with each further day off-work between those three categories, where effects are more pronounced for women. Although controlling for a rich set of covariates, unobserved heterogeneity may still induce bias in Hesselius' estimates. Similar correlations in terms of unemployment are reported by Amilon and Wallette (2009).

1.2 Institutional Background

Austria has a Bismarckian welfare system with almost universal health care access. Social pension, health, and work accident insurance are covered by a total of 22 social insurance institutions organized through an umbrella organization called *"Main Association of Austrian Social Security Organisations"*. Once employed, workers are automatically insured at one of these 22 institutions depending on their industry affiliation, their place of residence, and whether they are employed in the private or in the public sector. In this paper, I focus on employees insured at the *Upper Austrian Sickness Fund*, which covers around one million members representing roughly 75 percent of the population in Upper Austria, one of the nine Austrian provinces.

Sick leave insurance in Austria is designed to compensate workers for lost earnings due to both occupational and non-occupational diseases. Depending on their job tenure, employees receive full salary during the first six weeks (for workers with less than five years of tenure) to twelve weeks (for workers with more than 26 years of tenure). After this period of full reimbursement, workers receive half their salary for another six to twelve weeks (again, depending on job tenure) and then one quarter of the full salary for another four weeks (Federal Ministry of Social Affairs, 2014).

Sick employees are obliged to inform their employer as soon as they become incapacitated for work. In most cases, sickness certificates are issued by general practitioners, who act as gatekeepers in the Austrian health care system. Hospitals or specialists certify sick leaves only in rare circumstances. The certificate itself contains mainly the starting date of the sick leave as well as its expected duration as declared by the GP. The latter is only binding in one way, meaning that the actual absence must not exceed the recommended duration, but may fall short of it in case the employee decides to return to work earlier. If this is the case, the firm has to notify the insurance fund immediately. Sickness certificates do not reveal a specific diagnosis, as law does not grant employers the right to learn about diagnoses.

One particularity in the Austrian system is that, by law, no certificate is required for absences of less than three days, unless the firm explicitly requires it. This induces measurement error in my estimations, because I do not observe very short sick leave spells for some firms in the data. As long as a firm's personnel planning is unrelated to its decision whether to request certificates for short sick leaves or not, however, estimations should not be affected by this kind of sample truncation.

In principal, employment contracts can be terminated at any point of time by either the employer and the employee observing the period of notice. In case both parties reach a consensual agreement about the termination, the contract ends according to the agreement. Whenever no agreement is reached, however, there is a cancellation period which usually lasts one month. Certain groups are protected against dismissal by law, most notably apprentices or workers who go on maternity leave. In my empirical analysis, I decided to drop them altogether. It is important to stress that workers are *not* protected against dismissal whilst being on sickness absence.

2 Methodology

Let $S_k = [t_{-n_k}, t_0]$ denote a sick leave spell, where S_k is a finite interval with cardinality n_k (hence, n_k is the actual duration of sick leave k) and let $E_k = [t_0, t_{e_k}]$ be the total remaining employment spell after t_0 with cardinality e_k. Each worker $i = 1, \ldots, N$ in the sample may have $k = 1, \ldots, K_i$ different non-intersecting sick leaves ordered by t_{-n_k}, followed by another k different remaining employment spells. To account for the time dimension of the employment outcome, I partition the first M months of E_k, say $\tilde{E}_k = [t_0, t_M]$, into $m = 1, \ldots, M$ disjoint subintervals, each of equal length. In order to ease notation, I denote each subinterval by its endpoint, for example, $t_m = (t_{m-1}, t_m]$ for some $t_m \subset \tilde{E}_k$. Although I do not observe whether workers are being laid-off or terminate their contract themselves, I do observe whether the subsequent spell following E_k is an unemployment spell or another employment spell at a different firm (i.e., a firm-to-firm transition). Retirees or workers who go on maternity leave (thereby enjoying protection against dismissal) are dropped from the sample.

Define M binary outcomes equal to unity if i is still employed at the end of t_m for each spell k of observation i, and define another set of M binary outcomes indicating whether i became unemployed between t_0 and t_m, $m = 1, \ldots, M$. Let $M = 24$ and each interval span thirty days (that is, I analyze employment status up until two years after the end of the sick leave spell). Formally,

$$y^e_{ikm} \equiv \mathbb{P}[i \text{ is still employed at the end of } t_m], \qquad m = 1, \ldots, 24 \tag{1}$$

$$y^u_{ikm} \equiv \mathbb{P}[i \text{ became unemployed between } t_0 \text{ and } t_m], \qquad m = 1, \ldots, 24 \tag{2}$$

Consider the following linear two-stage regression model:

$$\begin{aligned} y_{ikm} &= \rho_m \hat{n}_{ik} + \mathbf{x}'_{ik}\mathbf{\Theta}_m + \omega_i + \varepsilon_{ikm}, \qquad m = 1, \ldots, 24 \\ n_{ik} &= \delta \Lambda_{d(ik)} + \mathbf{x}'_{ik}\mathbf{\Gamma} + \omega_i + \xi_{ik}, \end{aligned} \tag{3}$$

where $y_{ikm} \in \{y^e_{ikm}, y^u_{ikm}\}$ is the outcome variable of interest, n_{ik} is the length of sick leave spell k, $\Lambda_{d(ik)}$ is a binary instrumental variable indicating whether GP d who certifies observation i's sick leave k has an above-average certification propensity (see Section 2.1 for more details), $\mathbf{x}_{ik}$ is a vector of exogenous control variables, $\boldsymbol{\omega}_i$ is a $(N-1) \times 1$ vector of worker fixed-effects, and ε_{ikm} and ξ_{ik} are *i.i.d.* error terms with mean zero and finite variance. The model amounts to $M = 24$ separate second-stage regressions, where the coefficients $(\rho_m, \boldsymbol{\Theta}_m)$ are indexed by m indicating that they are allowed to vary over time.

The vector of control variables $\mathbf{x}_{ik}$ comprises age squared, initial wages, tenure, experience, log firm size, and indicator variables indicating if the worker is a part-time worker and if the worker is a blue collar worker, all measured at t_{-n_k}. As a proxy of health status, I use the total amount of drug expenses two years prior to t_{-n_k} in logarithmic form, along with total days spent in hospital two years prior to t_{-n_k}. Finally, I use industry-specific unemployment rates (214 sectors), as well as full sets of region and year dummies to capture macroeconomic fluctuations. Note that, due to the lack of data, I do not model the probability of getting a sick leave certificate in the first place. Thus, my results have to be interpreted as marginal increases in sick leave duration, conditional on receiving a sick leave certificate.

I use two-stage least squares (2SLS) in order to obtain estimates for ρ_m, $m = 1, \ldots, 24$.[2] These capture precisely the effect of a *marginal* day of sick leave on y_m, where the additional day is only granted because i's GP who is responsible for sick leave k has an above-average certification propensity. Because I allow for heterogeneous treatment effects both across observations i and levels of sick leave duration n_{ik}, each ρ_m is in fact a weighted average of unit causal effects evaluated at different units of n_{ik}, where weights depend on the location of compliers over the support of n_{ik}. In Section 2.2, I discuss the assumptions which are necessary for instrument validity and derive an analytical expression for the weighting function which partly determines $\hat{\rho}_m$. It turns out that individuals who contribute most to the estimated $\hat{\rho}_m$ are compliers with counterfactual sick leave durations between three days and nine days. Inference throughout the

[2] Although E_k is naturally a duration outcome, I refrain from using survival analysis in the paper. The reason is that I am unaware of estimators which deal with endogeneity in a survival analysis framework when the endogenous variable is continuous or discrete as in my case. A notable exception is Li *et al.* (2015), who essentially propose a control function approach where the second stage is specified as an additive hazards model. This is not practible, however, because (1) it requires assumptions on the underlying hazard function that are doubtable at best, and (2) incorporating a large set of fixed-effects makes its computation infeasible. Apart from that, the LATE interpretation which is crucial for my research design requires estimating the model via 2SLS (Angrist and Imbens, 1995).

paper is based on heteroskedasticity-robust and worker-level clustered standard errors.[3]

2.1 Estimating the Instrumental Variable

To obtain a certification propensity measure from the data, I decompose aggregated certified days of sick leave into time-varying observable patient characteristics and time-invariant patient and general practitioner fixed-effects. Consider the following two-way additive fixed-effects model proposed by Abowd *et al.* (1999, AKM hereafter):[4]

$$\tilde{n}_{it} = \mathbf{x}_{it}\mathbf{\Pi}' + \theta_i + \psi_{d(it)} + r_{it}, \tag{4}$$

where subscripts $i = 1, \ldots, N$ again denote patients, $d = 1, \ldots, D$ denote GPs with $d(it)$ being the dominant GP of patient i in year $t = 1, \ldots, T_i$,[5] and $\tilde{n}_{it}$ are medical expenses induced by doctor d for patient i in year t. Time-invariant effects are split into a patient-specific effect θ_i and a GP fixed-effect ψ_d. While θ_i is some sort of time-invariant health-stock unique to patient i, I interpret the GP fixed-effect ψ_d as an inherent propensity to certify sick leaves.[6] Observable time-varying health characteristics, including a qubic in age, a binary variable equal to unity if i was pregnant in year t, the number of days spent in hospitals where referral was not initiated by a GP in $t-1$, along with a vector of region binary variables are captured within the vector $\mathbf{x}_{it}$.

Following Card *et al.* (2013), I assume the residual r_{it} to be comprised of a random match component η_{idt}, a unit root component m_{it}, and an idiosyncratic error ν_{it}. That is,

$$r_{it} = f(\eta_{idt}, m_{it}, \nu_{it}), \tag{5}$$

with $f(\bullet)$ being some function and each of its components having zero conditional mean and

[3] Bootstrapped standard errors which account for the variance of the instrumental variable are similar to the analytical ones reported here and are available upon request.

[4] The idea of using the AKM model to estimate an instrumental variable from the data is based on Ahammer *et al.* (2015), who analyze the effect of labor income on mortality in Austria. In a similar vein, Markussen (2012) uses fixed-effects obtained from competing risks survival models as instrumental variables for sick leave take-up (see Section 1.1 for further details).

[5] The "dominant" GP is defined as the GP who billed the highest amount of fees to the health insurance for patient i in year t.

[6] Markussen (2012) terms this the *leniency* in prescribing sick leaves. This interpretation is somehow misleading, however, because GPs with a low fixed-effect may not necessarily be more lenient than others *per se*. Instead, other factors such as better knowledge about certain treatments or simply diverging preferences (e.g., with regard to the substutability of prescribing medication versus certifying sick leaves) may come into play.

finite variance. Note that consistent estimation of the AKM model requires that observables, the patient fixed-effect, the GP fixed-effect, and the residual contribute additively separably to prescribed days of sick leave. This implies that mobility between patients and GPs is exogenous conditional on these factors. In particular, it implies that motives for transitions of patients to new GPs are orthogonal to the random match component η_{idt}. In Ahammer and Schober (2016) we provide a battery of tests which uniformly support these assumptions.[7]

Markussen (2012) raises another important point, namely that estimates of the GP fixed-effect may suffer from a reflection problem whenever single patients influence the GP's fixed-effect, which would pose problems to identification later on. GPs in Austria, however, have large patient stocks (on average 898 patients), so it is unlikely that a single patient influences its GP's fixed-effect significantly. However, I follow Markussen and split every GP's patient stock in two halves in order to estimate ψ on one half of the sample while using it as an instrumental variable for the other half. The results from this exercise are highly similar to the ones reported here, and are available upon request.

In order to estimate equation (4), I build a panel spanning 2005–2012 comprising 1,294,460 patients at 857 GPs, which gives a total of 8,743,451 observations. This sample is larger than the one used to estimate (3), because it contains non-employed individuals (for instance, pensioners, students, or unemployed people) and children as well. Additionally, patients having zero days of sick leave in a given year are included as well, as long as I know that they were still insured in this year.

Using the estimated GP fixed-effects $\hat{\psi}_d$, define the instrument for GP d as a binary variable equal to unity if $\hat{\psi}_d$ is above its sample mean, that is,

$$\Lambda_d \equiv \mathbf{1}\{\hat{\psi}_d > \bar{\hat{\psi}}_d\}, \tag{6}$$

[7]In Ahammer and Schober (2016) we replicate suggestive tests proposed by Card *et al.* (2013) on the exogenous mobility assumption. Note that conditions for identification of the AKM model are somewhat weaker compared to those necessary for the instrumental variables framework in equation (3) to be valid. For (4) to be identified, mobility between patients and GPs can also be conditioned on the GP fixed-effect ψ_d, while this is obviously not the case for (3), where ψ (or, rather, its binary-coded counterpart Λ_d) is excluded from the second-stage regression.

where $\mathbf{1}\{\bullet\}$ denotes an indicator function and

$$\bar{\hat{\psi}}_d = D^{-1} \sum_{d=1}^{D} \hat{\psi}_d, \tag{7}$$

is the sample mean of the estimated GP fixed-effects.

Note that different specifications of the instrument, for instance, defining Λ_d to be equal to one if $\hat{\psi}_d$ is above its sample median or above the 90th percentile of the GP fixed-effect distribution, or simply using $\hat{\psi}_d$ as a continuous instrument, yield similar results.[8]

2.2 Identification and Treatment Effect Heterogeneity

Imposing linearity on the second-stage equation in (3) constrains the treatment effect ρ to be constant across individuals i and levels of sick leave duration n.[9] Assuming that ρ is the same regardless of the initial level of n, is perhaps an unrealistic assumption. Under weak regularity conditions outlined in Angrist and Imbens (1995), however, this effect can be interpreted as a weighted average of unit causal responses. This allows for treatment heterogeneities both across individuals and different initial levels of the endogenous variable.

Consider again the linear regression model from equation (3),

$$\begin{aligned} y_{ik} &= \rho \hat{n}_{ik} + \mathbf{x}'_{ik}\mathbf{\Theta} + \omega_i + \varepsilon_{ik}, \\ n_{ik} &= \delta \Lambda_{ik} + \mathbf{x}'_{ik}\mathbf{\Gamma} + \omega_i + \xi_{ik}, \end{aligned} \tag{8}$$

which incorporates, amongst others, a multi-valued endogenous variable $n_{ik} \in \{1, 2, \ldots, \bar{n}\}$ and a binary instrument Λ_{ik}. Note that I dropped subscripts m in order to simplify notation here, hence equation (8) is a special case of (3), with m being fixed at some month $\bar{m}$. This has one important consequence which is discussed below. Throughout this section, I study the properties of $\hat{\rho}$ using the potential outcomes framework (see, e.g., Rubin, 1974). Let $y_{ni} \equiv f_i(n)$ denote the potential outcome of observation i for any sick leave duration n, and let n_{1i} (n_{0i}) be i's potential sick leave

[8]These results are available upon request.

[9]This is of course an important limitation in my empirical analysis. Generalizing the model by allowing for random coefficients and non-linear covariate effects is theoretically possible as well, but makes computation infeasible. Note, however, that marginal treatment effects at the mean estimated from a bivariate probit are remarkably similar to the linear effects reported here (these are available upon request), hence I stick to the latter in order to estimate my main results. Random coefficients, however, are difficult to implement in a non-linear setting.

duration when $\Lambda_i = 1$ ($\Lambda_i = 0$).[10] In order to be able to interpret $\hat{\rho}$ as a weighted average of unit causal responses, three important assumptions are required. These can be written as (Angrist and Imbens, 1995)

(A1)	$\mathbb{E}[n_{1i} - n_{0i} \,|\, \mathbf{x}_i, \omega_i] \neq 0$	(first-stage)
(A2)	$\{y_{0i}, y_{1i}, \ldots, y_{\bar{n}i}, n_{0i}, n_{1i}\} \perp\!\!\!\perp \Lambda_i \,|\, \mathbf{x}_i, \omega_i$	(independence and exclusion)
(A3)	$n_{1i} - n_{0i} \geq 0 \;\forall i$ or vice versa	(monotonicity)

where $\perp\!\!\!\perp$ denotes statistical independence.

Assumption (A1) requires the existence of a first-stage, which is trivially met whenever $\delta \neq 0$ in (3). First stage regression results are given in Section 4, Table 4 – the null hypothesis that $\delta = 0$ can easily be rejected at $p < 0.01$. Assumption (A2) is commonly referred to as the *exclusion restriction*, sufficient conditions for it to hold are (1) random assignment of Λ_i conditional on covariates $\mathbf{x}_i$ and worker fixed-effects ω_i, and (2) that GPs' certification propensities affect patients' labor market outcomes only indirectly through their effect on sick leave durations. Here the biggest threat to identification is endogenous matching between patients and GPs. If patients select GPs based on their propensity to certify sick leaves, and this mobility decision is correlated with unobserved characteristics affecting employment or wages as well, (A2) would be violated, possibly biasing the estimates. I address this issue by providing various robustness checks in Section 4.2. Additionally, identification requires that doctors' time-invariant propensities to certify sick leaves are independent of their patients' employment outcomes. Since these propensities could be seen as an inherent trait, something doctors are born with or develop during their studies, this assumption seems reasonable. On a related issue, it is crucial to control properly for health status of patients in order to avoid omitted variable bias.

Principal-agent problems between the patient and the GP, as mentioned in the introduction, do not play a role for my analysis. The instrumental variable in (6) is a measure of prescription behavior fully orthogonal to patients' observed time-variant and observed as well as unobserved time-invariant characteristics, which include also traits such as motivation that determine the tendency to bargain about sick leave durations. In the 2SLS framework, the LATE induced by

[10] The function $f_i(n)$ gives the potential labor market outcome of individual i when sick leave duration is n. Note that the function f_i has subscript i, indicating that I allow for different responses to the treatment n across individuals. Furthermore, it is important to stress that $f_i(n)$ gives the potential outcome for *any* sick leave duration n, not just for the realized value n_i of individual i.

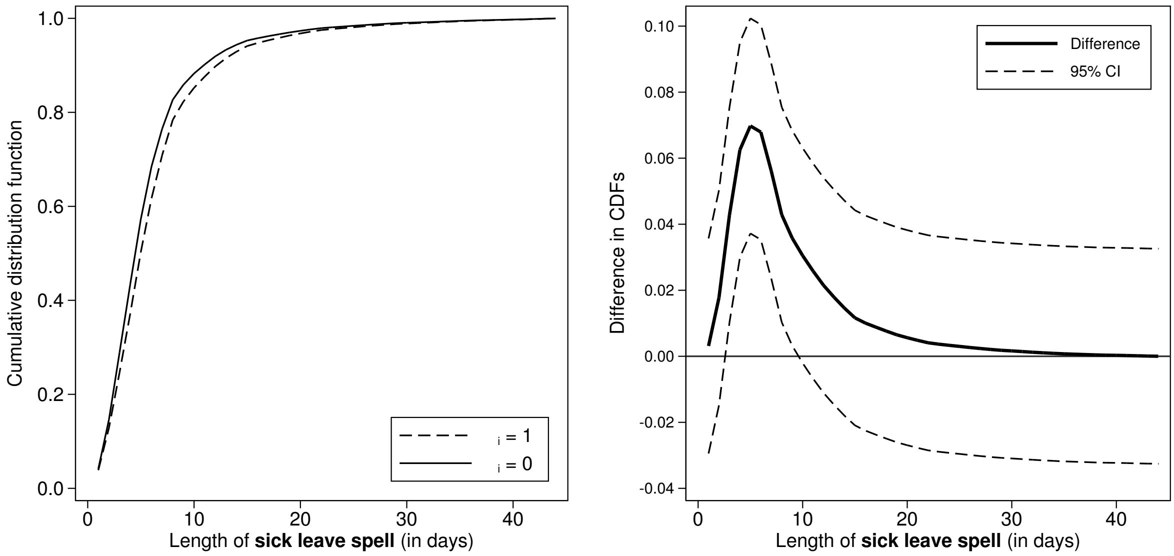

Figure 1 — The left graph illustrates the cumulative distribution functions (CDF) of sick leave duration for both realizations of the instrument ($\Lambda_i = 1$ and $\Lambda_i = 0$). The right graph plots the difference of those CDFs, i.e., the differences in the probability that sick leave duration is greater or equal to the respective level on the horizontal axis. Additionally, the 95% confidence interval for the difference function is given.

this instrumental variable captures *only* the effect of GPs' propensities to prescribe on sick leave duration, while possibly bargained days of sick leave are out of the equation.

Lastly, assumption (A3) constrains the instrument to shift all individuals' sick leave durations in the same direction (following Angrist *et al.*'s (1996) jargon, this rules out the existence of *defiers*). Since it is implausible that a higher propensity to certify sick leaves increases actual sick leave duration for some patients but decreases it for others, this assumption is likely fulfilled.

Angrist and Imbens (1995) show that, under assumptions (A1) through (A3), $\hat{\rho}$ estimated by 2SLS is a weighted average of unit causal responses. Let $\rho_i(n) \equiv y_{ni} - y_{n-1,i}$ be the causal response of a change in sick leave duration by one day for individual i at point n. Note that treatment effects $\rho_i(n)$ are allowed to vary across individuals. Moreover, since n is multi-valued with possible realizations $\{1, 2, \ldots, \bar{n}\}$ there are $\bar{n}$ different unit causal effects.[11] Neglecting covariates,[12] I can write the 2SLS estimate of ρ according to Angrist and Imbens (1995) in potential outcomes

[11]The sample value of $\bar{n}$ is 44.

[12]Neglecting covariates greatly simplifies the derivation of the LATE in a model with variable treatment intensity. To be fully correct, by Theorem 3 in Angrist and Imbens (1995), estimating $\hat{\rho}$ by 2SLS gives a weighted average of unit causal responses which are again weighted averages of covariate-specific causal responses. The covariate-specific weights depend on the variance of $\mathbb{E}[n_i \,|\, \mathbf{x}_i, \Lambda_i]$.

TABLE 1 — Most common medical conditions with average sick leave durations between 4 and 6 days.

		Occurences		Sick leave durations	
ICD-10 code	Description	No. of cases	in %	Mean of n_k	Std. dev.
J06.9	Acute upper respiratory infection	957665	31.42%	5.39	(3.32)
J02	Acute pharyngitis	81368	2.67%	4.90	(3.04)
J01	Acute sinusitis	67769	2.22%	5.75	(3.65)
B34.8	Other viral infections of unspecified site	51999	1.71%	4.50	(2.95)
J03	Acute tonsillitis	40135	1.32%	5.34	(3.15)

Notes: This table presents the five most common ICD-10 codes whose average sick leave duration in the sample is between 4 and 6 days.

notation as

$$\hat{\rho} = \sum_{n=1}^{\bar{n}} \omega_n \mathbb{E}[y_{ni} - y_{n-1,i} \,|\, n_{1i} \geq n \geq n_{0i}], \tag{9}$$

where each unit causal response $\mathbb{E}[\rho_i(n) \,|\, n_{1i} \geq n \geq n_{0i}]$ is the difference in potential outcomes $y_{ni} - y_{n-1,i}$ for compliers at point n (i.e., individuals whose treatment intensity changes from less than n to at least n when Λ_i switches to one). The weight is given by

$$\omega_n = \frac{\mathbb{P}[n_{1i} \geq n \geq n_{0i}]}{\sum_{j=1}^{\bar{n}} \mathbb{P}[n_{1i} \geq j \geq n_{0i}]}, \tag{10}$$

where $\mathbb{P}[n_{1i} \geq n \geq n_{0i}]$ is the relative size of the complier subpopulation at point n. Note that $\omega_n \geq 0$ for all n and $\sum_{n=1}^{\bar{n}} \omega_n = 1$. Thus, $\hat{\rho}$ is not a LATE in the traditional sense, but rather an average over multiple LATEs evaluated at different values of the endogenous variable, weighted by some function ω_n which depends upon the location of compliers across the support of n.

Following Angrist and Imbens (1995), I examine compliance (which crucially determines both the weighting function and unit causal responses) by comparing the cumulative distribution function (CDF) of sick leave duration n_i when the instrument Λ_i is switched on and off. Figure 1 plots these CDFs in the left-hand graph, their difference (i.e., the difference in probabilities that n_i is greater or equal to the respective level on the horizontal axis when $\Lambda_i = 0$ and $\Lambda_i = 1$) is illustrated in the right-hand graph. Compliers are located almost exclusively between 3 and 9 days of sick leave along the support of n_i, with the maximum being at 5 days. Thus, $\hat{\rho}$ is identified primarily through patients whose counterfactual sick leave duration of 5 to 9 days is extended by consulting a lenient physician.

Listing the five most common ICD-10 codes whose average sick leave duration in the sample lies between 4 and 6 days (which is ± 1 day around the maximum, see Table 1) shows an interesting pattern: These are outright diseases of the respiratory system.[13] Thus, it is mostly acute colds and flus for which doctors who have a high certification propensity grant an additional day of sick leave.

Although compliers cannot be identified individually, we can learn about distributional features of their demographic and occupational characteristics using simple calculations proposed by Angrist and Fernández-Val (2013). For simplicity, I recode the discrete treatment status n_i into a binary variable equal to unity if n_i is above the mean sick leave duration within its diagnosis group, where 26 groups are defined according to the first letter of the ICD-10 code. Denote by $\mathfrak{n}_i \in \{0,1\}$ the resulting treatment indicator, and let $x_i \in \{0,1\}$ be a *Bernoulli* distributed characteristic such as being female or being a migrant. By Bayes' rule, the relative likelihood a complier has $x_i = 1$ can be written as

$$\begin{aligned}\frac{\mathbb{P}[x_i = 1 \,|\, \mathfrak{n}_{1i} > \mathfrak{n}_{0i}]}{\mathbb{P}[x_i = 1]} &= \frac{\mathbb{P}[\mathfrak{n}_{1i} > \mathfrak{n}_{0i} \,|\, x_i = 1]}{\mathbb{P}[\mathfrak{n}_{1i} > \mathfrak{n}_{0i}]} \\ &= \frac{\mathbb{E}[\mathfrak{n}_i \,|\, \Lambda_i = 1, x_i = 1] - \mathbb{E}[\mathfrak{n}_i \,|\, \Lambda_i = 0, x_i = 1]}{\mathbb{E}[\mathfrak{n}_i \,|\, \Lambda_i = 1] - \mathbb{E}[\mathfrak{n}_i \,|\, \Lambda_i = 0]} \qquad (11)\end{aligned}$$

where the numerator is the first-stage for a subsample for which $x_i = 1$ and the denominator is the overall first-stage. Conditional expectations in (11) are approximated by ordinary least squares (OLS). Additionally, moments of the distribution of continuous covariates $x_i \in \mathbb{R}$ can be obtained by making use of Abadie's (2003) kappa:

$$\mathbb{E}[x_i \,|\, \mathfrak{n}_{1i} > \mathfrak{n}_{0i}] = \frac{\mathbb{E}[\kappa_i x_i]}{\mathbb{E}[\kappa_i]}, \qquad (12)$$

where

$$\kappa_i = 1 - \frac{\mathfrak{n}_i(1 - \Lambda_i)}{1 - \mathbb{P}[\Lambda_i = 1 \,|\, x_i]} - \frac{(1 - \mathfrak{n}_i)\Lambda_i}{\mathbb{P}[\Lambda_i = 1 \,|\, x_i]}. \qquad (13)$$

Results of these calculations are reported in Table 2. Conditional probabilities $\mathbb{P}[\Lambda_i = 1 \,|\, x_i]$ in equation (13) are estimated parametrically by Probit and the resulting estimates are plugged into the sample analogue of (12). Compliers – i.e., workers whose sick leave duration is increased

[13] Acute upper respiratory infection is the common diagnosis for typical colds, pharyngitis is the inflammation of the pharynx, sinusitis is inflammation of sinuses, and tonsillitis is inflammation of the tonsils. All of these are viral or bacterial infections, with symptoms possibly including plugged nose, sore throat, headache, or fever. Finally, B34.8 ("other viral infections of unspecified site") comprises other (unspecified) diseases caused by the rhinovirus.

TABLE 2 — Characteristics of compliers.

Binary covariates	Female	Migrant	High educ.	Part-time	Blue-collar
$\mathbb{P}[x_i = 1]$	0.397	0.197	0.342	0.169	0.612
$\mathbb{P}[x_i = 1 \mid \pi_{1i} > \pi_{0i}] \ / \ \mathbb{P}[x_i = 1]$	1.000	1.188	0.889	0.937	1.111
Continuous covariates	Age	Wage	Experience	Tenure	
$\mathbb{E}[x_i]$	36.8	26077.5	15.2	5.3	
$\mathbb{E}[x_i \mid \pi_{1i} > \pi_{0i}]$	35.7	26715.0	15.3	5.3	

Notes: This table reports characteristics of compliers based on calculations derived in Angrist and Fernández-Val (2013). All covariates are measured at t_{-n_k}. High education is a binary variable indicating whether the observation has at least an A-level degree. Age, wage, and tenure are given in years, wage is given in Euros. The number of observations is 3,125,759 in all cells.

because they consult a high-propensity physician – are 18.8% more likely to be migrants, 11.1% less likely to have at least an A-level degree, 6.3% less likely to be part-time workers, and 11.1% more likely to be blue-collar workers. In terms of gender, compliers are equally likely male or female. Furthermore, compliers are on average 35.7 years old, earn 26,715 Euros, have around 15.3 years of experience, and roughly 5.3 years of tenure. It seems as though compliers are largely located near the means of the independent variables in the model, which is highly beneficial in terms of external validity. Also, 2SLS coefficients can easily be compared to OLS coefficients under these circumstances.

As a final remark, recall that I derived $\hat{\rho}$ by holding the month of the outcome m fixed. By allowing ρ to vary over m, I consequently have to assume that the composition of the treatment effect (in particular compliance and the weighting function) is independent of the month after which the patient's labor market outcome is evaluated. Note that, even if GPs take future employment status of their patients into account when deciding about sick leave duration, this particular assumption is not necessarily violated. It would be violated, however, if GPs decided differently depending on whether they consider their potential employment status in one month t_m, or in another month $t_{m'}$, $m' \neq m$ – which is rather unlikely.

3 Data

I combine data from the *Upper Austrian Sickness Fund*, the *Austrian Social Security Database*, and tax data from the *Austrian Ministry of Finance*. The *Upper Austrian Sickness Fund* database comprises individual-level information on health-care service utilization in both the inpatient and

Table 3 — Descriptive statistics.

	Entire sample		$\hat{\psi}_{d(it)} > \bar{\hat{\psi}}_d$		$\hat{\psi}_{d(it)} \leq \bar{\hat{\psi}}_d$		
	Mean (1)	Std. dev. (2)	Mean (3)	Std. dev. (4)	Mean (5)	Std. dev. (6)	Difference (7)
Spells							
Length of sick leave spell (in days)	5.99	(5.34)	6.33	(5.51)	5.73	(5.18)	-0.597***
Length of total employment spell (in years)	8.00	(7.62)	7.90	(7.60)	8.07	(7.63)	0.164***
Length of remaining employment spell (in years)	2.69	(2.42)	2.65	(2.43)	2.71	(2.41)	0.059***
Subsequent spell is unemployment spell	0.27	(0.45)	0.28	(0.45)	0.27	(0.44)	-0.016***
Outcome variables							
$y^e_{ik,12} \equiv \mathbb{P}[i$ is still employed at the end of $t_{24}]$	0.51		0.51		0.52		0.015***
$y^u_{ik,12} \equiv \mathbb{P}[i$ became unemployed between t_0 and $t_{24}]$	0.22		0.23		0.21		-0.014***
Instrumental Variable							
Estimated GP fixed-effect ($\hat{\psi}_{d(it)}$)	0.11	(1.27)	0.98	(1.41)	-0.57	(0.50)	-1.550***
Binary instrument ($\Lambda_i \equiv \mathbf{1}\{\hat{\psi}_{d(it)} > \bar{\hat{\psi}}_d\}$)	0.44						
Control variables							
Part-time worker	0.17		0.17		0.17		-0.007***
Female	0.40		0.40		0.39		-0.013***
Migrant	0.20		0.21		0.19		-0.022***
At least A-level degree	0.34		0.35		0.34		-0.006***
log(annual wage at t_{-n_k})	9.94	(0.84)	9.92	(0.86)	9.96	(0.83)	0.032***
Experience until t_{-n_k} (in years)	15.25	(5.78)	15.20	(5.79)	15.28	(5.78)	0.084***
Tenure until t_{-n_k} (in years)	5.30	(6.57)	5.24	(6.53)	5.34	(6.60)	0.107***
Part-time worker	0.17		0.17		0.17		-0.007***
Blue collar worker	0.61		0.61		0.62		0.009***
log(drug expenses 2 years prior to t_{-n_k})	4.94	(1.99)	4.99	(1.98)	4.89	(2.00)	-0.103***
Days of hospitalization 2 years prior to t_{-n_k}	3.42	(2.76)	3.47	(2.76)	3.39	(2.76)	-0.074***
log(firm size)	0.73	(6.20)	0.73	(6.28)	0.72	(6.13)	-0.005
Physician density within community[a]	0.85	(0.35)	0.81	(0.35)	0.88	(0.35)	0.063***
Unemployment rate at industry sector level[b]	8.28	(4.17)	8.39	(4.26)	8.18	(4.09)	-0.206***
Number of observations (N^*)	3,125,759		1,373,340		1,752,419		
Number of different workers (N)	423,352		250,976		317,404		
Number of different firms (J)	43,297		31,373		36,293		
Number of different general practitioners (D)	1,078		350		728		

Notes: This table reports descriptive statistics for all variables used throughout the empirical analysis. In columns (3) to (6) the sample is split into sick leaves certified by physicians having an above-average propensity to certify sick leaves [(3) and (4)] and those certified by physicians having a below-average propensity [(5) and (6)]. In column (7) the differences in means between (3) and (5) are tested for statistical significance using Welch's t-test. *** denotes statistical significance at the 5% level ($p < 0.05$).

[a] Number of GPs per 10,000 inhabitants within a community.
[b] Number of unemployed workers divided by the total work force for each NACE95 two-digit industry sector.

outpatient sector for roughly one million members of the sickness fund. These members represent around 75 percent of the population in Upper Austria, which is one of nine provinces in Austria and, in turn, comprises around one-sixth of the entire Austrian population. I extract sick leave durations, diagnoses, and certain health indicators from these data. Information on employment histories, wages, as well as certain demographic information are taken from the *Austrian Social Security Database* (ASSD), which is a longitudinal matched employer-employee dataset covering the universe of Austrian workers from the 1970s onwards (Zweimüller *et al.*, 2009). Since wages are right-censored up to a tax cap, I augment the ASSD with income data from the *Austrian*

Ministry of Finance.

The data used for my empirical analysis cover all sick leaves certified in Upper Austria between 2005 and 2012 by general practitioners who have a contract with the sickness fund and have at least 50 patients on average during this time. I construct a panel where each observation is a single sick leave spell. Because each worker may have multiple (non-intersecting) sick leaves during the sampling period, I use worker fixed-effects and clustered standard errors to account for autocorrelation amongst the observations. Starting with 3,920,075 observations, I drop 445,807 apprenticeship spells, 230,481 spells whose subsequent spell is either a retirement or a maternity leave spell (workers belonging to these three groups are protected against dismissal by law), and 19,904 spells of employees who are either younger than 18 years or older than 65 years. Another 37,983 observations whose sick leave duration is above the 99$^{\text{th}}$ percentile at 44 days are dropped as well. Finally, I follow Correia (2015) and drop 97,898 singleton observations (i.e., workers for whom I have only one observation) in order to ensure proper inference and improve computational efficiency in my fixed-effect regressions. After all I am left with a total of $N^* = 3{,}125{,}759$ sick leave spells granted to $N = 423{,}352$ workers in $J = 43{,}297$ firms. Each worker has on average 7.19 distinct sick leave spells during the observation period of 8 years.

Detailed descriptive statistics are provided in Table 3. The mean sick leave spell (S_k) lasts around 6 days (the median is 5 days), while the mean employment spell lasts 8 years. Figure A.1 depicts their distributions, which both are right-skewed. After a sick leave, the average *remaining* employment spell (E_k) lasts 2.69 years (here, the median is 2.1 years). Surprisingly, a small yet negative reduced-form relationship can be observed in the raw data: Sick leaves certified by below-average propensity-to-certify doctors are followed by employment spells that last around 0.059 years ($\approx$ 22 days) longer than those following absences certified by more lenient doctors (this difference is statistically significant at the 5% level).

With a probability of 27%, the subsequent spell after E_k is an unemployment spell rather than a firm-to-firm transition. As expected, sick leaves certified by physicians with an above-average propensity to certify are on average 0.597 days longer ($p < 0.05$). After two years, 51% of all workers still belong to the same firm in which they worked in at t_0, while 22% registered at the unemployment office and 27% transitioned to a different firm.

Lenient GPs seem to be more often consulted by females, older patients, migrants (as compared to Austrian citizens), and lower income workers. Average levels of both tenure and experience in the sample are at 15.25 and 5.30 years, respectively. This can be interpreted as a sign that more sick leaves are taken towards the end of one's career, which is reasonable because workers health status decreases with age. Another reason, however, is simply that I dropped all apprentices from the sample, who indeed account for a large share of the young workforce in Austria.

Health proxies such as the amount of drug expenses aggregated over two years prior to the start of the sick leave, as well as aggregate days of hospitalization two years prior to the sick leave both seem to be higher for patients who consult more lenient doctors, although the difference in means is non-significant for the latter at any conventional level. Also, there seems to be a negative relationship between physician density and doctors' certification propensities. Finally, it is worthwhile to note that patients consulting high-propensity doctors tend to live in areas with higher unemployment rates.

The most common diagnoses for sick leaves are given in Table A.1. Typical flus (for instance, J06.9, "acute upper respiratory infection") make up a considerable portion of all sick leaves. In total, 62% of all diagnoses can be attributed loosely to this category, with ICD-10 code J06.9 ("acute upper respiratory infection") being the biggest contributor. The means of sick leave durations for such diagnoses lie between three and six days. Musculoskeletal diseases (indexed by the letter M), including conditions involving acute pain, account for another 13.6% of all diagnoses with mean sick leave durations being higher at seven to ten days. Potentially stress-related conditions, such as headaches (R51), migraine (G43), and major depressive disorders (F32), make up for 2.9% of cases. Burn-outs (Z73.0) are diagnosed 1,489 times during the observation period. Bear in mind that GPs are required to disclose diagnoses solely to the sickness fund, but *not* to the employer, therefore I do not control for them in my regressions.

4 Results

In order to identify a causal effect of supply-variation in sick leaves on employment, I proceed by estimating the IV model outlined in Section 2, which uses exogenous supply-side variation in sick leave certifications to identify the duration of single absences. Figure A.2 depicts the first-

Table 4 — Summary of first-stage regression results for different choices of Λ_d

Instrumental variable	Explanation	$\hat{\delta}$	Std. err.	F-statistic[a]	partial R^2
$\Lambda_d \equiv \hat{\psi}_d$	$\hat{\psi}_d$ is continuous	0.2444	(0.005)***	2275.3	0.00084
$\Lambda_d \equiv \mathbf{1}\{\hat{\psi}_d > \bar{\hat{\psi}}_d\}$	$\hat{\psi}_d$ is above its sample mean	0.4583	(0.011)***	1750.8	0.00065
$\Lambda_d \equiv \mathbf{1}\{\hat{\psi}_d > \hat{\psi}_{d,50}\}$	$\hat{\psi}_d$ is above its sample median	0.4701	(0.011)***	1939.0	0.00072
$\Lambda_d \equiv \mathbf{1}\{\hat{\psi}_d > \hat{\psi}_{d,90}\}$	$\hat{\psi}_d$ is above its 90th percentile	0.5833	(0.020)***	892.4	0.00033

Notes: This table summarizes results of estimating the first-stage equation in (3) with different instrumental variables. Each row represents a separate regression where sick leave duration n_{ik} is regressed on the instrumental variable $\Lambda_{d(it)}$, a vector $\mathbf{x}_{ik}$ of control variables described in Section 2, and a full set of worker-level fixed-effects. The number of observations in all regressions is 3,125,759. Standard errors are heteroskedasticity-robust and clustered on the worker-level. *** indicates statistical significance at the 1% level ($p < 0.01$).

[a] Kleinbergen Paap *rk* F-statistic.

stage relationship graphically, plotting the duration of individual sick leaves against estimated GP fixed-effects from the AKM model in equation (4). We see a positive relationship between these two variables in the raw data: The higher a GP's propensity to certify sick leaves, the longer their actual durations. As discussed in Section 2, however, I refrain from using $\hat{\psi}_d$ in continuous form as an instrumental variable, simply because a binary instrument considerably eases interpretation of the treatment effect $\hat{\rho}$ derived in Section 2.2. Therefore, I use the sample mean of the GP fixed-effect distribution as a cut-off point to construct the instrumental variable.

The results from estimating the first-stage are summarized in Table 4. Judging from the coefficient $\hat{\delta}$, consulting a GP who has an above-average propensity to certify sick leaves increases the duration of a single sick leave spell by roughly half a day ($p < 0.01$). Using other cut-off points, e.g., the median or the 90th percentile of the fixed-effect distribution, yields almost identical results, with F-statistics being far beyond the conventional rule-of-thumb level of 10. Likewise, second-stage estimates change only little with the choice of the cut-off point as well.[14]

Main results are presented in Figure 2, where 2SLS estimates of the local average treatment effects $\hat{\rho}_m$ along with their 95% confidence intervals are plotted against time. Each point along the line is obtained from a separate regression. In the left-hand graph, the employment probability after month $m = 1, \ldots, 24$ is regressed on sick leave duration. In the right-hand graph, unemployment probabilities after months $m = 1, \ldots, 24$ are the outcome variables. For comparison, OLS estimates are plotted as dashed lines.

I find that a marginal day of sick leave decreases employment probabilities persistently during

[14] All IV estimations reported in the remainder of this paper are available using one of the instrumental variables specified in Table 4 upon request.

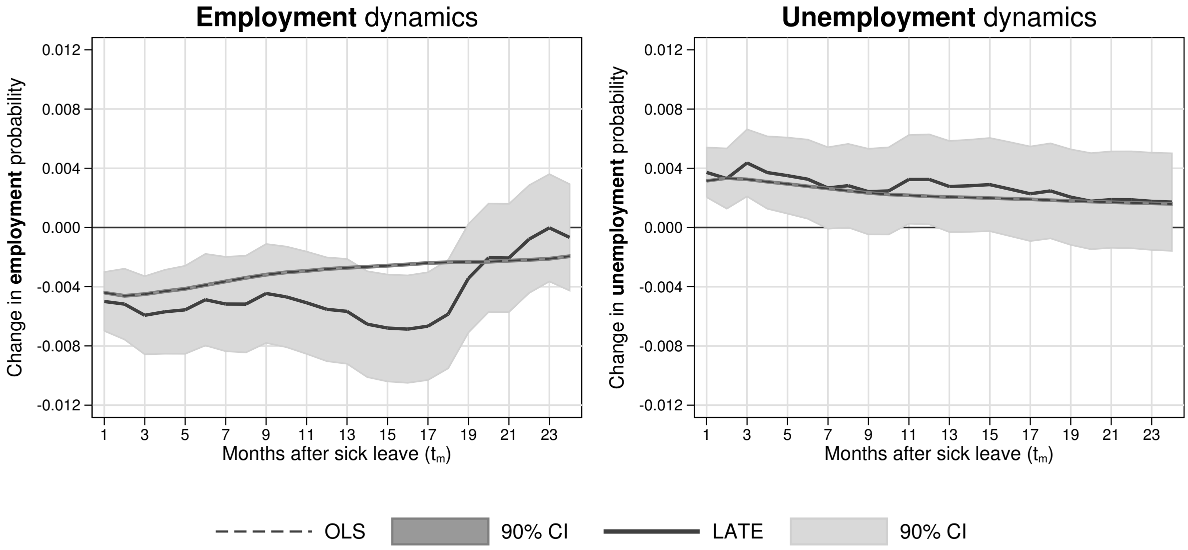

FIGURE 2 — These figures plot the estimated local average treatment effects $\hat{\rho}_m$, $m = 1, \ldots, 24$, of a marginal day of sick leave on employment probabilities (left-hand graph) and unemployment probabilities (right-hand graph) for the full sample (N = 3,125,759). Each coefficient for $t_1, \ldots, t_{24}$ is estimated from a separate regression based on the model in (8). LATEs are estimated by 2SLS, OLS effects (where the length of sick leaves n_k is treated as exogenous) are plotted as dashed lines for comparison.

the first 18 months with lows hitting at 3 months and 16 months. From month 18 onwards, effects decrease sharply and become statistically indistinguishable from zero. Conversely, the LATE on unemployment probabilities peaks in month 3 and then slowly converges to zero. After month 6, the effect remains non-significant at the 5% confidence level until the end of the observation period.

In terms of magnitudes, the LATE on employment probabilities varies between -0.0045 (week 9, $p = 0.03$) and -0.0069 (week 16, $p < 0.01$), whereas it ranges between 0.0028 (week 12, $p = 0.10$) and 0.0044 (week 3, $p < 0.01$) for unemployment probabilites. Thus, each marginal day of sick leave leads, *ceteris paribus*, to a decrease in employment probabilities between 0.45 percentage points (pps.) and 0.69 pps., and to an increase in unemployment probabilities between 0.28 pps. and 0.44 pps. Although not directly comparable, my results seem to be somewhat smaller and less persistent than those Markussen (2012) found for Sweden.

In order to gain a more comprehensive picture – especially with regard to coefficients of control variables and test statistics – I show full regression results for t_3 (where the LATE is strongest in magnitude for both employment and unemployment probabilities) in Table 5. Notice that F-statistics of the excluded instrument are well above 1,000 in all specifications. The outcome

Table 5 — Linear fixed-effects regressions for t_3.

	Dependent variable: Pr[i is still employed at the end of t_3]						Dependent variable: Pr[i became unemployed between t_0 and t_3]					
	OLS		IV-LATE (instrument: $\Lambda_{d(it)}$)				OLS		IV-LATE (instrument: $\Lambda_{d(it)}$)			
	(E.1)	(E.2)	(E.3)	(E.4)	(E.5)	(E.6)	(U.1)	(U.2)	(U.3)	(U.4)	(U.5)	(U.6)
Length of sick leave spell (in days)	-0.0051	-0.0045	-0.0063	-0.0048	-0.0049	-0.0059	0.0035	0.0033	0.0039	0.0036	0.0036	0.0044
	(0.000)***	(0.000)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.000)***	(0.000)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Age		-0.0492		0.0038	0.0037	-0.0504		0.0434		0.0041	0.0041	0.0443
		(0.001)***		(0.000)***	(0.000)***	(0.001)***		(0.001)***		(0.000)***	(0.000)***	(0.001)***
Age^2		-0.0001		-0.0001	-0.0001	-0.0001		-0.0000		-0.0000	-0.0000	-0.0000
		(0.000)***		(0.000)***	(0.000)***	(0.000)***		(0.000)***		(0.000)	(0.000)	(0.000)***
log(annual wage at t_{-n_k})		0.1860		0.1874	0.1873	0.1855		-0.0954		-0.0965	-0.0965	-0.0950
		(0.000)***		(0.001)***	(0.001)***	(0.001)***		(0.000)***		(0.000)***	(0.000)***	(0.000)***
Part-time worker		0.0485		0.0465	0.0465	0.0483		-0.0261		-0.0253	-0.0253	-0.0260
		(0.001)***		(0.001)***	(0.001)***	(0.001)***		(0.001)***		(0.001)***	(0.001)***	(0.001)***
Tenure until t_{-n_k} (in years)		-0.0100		-0.0095	-0.0095	-0.0099		0.0030		0.0026	0.0026	0.0029
		(0.000)***		(0.000)***	(0.000)***	(0.000)***		(0.000)***		(0.000)***	(0.000)***	(0.000)***
Experience until t_{-n_k} (in years)		-0.0002		-0.0002	-0.0002	-0.0002		0.0001		0.0001	0.0001	0.0001
		(0.000)***		(0.000)***	(0.000)***	(0.000)***		(0.000)**		(0.000)**	(0.000)**	(0.000)**
Blue collar worker		-0.0402		-0.0491	-0.0491	-0.0401		0.0230		0.0277	0.0277	0.0229
		(0.001)***		(0.001)***	(0.001)***	(0.001)***		(0.001)***		(0.001)***	(0.001)***	(0.001)***
log(drug expenses two years prior to t_{-n_k})		0.0003			0.0005	0.0006		-0.0001			-0.0003	-0.0003
		(0.000)***			(0.000)**	(0.000)**		(0.000)*			(0.000)	(0.000)
Days of hospitalization two years prior to t_{n_k}		-0.0001			-0.0000	0.0000		-0.0000			-0.0001	-0.0001
		(0.000)**			(0.000)	(0.000)		(0.000)			(0.000)	(0.000)
log(firm size)		0.0029				0.0029		-0.0053				-0.0053
		(0.000)***				(0.000)***		(0.000)***				(0.000)***
GP density at community level[a]		-0.0012				-0.0012		0.0006				0.0006
		(0.001)				(0.001)		(0.001)				(0.001)
Unemployment rate at industry level[b]		(0.001)				(0.001)		(0.001)				(0.001)
		(0.000)***				(0.000)***		(0.000)***				(0.000)***
Region dummies	No	Yes	No	No	No	Yes	No	Yes	No	No	No	Yes
Year dummies	No	Yes	No	No	No	Yes	No	Yes	No	No	No	Yes
N	3125759	3125759	3125759	3125759	3125759	3125759	3125759	3125759	3125759	3125759	3125759	3125759
Mean of outcome	0.86	0.86	0.86	0.86	0.86	0.86	0.08	0.08	0.08	0.08	0.08	0.08
First-stage F-statistic			2100.64	1854.19	1776.94	1750.84			2100.64	1854.19	1776.94	1750.84

Notes: Columns E.1, E.2, U.1, and U.2 are estimated via ordinary least squares (OLS), columns E.3 through E.6 as well as columns U.3 through U.6 are estimated via two-stage least squares (2SLS) where the instrumental variable is defined in equation (6). All regressions incorporating worker-level fixed-effects as well. The outcome is a binary variable equal to unity if the worker is still employed in the same firm as in t_0 (i.e., the end of the sick leave) after three months for columns E.1 – E.6, and a binary variable equal to unity if the worker became unemployed at some point of time between t_0 (i.e., the end of the sick leave) and t_3 (i.e., three months after the sick leave) for columns U.1 – U.6. The coefficient on "length of sick leave spell" represents the local average treatment effect (LATE) of a marginal day of sick leave. Heteroskedasticity-robust and worker-level clustered standard errors are given in parentheses below coefficients, stars indicate significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

[a] Measured as the number of GPs per 10,000 inhabitants within a community.

[b] Measured as the number of unemployed workers divided by the total work force for each NACE95 two-digit industry sector.

variable in columns E.1 – E.6 is the probability of still being employed in the same firm three months after the sick leave, whereas the outcome in columns U.1 – U.6 is the probability of becoming unemployed during the first three months after the sick leave. The table is organized such that the model is extended in various steps.

Columns E.1 and E.2 show the estimated coefficients for a model without any covariates. Here, the LATE is estimated to be -0.0045 ($p < 0.01$), indicating that a marginal day of sick leave decreases employment probability during the first three months by roughly 0.45 pps. In column E.4, the model is augmented with worker-level characteristics such as age, wages, and occupational characteristics. All of them have a significant impact on the employment probability: The age effect is inverted U-shaped, higher initial income leads to higher employment probabilities; tenure, experience, and being a blue collar worker are associated with lower employment prospects. Part-time workers have on average higher employment probabilities (the probability that part-time workers are still employed three months after the sick leave is 4.85 pps. greater than for full-time workers). The LATE is slightly smaller than before at -0.0048 ($p < 0.01$). Incorporating health proxies does not change the estimated coefficients by much. In contrast, firm size and macroeconomic conditions have a sizable effect on employment probability. Interestingly, competition amongst doctors (measured through the GP density at the community level) does not appear to have a significant effect on employment.

My preferred specification is the full model in column E.6, where the LATE is -0.0059 ($p < 0.01$), suggesting that employment probability three months after the absence spell is reduced by 0.59 pps. for each marginal day of sick leave. OLS effects are similar to the instrumental variable estimates across specifications. The results for unemployment probabilities mirror the employment effects. Again, the F-statistics are above 1,000 in all estimated models. In my preferred specification (column U.6), the LATE is estimated as 0.0044 ($p < 0.05$), implying that a marginal day of sick leave increases the probability of becoming unemployed after three months by 0.44 pps.

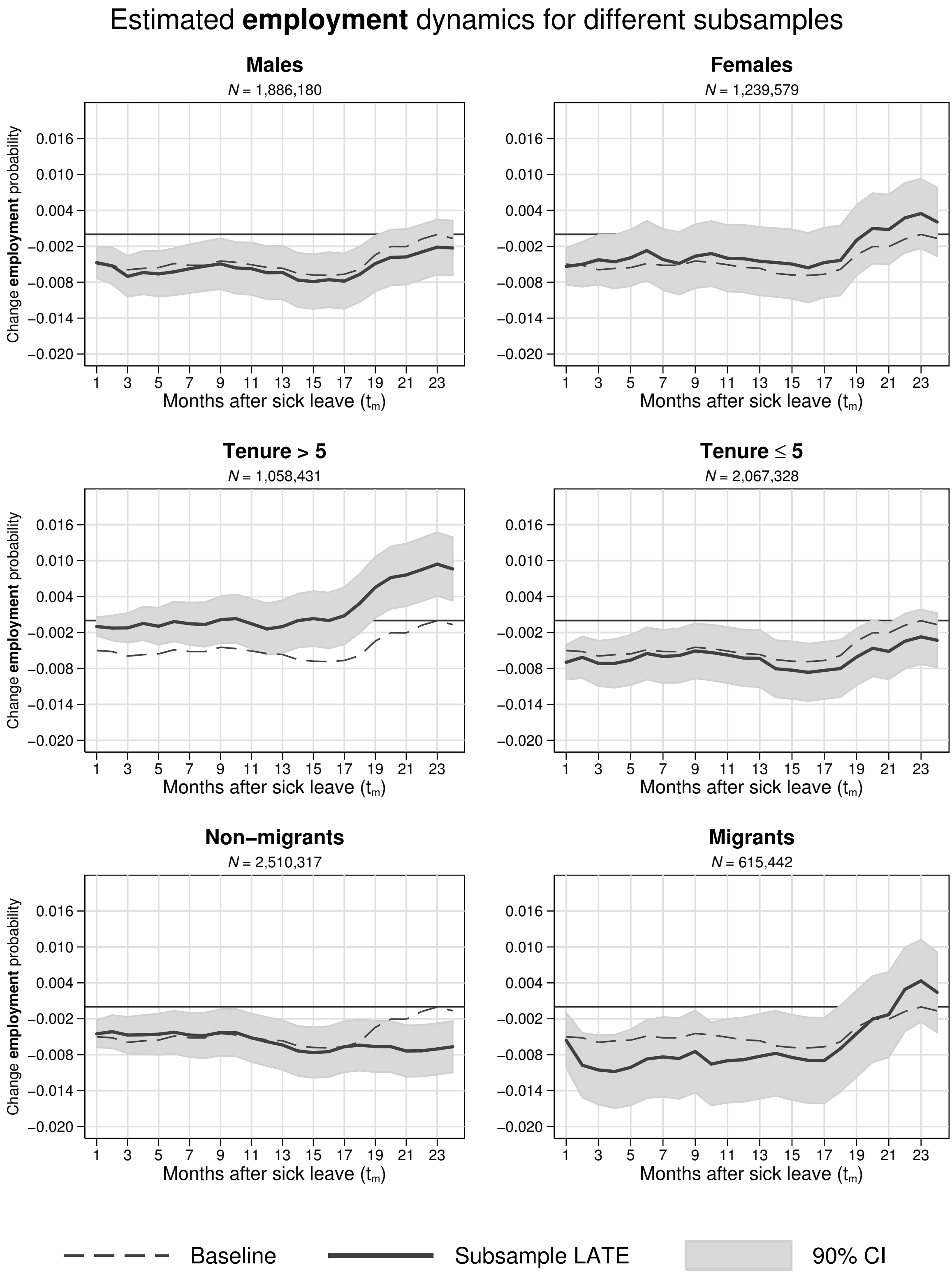

FIGURE 3 — These figures plot the estimated local average treatment effects $\hat{\rho}_m$, $m = 1, \ldots, 24$, of a marginal day of sick leave on employment probabilities for different subsamples of the population. Each coefficient for $t_1, \ldots, t_{24}$ is estimated from a separate regression based on the model in (8). The dashed line plots baseline results from Figure 2.

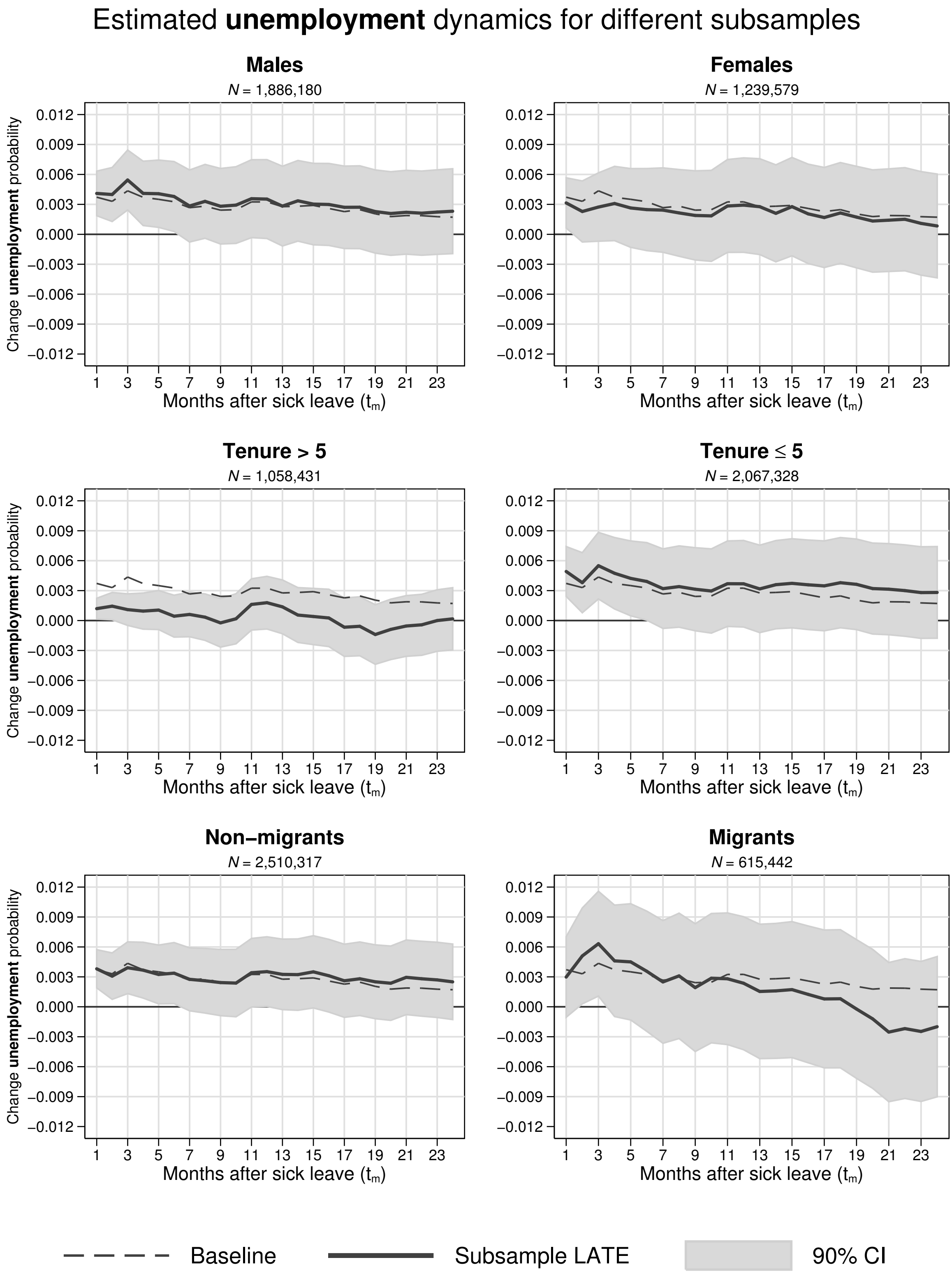

Figure 4 — These figures plot the estimated local average treatment effects $\hat{\rho}_m$, $m = 1, \dots, 24$, of a marginal day of sick leave on unemployment probabilities for different subsamples of the population. Each coefficient for $t_1, \dots, t_{24}$ is estimated from a separate regression based on the model in (8). The dashed line plots baseline results from Figure 2.

4.1 Heterogeneous Effects

In a next step, I compare the effects for different subsamples of the population. Dynamic effects are provided in Figures 3 (employment probabilities) and 4 (unemployment probabilities). Again, solid lines show the evolution of the LATE coefficient up to 24 months after the end of the sick leave, whereas the dashed line provides the baseline estimates from Figure 2 for comparison.

Firstly, I split the sample by gender. Estimated employment probabilities are close to the baseline estimates. The coefficients are somewhat greater for men than for women, so the overall effect seems to be driven relatively more by men. For women, the initial effect is almost identical to the baseline, but quickly approaches zero and is insignificant after month three. For men, the effect is similarly persistent as in the combined sample. The LATE on unemployment, on the other hand, seems to be driven only by men. For women, coefficients are statistically insignificant across the entire observation period.

Secondly, I stratify by tenure levels. One might suspect that workers with lower job tenure get punished harder for longer absences, because they had less time to reveal their inherent productivity or to convince the employer about their trustfulness (in case longer sick leaves are really perceived as a signal of absenteeism). On the other hand, firms may have a preference for younger workers which could lead to the opposite effect. In fact, I find that the LATE is insignificant for workers with more than five years of tenure, and is positive and significant after month 18. One explanation could be that high tenure workers do not get punished for a marginal day of sick leave, but eventually a positive health effect kicks in and increases employment probabilities. The unemployment effect is insignificant throughout the observation period. For workers with tenure of less than five years, the LATE is similar to the baseline effect, but slightly larger in magnitude. In terms of unemployment probabilities, I estimate that low tenure workers have initially a high positive initial effect, which is insignificant after four months.

Thirdly, the estimated effects are stronger for migrants than for Austrian citizens. However, it seems that for Austrians, the negative employment effect is more persistent, whereas it is statistically insignificant for migrants. Similar to results discussed before, Austrians initially experience a positive effect in terms of unemployment risk, which then deteriorates over time. For migrants I find a positive effect three months after the sick leave, which quickly becomes insignificant and stays at zero throughout the observation period.

4.2 Robustness

As discussed in Section 2, the main threat to identification is endogenous matching between patients and doctors. Whenever patients select GPs based on their propensity to certify sick leaves, and this mobility decision is also correlated with unobserved characteristics affecting employment and wages, the exclusion restriction is violated and estimates will be biased. In this section, I analyze different subsamples of the population where either mobility is restricted, or where motives of transitions can be assumed to be caused by factors other than the prescription behavior of the new GP. Whenever results hold, it is likely that – even if there is sorting on unobservables – its quantitative effect is negligible. Additionally, another important requirement for identification is that health status of the patient is adequately controlled for. Thus, I follow Halla *et al.* (2016) and estimate my main regressions on a specific subsample which can be considered as homogeneous with regard to health status. For these individuals, GP consultations can be considered more or less random.

First, I restrict the sample to sick leaves that start either on weekends or public holidays when doctors typically close their practices. In order to maintain the provision of basic health care on such days, each district in Upper Austria has a schedule of rotating GPs who provide out-of-hours services. Thus, assignment between patients and GPs is more or less random on weekends and holidays, because it depends solely on the rotation schedule.[15] Although the purpose of such services is to offer assistance in medical *emergencies*, patients may avail them irrespective of the actual condition they suffer from. In fact, the first six most common diagnoses certified on weekends or holidays are identical to those for the full sample shown in Table A.1.

[15]Note, however, that there are some problems associated with this assignment mechanism: Firstly, the resulting sample might be selected, insofar as patients will typically wait until their family doctor's practice is open again unless they suffer from an acute condition which requires immediate treatment. Furthermore, ambulances are open on weekends and holidays as well – thus, in areas where hospitals are reachable in a few minutes, patients will likely prefer going to the ambulance rather than consulting an emergency GP. Supposedly, workers living in rural areas will therefore be overrepresented in this subsample. Thirdly, I do not observe the actual day of consultation. Although law prohibits sick leaves being certified retroactively, it is possible that consultations preceding spells which start on weekends or holidays in fact took place during the week. However, this can only apply to employees who work on weekends but not during the week, which is indeed a rather unusual type of working contract. Hence, bias induced by such observations should be rather small. Finally, I cannot use worker-level fixed-effects in this specification, because only few observations in the sample consult a doctor twice or more on weekends or holidays.

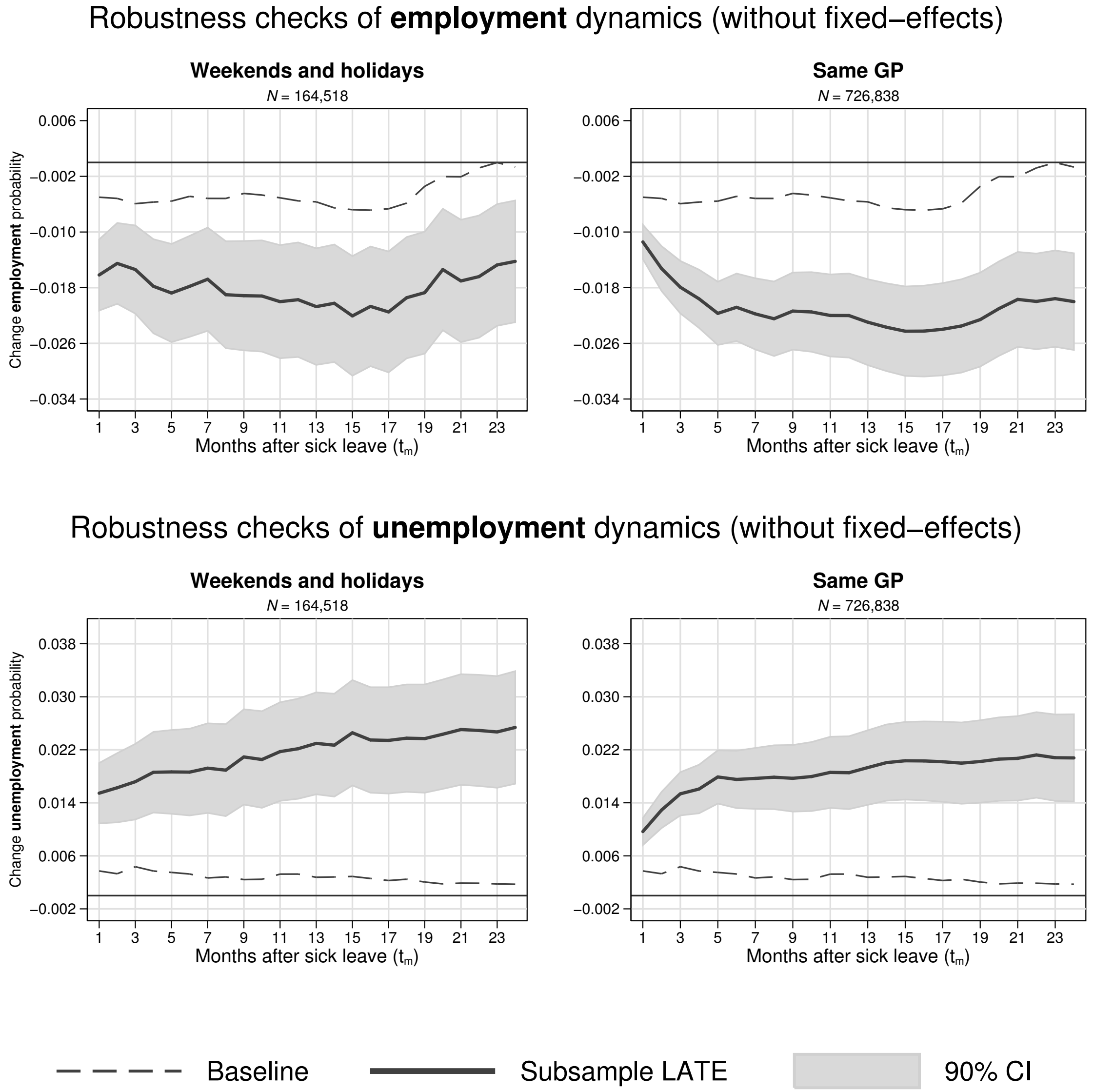

FIGURE 5 — These figures plot the estimated local average treatment effects $\hat{\varsigma}_m$, $m = 1, \ldots, 24$, obtained from the model in (14) which does not incorporate fixed-effects but controls for gender, education, and migratory status instead. The outcomes in the upper two graphs are employment probabilities estimated from separate regressions for each month $t_1, \ldots, t_{24}$, while the outcomes in the lower two graphs are unemployment probabilities for each month $t_1, \ldots, t_{24}$. *Weekends and holidays:* the sample is restricted to sick leaves certified on weekends or public holidays, *same GP:* the sample is restricted to workers who never change GPs during the observation period. The dashed lines show baseline results from Figure 2.

Without fixed-effects, the main regression model in (3) translates to

$$\begin{aligned} y_{ikm} &= \varsigma_m \hat{n}_{ik} + \mathbf{x}_{ik}' \mathbf{\Omega}_m + \mathbf{z}_i' \mathbf{\Xi}_m + \upsilon_{ikm}, \qquad\qquad m = 1, \ldots, 24 \\ n_{ik} &= \iota \Lambda_{d(ik)} + \mathbf{x}_{ik}' \mathbf{\Phi} + \mathbf{z}_i' \mathbf{\Psi} + \zeta_{ik}, \end{aligned} \tag{14}$$

where I incorporate a vector $\mathbf{z}_i$ of time-invariant control variables comprising a female dummy, a migrant dummy, and education in categorical form in place of the fixed-effect. Note, however, that treatment effects $\hat{\varsigma}_m$, have the same properties discussed in Section 2.2 as the $\hat{\rho}_m$ estimated from the fixed-effects model in (3).

The distribution of weekdays, the sick leave spells start and end on is shown in Figure A.3. Most sick leaves start on Mondays and end on Fridays. A total of 159,856 spells (approximately 5.2% of the full sample) start either on a Saturday, a Sunday, or on a public holiday. The estimated employment and unemployment dynamics for this sample are illustrated in Figure 5. The main conclusions hold also for this sample of randomly assigned patient-GP matches: Coefficients have the expected sign and are roughly three times as high as those obtained from the baseline model. Three months after a spell, the employment probability is decreased by 1.86 pps. ($p < 0.01$) through a marginal day of sick leave. The estimated coefficient for the unemployment probability is even higher at 1.93 pps.

As a next step, I consider only workers who never change their GP during the observation period. For these patients, endogenous matching is obviously only a problem if it happened before 2005. The sample is reduced to 707,624 observations, which amounts to roughly 23% of the original data. Again, because worker fixed-effects coincide with the instrumental variable in this subsample where patients stick to one GP over time, I estimate the model in (14) instead, where I control for gender, migratory status, and education. Results are provided in Figure 5. Once again, each additional day of sick leave has a strong negative effect on employment and a positive effect on unemployment, with coefficients being relatively large in magnitude. Both effects even appear to persist well beyond the observation period of two years. This is in contrast to the evolution of the baseline estimates approaching zero towards the end of the observed time horizon. At t_3, the estimate of ς_3 suggest a 1.99 pps. ($p < 0.01$) decrease in employment probability and a 1.72 pps. ($p < 0.01$) increase in unemployment probability for each marginal day of sick leave.

An important restriction for matching is certainly competition among doctors. In areas with high competition, patients can easily change doctors if they encounter one who refuses to match their demands. In low-density areas, on the other hand, patients face only a small set of different doctors they can choose from. As another robustness check, I therefore restrict the sample to areas with a density of less than 0.63 doctors per 100,000 inhabitants at the community level (this roughly corresponds to the 25$^{\text{th}}$ percentile of the density distribution). Results are shown in Figure 5. In this subsample, the initial effect found between the first seven months is robust for

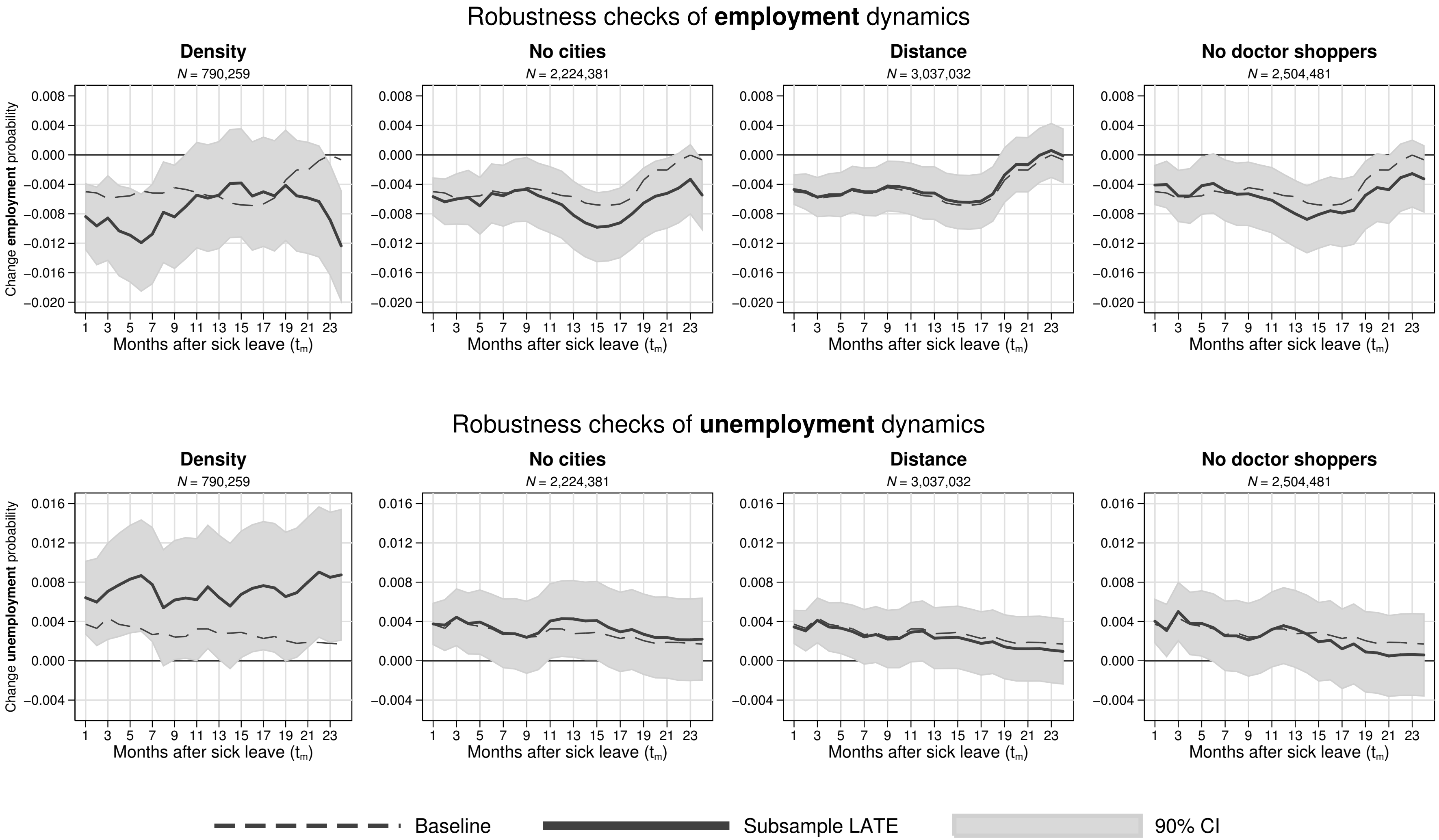

FIGURE 6 — These figures plot the estimated local average treatment effects $\hat{\rho}_m$, $m = 1, \ldots, 24$, of a marginal day of sick leave on both employment (upper four graphs) and unemployment (lower four graphs) probabilities for different subsamples of the population. *Density:* sample is restricted to areas with a GP density of less than 0.63 doctors per 100,000 inhabitants at the community level, *no cities:* sample is restricted to areas with fewer than 18,705 inhabitants, *distance:* sample is restricted to patient-GP matches where geographical distance is less than 10 kilometers, *no inside movers:* observations who change GP but do not change their living place are dropped, *same GP:* only observations who do not change their GP during the observational period are kept. Each coefficient for $t_1, \ldots, t_{24}$ is estimated from a separate regression based on the model in (8). The dashed line shows baseline results from Figure 2.

both employment and unemployment probabilities.

Next, I drop areas with more than 18,705 inhabitants (which is the population size of the smallest city in Austria in 2016) from the data. This approach is based on the idea that workers who live in rural areas face a limited variety of different doctors and thus are restricted in their mobility. Roughly 28% of all workers live in the sample live in cities, thus the sample size remains relatively stable after dropping these. Results are given in Figure 6. These estimates are similar to the baseline specification. In a similar vein, I keep only patient-GP pairs between which geographical distance is low. I define the distance between patients and GPs as the minimum of either the distance between a patient's place of residence and her doctor's practice, or the distance between a patient's working place and the GP practice. Arguing that if the distance is shorter than 10 kilometers, the patient likely selected his GP based on close proximity rather than because of the doctor's practice style, I estimate the model in equation (3) for this subsample (see Figure 6). The evolution of effects over time as well es their magnitudes are roughly the same as for the full sample.

As another robustness check, I drop patients from the sample who move to a new GP, but do not change their area of residence at the same time. This eliminates "doctor hoppers" from the analysis, i.e., patients who alternate between doctors until they encounter one who provides them with the treatment they seek. These results, which are shown in Figure 6, indicate that effects are larger in magnitude compared to the baseline and keep their statistical significance. Three months after the sick leave, the LATE for the employment probability is estimated as -0.0075 ($p < 0.01$).

I conclude that sorting between patients and doctors does not pose a significant problem for my empirical analysis. This is perhaps not surprising: Although patients are free to select among the set of available GPs, 73.7% of Upper Austrians choose a GP within their zip code (Hackl *et al.*, 2015). Thus, patients presumably tend to select the nearest GP in terms of geographic proximity, rather than one whose prescription behavior fits them best. This impression is confirmed by Ahammer and Schober (2016), who adapt tests on the exogenous mobility assumption proposed in the empirical labor literature and do not find evidence of sorting on observables. A similar conclusion has been made by Markussen *et al.* (2011) for Sweden.

Finally, similar to Halla *et al.* (2016), I restrict the sample to patients who have not been admitted to hospital and have less than 330 Euros of medical expenses on aggregate two year

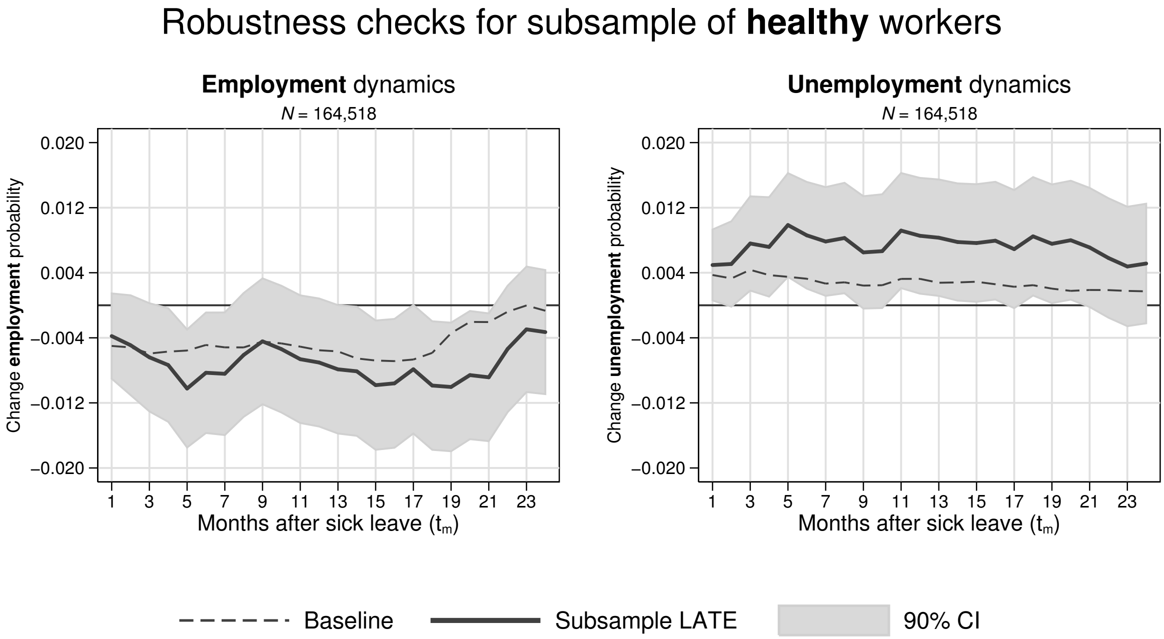

FIGURE 7 — These figures plot the estimated local average treatment effects $\hat{\rho}_m$, $m = 1, \ldots, 24$, of a marginal day of sick leave on both employment (left-hand graph) and unemployment (right-hand graph) probabilities for a subsample of the population which has zero days of hospitalization and less than 330 Euros of aggregate medical expenses two years prior to t_{-n_k}. Each coefficient for $t_1, \ldots, t_{24}$ is estimated from a separate regression based on the model in (8). The dashed line shows baseline results from Figure 2.

prior to the start of the sick leave. This leaves me with a sample of 667,706 observations. Here, estimates become rather imprecise due to the comparably low sample size, resulting in non-significant coefficients whenever effects are small (for instance in month one or month nine). However, estimates are almost uniformly higher in magnitude compared to the full sample, so the statistical non-significance should not be overemphasized. In general, most robustness checks yield effects that are larger in magnitude compared to the baseline, so I am inclined to think that the initial estimate of the LATE is in fact a lower bound for the actual effect.

5 Discussion

I quantify the impact physicians have on employment prospects of their patients by granting sick leaves longer than necessary. In order to isolate this channel, I establish a LATE framework where supply-side variation in sick leave certifications is used to instrument for actual sick leave durations. The resulting effect on employment probabilities is identified solely through workers whose sick leave duration is extended due to consulting a GP who has an above-average propensity to certify sick leaves. Thus, I estimate the effect of a *marginal* day of sick leave, namely one that is only granted because the certifying doctor tends to prescribe longer sickness absences, not because health status of the patient requires it. I find that this marginal day of sick leave has a persistent negative effect on employment probabilities and a positive effect on unemployment probabilities, especially for men with low job tenure and migratory background. Crucial for the identification of the causal effect is that sorting between patients and GPs is conditionally exogenous. I devote a substantial part of the paper to sensitivity analyses which all show that, even if there is sorting on unobservables, any bias induced by it is quantitatively negligible.

Assessing the economic size of effects I estimate is difficult. In fact, I find that employment probability three months after the sick leave, for instance, decreases by 0.59 pps. due to the additional day of sick leave, while the unemployment probability increases by 0.44 pps. Note, however, that these effects are solely caused by supply-variation in the certification of sick leaves – overall effects could be much greater. As it turns out, high income is a strong protective factor in terms of employment. This coefficient, however, does not have a causal interpretation on its own, because it might be driven by unobserved variables which simultaneously determine sick leave durations.

An important question which remains to be answered is why sick leaves entail adverse employment effects despite being designed as an institution in fact supposed to protect workers. The most obvious explanation is that workers are somehow penalized by their employer for being off work. If used as a screening device, absences may be interpreted as low work effort or motivation. Additionally, there is a definite link between health and productivity. Employers who are aware of this link may therefore discriminate against workers with longer absences. Besides employer-side penalization, the second plausible explanation is that the sickness absence *itself* drives this negative effect through preventing the worker from regular activity. Interestingly, applying the model in (3) on the probability of going into disability pension eventually (this concerns around 2.7% of workers in the sample) yields a positive coefficient: Lenient physicians are estimated to *increase* the likelihood that the patient goes into invalidity pension through granting only one additional day of sick leave by 0.17%, yet this effect is fairly imprecisely estimated (p = 0.186). Nevertheless, it indicates that this other channel may also exist. Unfortunately, it is difficult to pin down such mechanisms empirically with the data at hand.

My results raise one important recommendation for doctors: In case of doubt, it may be beneficial to certify shorter sick leaves whenever it is medically justifiable. Additionally, policy makers may consider introducing upper bounds of possible absence spell durations for certain groups of diagnoses.

6 Bibliography

Aakvik, A., Holmås, T. H., and Islam, M. K. (2010). Does Variation in General Practitioner (GP) Practice Matter for the Length of Sick Leave? A Multilevel Analysis Based on Norwegian GP-Patient Data. *Social Science & Medicine*, **70**(10), 1590 – 1598.

Abadie, A. (2003). Semiparametric Instrumental Variable Estimation of Treatment Response Models. *Journal of Econometrics*, **113**(2), 231–263.

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High Wage Workers and High Wage Firms. *Econometrica*, **67**(2), 251–333.

Ahammer, A., Horvath, G. T., and Winter-Ebmer, R. (2015). The Effect of Income on Mortality – New Evidence for the Absence of a Causal Link. Working Paper 1504, Johannes Kepler University Linz, Department of Economics.

Ahammer, A. and Schober, T. (2016). Explaining Variations in Health Care Expenditures – What is the Role of Practice Styles? Unpublished manuscript, Johannes Kepler University Linz, Department of Economics. Presentation slides including methodological details and main results are available on ResearchGate: http://dx.doi.org/10.13140/RG.2.1.2303.4645.

Amilon, A. and Wallette, M. (2009). Work Absence – A Signalling Factor for Temporary Workers? *Labour*, **23**(1), 171–194.

Andersen, S. H. (2010). The Cost of Sickness: On the Effect of the Duration of Sick Leave on Post-Sick Leave Earnings. *Social Science and Medicine*, **70**(10), 1581–1589.

Angrist, J. D. and Fernández-Val, I. (2013). ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework. In Acemoğlu, D., Arellano, M., and Dekel, E. (editors), *Advances in Economics and Econometrics: Volume 3, Econometrics: Tenth World Congress*, volume 51, page 401. Cambridge University Press.

Angrist, J. D. and Imbens, G. W. (1995). Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity. *Journal of the American Statistical Association*, **90**(430), 431–442.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, **91**(434), 444–455.

Böheim, R. and Leoni, T. (2014). Firms' Sickness Costs and Workers' Sickness Absences. Working Paper 20305, National Bureau of Economic Research.

Brown, S. and Sessions, J. G. (1996). The Economics of Absence: Theory and Evidence. *Journal of Economic Surveys*, **10**(1), 23–53.

Card, D., Heining, J., and Kline, P. (2013). Workplace Heterogeneity and the Rise of West German Wage Inequality. *Quarterly Journal of Economics*, **128**(3), 967–1015.

Correia, S. (2015). Singletons, Cluster-Robust Standard Errors and Fixed Effects: A Bad Mix. Mimeo, Duke University.

Federal Ministry of Social Affairs (2014). *Social Protection in Austria 2014*. Available online at https://broschuerenservice.sozialministerium.at, accessed: Wednesday 7$^{\text{th}}$ December, 2016.

Grytten, J. and Sørensen, R. (2003). Practice Variation and Physician-Specific Effects. *Journal of Health Economics*, **22**(3), 403–418.

Hackl, F., Hummer, M., and Pruckner, G. J. (2015). Old Boys' Network in General Practitioners' Referral Behavior? *Journal of Health Economics*, **43**, 56–73.

Halla, M., Mayr, H., Pruckner, G. J., and García-Gómez, P. (2016). Cutting Fertility? The Causal Effect of Supply-Determined Cesarean Section on Fertility. Unpublished manuscript, Johannes Kepler University Linz, Department of Economics.

Halla, M., Pech, S., and Zweimüller, M. (2015). The Effect of Statutory Sick Pay Regulations on Workers' Health. Working Paper 1502, Johannes Kepler University Linz, Department of Economics.

Hansen, J. (2000). The Effect of Work Absence on Wages and Wage Gaps in Sweden. *Journal of Population Economics*, **13**(1), 45–55.

Hesselius, P. (2007). Does Sickness Absence Increase the Risk of Unemployment? *Journal of Socio-Economics*, **36**(2), 288–310.

Imbens, G. W. and Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, **62**(2), 467–475.

Li, J., Fine, J., and Brookhart, A. (2015). Instrumental Variable Additive Hazards Models. *Biometrics*, **71**(1), 122–130.

Markussen, S. (2012). The Individual Cost of Sick Leave. *Journal of Population Economics*, **25**(4), 1287–1306.

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

Nilsen, S., Werner, E. L., Maeland, S., Eriksen, H. R., and Magnussen, L. H. (2011). Considerations made by the general practitioner when dealing with sick-listing of patients suffering from subjective and composite health complaints. *Scandinavian Journal of Primary Health Care*, **29**(1), 7.

Phelps, C. E., Mooney, C., Mushlin, A. I., and Perkins, N. A. K. (1994). Doctors Have Styles, and They Matter! *University of Rochester Working Paper*.

Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, **66**(5), 688–701.

Zweimüller, J., Winter-Ebmer, R., Lalive, R., Kuhn, A., Wuellrich, J.-P., Ruf, O., and Büchi, S. (2009). Austrian Social Security Database. Working Paper 0903, NRN: The Austrian Center for Labor Economics and the Analysis of the Welfare State.

A Additional Tables and Figures

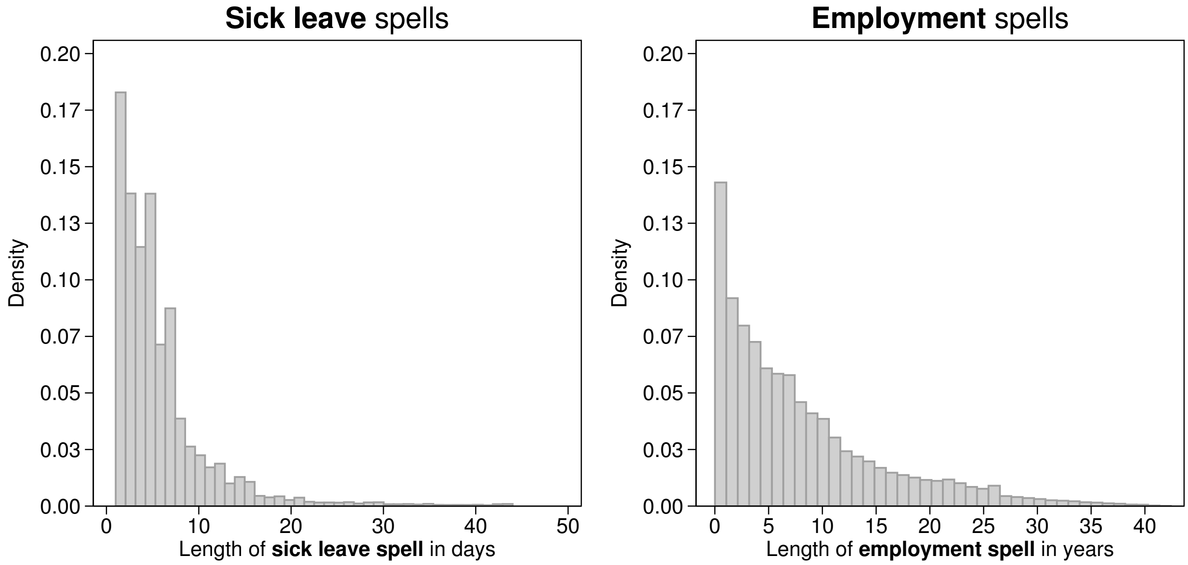

FIGURE A.1 — These graphs depict the distribution of sick leave spell durations (n_k, left figure) and total employment spell durations (right figure) in the data.

TABLE A.1 — Most common medical conditions, only diagnoses with more than 20,000 cases in the data.

		Occurences		Sick leave durations	
ICD-10 code	Description	No. of cases	in %	Mean of n_k	Std. dev.
J06.9	Acute upper respiratory infection	957665	31.42%	5.39	(3.32)
A09	Infectious gastroenteritis and colitis	329143	10.80%	3.83	(2.86)
M53	Other and unspecified dorsopathies	169131	5.55%	8.18	(6.94)
J40	Bronchitis, not specified as acute or chronic	111004	3.64%	6.10	(3.87)
J02	Streptococcal pharyngitis	81368	2.67%	4.90	(3.04)
J01	Acute sinusitis	67769	2.22%	5.75	(3.65)
J20	Acute bronchitis	59194	1.94%	6.06	(3.80)
M54.5	Low back pain	58922	1.93%	7.16	(6.28)
B34.8	Other viral infections of unspecified site	51999	1.71%	4.50	(2.95)
J03	Acute tonsillitis	40135	1.32%	5.34	(3.15)
M54.4	Lumbago with sciatica	37977	1.25%	9.60	(7.85)
J06	Acute upper respiratory infections	37363	1.23%	5.63	(3.62)
A08.5	Other specified intestinal infections	37190	1.22%	3.47	(2.69)
K29	Gastritis and duodenitis	27216	0.89%	4.63	(4.49)
R51	Headache	26752	0.88%	3.95	(4.42)
G43	Migraine	26216	0.86%	2.50	(2.96)
F32	Major depressive disorder, single episode	24228	0.79%	11.62	(9.80)
J04	Acute laryngitis and tracheitis	23862	0.78%	5.45	(3.45)
N39.0	Urinary tract infection, site not specified	20710	0.68%	4.65	(4.11)

Notes: This table presents all ICD-10 codes with more than 20,000 cases in the data, including sample means and standard deviations of sick leaves certified based on these diagnoses.

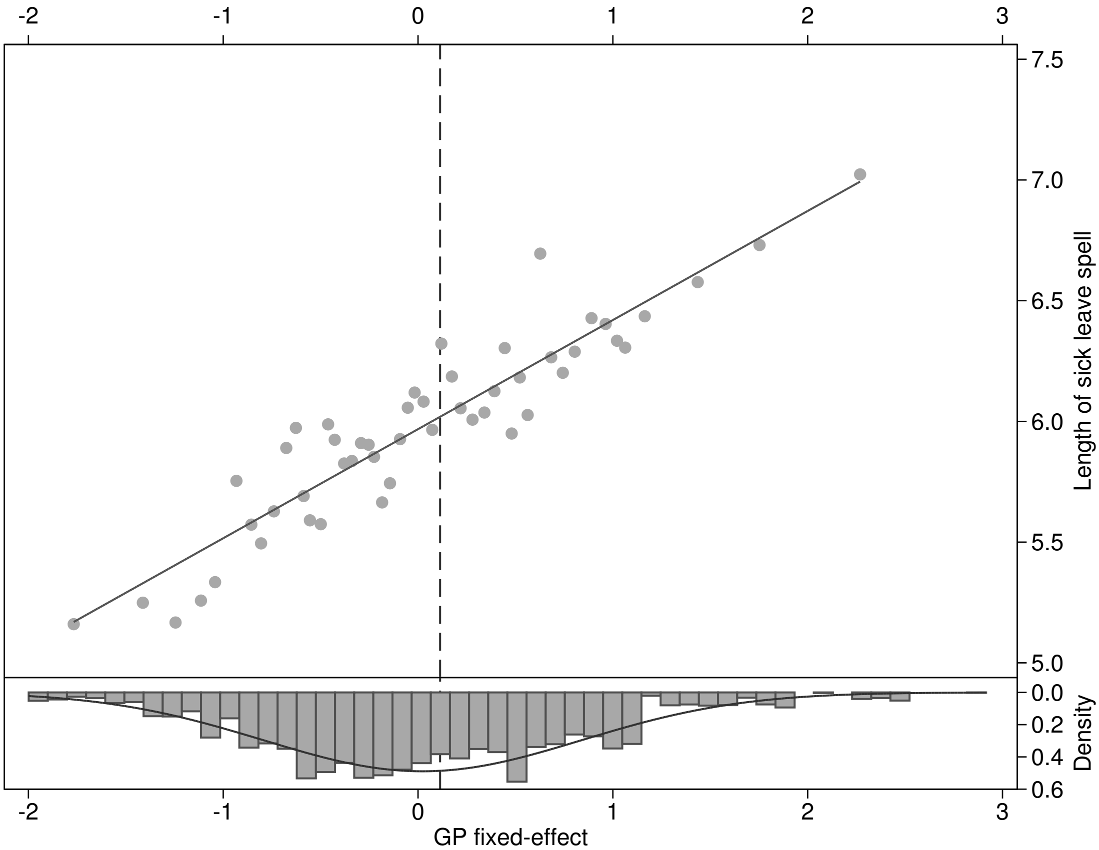

FIGURE A.2 — This graph illustrates the relationship between estimated GP fixed-effects $\hat{\psi}_d$ on the horizontal axis and sick leave durations n_k on the vertical axis. Due to the large sample size (N = 3,125,759), observations are grouped into 100 equally sized bins. Within each bin, means of E_k and n_k are calculated and then plotted in the upper graph. The solid line indicating fitted values is calculated based upon all observations in the data. Furthermore, the distribution of estimated GP fixed-effects $\hat{\psi}_d$ along with a hypothetical normal distribution are plotted underneath the graph (65,401 observations whose fixed-effect lies outside the interval $[-2, 3]$ are not shown for presentational reasons). The dashed line indicates the sample mean of $\hat{\psi}_d$.

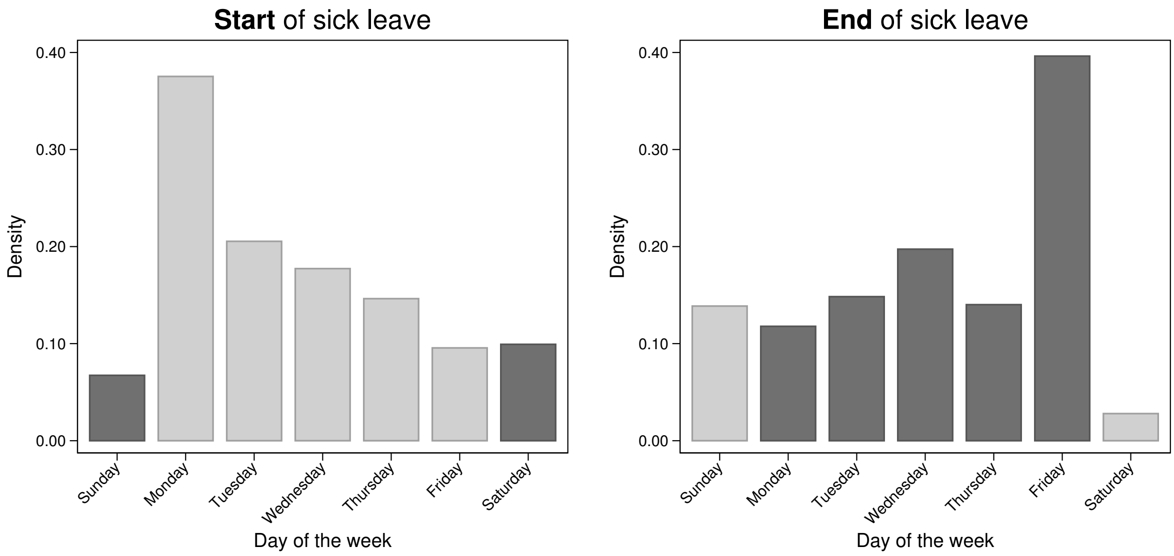

Figure A.3 — These graphs illustrate the distribution of both the first week day of the sick leave (left graph) and the last week day of the sick leave (right graph).