On the Design of Paid Sick Leave: A Structural Approach

Victoria Barone*
University of California Los Angeles

November 1, 2022

JOB MARKET PAPER
[Click here for the latest version]

Abstract

What is the optimal paid sick leave system? To answer this question, I combine individual-level data on paid sick leave claims with a model of sick pay insurance provision. I start by providing evidence that workers respond to the monetary incentives induced by the benefit scheme—i.e., they engage in moral hazard behavior. I also document that workers respond to nonmonetary shifts in the temptation to shirk induced by variation in the day of the week on which a worker falls sick. I use these patterns to inform a model of sick pay insurance. In the model, risk-averse workers face a health risk and decide how many days to be on leave. Workers are insured by a risk-neutral social planner who chooses the optimal contract to maximize social welfare considering workers' behavioral responses. Social welfare is a function of workers' utility and the potential production losses induced by sick pay provision. The main empirical challenge in estimating the model is to disentangle the underlying distribution of health from workers' preferences. To overcome this challenge, I combine the individual-level data on sick pay utilization with detailed medical assessments of recovery times associated with each health condition. This strategy allows me to construct the underlying distribution of health without imposing parametric assumptions. To estimate workers' preference parameters, I exploit the day of the week on which a sick leave claim is filed as a quasi-exogenous shifter of the temptation to shirk as the main source of variation. Finally, I use the estimated model to derive the optimal sick pay contract and estimate the welfare gains from its implementation. I find that relative to the current system, the optimal system would provide more insurance for short-term sickness and less insurance, i.e., lower replacement rates, for longer sickness spells. I estimate that workers are willing to give up 1.53% of their earnings to be insured under the optimal policy.

---

*Department of Economics, University of California Los Angeles, 8276 Bunche Hall, LA, CA 90095. Email: mvbarone@g.ucla.edu. Website: baronevictoria.com. I am very grateful to Maurizio Mazzocco, Adriana Lleras-Muney, Michela Giorcelli, Martin Hackman, Daniel Haanwinckel, and Laura Wherry for their continual support and guidance throughout this research project. I thank Martha Bailey, Alexander Bloedel, Natalie Bau, Dora Costa, Felipe Goncalves, Juliana Londoño-Velez, Yotam Shem-Tov, and Bernardo Silveira for their comments and helpful conversations. I thank Fernanda Rojas-Ampuero, Kirill Ponomarev, Nicole Gorton, Jacob Kohlhepp, Matías Morales, Ariadna Jou, Carolina Arteaga, Shooshan Danagoulian, Rajeev Dehejia, Christopher Flinn, Sherry Glied, Dmytro Hryshko, Andrew Jordan, Tim Layton, Catherine Maclean, Daniel Waldinger, Nicolas Ziebarth, and participants in the Albert Family Fund Proseminar in Applied Microeconomics at UCLA, Micro Lunch at NYU Econ, Wagner Research Colloquium, WEAI Graduate Student Workshop, 2022 ASHEcon Annual Conference, Young Economist Symposium (YES), EGSC at Washington University in St. Louis for insightful discussions. This project was supported by the Lewis L. Clarke Fellowship Fund and the Stan Ornstein Summer Mentored Research Fellowship Program. I benefited from facilities and resources provided by the California Center for Population Research at UCLA (CCPR), which receives core support (P2C-HD041022, NICHD); and the Wagner School of Public Service.

I Introduction

Social insurance programs offer valuable protection against a broad range of risks that could be detrimental to individuals' well-being, such as health deterioration that limits one's ability to work. In particular, paid sick leave provides income replacement for workers who suffer from short-term impairments caused by non-work-related sickness (e.g., common flu).[1] If adequately designed, paid sick leave can be greatly beneficial. It allows workers to meet their personal health needs and smooth consumption. Nonetheless, the availability of sick pay could induce workers to request more sick days than should be assigned based on their health. This raises the following question: What is the optimal paid sick leave system?

The main contribution of this paper is to answer this question. To do so, I proceed in three steps. First, I use detailed data on paid sick leave claims to document how workers' characteristics and the institutional details of the sick pay system—e.g., the presence of deductibles—determine sick pay utilization. Second, I proposed and estimated a model of sick pay provision. This exercise gives me key inputs to derive the optimal paid sick leave system: the value of risk protection, the costs of insurance provision in terms of moral hazard, and the underlying distribution of health shocks. Finally, I use these estimates to determine the replacement rates that characterize the optimal sick paid system. I find that the optimal system features a low replacement rate for short claims, i.e., up to three days long claims are partially insured, with most of the cost on the worker side. Longer sickness spells are covered at a higher rate. The optimal system offers less coverage relative to the current system.

I study this question in Chile, which is an ideal setting for this research for several reasons. First, it has a comprehensive paid sick leave system that covers all workers and features only one plan designed by the central government. Thus, workers do not choose their sick leave coverage.[2] The Chilean system provides no coverage for sickness lasting three days or fewer. This nonpayable period works like a deductible that resets with every new sick leave spell and is a common mechanism among sick leave programs.[3] Starting

[1]While closely related to worker's compensation programs, which provide income replacement and medical benefits in case of work-related sickness, and disability insurance programs, which provide income replacement in case of permanent or long-term impairments to working ability, paid sick leave programs offer protection against the risk of contracting a disease that impairs workers for a short period and has a foreseeable recovery.

[2]This alleviates adverse selection concerns. If workers could choose their sick pay coverage, we could expect individuals with preferences for more absences to self-select into plans with more generous provisions. The presence of adverse selection would result in an upward bias in the estimates of the moral hazard responses.

[3]These resettable deductibles are similar to those used in automobile or homeowners insurance: Sepa-

on the fourth day, there is full coverage of each missed day, i.e., the replacement rate is one. If the sick leave spans 11 days or more, the nonpayable period is reimbursed; this implies that the average replacement rate varies with the duration of a claim and jumps discretely at 11 days.[4]

The second advantage of this setting is that Chile has greatly detailed administrative data. I observe the universe of workers insured by the government eligible to file a sick leave claim between 2015 and 2019 and their utilization of sick leave benefits.[5] This database includes rich demographic information at the worker and sick-leave claim levels. In particular, I observe the exact beginning and end dates and the primary diagnosis related to a sick leave claim at the International Classification of Diseases 10th revision (ICD-10) four-digit level. I combine the claims data with medical assessments from the Peruvian Handbook of Recovery Times (EsSalud, 2014). This handbook specifies the average recovery times for 2,763 unique disease codes at the ICD-10 four-digit level. These recommendations are adjusted based on workers' gender, age, and occupation.

Exploiting these data, I document three facts that provide qualitative motivation for the model and serve as quantitative targets in the estimation. First, workers' sick leave claim utilization varies with age and occupation. For example, on average, workers aged between 55 and 64 use an extra 2.72 days per year relative to their younger counterparts. Similarly, compared to white-collar workers, workers in blue-collar occupations use, on average, 1.15 more days. These patterns could reflect differences in the underlying distribution of health shocks. Thus, in the estimation of the model, I allow the distribution of health shocks to vary with age and occupation.

Second, I provide evidence that workers respond to the financial incentives induced by the benefit scheme. To do so, I construct an underlying distribution of recovery times—i.e., an underlying health distribution—and use it as a counterfactual to estimate the excess mass on an 11-day-long claim. The underlying distribution of health relies on the handbook's recommendation of how many rest days a worker needs to recover from a disease. I estimate that 11-day-long sick leave claims are 4.55 percentage points more likely than what the underlying distribution of health predicts. I use this discontinuity to assess the model performance and find that the proposed model can reproduce the excess mass observed at 11 days.

rate deductibles apply to each loss. Many European paid sick leave systems have a similar deductible. See Marie and Castello (2022) on the case of Spain and Pollak (2017) on the Italian experience.

[4]Panel (a) of Figure 1 presents days paid as a function of days on leave for claims of different duration. Panel (b) shows the average replacement rate.

[5]This group accounts for about 70% of the Chilean workforce and is representative of the average Chilean worker.

Third, I show that workers respond to nonmonetary shifts in the temptation to engage in shirking behavior.[6] To do so, I exploit the data on exact dates when sick leave claims are filed. I argue that the temptation to shirk varies with the day of the week when a worker falls sick. For example, the incentives to file a two-day-long sick leave claim on a Thursday differ from the incentives to file on a Tuesday. I propose the following exercise: I fix the duration of a sick leave claim and inspect the share of claims filed on each day of the week. I find an excess mass on combinations of days of the week and durations that allow the worker to extend her leave through the weekend. I refer to such combinations of durations and start days as "weekend-streak combinations". I document that workers are 12% more likely to file a weekend-streak claim than to file a sick leave claim of the same duration on any other day of the week. To capture this empirical regularity, the model allows workers' behavior to vary with the day of the week of a sick leave claim.

I use this evidence to develop and estimate a model of sick pay provision. The model has two main agents: workers and a social planner. Workers are risk-averse expected utility maximizers and choose their sick leave utilization. When choosing the demand for sick days, the worker trades off the utility cost of working while sick with the consumption loss from missing work when taking up sick leave. The latter depends on sick leave generosity. The former is a function of two key parameters: (i) workers' valuation of time outside work to recover from health shocks or engage in leisure and (ii) workers' propensity to engage in moral hazard behavior. Additionally, utility varies with the day of the week when a sick leave claim is filed to capture the empirical fact that the temptation to shirk varies at this level. Since, the provision of sick pay impacts workers' incentives it changes the pool of workers showing up to work. For example, a more generous sick coverage policy could reduce labor supply and have a negative impact on aggregate welfare via production losses. The optimal policy design must take these effects into account. I capture this channel by adding a production side to the model where I (i) differentiate between healthy days worked and sick days worked and (ii) assume that sickness affects workers' productivity.

A risk-neutral social planner offers a sick pay contract to maximize aggregate welfare taking workers' choices as given. The optimal contract balances the benefits of risk protection with the cost associated with moral hazard. A more generous sick pay scheme would increase workers' well-being but provide incentives to overstate sickness, resulting in lost production and increased program costs. On the other hand, if a contract is not generous enough, some workers would work while sick, facing a utility cost and reducing

[6] In the sick pay literature, shirking refers to innapropiate leave-taking [expand]. (Cronin, Harris and Ziebarth, 2022) uses this definition.

aggregate welfare. Thus, the optimal level of benefits depends on risk preferences, workers' behavioral responses, production losses, and the distribution of risks. The principal empirical focus of this paper is to quantify these elements.

To determine the optimal system, I proceed in two steps. The first step concerns workers' sick pay utilization choices. In this step, I recover a vector of preference parameters—time valuation and compliance costs—from workers' behavior. The second step determines the optimal paid sick leave system—the replacement rates that maximize total welfare. The validity of this approach relies on the fact that the worker's problem can be viewed as a two-stage problem. Once the health shock is realized, workers optimally choose their sick pay utilization. Risk preferences do not affect the utilization decision since the uncertainty has been resolved when this decision is made. Thus, the focus of the first step. Nonetheless, workers' expected utility depends on their risk preferences: more risk-averse workers would prefer a contract with more coverage. This is accounted for in the second step. The main strength of this approach is that I only rely on workers' observed decisions to estimate the model and do not need to assume that the current sick pay plan is optimal.

I estimate the model of workers' behavior by the simulated method of moments (SMM). The main empirical challenge is to disentangle the underlying distribution of health from the distribution of workers' preferences. To overcome this challenge, I build the underlying distribution of health exploiting the Peruvian Handbook of Recovery Times recommendations and the observed diagnoses. This approach has two advantages: (i) it provides an objective measure of recovery times constructed outside the structure of the Chilean system, and (ii) it does not impose parametric assumptions on this distribution. The empirical distribution of health states incorporates observed heterogeneity across workers: I allow for variation in age and occupation. That is, the same diagnosis has age- and occupation-specific associated recovery times. On average, older workers and workers employed in blue-collar occupations are assigned longer recovery times. In estimating the model, I also allow an arbitrary correlation between health states and workers' income to capture that wealthier workers tend to have better health and could require shorter absences.

To recover the distribution of workers' preference parameters, I employ the day of the week when a sick leave claim is filed as a quasi-exogenous shifter of the temptation to shirk as the main source of variation. First, the excess of weekend-streak sick leave claims informs how workers' utility from a sick leave claim of the same duration varies with the day of the week on which the claim is filed. Second, I consider workers with similar characteristics and the same *assigned* recovery time—i.e., I hold workers' health, age, and occupation fixed—and compare their demand for sick pay across days of the week. I start by computing the share of claims filed for a duration that matches the assigned recovery time and compare this figure with the share of claims filed for an extra day. This difference is informative on how costly it is for individuals to ask for an extra day of leave. I restrict this comparison to claims filed for a combination of durations and days of the week representing a weekend streak. For example, I compare the share of two-day-long claims filed on a Thursday with that of three-day-long claims filed on a Wednesday for workers assigned two days of recovery. The larger the difference is, the more costly it is for workers to ask for an extra day of leave. Comparing these shares keeps the incentives for shirking behavior fixed since every combination implies that workers would be on leave through the weekend. These comparisons inform the distribution of compliance costs. Lastly, to learn the distribution of workers' time valuation, I construct the ratio of leisure to (a measure of) consumption from the claims data.

The estimation of the model incorporates observed heterogeneity across workers in the valuation of time outside work and the propensity for moral hazard behavior. Heterogeneity across the valuation of time outside work reflects variation in opportunity costs from missing work to recover from a disease—e.g., due to workers' role in the firm—or variation in tastes for leisure relative to consumption. Heterogeneity in the propensity for moral hazard behavior reflects variation in workers' preferences over behaving as expected or revealing their "true" health status. Additionally, the model of workers' behavior allows for heterogeneity in how workers perceive their sickness. That is, workers who suffer the same health shock can be affected by it differently. While the parameter that governs this perception is not separately identified, the derivation of the optimal policy does not require its identification. This derivation rather relies on workers' responses to the incentives generated by the provision of sick pay.

The estimated model provides a good fit for the targeted and nontargeted moments. Regarding the model's internal validity, [add a short mention on the model's match] . I exploit the discontinuity at 11 days—a nontargeted moment—to assess the model performance. The model predicts that if a worker realizes a health state just under 11 days, she will take advantage of the proximity to the full-coverage region and fake her type to gain full coverage. That is, the proposed model can reproduce the excess mass at 11 days.

I use the estimated model to derive the optimal sick pay policy, i.e., to determine the replacement rates that maximize aggregate welfare. The optimal policy differs from the current system in three key ways. First, it offers partial replacement, with an average replacement rate of 0.36, for claims of up to three days. This shift increases the utility of sick workers who would not take sick leave under the current system but do under the optimal policy. At the same time, partial coverage constrains moral hazard since most of the cost of those absences is faced by workers.

Second, the optimal policy eliminates the discontinuity at 11 days and exhibits a higher average replacement rate between 4 and 10 days. Doing so curbs the cost of the behavioral responses to the program incentives and provides more risk protection. Implementing the optimal scheme would shift the distribution of sick leave duration: workers would be more likely to file sick leave claims of between 8 and 10 days and less likely to file claims for 11 days relative to the corresponding probabilities under the current Chilean system.

Third, the optimal policy does not offer full replacement for sick leave claims longer than 11 days. The average replacement rate is increasing, as in the current system, but it is less generous for longer claims. Taken together, these changes in the replacement rate reflect that the workers value a contract that offers more protection for shorter claims to smooth consumption across different health states. I estimate that workers are willing to give up 1.53% of their earnings to be insured under the optimal policy.

This paper contributes to several areas of the economics literature. First, it contributes to a large body of literature on public insurance programs. This literature has modeled the trade-offs between protection against risk and moral hazard present in unemployment risks (Hopenhayn and Nicolini, 1997; Chetty, 2008; Hendren, 2017), disability and retirement risks (Gruber, 2000; Low and Pistaferri, 2015), healthcare risks (Cutler and Zeckhauser, 2000; Einav, Finkelstein and Cullen, 2010; Handel, Hendel and Whinston, 2015; Marone and Sabety, 2022), and work-related injuries (Powell and Seabury, 2018; Cabral and Dillender, 2020). This paper contributes to this literature by being the first study to propose a theoretical framework for designing the provision of paid sick leave and quantifying the welfare gains from its implementation.

The closest programs to paid sick leave are disability insurance and workers' compensation. These three programs condition benefits on a difficult-to-verify state: the true impairment of working ability due to health deterioration. Thus, the trade-offs considered in the design of disability insurance and workers' compensation are relevant to the design of paid sick leave. Nonetheless, disability insurance and workers' compensation programs target specific groups of workers—elderly people and workers especially vulnerable to accidents—and provide protection for different set of health shocks—more permanent and more severe shocks. The nature of the health shock insured by paid sick leave provision—short spells of non–work-related illness—implies that virtually every worker could benefit from the risk protection under the program and demand paid sick leave.

This paper is close to Maclean, Pichler and Ziebarth (2020), who evaluate the labor market effects of sick pay mandates in the United States and extend the Baily–Chetty framework of optimal social insurance to assess the welfare consequences of mandating sick pay. Their framework allows researchers to study the effects of policies that vary the share of employees eligible for the benefit. This paper differs in two critical dimensions. To begin with, I propose a structural approach to conducting welfare analysis. This approach does not rely on the assumption that policy changes are marginal.[7] Relaxing this assumption is important since it allows the optimal policy to differ freely from the actually implemented policy, i.e., it allows for non-marginal policy changes.

Second, this paper relates to the empirical literature on sick pay insurance. Exploiting arguably exogenous variations, the literature has documented a positive response of sick pay utilization to increases in benefit levels (Johansson and Palme, 2005; Ziebarth, 2013; De Paola, Scoppa and Pupo, 2014; Ziebarth and Karlsson, 2014; Pollak, 2017; Cronin, Harris and Ziebarth, 2022; Marie and Castello, 2022). This paper goes beyond workers' responses to policy changes and proposes a framework and an empirical strategy to quantify the welfare effects of policy changes.

Additionally, this paper is the first to use administrative data on sick leave claims at the individual level.[8] These data allow me to study daily leave-taking behavior and estimate an individual demand for sick pay. In addition, these data are less prone to measurement error. Many papers have used survey questions that ask respondents how many days of work they have missed due to illness in a reference period. The use of survey data raises the usual measurement error issues with self-reported recall data and prevents researchers from distinguishing the incidence of absences from their length. Observing the length of absences is a crucial input for quantifying moral hazard responses, as workers could extend their absences to obtain more sick pay.

This paper also contributes to the literature on variation in shirking behavior across days of the week. Card and McCall (1996) and Campolieti and Hyatt (2006) provide evidence of a "Monday effect"—which refers to a spike in back injury and sprain claims on Mondays—among workers' compensation claimants. Thoursie (2004) shows that Swedish men were likelier to call in sick the day after popular skiing competitions were broadcast at night during the Winter Olympics in Calgary. Implementing a similar test, Cronin,

[7] For the validity of the sufficient statistics approach, the analyzed policy changes should be infinitesimal or at least close enough to infinitesimal for first-order approximations to be precise Kleven, 2021.

[8] Cronin, Harris and Ziebarth (2022) constructs a similar dataset for the Scott County School District (SCSD) in Kentucky, which allows a detailed study of teachers' use of paid sick leave. While the data structure is similar to the one used in this paper, I observe sick leave utilization regardless of workers' occupations. Marie and Castello (2022) also exploit administrative data for Spain, though their data are at the spell rather than the individual level. The data in this paper capture the sick leave choices of 70% of the universe of Chilean workers.

Harris and Ziebarth (2022) document that teachers in Kentucky are not more likely to use sick leave while Keeneland is in session, on Mondays following Super Bowls, or on days when the University of Kentucky men's basketball team plays in the NCAA tournament. I provide new evidence on shirking behavior across weekdays by exploiting the exact dates of sick leave spans.

This paper proceeds as follows. Section II describes the empirical setting that I study and the data. Section III presents the theoretical framework and discusses the optimal design of a paid sick leave system. Section IV presents the empirical implementation of the model. Section V presents the model estimates and main results. Section VI discusses the optimal policy. Section VII concludes.

II Background and Data

In this section, I discuss the Chilean health care and sickness insurance systems, focusing on the institutional features relevant to my analysis. I then present the data and patterns in the data that motivate my modeling choices.

II.A The Chilean Health Insurance System

In Chile, healthcare insurance providers serve two functions: (i) offer healthcare insurance contracts and (ii) administer the paid sick leave system. The healthcare insurance system is composed of a government-run healthcare insurance provider and a handful of private insurance providers.[9] The government-run healthcare insurance provider offers four plans, whose eligibility is based on monthly salary and household composition. The lowest-tier plan provides coverage for individuals with no income at no cost in public system hospitals. As income increases, workers would qualify for a higher-tier plan. This plan provides healthcare coverage in public system hospitals with low copayments and access to private healthcare institutions with high copayments.[10] Private insurance companies provide tiered plans with financial vertically differentiated coverage levels—similar to the Gold, Silver, and Bronze plans offered by Affordable Care Act exchanges in the US. The plans offered by private insurance companies allow beneficiaries to obtain healthcare from private healthcare institutions, which provide a higher quality of care than public institutions.

[9] These are called FONASA, for the Spanish name—*Fondo Nacional de Salud*—and ISAPRES—*Instituciones de Salud Previsional*—respectively.

[10] Plans are indexed by letters, where A is the lowest-tier plan and D is the highest-tier plan. The highest-tier plan has a 20% copayment in public system hospitals and vouchers to use health care providers who participate in the plan's network at a discounted price.

Workers are mandated to purchase health insurance and allocate at least 7% of their salary to a healthcare plan. Workers can choose their health insurance provider. In order to select one of the private providers workers might need to contribute a higher proportion of their salary to qualify for the healthcare plan of their choice.[11] In contrast, with the mandatory contribution a worker could select government-run healthcare insurance, in such case they would be enrolled in one of the four plans based on their monthly salary and household composition. For example, a single worker who earns USD $693 a month—the median salary in 2017—and chooses the government-run insurance system would be enrolled in the highest-tier plan; and could not choose lower tier plans with lower copayments.

In 2017, 73% of workers enrolled in plans offered by the government-run healthcare insurance system; the remaining 27% enrolled in plans offered by one of the private providers (see Panel A of Table A1). Workers enrolled in government-run plans have observable characteristics that would predict that they are more costly to insure: they are older, more likely to be women, and have lower salaries. In contrast, those enrolled in government-run plans are older, and their age distribution has a heavier right tail. For example, 31% of the workers insured by the private provider are between 25 and 34 years, and only 25% of workers insured by the government belong to this age group.

The second function of healthcare insurance providers is to *administer* the paid sick leave system. Insurers are in charge of receiving and screening sick leave claims and disbursing sick leave benefits. Insurers cannot design sick pay plans and must follow the rules set by the central government regarding eligibility criteria and benefits. This implies that workers do not choose their sick leave plan. Thus, the Chilean context provides a unique setting to study workers' choices over the utilization of sick pay benefits since it alleviates adverse selection concerns. If workers could choose their sick pay coverage, we could expect sicker individuals to choose plans offering more generous insurance coverage. Whereas this mechanism could be at play in the choice of healthcare insurance provider, conditional on this decision, sicker and healthier individuals, face the same sick pay coverage.

While the eligibility rules and benefits structure are independent of insurer choice, there are differences in how each provider applies these rules in practice. For example, panel B of Table A1 shows that the rejection rate by private insurers is almost three times as high as that of the government-run insurer. I consider these differences in leniency to be suggestive evidence that private insurers might have different motives—such as min-

[11] Plans offered by private insurers are highly regulated. These insurers can set prices based on observable characteristics—including age and until April 2020 sex—and risk factors.

imizing sick leave payments—when screening sick leave claims. My empirical analysis focuses on workers enrolled in the government-run health insurance system—they represent about 73% of all Chilean workers. The main reason for this choice is that this paper focuses on the provision of paid sick leave as a social insurance system, which is closer to the behavior of the government-run healthcare provider.

II.B The Chilean Paid Sick Leave System

The Chilean paid sick leave system gives employees the right to call in sick and receive sick pay due to short-term, non-work-related sickness—e.g., the common flu or back pain.[12] Workers can use sick leave to meet their health needs but not to care for family members. The eligibility criteria to claim a paid sick leave requires that workers (i) have been enrolled in the social security system for six months and (ii) have made contributions to the health insurance system for three months.[13] Upon falling sick, workers must get a physician's certificate of their sickness, which states the primary diagnosis and the number of days the physician considers the worker will need to recover from the disease. This certificate is reviewed by an insurance office that decides whether the sick leave claim is (i) approved with no changes, (ii) approved with a reduction in its length, or (iii) denied.

Sick leave payments are a function of sick leave duration subject to a maximum salary.[14] For workers with a salary below the maximum, benefits are computed as follows: The benefit scheme exhibits a non-payable period of 3 days, i.e., the replacement rate for the first three days of a sick leave span is zero.[15] This non-payable period works like a deductible that resets for every new sick leave span.[16] Starting on the fourth day, there is full coverage of each missed day—i.e., the replacement rate is one. If the sick leave spans 11 days or more, the non-payable period is reimbursed. That is, claims with an 11-day or longer duration are fully covered.[17] Panel (a) of Figure 1 presents days paid as a function

[12]The sickness insurance system aims to provide risk protection from impairments to work that are temporary and where full recovery is foreseeable. A separate program provides disability insurance to workers in the case of permanent impairments to work.

[13]These restrictions are independent of job tenure.

[14]In my sample, less than 1% of workers earn above this threshold, see Figure A1. I exclude these workers from the analysis.

[15]Cid (2006) documents that the origin of the 3-day non-payable period dates to a regulation implemented in 1952 that aimed to prevent abusive behavior, which has not been revised since.

[16]These resettable deductibles are similar to those used in automobile or homeowners insurance: Separate deductibles apply to each loss.

[17]If a worker files two (or more) consecutive claims, they are treated as one claim for the computation of benefits. To be consistent, I treat these claims as one in the analysis. Appendix table A2 presents counts and summary statistics of sick leave claims and sick leave spells.

of days on leave for claims of different duration. Reimbursement of the non-payable period after 11 days implies that the average replacement rate jumps discretely at 11 days, and it is non-constant (see panel (b) of Figure 1).

II.C Data

I exploit unique administrative data on sick leave claims matched to enrollment data for workers insured by the government-run healthcare system. These restricted-access data were provided directly by the government-run healthcare insurance office and cover the period 2015-2019. I combine these data with medical assessments from the Peruvian Handbook of Recovery Times.

The enrollment dataset covers the universe of individuals enrolled in government-run healthcare insurance regardless of whether they have filed a sick leave claim. I observe individuals' demographic and economic characteristics: sex, age, annual earnings, and health indicators for chronic conditions.[18] Additionally, I observe individuals' residence ZIP codes and the health insurance plan assigned to them. The latter allows me to exclude individuals enrolled in the lowest-tier plan from the analysis, as these individuals are not active in the labor market.

In the claim dataset, I observe detailed information about each sick leave claim: start and end dates, prescribed days on leave, the primary diagnosis—coded following the 10th revision of the International Classification of Diseases (ICD-10)—physician identifiers, and the amount received for paid sick leave. I also observed the occupation the workers is employed at the moment of filling a sick leave claim.

The Peruvian Handbook of Recovery Times (EsSalud, 2014) specifies an average recovery time for 2,763 unique disease codes at the fourth digit level of the ICD-10. Crucially, these recommendations are adjusted based on workers' gender, age, and occupation. Table A3 provides an example of the average recovery times for three common diagnoses—lumbago with sciatica, common cold, and infectious gastroenteritis—and the correction factors proposed by the handbook. The main advantage of exploiting this external source of data is that it provides an objective measure of recovery times constructed outside the structure of the Chilean system. That is, it is not affected by the brackets used in the paid sick leave benefit function.

I combine these three sources to construct a claim-level dataset with detailed infor-

[18]These conditions are: cerebral vascular accident, Alzheimer, juvenile arthritis, rheumatoid arthritis, bronchial asthma, lung cancer, diabetes, chronic obstructive pulmonary disease, chronic kidney disease, arterial hypertension, acute myocardial infarction, leukemia, lymphoma, lupus, multiple myeloma, and HIV.

mation on workers' demographic characteristics, leave-taking behavior, and the average recovery time. My primary measure of leave-taking behavior is the duration of a sick leave claim filed on a given day of the week. I use the disease code recorded in each claim to assign average recovery days based on the handbook. I construct these measures at the sick-leave-spell level, i.e., I consider consecutive claims as one claim. Thus, the unit of analysis is the same as the one used to compute sick leave benefits. Appendix table A2 presents counts and summary statistics of sick leave claims and sick leave spells.

To arrive at the analytic sample, I impose two restrictions. First, the estimation sample includes claims from private-sector male workers aged 25 to 64. This is a demographic group with high labor market participation rates. Although women sick leave-taking behavior is of high interest in the design of sick leave programs, women have much lower participation rates than men. For example, in Chile, women have more than 20 percentage points lower labor force participation rates—52.6% for women at the beginning of the sample period and 73.2% for men. Thus, a model of sick leave-taking behavior that explains women choices would also require incorporating their decision to participate in the labor market. Nonetheless, abstracting from the participation decision simplifies the model estimation.

Second, the estimation sample includes claims from a subset of diseases. I exclude mental-health diagnoses because their filing process is more cumbersome than the one for non-mental-health claims.[19] That is, in this paper I focus on non-mental-health sick leave claims. Among non-mental-health diagnoses, I exclude diagnoses for which is hard to assign a recovery time. For example, I exclude claims with codes corresponding to neoplasms. Table A4 lists the conditions included in the analysis and the share of claims recorded under each diagnosis. Appendix B. provides additional details. The final sample includes 90.19% of all non-mental-health sick leave claims.

II.D Descriptive Evidence

Summary statistics. Table 1 presents summary statistics for all the workers in the sample and for those who have used sick pay during 2017. I split the last group based on the type of disease—included in the analysis or not—and duration of the sick leave claim. Almost 20% of Chilean workers filed a non-mental-health-related sick leave claim in 2017. The average worker in the sample is 44 years old, and the average worker who has used sick pay is about the same age (column 1 vs. column 2 comparison). Nonetheless, the average claimant has a higher salary than the average worker, and this difference is statistically

[19] For example, these claims must be certified by a psychiatrist and require a comprehensive medical assessment at the time of filing.

and economically significant.

To better understand the differences between workers who have filed sick leave claims and those who have not, Table 2 presents characteristics of workers who have used sick leave benefits based on the duration of the claims. I group workers that had filed (i) at least one claim with a duration of up to 3 days, (ii) at least one claim with a duration between 4 and 10 days, and (iii) at least one claim with a duration of 11 days or longer. Shorter sick leave claims are associated with younger workers with higher average wages, who are also less likely to have chronic conditions. This pattern is compatible with the 3-day waiting period reducing the likelihood that lower-earning workers file a sick leave claim. Additionally, the association between workers' age and chronic conditions prevalence is consistent with older workers experiencing more severe conditions than their younger counterparts.

Determinants of sick leave duration. Figure 2 shows the distribution of the duration of sick leave claims of up to 29 days. Three main patterns characterize the distribution of days on leave. First, about 26.54% of sick leave claims have a duration of up to 3 days. Sick leave claims lasting between 4 and 10 days explain 41.06% of claims. Second, there is an excess of mass or bunching at 11 days. This coincides with the most significant jump in the average replacement rate, starting at 11 days long claims, workers are fully reimbursed for the time off work. This jump incentivizes workers to extend their leaves to enter the "full" insurance region. At the same time, it provides more generous coverage for more severe health shocks. Third, there is rounding at multiples of 5 and 7 days. This rounding is consistent with physicians being more likely to write recovery times that match with a work-week—five days—or a calendar week.

Figure 3 shows the histogram of sick leave claims duration by workers' characteristics. I group workers into eight groups or bins defined based on age and occupation type: blue-collar and white-collar occupations.[20] Conditional on workers' occupation, older workers require a higher proportion of long sick leave claims. Their distribution of sick leave claims is shifted toward the right relative to the distribution of younger workers (comparison across rows of Figure 3). This pattern is consistent with workers requiring more time to recover from the same conditions as their age and workers suffering more severe underlying conditions. Comparisons across occupations for workers in the same age group indicate that claims from blue-collar workers are longer on average, with a

[20]Blue-collar worker refers to an individual who performs manual labor. For example, operators, assemblers, and laborers are considered blue-collar workers. White-collar worker refers to an individual who performs professional, desk, managerial or administrative work. For example, sales representatives are considered white-collar workers. Table A6 details the occupations classified as blue-collar and white-collar.

smaller share of claims or up to 3 days. This comparison suggests that differences in the underlying distribution of health could be correlated with occupation type. Motivated on this results, I allow the underlying distribution of health to vary with workers' age and occupation in the estimation of the model.

Workers' responses to the benefit scheme. I provide evidence that workers respond to the discontinuity in the replacement rate at 11 days. I start by inspecting the raw data; Figure 4 reproduces the distribution of days on leave from the data and the counterfactual distribution of days on leave. The latter is constructed assigning to each sick leave the recovery time suggested by the Peruvian Handbook of Recovery Times, adjusted by workers' age and occupation. This exercise assumes that the diagnoses recorded for each claim reflect workers' "true" condition. That is, I abstract from potential moral hazard responses in the diagnoses.

Workers' behavior by day of the week. The temptation for shirking behavior varies with the day of the week a worker falls sick. For example, the incentives for filing a two-day-long sick leave claim on a Thursday differ from the incentives induced by filing a two-days-long claim on a Tuesday. The first combination implies four continuous days on leave while the second combination implies two days. I refer to the first type as "weekend-streak" combination.

Figure 5 shows the share of sick leave claims filed each day of the week. For each day of the week, I compute the share of sick leave claims, indexed by j, of duration s that are filed that day . That is:

$$\text{share}_s^{day} = \frac{\sum_j \mathbb{1}\{dow_j = day, s_j = s\}}{\sum_j \mathbb{1}\{dow_j = day\}} .$$

Consider 1-day-long sick leave claims; the share of claims filed on Friday is about three times higher than the share of claims filed on any other day of the week (see Panel (a) of Figure 5). This pattern is present for one- to five-day-long claims. Crucially, when inspecting 7-day-long claims, the share is constant across days of the week.[21] I document that workers are 12% more likely to file a "weekend-streak combination" relative to sick leave claims of the comparable duration on any other day of the week. I leverage this

[21]Claims of duration longer than six days exhibit a similar pattern. I use seven days as a reference point since the share of these claims in the data is greater than the share of 6-day-long claims. Appendix Figure A4 presents the distribution of the share of sick leave claims by day of the week for claims with a duration between 8 and 15 days, pooled in 2-day groups.

variation in the temptation for shirking to estimate workers' preference parameters.

III Theoretical Framework

In this section I present the model of paid sick leave provision that I use to derive the optimal sick leave insurance contract. First, I model the choices of an expected utility maximizer worker who faces uncertainty about her health and her ability to work. Second, I describe how workers' choices and provision of sick pay affect production. Third, I discuss the social planner's problem. In the rest of this section, I omit i subscripts to simplify notation and present the baseline version of the model. I later describe how individuals might vary across (i) their distribution of health shocks, (ii) preferences over time outside work, and (iii) their propensity for moral hazard.

III.A Workers

Workers are subject to a stochastic health shock (θ,dow), drawn from a distribution $G(\theta,dow)$, where θ represents the number of days a worker is sick, and dow indicates the day of the week a worker falls in sick. I assume that θ is discrete and bounded between zero and M and that higher values of θ are associated with longer sickness spans.[22] The sickness distribution $G(\theta)$ accumulates positive mass in the no-sickness realization; i.e., the value of zero for θ corresponds to the healthy state.

Sick pay utilization. Upon realizing the health shock (θ,dow), the worker decides her sick pay utilization to maximize her utility. I assume the worker derives utility over consumption (c) and time outside of work (s), given her budget constraint. The budget constraint is: $c = w(M - s) + wB(s)+$, where w is the daily wage rate, M is the number of workable days in a month which the worker takes as given and $B(s)$ represents the sick pay transfer function.[23] I assume that $B(s)$ is a piece-wise linear function, with marginal replacement rates (b_j) constant for sick leave claims in a duration bracket $[\underline{s},\bar{s}]$. Worker's utility takes the following form:

$$u(s;\theta,dow) = w(M - s) + wB(s) + \phi(s_l(s;dow) - \theta) - \phi f(s - \theta) + \phi q \, \mathbb{1}\{\text{weekend}\}.$$

The last three terms represent utility from time outside work. The preference parame-

[22]The sickness level is bounded to capture the fact that paid sick leave insurance aims to provide risk protection from impairments to work when full recovery is foreseeable. I focus on sick leave claims for up to 30 days in the empirical application.

[23]For individuals working full time, M is the number of workdays in a month.

ter ϕ reflects the opportunity cost of time away from work relative to the time allocated to consumption. The term $(s_l(s; dow) - \theta)$ captures the utility cost of sick work, when $s_l < \theta$, or the gains from days away from work, when $s_l > \theta$. In this expression, s_l indicates business days, a function of total days on leave and the day of the week a sick leave claim starts.[24]

The function $f(s - \theta)$ equals zero if the difference between s and θ is non-positive and it takes positive values otherwise. This function increases in ($s - \theta$) and captures the cost of filing a sick leave claim for duration above the worker's health state. Thus, the parameters of the compliance cost function govern a worker's moral hazard behavior. For example, they reflect the worker's willingness to engage in doctor shopping—the effort the worker could exert to find a physician who would sign off on a longer leave—or the cost of being caught filing an unjustified sick leave claim. These mechanisms are captured in a reduced form, i.e., I do not model the specific action that workers take to engage in moral hazard.

The term $\phi q \ \mathbb{1}\{\text{weekend}\}$ captures the extra utility a worker derives when the sick leave claim has a duration that allows the worker to not return to work until after the weekend. The indicator variable $\mathbb{1}\{\text{weekend}\}$ is defined as follows[25]:

$$\begin{aligned}
\mathbb{1}\{\text{weekend}\} &= 1 \text{ if } dow = \text{Monday and } \ s = 5 \\
&= 1 \text{ if } dow = \text{Tuesday and } \ s = 4 \\
&= 1 \text{ if } dow = \text{Wednesday and } \ s = 3 \\
&= 1 \text{ if } dow = \text{Thursday and } \ s = 2 \\
&= 1 \text{ if } dow = \text{Friday and } \ s = 1 \\
&= 0 \text{ otherwise}
\end{aligned}$$

Workers choose sick pay utilization by trading off the cost of a day away from working $w(1 - B'(s_c))$ with its net gain. This net gain depends on the day of the week and the duration of the claim. An additional day on leave beyond the worker's sickness level (i) lowers utility by increasing the compliance cost term in $\phi f'(s - \theta)$, (ii) increases utility

[24] For example, two days long sick leave claim that starts on a Monday represents two business days away from work, while a sick leave claim that starts on Friday implies one day away from work. See Appendix Table A10

[25] The definition of the indicator variable $\mathbb{1}\{\text{weekend}\}$ only considers the extra utility for sick leave claims with a duration of up to 5 days. A more general definition will be to have $\mathbb{1}\{\text{weekend}\}$ equal to one for each sick leave claim that ends on a Friday and assume different values of q for the first and second weekends. I argue that the extra utility from the first weekend is more salient in a worker's decision when filing a sick leave claim. The data presented in Figures 5 and A4 support this assumption.

in ϕ if s_l increases by a unit, i.e., if $s_l'(s) = 1$, and (iii) increases utility in q if the sick leave claim ends on a Friday. Thus, the net gain in terms of time away from work is given by the term $\phi\,[s_l' - f'(s-\theta) + q\mathbb{1}\{\text{weekend}\}]$. The optimal sick pay utilization is $s^*(\theta{,}dow) = \text{argmax}_s\, u(s;\theta{,}dow)$.

Insurance provision lowers the marginal cost of sick leave and increases sick pay utilization. That is, $s^*(\cdot)$ is non-decreasing in sick leave benefits $B(s)$. Following the health literature, I refer to this responsiveness of sick pay demand to insurance coverage as moral hazard. This definition of moral hazard refers to "ex-post moral hazard"; i.e., it focuses on the responsiveness of workers' demand for sick pay to varying levels of insurance generosity. It abstracts from "ex-ante moral hazard", i.e., actions workers can take to prevent deterioration of their health.[26]

This model aims to understand how illness affects the absence behavior of employed individuals; consequently (i) it abstracts from how illness affects the labor force participation decision and (ii) does not explicitly consider the disutility of effort when the agent is employed. These modeling choices have no consequences for the qualitative properties of the optimal contract derived below.

The model abstracts from the behavior of "when" to file a sick leave claim—that is, in the model, a worker cannot choose the day she files a claim. Nonetheless, the model allows for strategic behavior on the duration margin of a sick leave claim, which is the main focus of this paper. While both margins play a role in the worker's choices, incorporating a filling-day choice in the model would require detailed data on when a worker falls sick and when she files a claim. Absent such data, I assume that workers claim on the day they fall sick.

Optimal utilization under linear contracts and quadratic penalties. To facilitate intuition, I impose the following functional form assumptions: (i) a linear benefit scheme $B(s) = bs$ where $b \in [0{,}1]$ and (ii) a quadratic compliance cost function. Under these assumptions and not taking into account the day of the week a sick leave claim is filed, the worker's utility function boils down to:

$$u(s;\theta) = w(M-s) + wbs + \phi(s-\theta) - \phi\frac{1}{\kappa}(s-\theta)^2 \times \mathbb{1}\{(s-\theta) > 0\}.$$

[26]This definition follows the conventional use of the term "moral hazard" in the health insurance literature. A fuller discussion of this (ab)use of terminology in the health insurance literature can be found in Section I.B. of Einav et al. (2013).

Thus, the optimal choice of sick leave duration, conditional to $s > \theta$ is:

$$s^*(\theta,\phi,\kappa) = \theta + \kappa \left(1 - \frac{w}{\phi}(1-b)\right) .$$

In the case of full coverage ($b = 1$) the worker optimally chooses $s^* = \theta + \kappa$. This case is presented in panel (a) of Figure A13, for strictly positive values of κ sick pay utilization is above the worker's health state. Lower compliance costs (higher κ) are associated with greater deviations from the worker's health status. The non-paid sick leave contract ($b = 0$) is presented in panel (b) of Figure A13. If the worker's valuation of time outside work ϕ is greater than the wage rate, the worker would optimally choose to ask for longer sick leave claims. Panels (c) and (d) of Figure A13 consider the case of partial coverage—i.e., a strictly positive replacement rate less than one $b \in (0,1)$ —for different values of κ and ϕ.[27] All else equal, a greater valuation of time outside work (a higher ϕ) is associated with a longer sick leave claim. Similarly, lower compliance costs (higher κ), are associated with longer sick leave claims. That is, moral hazard is increasing in the valuation of time outside work and decreasing in compliance costs.

Additionally, given the worker's preferences, the previous expression shows that the duration of a sick leave claim is increasing in the replacement rate b. This positive relation between the demand for sick days and the replacement rate is empirically verified by previous research (see for example Ziebarth, 2013) and by the evidence presented in this paper.

Expected utility. Ex ante, the worker aims to maximize her expected utility, taken over the distribution of health shocks $G(\theta,dow)$. I assume that the worker is risk averse with a von Neumann Morgenstern (vNM) utility function of the constant relative risk aversion (CRRA) type: $v(y) = y^{1-\gamma}/(1-\gamma)$ where y corresponds to the realized utility $u^*(\theta, dow)$. Thus, expected utility is given by

$$U = \mathrm{E}\left[v(u^*(\theta,dow))\right] = \int v(u^*(\theta,dow))\; dG(\theta,dow) ,$$

The utility maximization problem can be viewed as a two-stage problem (Einav et al., 2013). Once the health shock is realized, there is no source of uncertainty and workers aim to maximize the contribution of the *state* utility $u^*(\theta, dow)$ to their expected utility $\mathrm{E}_{\theta,dow}\left[v(u^*(\theta,dow))\right]$ by optimally choosing the duration of a claim s. Put another way,

[27] In this case, workers care about the effective valuation of time, i.e., $\frac{w(1-b)}{\phi}$.

given the health shock, risk preferences become irrelevant. That is, risk aversion does not affect workers' decision over sick pay utilization and, *all else equal*, variation in the utilization of paid leave across workers reflects variation in their preference parameters (ϕ, f and q).

III.B Production

Sick leave insurance provision affects workers' incentives. Given a level of coverage, on any given day, workers could decide to (i) show up to work sick or (ii) be absent from work. Thus, changes to the generosity of sick pay could affect the pool of workers showing up to work. The design of an optimal policy requires incorporating this channel into aggregated welfare. To do so, I assume that there is a firm in the economy that hires workers and produces a consumption good. While this is a stylized version of a model of firm behavior, it captures the production losses (or gains) of changes in the sick pay policy.

I assume that sickness is detrimental to workers' productivity. When sick, a worker is less productive than in her healthy state. For each worker, I define "effective days worked" as the sum of days worked when healthy ($d_{healthy}$) and days worked when sick (d_{sick}) adjusted by the parameter ν that reflects the lost productivity from sickness. Thus, $\nu \in (-\inf, 1]$.[28] That is, effective days worked (d_e) are: $d_e(s^*) = d_{healthy}(s^*) + \nu d_{sick}(s^*)$. The firm can not adjust workers' pay based on their health state. Thus, labor costs are a function of total days worked. Using these definitions, the profits generated by worker i are:

$$\pi^i(s^{i*}) = g(d_{sick}(s^{i*}) + \nu d_{healthy}(s^{i*})) - w(d_{healthy}(s^{i*}) + d_{sick}(s^{i*}))$$

I assume that the function $g(\cdot)$ exhibits diminishing returns on effective days worked, that capital is fixed, and normalized the price of the good to one. Total profits equals the sum of individual workers' profit: $\Pi = \sum_{i=1}^{I} \pi^i(s^{i*})$. I assume that every worker i gets the same share of total profits π.

III.C Social Planner

In this section I discuss the definition of social welfare and how the social planner chooses the optimal sick pay contract. For this exposition, it is helpful to consider an economy populated by I workers and let $U^i(\theta^i, dow^i)$ represent the expected utility of worker i.

[28] If $\nu = 1$ the worker is equally productive when healthy and sick. I let ν take negative values to reflect that productivity losses from sickness could imply product losses.

This paper's main question is what is the transfer function $B(s)$ that maximizes social welfare. Thus, I assume that the planner offers only one contract and ask what is the level of replacement rates b_s that maximize social welfare. Additionally, I assume that the social planner only observes the duration of a sick leave. The replacement rates therefore depend only on them.

Aggregate welfare can be written as follows:

$$W(B(s)) = \sum_{i}^{I} \omega^i U^i(B(s); \theta^i, dow^i)$$

where $U^i(B(s); \theta^i, dow^i) = \mathrm{E}_{\theta,dow}\left[v(u^*(\theta,dow) + \pi)\right]$ and ω^i represents the pareto weight assigned to worker i.

The social planner chooses $B(s)$ to maximize the sum of individual welfare:

$$\max_{B(s)} \quad W(B(s)) = \sum_{i}^{I} \omega^i U^i(B(s); \theta^i, dow^i) \;\; \text{s.t.} \;\; \sum_{i}^{I} s^{i*} B(s) \leq S \; ,$$

where S represents the allocated funds to cover the cost of the sick paid system. This constraint allows comparisons across policies that have the same cost. The set of contracts that I consider are the ones characterized by the following transfer function:

$$\begin{aligned} B(s) &= b_1 \, s & &\text{for } s \in [1, \underline{s}] \\ &= b_2 \, (s - \underline{s}) & &\text{for } s \in (\underline{s}, \bar{s}] \\ &= b_3 \, s & &\text{for } s > \bar{s} \end{aligned}$$

This function features a deductible of $\underline{s}$ days and it can reproduce the current Chilean system. Additionally I assume that transfers are non-decreasing, i.e., $B(s+1) \geq B(s)$ and the system is at most as generous as the full coverage case: $B(s) \leq s$.

Absent of moral hazard, the optimal sick leave contract would be one featuring full coverage for any duration, i.e., $B_s = b \;\forall\; s$. It would be socially optimal for all workers to be fully insured against health risks since leave's duration would equal their health state. In the presence of moral hazard, the optimal contract would feature some incomplete coverage to deter unjustified leave taking. In what follows, I explore the more interesting (and more realistic) case in which workers' behavior exhibit moral hazard.

IV Model Estimation and Identification Discussion

This section is organized as follows: first, I discuss the parametric assumptions and the procedure I follow to estimate the model of workers' behavior and present heuristic arguments regarding the identification of the model. Second, I discuss the solution to the planners' problem and the calibration of the remaining parameters, e.g., the productivity loss parameter.

IV.A Model Estimation: Workers' behavior

I exploit workers' responses to the incentives induced by sick insurance to estimate a vector of preferences parameters. The estimation of these parameters relies only on workers' observed decisions. That is, I do not impose optimality of the current policy. This result relies on the fact that the worker's problem can be viewed as a two-stage problem. Once the health shock is realized, workers optimally choose their sick pay utilization. Neither risk preferences nor production effects affect workers' utility. Risk preferences do not affect the utilization decision since the uncertainty has been resolved once the health shock is realized.

Parameterization. I parameterize workers' utility to represent the theoretical model fully in terms of parameters to estimate. I assume the following utility function:

$$\begin{aligned} u(s^i;\theta^i,dow^i) =& w^i(M-s^i) + w^iB(s^i) + \phi^i(s_l^i(s^i;dow^i)-\theta^i) + \phi^i q\ \mathbb{1}\{\text{weekend}\} \\ & - \phi^i\big[\kappa_0^i(s-\theta)^2\mathbb{1}\{s-\theta>0\} + \sum_{j=1}^{3}\kappa_j(s-\theta)\mathbb{1}\{s-\theta=j\}\big]\ , \end{aligned}$$

where i indexes workers and the compliance cost function takes a flexible functional form: it allows for quadratic penalties and specific costs of deviating one, two, or three days. Additionally, I assume the transfer function is captured by a piece-wise linear function with days brackets corresponding to the currently implemented in the Chilean system.

Worker i draws a health shock (θ^i,dow^i) from the distribution $G(\theta^i,dow^i) = P(\theta^i = m, dow^i = day|\ X)$, where X is a vector of observable characteristics (age and occupation). I group workers into eight groups or bins based on age and occupation type. Thus, all workers in a given bin b draw health shocks from the same distribution. This approach allows me to capture the empirical pattern that suggests that the diagnosis prevalence varies with the age and occupation of workers (see Figures 3, A2, A3).

Preference parameters. The preference parameter ϕ^i governs the valuation of time outside work. I assume that $\log(\phi^i)$ is drawn from a normal distribution with mean μ_ϕ and variance σ_ϕ^2 such that

$$\log(\phi^i) \sim N(\mu_\phi,\sigma_\phi^2) \ .$$

I allow for heterogeneity in the valuation of time outside work, but each worker draws a realization from the same distribution. Heterogeneity across this dimension reflects variation in the opportunity costs of missing work to recover from a disease or among tastes for leisure relative to consumption.

The term ϕq captures the extra utility a worker derives when the sick leave claim has a duration that allows the worker to not return to work until after the weekend. I assume that all of the variation in this term is governed by the parameter ϕ; thus, q does not vary across workers and is constant across sick leave duration. This assumption is justified by the fact that the valuation of time outside work is already captured by the parameter ϕ.

Variation in compliance cost parameter κ_0 reflects variation in the worker's preferences over behaving as expected or revealing their "true" health status. Additionally, job characteristics can justify variation in κ_0. For example, if a coworker can easily perform a worker's job, workers might face high compliance costs and ask for time outside work that closely follows their health status. I capture both of these mechanisms in a reduced-form manner. I assume that $\ln(\kappa^i)$ follows a normal distribution with mean and variance μ_κ and σ_κ^2:

$$\ln(\kappa^i) \sim N(\mu_\kappa,\sigma_\kappa^2) \ .$$

I interpret κ_1, κ_2, and κ_3 as shifters of the compliance cost of deviating for one, two, and three days respectively. Thus, heterogeneity in κ_0 implies that the cost of deviating for one day varies across individuals—a similar argument applies for two- and three-day long deviations. Nonetheless, this specification assumes that high (low) compliance costs workers face a high (low) cost of deviating either one, two, three, four, or any number of days.

Rounding and measurement error. I include two additional mechanisms when estimating the model to capture the behavior of physicians who prescribe sick leave claims in a reduced-form way.[29] First, I allow the duration of sick leave claims assigned to a worker to differ from the one optimally chosen by the worker. This discrepancy allows the model

[29] While explicitly modeling physician behavior is relevant for the design of paid sick leave, the lack of available data on physicians' characteristics limits the availability to address the question empirically.

to accommodate (i) informational frictions between a worker and a physician and (ii) observed sick leaves with a combination of duration and day of the week that the model does not predict. I assume that the duration of sick leave claims is measured with an additive error that has a mean zero and is uncorrelated with the "true" sickness level. That is, I assume that given the optimal sick leave duration s^*, the physician prescribes $\tilde{s}$:

$$\tilde{s} = s^* + \delta \ ,$$

where δ is a mean-zero random variable with support $[-3{,}3]$. With probability p_{me}, δ takes the values one or negative one, i.e., it shifts the duration of a sick leave claim in one day. With probability p_{me}^2 the duration of a sick leave claim is shifted two days. That is, δ takes the values two or negative two. Similarly

Second, I adjust sick leave duration to consider the rounding or heaping observed in the data. I interpret this pattern as coming from physicians being more likely to prescribe rest for a number of days that is a multiple of seven days. I assume that with some probability p_7 a sick leave claim of duration m is rounded up (dow) to 7 days.

Distribution of Health States. Workers face a distribution of health state, characterized by θ,dow, given by $G(\theta^i$,$dow^i) = P(\theta^i = m, dow^i = day|\ X)$, where X is a vector of observable characteristics (age and occupation). I observe the day a sick leave claim is filed from the data. When estimating the model, I assume that workers fall sick on the day they start a recovery span, i.e., the first day of a sick leave claim. Relaxing this assumption would require an additional source of data that distinguishes between the day a worker falls sick and the day she starts an absence spell or files a sick leave claim. Absent of such data, I use the starting day of a recovery span as the day the worker falls sick. I assume that workers file sick leave claims from Monday to Friday and that their work schedule is precisely Monday to Friday. In the data, less than 6% of sick leave claims are filed on weekends. Additionally, 83% of Chilean workers have a regular work schedule (Aguayo Ormeño, 2019).

I use the Peruvian Handbook of Recovery times to assign the average number of days a worker would need to recover from the condition reported in the sick leave claim data. I use these average days to construct the underlying distribution of health, as discussed in Section II. This approach allows me to construct an objective health measure without imposing parametric assumptions of the health distribution.

The model of workers' behavior allows for heterogeneity in how workers suffer a health shock. To see this, let α reflect how sickness affects a worker: workers with a higher α benefit more from time outside work. All else equal, if we consider two workers such that $\alpha^i > \alpha^j$, worker i is more affected by the symptoms of any disease than worker j. Thus, the utility function is given by:

$$u(s;\theta,dow) = w^i(M - s^i) + w^i B(s^i) + \tilde{\phi}^i \;\; \alpha^i(s^i_l(s^i;dow^i) - \theta^i) + \tilde{\phi}^i \tilde{q}\; \mathbb{1}\{\text{weekend}\}$$
$$- \tilde{\phi}^i \big[\tilde{\kappa}^i_0 (s-\theta)^2 \mathbb{1}\{s-\theta > 0\} + \sum_{j=1}^{3} \tilde{\kappa}_j (s-\theta)\mathbb{1}\{s-\theta = j\}\big] \, ,$$

where α^i is not identified. Nonetheless, it is not necessary to separately identify α to derive the optimal sick pay policy. What matters for the optimal design of the policy are workers' responses to the incentives generated by the provision of sick pay. These responses are a function of parameters: $\phi, q, \kappa_0, \kappa_1, \kappa_2, \kappa_3$.

Estimation procedure. I estimate a vector of ten parameters: $\Lambda = \{q, \mu_\phi, \sigma^2_\phi, \mu_{\kappa_0}, \sigma^2_{\kappa_0}, \kappa_1, \kappa_2, \kappa_3, p_{me}, p_7\}$. For this estimation, I select informative moments from the sick leave claims data and use the Simulated Method of Moments (SMM). Let $G(\Lambda)$ represent the vector of simulated moments and G^E their empirical counterpart. I aim to find the vector of parameters Λ that minimizes the squared distance between the simulated moments and the moments computed from the data:

$$\min_{\Lambda} \sum_{t=1}^{10} \left(\frac{G_t(\Lambda) - G^E}{G^E} \right)^2 \, .$$

To compute the simulated moments, I proceed as follows. First, I draw a representative sample of the data. This sample consists of a vector of wages, recovery times, and days of the week. The sample is stratified at the workers' group level.[30] The main strength of this approach is that it does not impose parametric assumptions on the distributions of wages and health shocks. Put another way, this strategy allows for arbitrary correlation between the health shocks and workers' wages to capture two empirical facts: (i) as discussed in II.D, the duration of days on leave varies with income; (ii) the diagnoses prevalence changes with age and occupation of the workers.

[30] Table A.11 verifies balance in terms of workers' characteristics and sick leave utilization between the sample drawn for estimation of the model and the sample used to document workers' behaviors and compute data moments.

IV.B Moments and Identification

In this section I provide a heuristic discussion of the most relevant moment for each parameter.

Weekend-streak utility (q). The term $\phi^i q 1\{weekend\}$ captures the extra utility a worker derives when the interaction of sick leave claim duration and day of the week implies a streak of days off work that includes the weekend, which I term a weekend-streak combination. To identify q I exploit variation across days of the week on which a sick leave claim of duration s is filed. That is, I rely on the fact that the temptation for shirking behavior varies between days of the week. For example, a 2-day-long sick leave claim is more attractive on a Thursday than a Tuesday. Figure 5 illustrates this variation. I argue that the share of sick leave claims observed on days outside the weekend-streak day provide a good counterfactual for the weekend-streak days.

The identification of q relies on the difference between the share of 1-to-5-day-long sick leave claims filed on a weekend-streak day and the share of 1-to-5-day-long claims filed any other day of the week. I pooled all the weekend-streak combinations to compute the average share of claims on those days and compare it with the average share of claims during the rest of the week. The model requires a higher q to rationalize the data if a larger difference is observed. This comparison relies on the idea that the share of sick leave claims of duration s on a non-weekend-streak day is a good counterfactual to estimate the effect of filling a sick leave claim of duration s on a weekend-streak day. The last panel of Figure 5 shows this moment graphically and Table A11 presents detailed computations.

Compliance cost function ($\mu_{\kappa_0}, \sigma^2_{\kappa_0}, \kappa_1, \kappa_2, \kappa_3$): I exploit variation across days of the week and sick leave claims duration conditional on workers' health to inform the distribution of compliance costs. I consider the pool of workers with similar characteristics and the same *assigned* recovery time—i.e., I hold fixed workers' health, age, and occupation—and compare their demand for sick pay across days of the week. For each day of the week and assigned recovery time, I compute the share of sick leave claims, indexed by j, of duration s filed by workers with health θ:

$$\text{share}^{day}_{s,\theta} = \frac{\sum_j \mathbb{1}\{dow_j = day, s_j = s, \theta_j = x\}}{\sum_j \mathbb{1}\{dow_j = day, \theta_j = x\}} ,$$

where the denominator counts the number of sick leave claims filed on the day of the week day with primary diagnoses that would require x days on leave, and the numerator counts how many of these claims have duration s. For example, the share of workers with a 1-day-long health shock on a Friday who asks for a one-day-long leave, i.e., is given by

$$\text{share}_{1,1}^{Friday} = \frac{\sum_j 1\{dow_j = Friday, s_j = 1, \theta_j = 1\}}{\sum_j 1\{dow_j = Friday, \theta_j = 1\}} \ .$$

Figure 6 illustrates this computation for sick leave claims with a health shock that requires a 1-day-long recovery. I start by computing the share of claims filed for a duration that matches the assigned recovery time and compare this figure with the share of claims filed for an extra day. That is, panel (a) vs. panel (b) of Figure 6. This difference is informative on how costly it is for individuals to ask for an extra day of leave. Panel (c) informs how costly it would be to ask for two extra days on leave. I restrict this comparison to claims filed for a combination of duration and day of the week representing a "weekend-streak". These are the darker columns in Figure 6. Thus, comparing these shares keeps the incentives for shirking behavior fixed. Every combination implies that workers would be on leave through the weekend. Panel (f) summarizes the probabilities of not-asking for extra days on leave, asking for one extra day on leave, two extra days on leave, and up to four extra days on leave conditional on filing a sick leave claim on a weekend-streak day.

I perform these comparisons for sick leave claims with assigned diagnoses of one, two, and three days of rest to inform the distribution of compliance costs. These shares are presented in Panel (a) of Figure 7. To compute the average share of claims with a given deviation, I average across claims from health states requiring up to 3 days on leave (see Figures A5 and A6). The pattern in the data suggest that a one day long deviation is not too costly relative to truth-telling, two days deviations are more costly as reflected by the lower share of sick leave claims in the third column of this graph.

Value of leisure (μ_ϕ, σ_ϕ^2): The parameter ϕ captures the taste for leisure relative to the taste for consumption. It can therefore be identified by the average ratio of leisure to consumption. I leverage data on wages, duration of sick leave claims, and sick pay to compute this ratio. I compute consumption as the net earnings in a month using data on wages and sick pay, that is, which is the consumption measure implied by the model. To compute leisure, I use the number of days a worker is on leave. For worker i this ratio is computed as follows:

$$LC_i = \frac{\text{leisure}}{\text{consumption}}_i = \frac{1}{N_i} \sum_m \frac{w_{i,m} \ \times \ Days\ on\ leave_{i,m}}{w_{i,m} \ \times Days\ worked_{i,m} + \ Sick\ pay_{i,m}} \ ,$$

where m indexes month of the year and N_i is the number of months in the year that worker i has used at least one sick leave claim. The numerator estimates worker i valuation of leisure in month m and the denominator estimates her consumption in month m. Thus, the ratio LC_i is the average relative valuation of leisure for individual i. Figure A12 shows the distribution of LC_i, the mean and standard deviation of this distribution inform the distribution of ϕ, which I assume log-normal with mean μ_ϕ and standard deviation σ_ϕ.

Rounding and measurement error. I use the difference between the share of 5-days-long sick leave claims filed on Monday relative to the share of claims filed on Tuesday, conditional on health shocks with a 1-day recovery, to pin down the success probability of the measurement error term δ (see panel (e) of Figure 6). Given the share of claims filed on Monday, a smaller difference implies that more sick leave claims have been moved away from the most-profitable duration. That is, the smaller the difference, the more likely the observed duration is not the optimally one in terms of workers' utility. To inform the probability of a sick leave claim to be rounded to duration that is a multiple of seven, I use the share of seven-day-long sick leave claims.

V Results

V.A Parameter Estimates

Table 3 presents the values of the estimated parameters. I use the spikes in the share of claims filed on weekend-streak days relative to non-weekend-streak days to identify the parameter governing the utility that workers derive from sick leave claims that end on a Friday (q). I estimate that, *all else equal* the utility of a worker increases in 0.79 for filing a sick leave claim weekend-streak combination. There are, on average, 12.33% more sick leave claims on weekend-streak days. The model's estimates follow a similar pattern: on average, I estimate 14.64% more sick leave claims on a weekend-streak day relative to non-weekend-streak days. The first row of Table 4 reports these moments. Figure **??** shows the simulated shares of sick leave claims by day of the week and duration. It reproduces Figure 5 using a model-simulated sample. Even if I only target the difference in the average shares, the model provides a good fit for the patterns observed by duration and days of the week.

Conditional on their health shocks, I exploit the share of sick leave claims observed on weekend-streak days to identify the parameters of the compliance cost function. Panel (b) of Figure 7 compares targeted moments from the data and a model-simulated sample.

The model matches the distribution of compliance costs—i.e., the cost of reporting the *true* health shock—reasonable well with $\mu_{\kappa_0} = 0.82$ and $\sigma_{\kappa_0} = 1.77$. I test the fit of the model in two ways. First, the model predicts that a one-day deviation is not too costly relative to truth-telling and that a two-day deviation is more costly, as reflected by the lower share of sick leave claims in this group. In particular, the model reproduces the main decay of the share of sick leave claims with positive deviations quite well.

I use the distribution of the ratio of leisure to consumption to identify the parameter ϕ. I estimate that, on average, workers value time off work, either to recover from disease or to engage in leisure, about 45% more than their wages. To put this estimate into context, consider that 26.54% of sick leave claims involve non-paid time off, and a total of 67.60% of claims involve partial paid for workers—, i.e., 67.60% of claims have a duration of up to 10 days for which the replacement rate is less than one.

V.B Model Fit

Matched moments. With the estimated parameters, the model matches the most relevant moments, presented in Table 4. There are, on average, 12.33% more sick leave claims on weekend-streak days. The share generated by the model is very close: on average, I estimate 14.64% more sick leave claims on weekend-streak days relative to non–weekend-streak days.

Panel (b) of Figure 7 compares the share of sick leave claims with non, one, and up to three-days long deviation implied by the data and by a model-simulated sample. I overestimate the share of claims with non deviations—i.e., claims with duration equal to the assigned diagnosis. I slightly underestimate the share of claims for a day above the assigned diagnosis. Nonetheless, the model replicates the decay in the share of sick leave claims with positive deviations quite well. For example, a two-day deviation is more costly than a one-day deviation, as reflected by the lower share of sick leave claims in this category.

The distribution of the ratio of leisure to consumption is assumed log-normal. Under this assumption, the mean generated by the model is slightly higher than the observed in the data, while the variance, on the other hand, is very close.

Specification Tests. I test how well the model matches data moments not used in the estimation. First, I construct the demand for days on leave as a function of the duration of the health shock. For each duration, I compute how many days; on average, workers request to be on leave. Figure 8 compares the average days on leave from the data and a model-simulated sample. This figure tests the model's ability to replicate workers' sick leave utilization choices and provides evidence that the model can replicate workers' responses to different health shocks. It is important that the model performs well on this dimension since the derivation of the optimal policy relies on estimates of workers' responses to changes in the paid sick leave policy.

Figure 9 compares the share of claims filed for a duration of 8 to 13 days from the data and a model-simulated sample. The model captures the main pattern observed in the data: sick leave claims spike at 11 days, with lower mass at 8, 9, and 10 days. In particular, using the measure of heaping proposed by Roberts and Brewer (2001), I estimate that, in the data, the 11-day duration accumulates an additional 4.50% mass than its neighbors. Using the model-simulated sample, I estimate an additional 4.03% mass relative to its neighbors.[31] The derivation of the optimal policy requires an estimate of the moral hazard costs associated with this discontinuity, thus the importance of a robust estimate of workers' responses to this feature of the paid sick leave contract.

I also propose an out-of-sample exercise exploiting data not used in the estimation of the model. Using data on sick leave claims filed in 2019, I compute the vector of moments used to estimate preference parameters and the share of claims with a duration in the neighborhood of 11 days. I compare these moments with their model-simulated counterparts to test the model performance. To obtain the latter, I simulate the model based on a representative sample drawn from the 2019 data and the estimated vector of preference parameters. Table 5 presents the results of this exercise. The results suggest that the model performs reasonably well out of the sample: preference parameters and the share of sick leave claims of selected durations are comparable in magnitude. Additionally, the model reproduces (i) the decay of the share of sick leave claims with positive deviations and (ii) the excess mass at 11 days.

V.C Robustness Checks: Moments' computation

Weekend-streak utility (q). The estimation of the model relies the computation of this moment using data on all sick leave claims filed during the 2017. In figures A7 to A10, I compute this moment restricting the sample to claims filed during each quarter of the year. These figures show the same qualitative pattern than Figure 5 providing evidence

[31] Roberts and Brewer (2001) proposes the following measure: $h_z = f(z) - \frac{f(z-1)+f(z+1)}{2}$, where z corresponds to 11 days, and $f(\cdot)$ indicates the frequency of sick leave claims with duration z. Thus, h_z gives the difference between a duration frequency and the average of the frequencies of the two immediately neighboring duration. It indicates how much a duration sticks out from the pattern suggested by its neighbors.

that claims from a particular time of the year, e.g., winter do not drive the patterns in the data.

Compliance cost function ($\mu_{\kappa_0}, \sigma^2_{\kappa_0}, \kappa_1, \kappa_2, \kappa_3$): Figure A11 proposes a similar exercise. I show the distribution of share of sick leave claims with none and positive deviations for each quarter of the year. This figure tells a similar story: the share of sick leave claims with no or one day deviations are almost the same, and the share of sick leave claims corresponding to longer deviations decreases monotonically.

VI The optimal paid sick leave contract

I use the estimated model to determine the optimal paid sick leave system—the system that maximizes aggregated welfare—and the corresponding replacement rates. In this section, I first present the set of assumptions I impose when solving this maximization problem. Then, I discuss what is the optimal system given the estimated distribution of health and moral hazard parameters.

I assume that contracts are piece-wise linear, i.e., the marginal replacement rate b is constant within a sick leave duration bracket $[\underline{s}, \bar{s}]$. When solving to welfare maximization problem, I consider three-brackets contracts. Thus, the determination of the optimal replacement rate ask what is the optimal replacement rates for each bracket.

The theoretical model of worker's behavior assumes that risk aversion only affects workers' expected utility, not their utilization choice. Identification of γ would require, for instance, variation in plan choices across workers. Nonetheless, the Chilean Paid Sick Leave System does not offer choice over sick pay plans. Absent this variation, I calibrate γ using results from the literature. A concern in calibrating γ is that risk aversion can vary significantly across the scale of shocks (Chetty and Szeidl, 2007). In the model, γ captures risk aversion relative to wage losses from short-term, non-work-related health shocks. The losses associated with such shocks on average are thought to be smaller than losses faced in the case of unemployment shocks. I assume that $\gamma = 2$ and present results with two alternative specifications that allow for preference heterogeneity.[32]

The optimal policy differs from the current system in three key ways. First, it offers a partial replacement, with an average replacement rate of 0.36, for claims of up to three days. This shift increases the utility of workers who would not take sick leave under

[32] I follow Herbst and Hendren (2021) and assume that the coefficient of relative risk aversion, γ, is drawn from a uniform distribution. I consider two alternative specifications (i) a uniform distribution between 1 and 3 and (ii) a uniform distribution between 0 and 4. Both specifications retain the mean risk aversion coefficient at 2, but introduce different degrees of heterogeneity.

the current system but do under the optimal policy. At the same time, partial coverage contains moral hazard since most of the cost of those absences is faced by workers.

Second, the optimal policy eliminates the discontinuity at 11 days and exhibits a higher average replacement rate between 4 and 10 days. Doing so curbs the cost of the behavioral responses to the program incentives and provides more risk protection. Implementing the optimal scheme would shift the distribution of sick leave duration: workers are more likely to file sick leave claims between 8 and 10 days and less likely to file claims for 11 days—relative to the distribution of sick leave claims under the current Chilean system. Third, the optimal policy does not offer full replacement for sick leave claims longer than 11 days. The average replacement rate is increasing, as in the current system, but it is less generous for longer claims. Taken together, these changes in the replacement rate reflect that the workers value a contract that offers more protection for shorter claims to smooth consumption across different health states. I estimate that workers are willing to give up 1.53% of their earnings to be insured under the optimal policy.

VI.A Counterfactual analysis

Changes in compliance cost function. In this section I ask how the optimal sick paid policy changes when workers' responses involve more (or less) moral hazard. To do so, I propose the following exercise. All else equal, I reduce (increase) the cost of filing a sick leave claim longer than the health state (θ). For example, the cost of filing a sick leave claim for an extra day (f_1)is given by:

$$f_1 = f(s = \theta + 1; \theta) = \kappa_0^i + \kappa_1 \ ;$$

I use the estimates of $(\mu_{\kappa_0}, \sigma_{\kappa_0}, \kappa_1, \kappa_2, \kappa_3)$ and construct a new distribution of compliance costs by shifting the mean of κ_j, such that $E(\kappa_j) = \mu_{\kappa_j} \times (1+\varepsilon)$ for either positive or negative values of ε. Using these counterfactual distributions of compliance costs, I first show workers' choices assuming they are insured under the Chilean paid sick leave system. Panel (a) of Figure 11 shows that when compliance costs are higher the average number of days on leave is closer to reflecting workers' health state. Similarly, when compliance costs are low, workers' as for more days on leave. Thus, workers' responses imply more moral hazard.

Turning into the implications for the optimal policy, panel (b) of Figure 11 presents the optimal policy for higher compliance costs (less moral hazard) and the policy for lower compliance costs (more moral hazard) and compare them to the benchmark case.

VII Conclusions

This paper addresses a relevant but poorly understood question in the provision of social insurance: the optimal design of paid sick leave. When designing such programs, policymakers face several choices: What level of benefits should be paid? Should benefits rise or fall over the utilization span? How should the programs be financed? This paper's main contribution is to answer these questions and provide evidence on how the adoption (or modification) of a paid sick leave system could help to preserve workers' health. This paper speaks directly to the rules regarding generosity of sick pay and cost-sharing mechanisms that would provide enough coverage and limit shrinking behavior.

This paper studies the incentive problem created by paid sick leave insurance. To do so, I develop a model of sick pay provision where risk-averse workers face health risks and decide their sick pay coverage and utilization. This model is helpful to (i) capture the trade-offs workers face when deciding their sick pay utilization, (ii) illustrate how workers respond to changes in the paid sick leave contract, and (iii) characterize the trade-off faced by the social planner. I obtain estimates of the model's parameters exploiting individual-level data on sick pay utilization and labor supply from the Chilean Paid Sick Leave System. I leverage exogenous variation in the incentives for shirking behavior induced by the day of the week a sick leave claim is filed. Finally, I analyze the impact on welfare and behavior of implementing the optimal contract.

The empirical application of this paper exploits the Chilean context, but the insights of my model and results are informative for other contexts, including the United States context. Currently, many US workers do not have access to paid leave time, and there is substantial inequality in coverage across jobs. For instance, 97% of private-sector employees in the finance and insurance industry have access to paid sick leave while 41% of employees in the accommodation and food services industry have access to paid sick leave (Maclean, Pichler and Ziebarth, 2020). While there is no federal mandate of paid leave, there has been traction on paid sick leave policies at the state and local levels. Over the past decade, thirty-six US jurisdictions have implemented sick pay mandates enacting laws that enable eligible workers to accrue paid time off from their employer (A Better Balance, 2021). Additionally, opinion polls show large bipartisan support for mandating paid sick leave.

This paper has the potential to inform the policy discussion in two ways. First, it quantifies the main trade-off faced by policymakers when designing a paid sick leave system. This trade-off applies to the US context: more generous sick pay contracts could reduce the number of workers showing up sick and help preserve workers' health at the cost of increased absenteeism, which entails product losses and increases program costs. Second, this paper shows that traditional cost-sharing rules—such as waiting periods—could contain moral-hazard behaviors, but contagious diseases undermine their effectiveness. Sick workers could spread diseases in the workplace, deteriorating public health. These externalities would justify more generous sick pay and highlight the relevance of enforcement mechanisms such as penalties for unjustified absenteeism.

References

Aguayo Ormeño, Irina. 2019. "Estadísticas de Jornada Laboral de Trabajadores del Sector Privado." Comisión Trabajo y Previsión Social de la Cámara de Diputados.

Balance, A Better. 2021. "Overview of Paid Sick Time Laws in the United States." *retrieve August 30, 2021.*

Cabral, Marika, and Marcus Dillender. 2020. "The Impact of Benefit Generosity on Workers' Compensation Claims: Evidence and Implications." NBER Working Paper No. 26976.

Campolieti, Michele, and Douglas E Hyatt. 2006. "Further evidence on the "Monday Effect" in workers' compensation." *ILR Review*, 59(3): 438–450.

Card, David, and Brian P McCall. 1996. "Is workers' compensation covering uninsured medical costs? Evidence from the "monday effect"." *ILR Review*, 49(4): 690–706.

Chetty, Raj. 2008. "Moral Hazard Versus Liquidity and Optimal Unemployment Insurance." *Journal of Political Economy*, 116(2): 173–234.

Chetty, Raj, and Adam Szeidl. 2007. "Consumption Commitments and Risk Preferences." *The Quarterly Journal of Economics*, 122(2): 831–877.

Cid, Camilo P. 2006. "Evaluación del Subsidio por Incapacidad Laboral por Enfermedad Común en Chile: 1990-2000." *Superintendencia de Salud. Documento de Trabajo*.

Cronin, Christopher J, Matthew C Harris, and Nicolas R Ziebarth. 2022. "The Anatomy of US Sick Leave Schemes: Evidence from Public School Teachers." NBER Working Paper No. 29956.

Cutler, David M, and Richard J Zeckhauser. 2000. "The Anatomy of Health Insurance." In *Handbook of Health Economics*. Vol. 1, 563–643. Elsevier.

De Paola, Maria, Vincenzo Scoppa, and Valeria Pupo. 2014. "Absenteeism in the Italian public sector: The effects of changes in sick leave policy." *Journal of Labor Economics*, 32(2): 337–360.

Einav, Liran, Amy Finkelstein, and Mark R Cullen. 2010. "Estimating Welfare in Insurance Markets Using Variation in Prices." *The Quarterly Journal of Economics*, 125(3): 877–921.

Einav, Liran, Amy Finkelstein, Stephen P Ryan, Paul Schrimpf, and Mark R Cullen. 2013. "Selection on Moral Hazard in Health Insurance." *American Economic Review*, 103(1): 178–219.

EsSalud. 2014. "Normas y Procedimientos Para la Emision, Registro Y Control De Las Certificaciones Médicas Por Incapacidad y Maternidad." Seguro Social de Salud.

Gruber, Jonathan. 2000. "Disability Insurance Benefits and Labor Supply." *Journal of Political Economy*, 108(6): 1162–1183.

Handel, Ben, Igal Hendel, and Michael D Whinston. 2015. "Equilibria in health exchanges: Adverse selection versus reclassification risk." *Econometrica*, 83(4): 1261–1313.

Hendren, Nathaniel. 2017. "Knowledge of Future Job Loss and Implications for Unemployment Insurance." *American Economic Review*, 107(7): 1778–1823.

Herbst, Daniel, and Nathaniel Hendren. 2021. "Opportunity Unraveled: Private Information and the Missing Markets for Financing Human Capital." NBER Working Paper No. 29214.

Hopenhayn, Hugo A, and Juan Pablo Nicolini. 1997. "Optimal Unemployment Insurance." *Journal of Political Economy*, 105(2): 412–438.

Johansson, Per, and Mårten Palme. 2005. "Moral Hazard and Sickness Insurance." *Journal of Public economics*, 89(9-10): 1879–1890.

Kleven, Henrik J. 2021. "Sufficient Statistics Revisited." *Annual Review of Economics*, 13: 515–538.

Low, Hamish, and Luigi Pistaferri. 2015. "Disability Insurance and the Dynamics of the Incentive Insurance Trade-Off." *American Economic Review*, 105(10): 2986–3029.

Maclean, Johanna Catherine, Stefan Pichler, and Nicolas R Ziebarth. 2020. "Mandated Sick Pay: Coverage, Utilization, and Welfare Effects." NBER Working Paper No. 26832.

Marie, Olivier, and Judit Vall Castello. 2022. "Sick Leave Cuts and (Unhealthy) Returns to Work." *accepted at Journal of Labor Economics*.

Marone, Victoria R, and Adrienne Sabety. 2022. "When Should There Be Vertical Choice in Health Insurance Markets?" *American Economic Review*, 112(1): 304–42.

Pollak, Catherine. 2017. "The Impact of A Sick Pay Waiting Period on Sick Leave Patterns." *The European Journal of Health Economics*, 18(1): 13–31.

Powell, David, and Seth Seabury. 2018. "Medical Care Spending and Labor Market Outcomes: Evidence from Workers' Compensation Reforms." *American Economic Review*, 108(10): 2995–3027.

Roberts, John M, and Devon D Brewer. 2001. "Measures and tests of heaping in discrete quantitative distributions." *Journal of Applied Statistics*, 28(7): 887–896.

Thoursie, Peter Skogman. 2004. "Reporting Sick: Are Sporting Events Contagious?" *Journal of Applied Econometrics*, 19(6): 809–823.

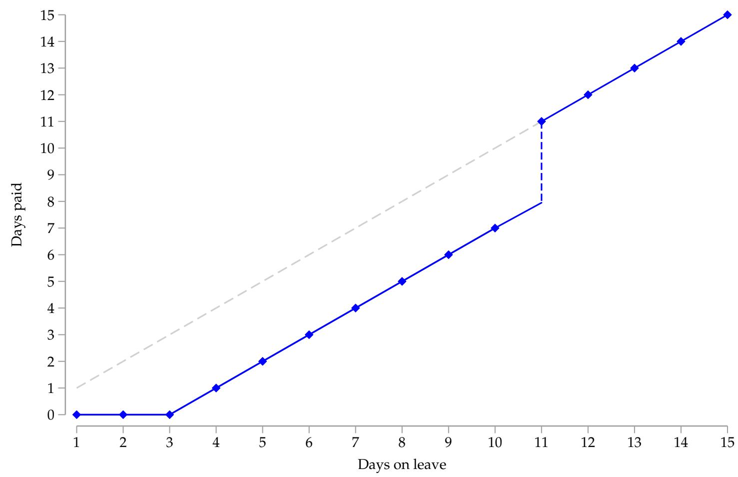
Ziebarth, Nicolas R. 2013. "Long-term Absenteeism and Moral Hazard - Evidence From A Natural Experiment." *Labour Economics*, 24: 277–292.

Ziebarth, Nicolas R, and Martin Karlsson. 2014. "The Effects of Expanding the Generosity of the Statutory Sickness Insurance System." *Journal of Applied Econometrics*, 29(2): 208–230.

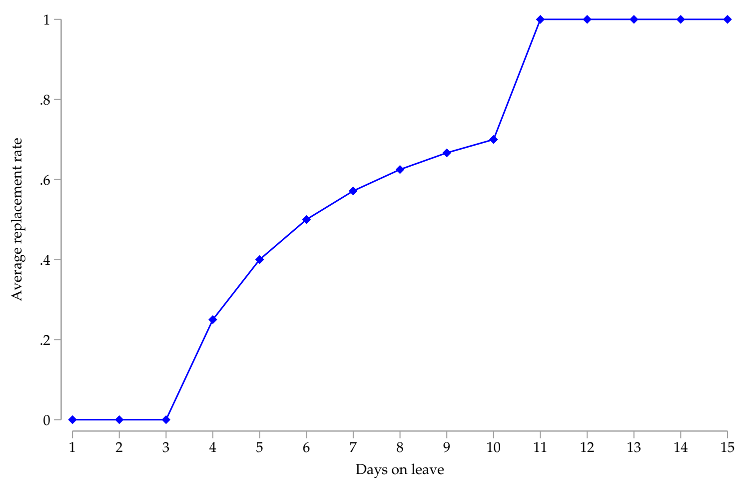
VIII Figures

Figure 1: Chilean paid sick leave system: benefits computation

(a) Days paid as a function of days on leave

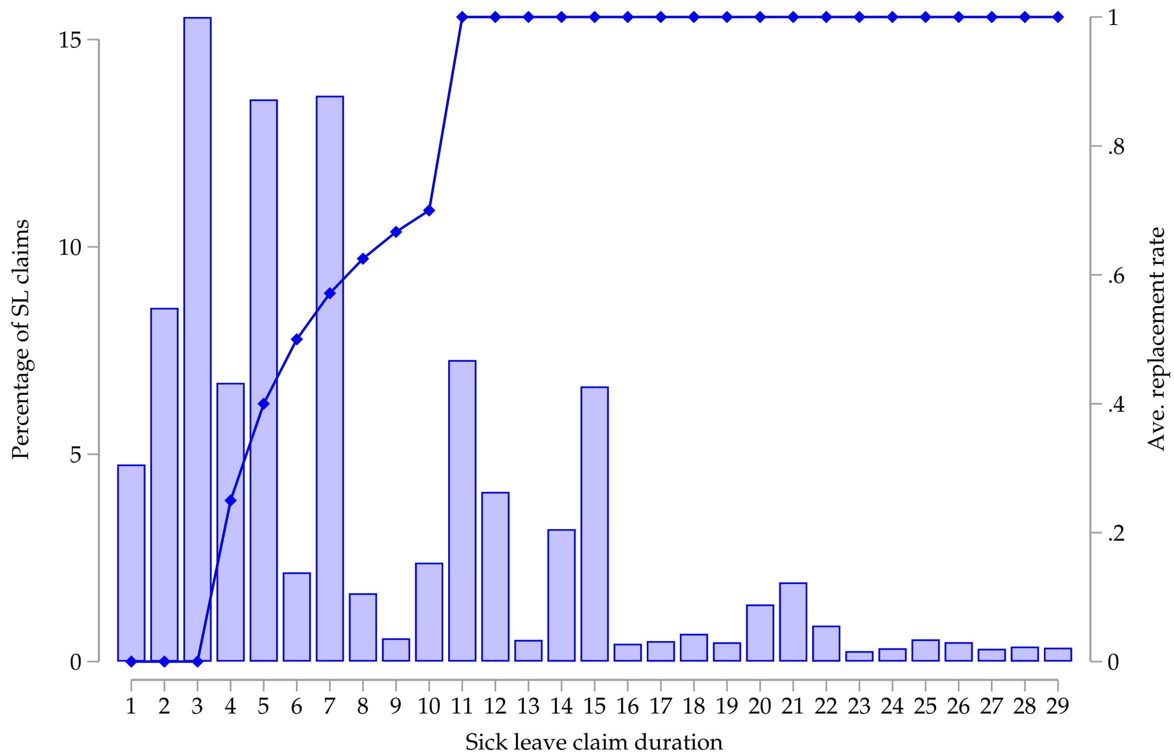


(b) Average replacement rate



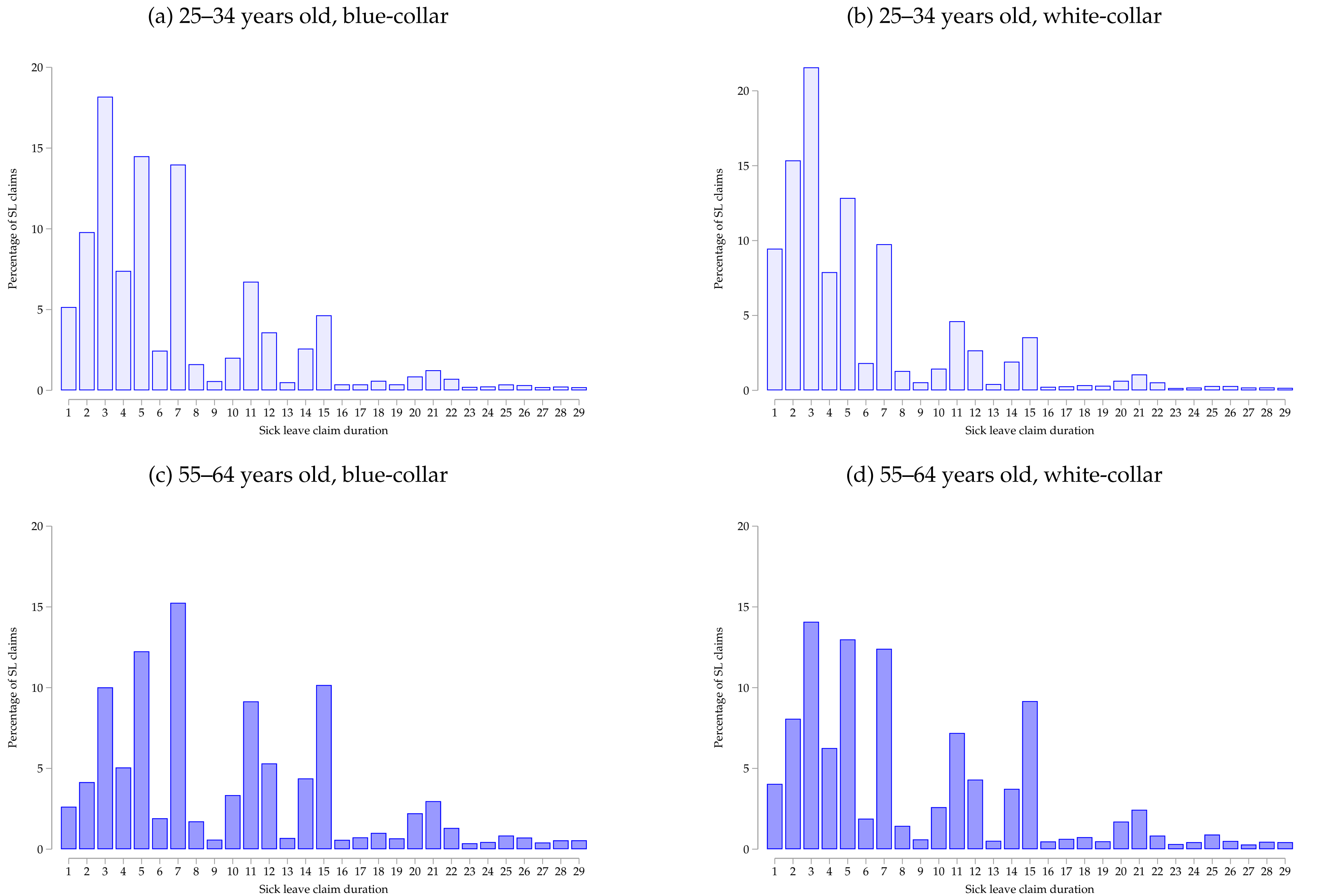
Notes: This figure shows the paid sick leave benefit scheme for private-sector employees. Panel (a) shows the number of days paid as a function of days on leave. The replacement rate for the first three days of a sick leave spell is zero. Starting on the fourth day, there is full coverage of each missed day—i.e., the replacement rate is one. If the sick leave lasts 11 days or more, the nonpayable period is reimbursed. Panel (b) shows the average replacement rate, i.e., the ratio between the number of days paid and the number of days on leave. This figure is referenced in Section II.B.

Figure 2: Duration of sick leave claims: Private-sector male workers



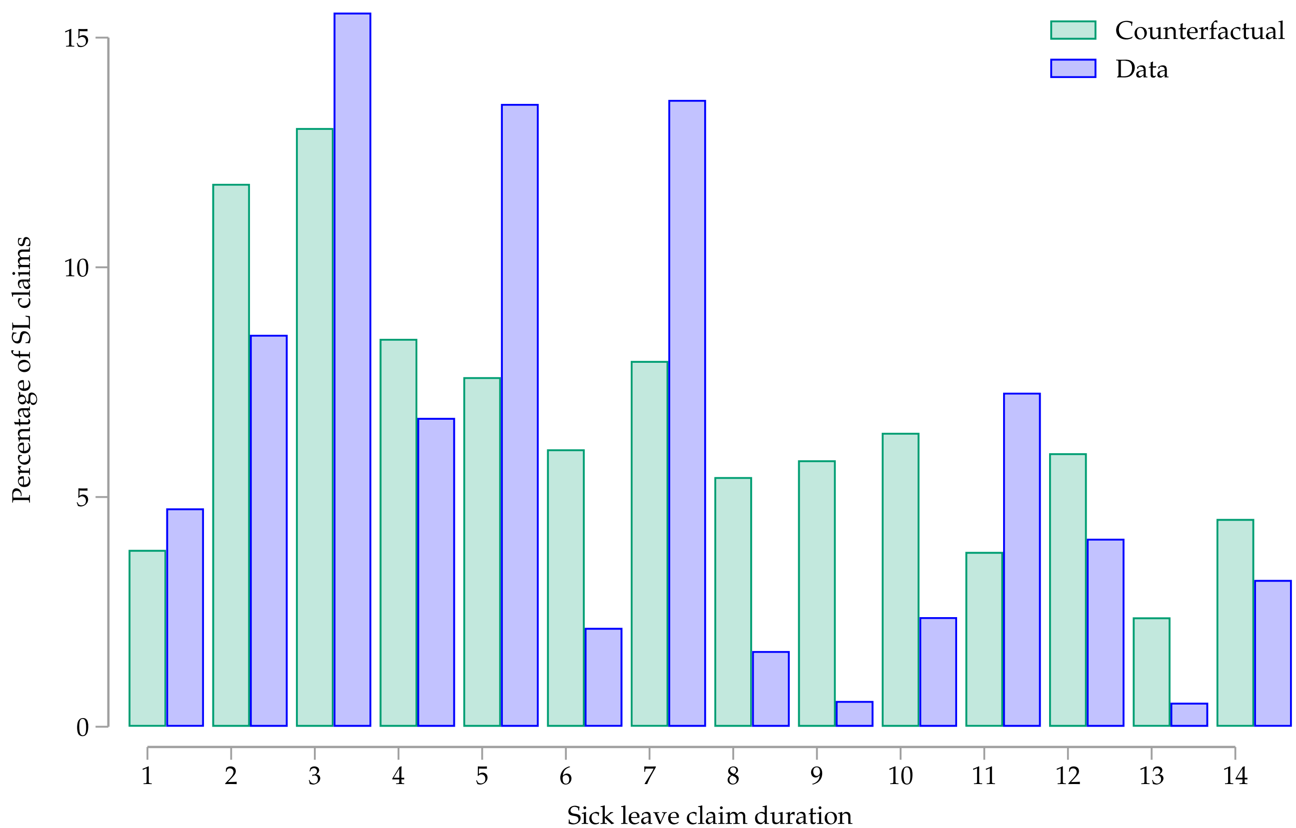
Notes: This figure shows the distribution of the duration of sick leave claims made by male workers on the left-hand-side vertical axis and the average replacement rate on the right-hand-side vertical axis. The figure includes only sick leave claims of up to 29 days; these represent 89% of all claims. This figure is referenced in Section II.D.

Figure 3: Histogram of days on leave by worker characteristics



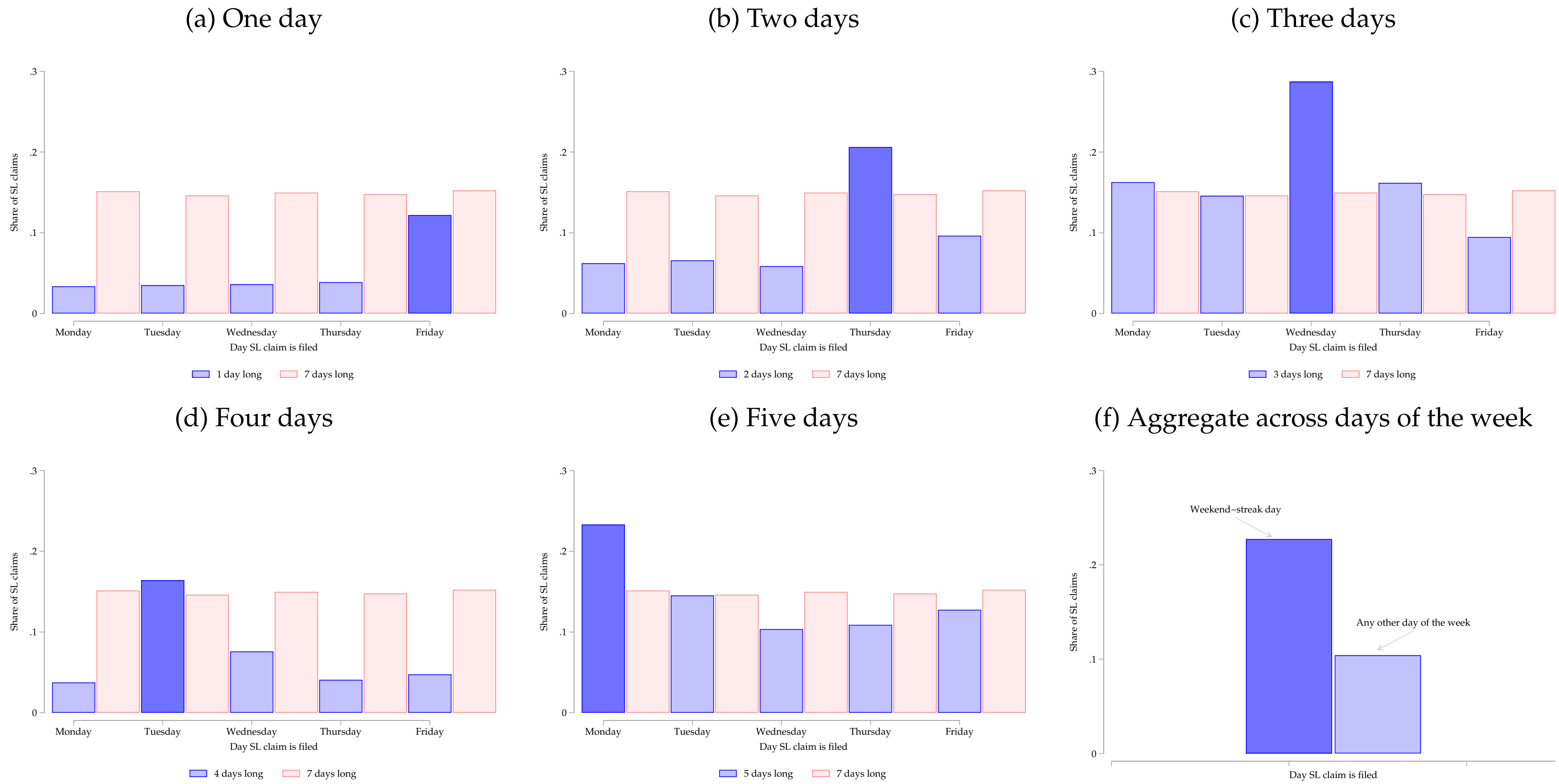
Notes: This figure shows the distribution of days on leave by worker age and occupation for the youngest and oldest group of workers. The sample includes male private-sector employees. Blue-collar workers refers to workers who engage in hard manual labor, typically agriculture, manufacturing, construction, mining, or maintenance. White-collar workers refers to workers whose daily work activities do not involve manual labor—e.g., teachers or administrative staff. Additional groups are presented in Appendix Figure A2. This figure is referenced in Section II.D and in Section IV.A.

Figure 4: Days on leave: Data and counterfactual distribution



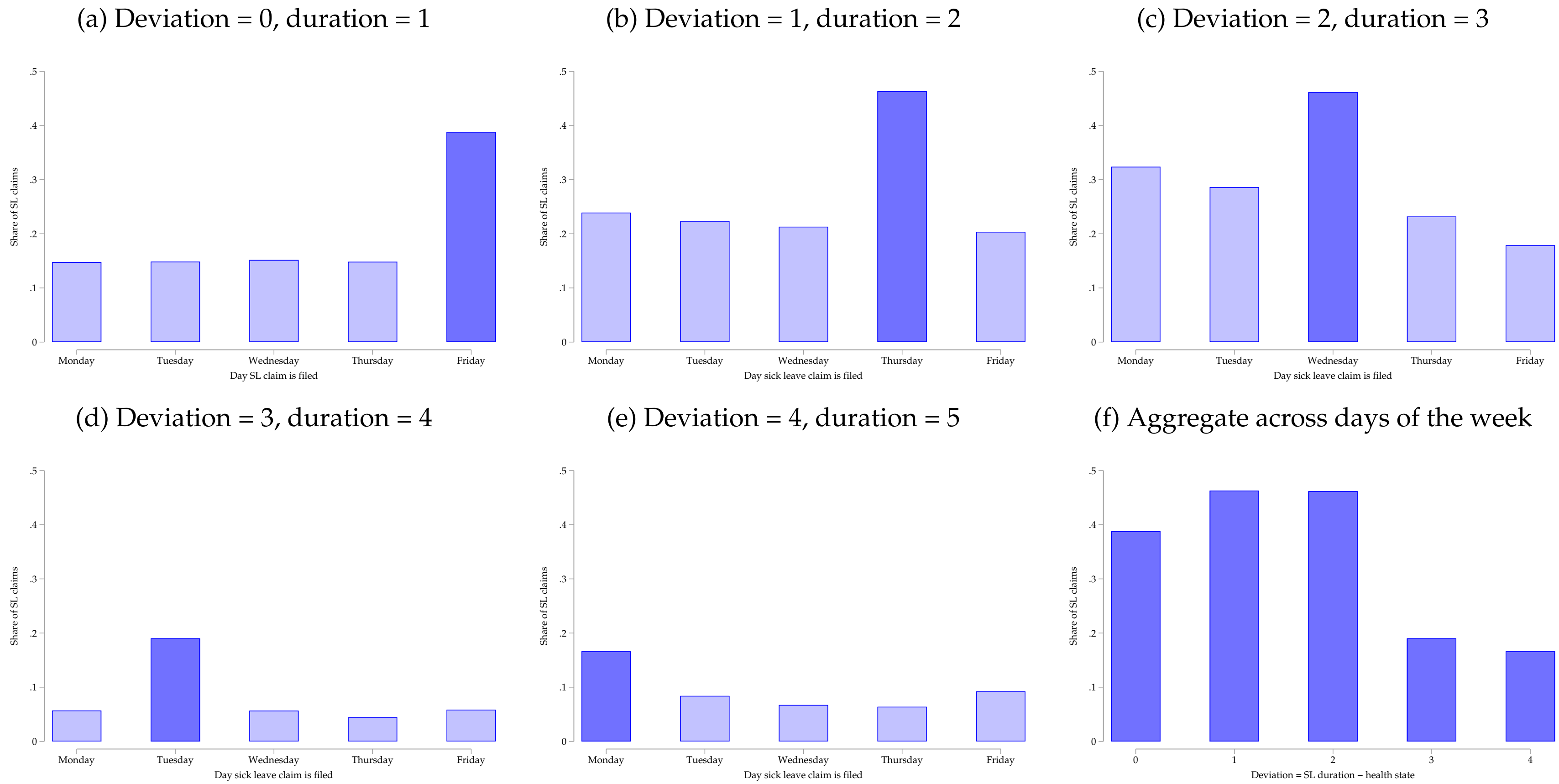
Notes: This figure shows the distribution of days on leave coming from the data, as shown in Figure 2, and the counterfactual distribution of days on leave. The latter is constructed assigning to each sick leave the recovery time suggested by the Peruvian Handbook of Recovery Times, adjusted by worker age and occupation. This figure is referenced in Section II.D.

Figure 5: Days of the week and sick leave claim duration



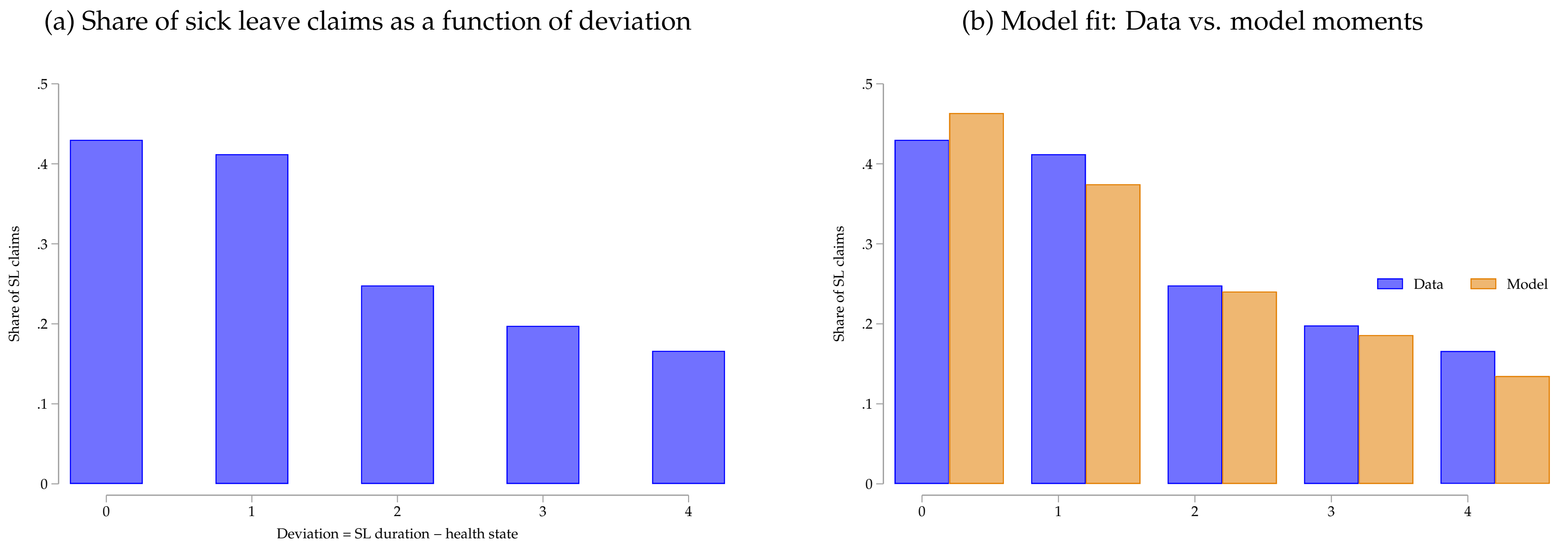
Notes: Panels (a) to (e) show the share of sick leave claims with duration s and the share of seven-day-long sick leave claims filed on each day of the week. Panel (f) aggregates across durations and days of the week: The first bar—labeled "weekend streak"—averages the share of one- to five-day-long sick leave claims that end on a Friday and are filed on any day of the week (for example, one-day-long claims filed on a Friday, two-day-long claims filed on a Thursday, and so on). The second bar—labeled "non–weekend streak"—averages the share one- to five-day-long sick leave claims filed on any other day of the week (for example, two-day-long claims filed on a Friday). Table A11 reports the estimated shares and moments. This figure is referenced in Sections III.A and IV.B.

Figure 6: Identification of compliance cost function: Sick leave claims by duration and day of the week (one-day recovery time)



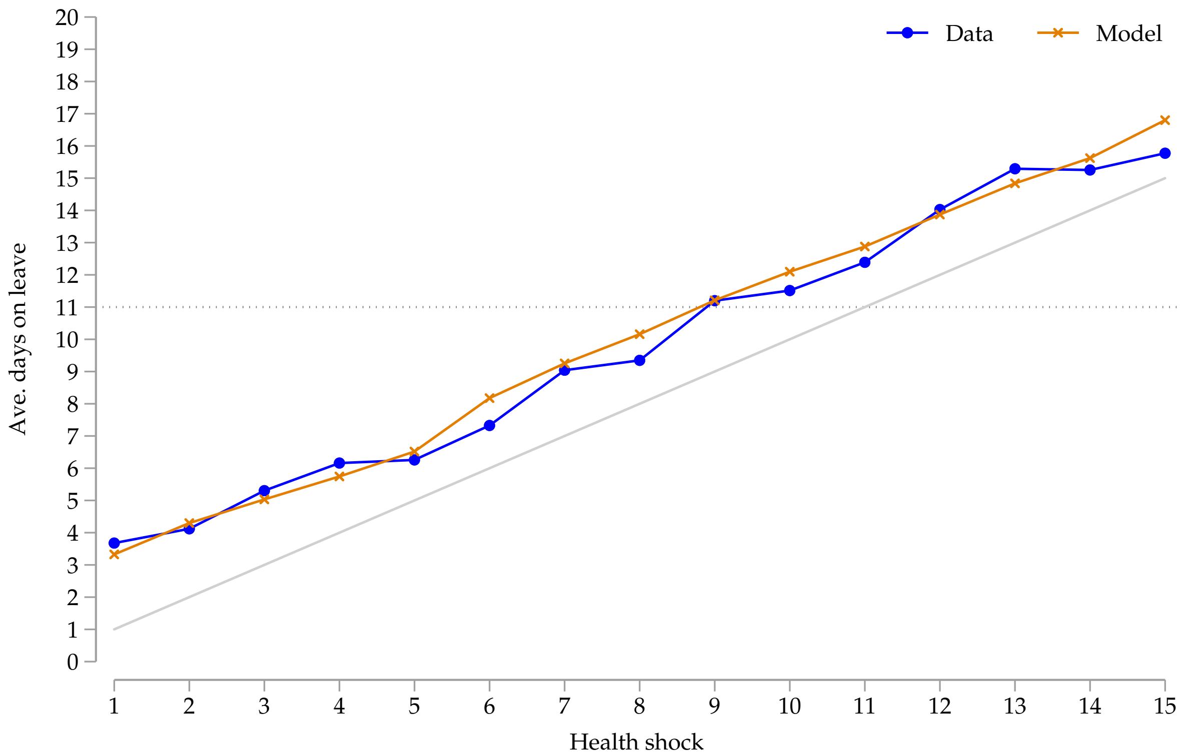
Notes: Panels (a) to (e) show the share of sick leave claims with duration s for workers whose main diagnosis would imply a health state of 1 day on leave. Panel (f) aggregates the share of sick leave claims across days of the week, including only weekend-streak combinations; e.g., from Panel (a), I consider only the share for Friday. This figure is referenced in Section IV.B.

Figure 7: Compliance cost function



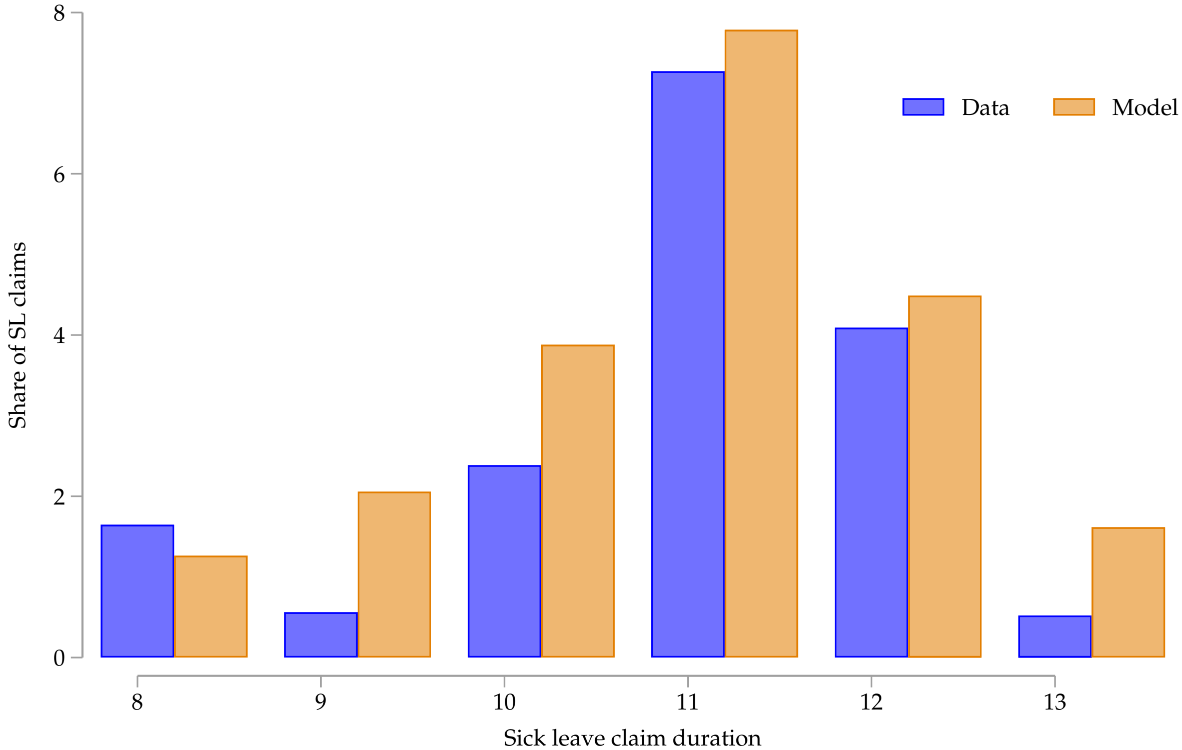
Notes: Panel (a) shows the average share of sick leave claims with deviations between 0 and 4 days. The average is computed over sick leave claims with primary diagnoses requiring 1, 2 or 3 days of rest filed on weekend-streak days. Each column is the weighted average of the probability for each health state. Panel (b) replicates this figure and adds the moments computed from the simulated data. This figure is referenced in Section IV.B.

Figure 8: Demand for days on leave as a function of health shock



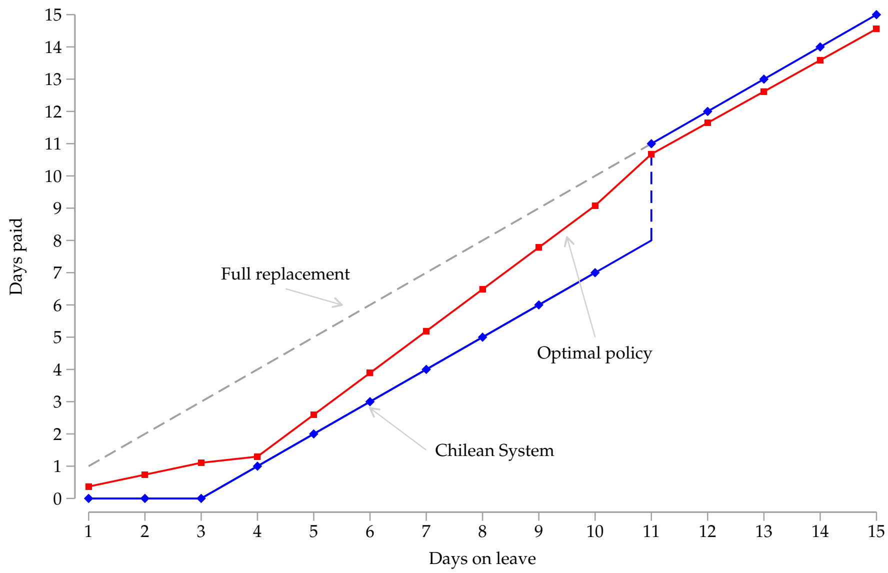
Notes: This figure shows the demand for days on leave as a function of the duration of the health shock from the data and a model-simulated sample. For each duration, I compute how many days; on average, workers request to be on leave. The 45 degrees line can be interpreted as the demand for days on leave when workers report their health. The horizontal line at 11 days indicates the position of the discontinuity in the sick pay scheme. This figure is referenced in Section V.B.

Figure 9: Model's fit: Distribution of sick leave claims with durations around 11 days



Notes: This figure compares the distribution of sick leave claims with a duration in the neighborhood of 11 days from the data and a model-simulated sample. Using the measure of heaping proposed by Roberts and Brewer (2001), I estimate that, in the data, the 11-day duration accumulates an additional 4.50% mass than its neighbors. Using the model-simulated sample, I estimate an additional 4.03% mass relative to its neighbors. This measure approximates how much a duration 'sticks out' from the pattern suggested by its neighbors. Roberts and Brewer (2001) proposes: $h_z = f(z) - \frac{f(z-1)+f(z+1)}{2}$, where z corresponds to 11 days, and $f(\cdot)$ indicates the frequency of sick leave claims with duration z. Thus, h_z gives the difference between a duration frequency and the average of the frequencies of the two immediately neighboring duration. This figure is referenced in Section V.B.

Figure 10: The Optimal Sick Pay System

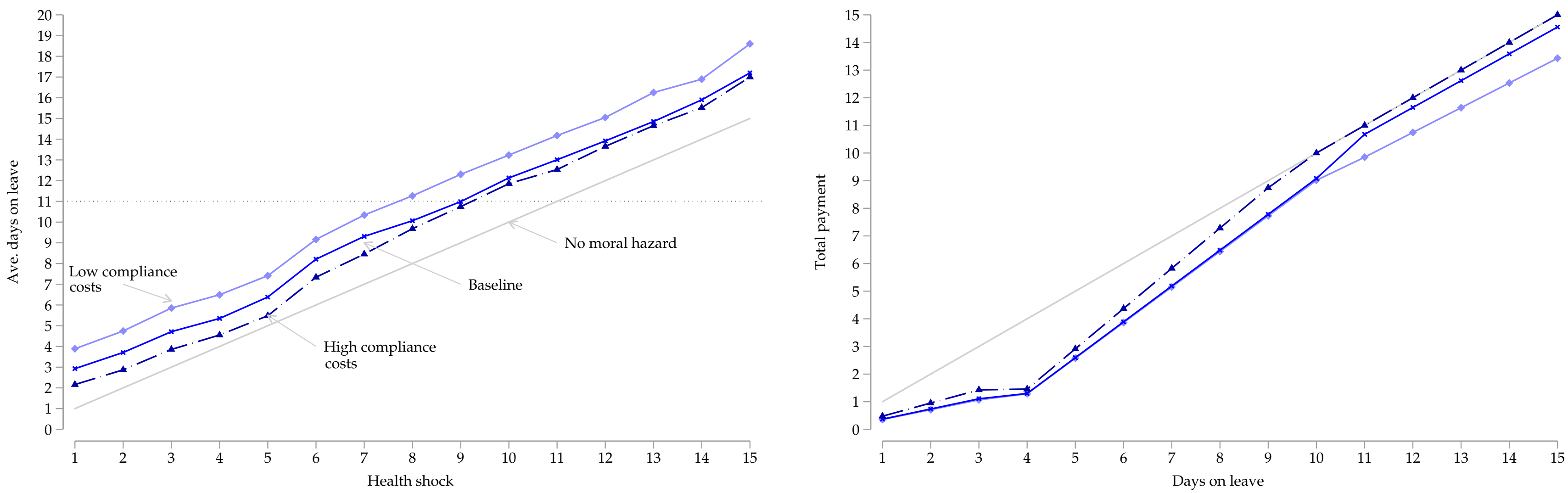


Notes: This figure compares the Chilean system with the optimal system. This figure is referenced in Section VI.

Figure 11: Counterfactual exercise: Changes in compliance cost

(a) Workers' responses: Ave. number of days on leave

(b) Optimal paid sick leave contract



Notes: This figure ... . This figure is referenced in Section VI.A.

IX Tables

Table 1: Summary statistics: all workers and workers who use sick leave insurance

	All workers	Workers who had used SL benefits		
		Any	Included conditions	
			All	Up to 30 days
	(1)	(2)	(3)	(4)
Age				
Mean	43.94	43.41	43.30	42.24
Share of workers aged (%)				
25 - 34 years old	26.35	28.90	29.14	32.11
35 - 44 years old	24.48	24.10	24.34	25.35
45 - 54 years old	26.70	24.71	24.66	23.73
55 - 64 years old	22.47	22.28	21.86	18.81
Income (monthly USD)				
Mean	772.00	904.70	909.42	918.02
Standard deviation	367.27	388.79	389.99	390.03
25th percentile	484.45	587.51	591.79	601.77
Median	682.15	829.84	835.53	845.74
75th percentile	997.97	1,146.82	1,152.48	1,161.29
90th percentile	1,328.04	1,483.17	1,489.20	1,496.41
Region (%)				
Central	34.97	40.76	41.22	41.92
Mining intensive	8.96	8.49	8.32	7.62
Health - chronic conditions (%)				
Hypertension	12.90	16.12	15.96	13.93
Diabetes	6.04	7.95	7.51	6.19
Share of workers (%)	100	18.50	17.12	13.78
Observations	1,916,138	354,469	328,053	263,951

Notes: This table presents summary statistics for all male workers in the sample (column 1) and for workers who have used sick leave benefits in the past year based on the conditions and duration of sick leave claims (columns 2 to 4). The sample includes private and public sector employees age 25 to 64 years old. Income statistics are based on the winsorized distribution where the lowest and highest 5% of the income values are excluded. Sick leave claims of up to 30 days account for 95% of all claims filed in a year. This table is referenced in Section II.D.

Table 2: Summary statistics: workers who use sick leave insurance by duration.

	All	Sick leave claims duration		
		1 to 3 days	4 to 10 days	11 to 29 days
	(1)	(2)	(3)	(4)
Age				
Mean	42.00	39.40	41.79	43.97
Share of workers aged (%)				
25 - 34 years old	33.07	41.74	33.74	26.38
35 - 44 years old	25.11	26.22	25.23	24.56
45 - 54 years old	23.50	19.82	23.15	26.10
55 - 64 years old	18.32	12.23	17.88	22.96
Income (monthly USD)				
Mean	870.30	953.50	852.48	849.98
Standard deviation	367.30	386.19	353.57	358.76
25th percentile	569.20	645.54	564.03	552.85
Median	797.88	892.11	782.44	776.38
75th percentile	1,103.05	1,201.53	1,075.22	1,072.83
90th percentile	1,418.81	1,523.37	1,380.31	1,390.51
Region (%)				
Central	48.35	59.10	48.05	43.78
Mining intensive	6.01	4.09	5.30	7.89
Health - chronic conditions (%)				
Hypertension	14.34	11.46	14.48	17.17
Diabetes	6.22	4.64	6.21	7.92
Share of workers (%)	100	32.00	49.76	38.08
Observations	177,531	56,803	88,345	67,599

Notes: This table presents summary statistics for all male workers who have used sick leave insurance in the past year. Column (1) presents characteristics of all workers who have filed at least one claim with duration of up to 30 days for conditions included in the analysis of this paper (see Table A4 for more details). Columns (2) to (4) present characteristics of workers by duration of the sick leave claims filed. Columns (2) to (4) are not exhaustive, that is, workers can be included in more than one category based on the claims they have filed. This table is referenced in Section II.C.

Table 3: Parameter Estimates

Parameter	Description	Value	Std. error
Preferences parameters			
q	Weekend-streak utility	0.7894	0.2110
μ_ϕ	Value of leisure relative to consumption, mean	56.8930	26.8567
σ_ϕ	Value of leisure relative to consumption, std. dev.	39.8246	13.1373
μ_{κ_0}	Compliance costs, mean	0.8290	0.2320
σ_{κ_0}	Compliance costs, standard deviation	1.7714	0.2616
κ_1	Cost of one day deviation	0.3594	0.0960
κ_2	Cost of two days deviation	0.3217	0.0998
κ_3	Cost of three days deviation	0.0041	0.0021
Measurement error			
p_{me}	Prob. physician assigns one day more (less) than asked	0.2992	0.0430
Rounding			
p_7	Prob. SL duration is round to the closest multiple of 7	0.5746	0.1751

Notes: This table presents the model's estimated parameters. The standard errors are based on 200 bootstrap simulations. This table is referenced in Section V.

Table 4: Moments used in the estimation.

Moments	Data (1)	Model (2)
Preferences parameters		
Weekend streak utility		
Weekend streak days relative to non-streak days	0.1233	0.1546
Compliance costs		
Sh. of SLC with 0 day deviation	0.4300	0.4635
Sh. of SLC with 1 day deviation	0.4120	0.3745
Sh. of SLC with 2 days deviation	0.2480	0.2401
Sh. of SLC with 3 days deviation	0.1975	0.1858
Sh. of SLC with 4 days deviation	0.1663	0.1371
Value of time outside work		
Mean time outside work to consumption ratio	0.2972	0.3342
SD time outside work to consumption ratio	0.1962	0.1965
Measurement error		
Sh. Monday SLC - sh Tuesday SLC conditional to 5-days-long and a day of recovery	0.0822	0.0802
Rounding		
Share of 7-days-long claims	0.1364	0.1050

Notes: This table presents the moments used to estimate the model's parameter. Column 2 reports the data moments. Column 3 reports simulated moments.This table is referenced in Section V.B.

Table 5: Out-of-sample: selected moments from 2019 data and simulated counterparts

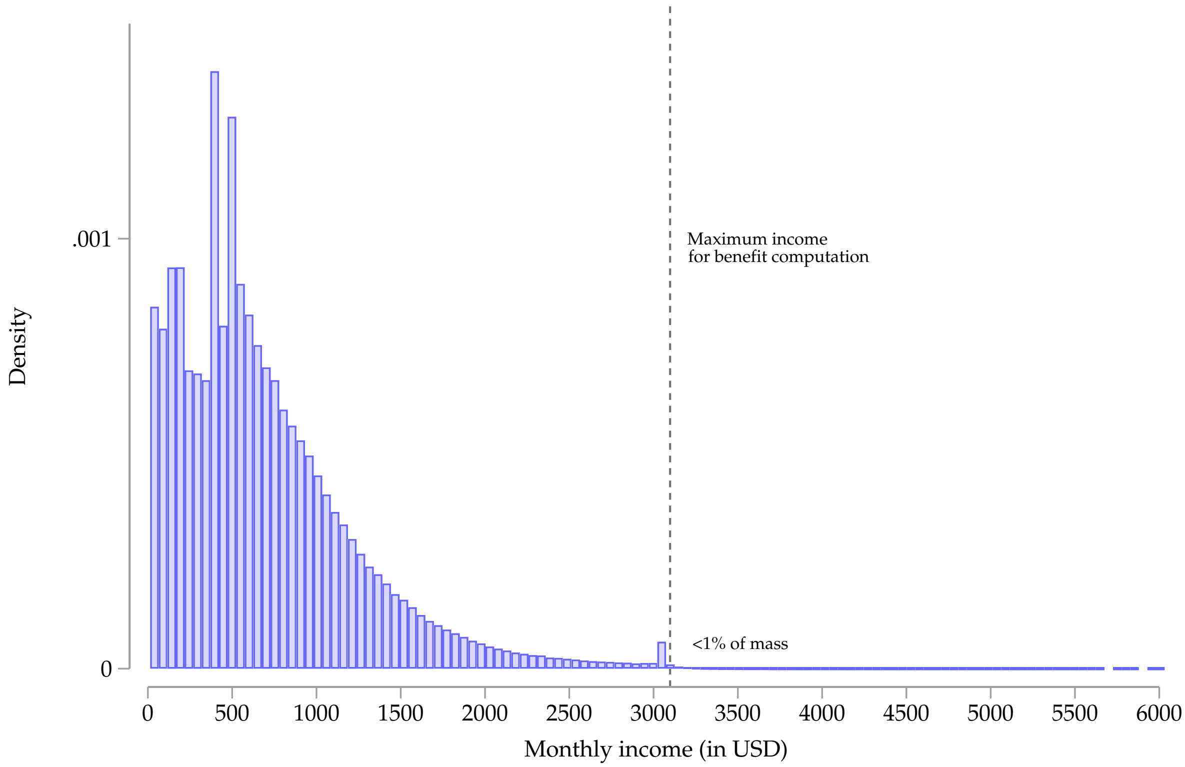
Moments	Data (1)	Model (2)
Preference parameters		
Weekend streak utility		
Weekend streak days relative to non-streak days	0.1158	0.1428
Compliance costs		
Sh. of SLC with 0 day deviation	0.4385	0.4611
Sh. of SLC with 1 day deviation	0.4223	0.3819
Sh. of SLC with 2 days deviation	0.2293	0.2330
Sh. of SLC with 3 days deviation	0.1843	0.1683
Sh. of SLC with 4 days deviation	0.1650	0.1232
Value of leisure		
Mean leisure to consumption ratio	0.2993	0.3297
SD leisure to consumption ratio	0.1985	0.1980
Share of SLC - selected durations		
7 days	0.1305	0.1053
8 days	0.0162	0.0162
9 days	0.0057	0.0262
10 days	0.0211	0.0461
11 days	0.0834	0.0939
12 days	0.0404	0.0535
13 days	0.0051	0.0172

Notes: This table presents results from an out-of-sample test of the model performance. Column (1) is constructed using data on sick leave claims filed in 2019. The moments included correspond to those used for the estimation of preference parameters and the shares of claims with duration in the neighborhood of 11 days. To test the performance of the model, I compare these moments with their model-simulated counterparts. These are presented in column (2). To construct column (2) I simulate the model based on a representative sample drawn from the 2019 data and the estimated vector of preference parameters. This table is referenced in Section V.B.

Appendix A. Additional Figures and Tables

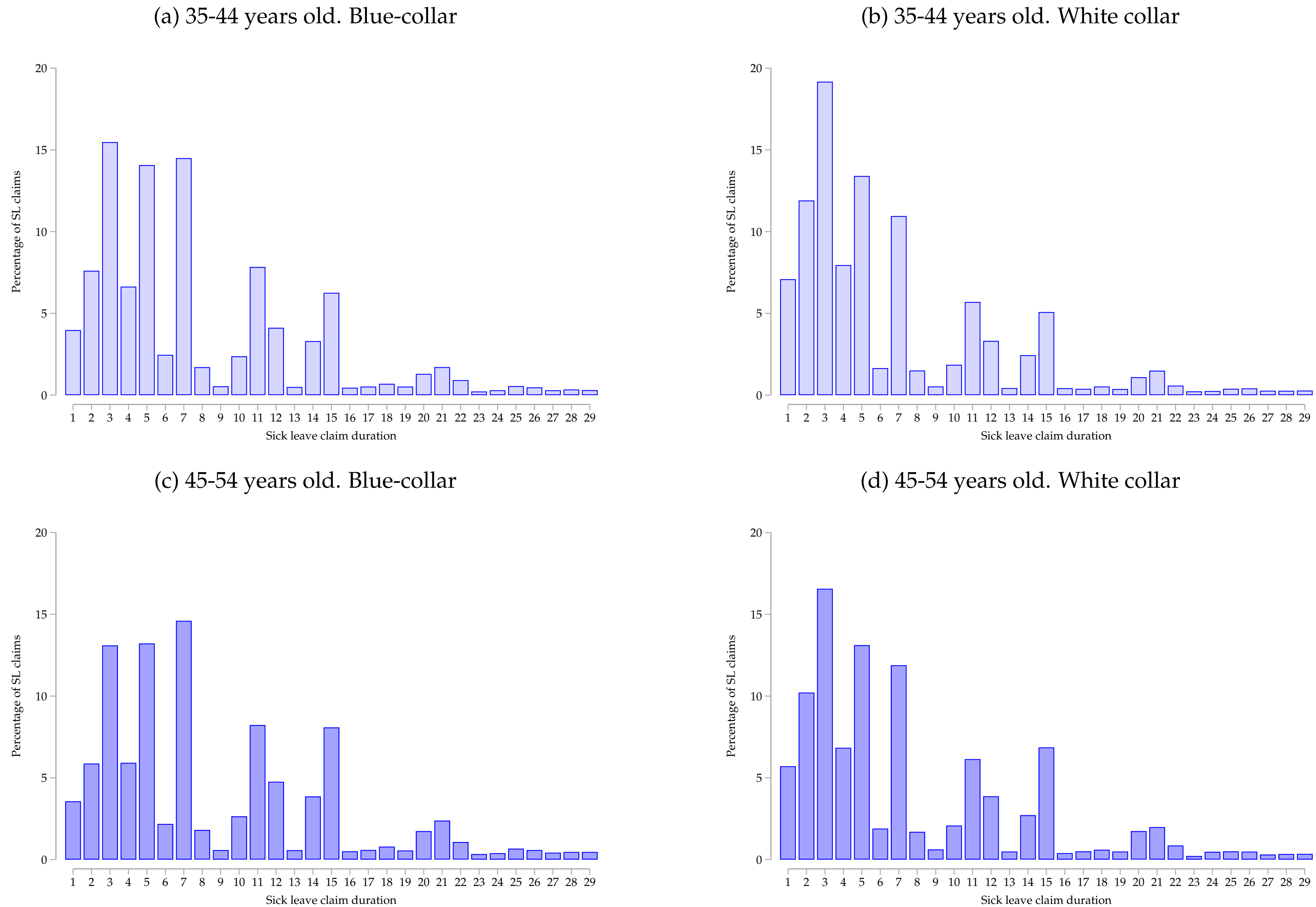
A.I Additional Figures

Figure A1: Distribution of monthly income (in USD). Workers eligible to file a sick leave claim



Notes: This figure shows the distribution of monthly income (in USD) for workers eligible to file a sick leave claim in 2017. The vertical line indicates the income level associated with the maximum benefit threshold. This figure is referenced in Section II.B.

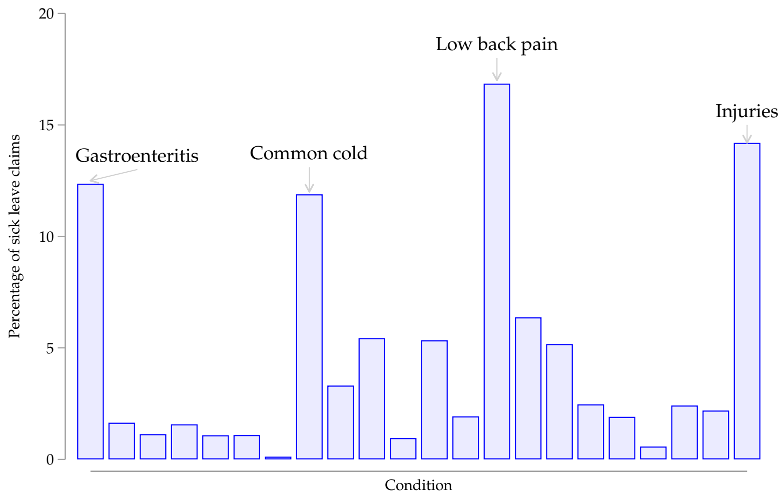
Figure A2: Histogram of days on leave by workers characteristics



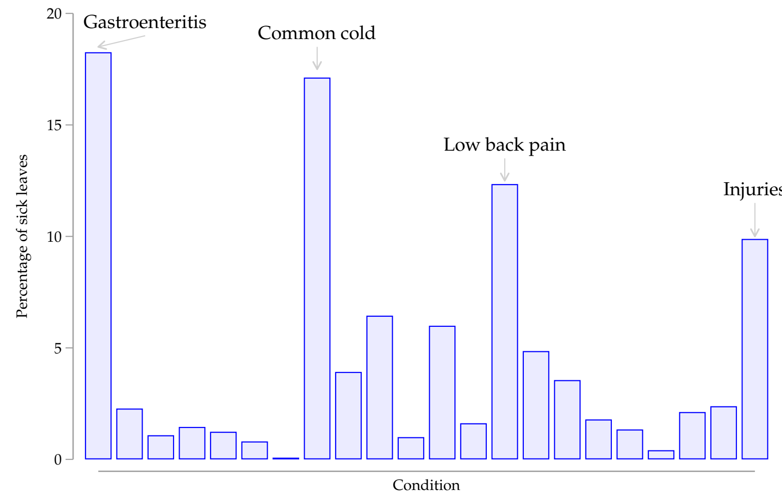
Notes: This figure shows the distribution of days on leave by workers' age and occupation for the youngest and oldest group of workers. Sample includes male private-sector employees. Blue-collar worker refers to workers who engage in hard manual labor, typically agriculture, manufacturing, construction, mining, or maintenance. White-collar worker refers to workers whose daily work activities do not involve manual labor—e.g., teachers or administrative staff. Additional groups are presented in Appendix Figure 3. This figure is referenced in Section II.C and in Section IV.A.

Figure A3: Histogram of days on leave by workers characteristics

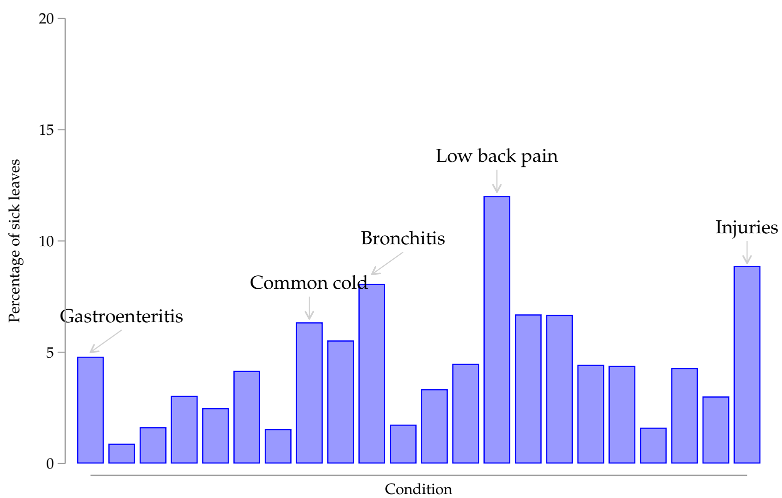
(a) 25-34 years old. Blue-collar



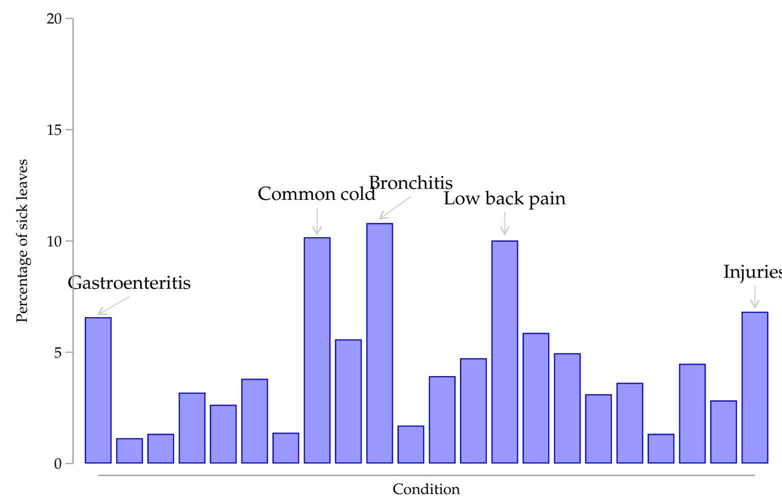
(b) 25-34 years old. White collar



(c) 55-64 years old. Blue-collar

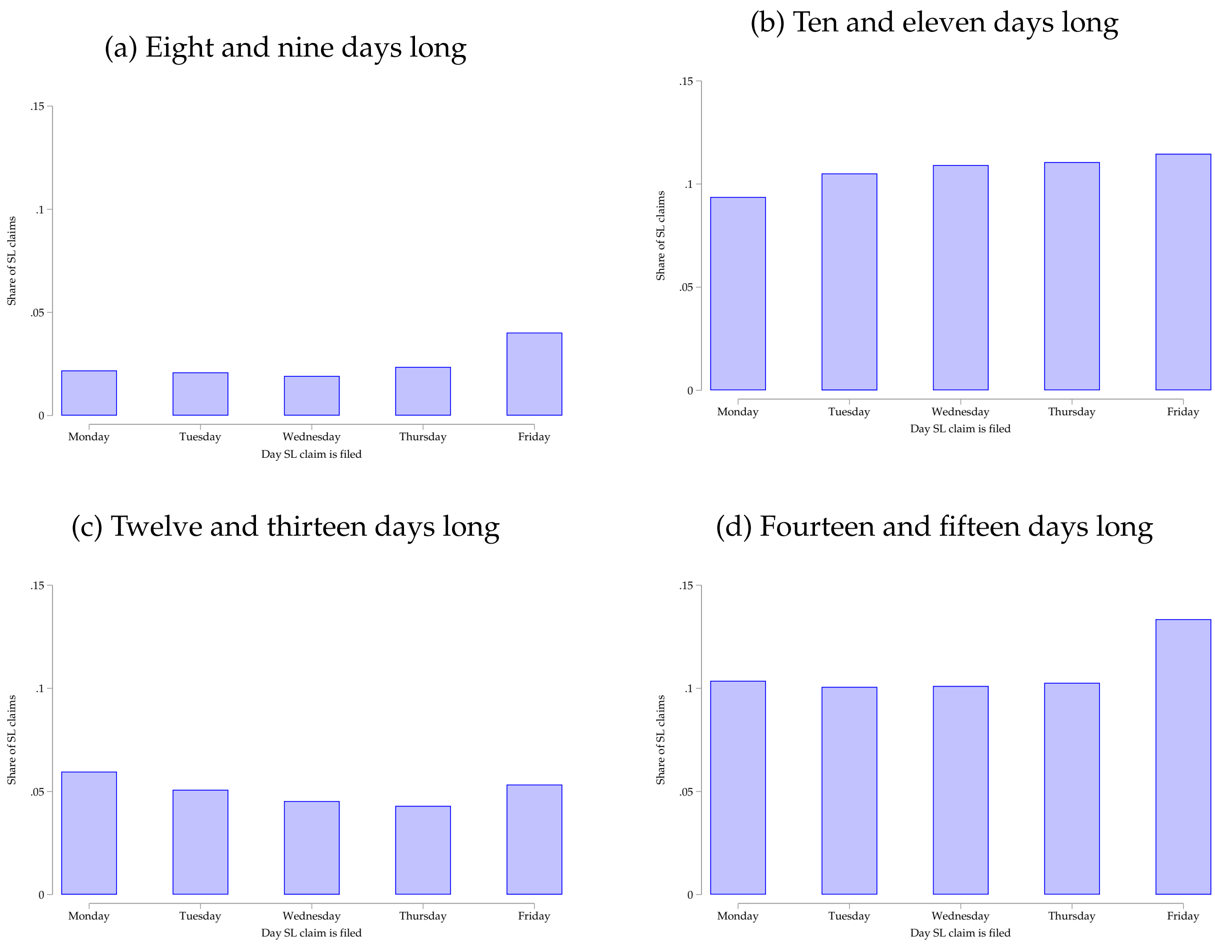


(d) 55-64 years old. White collar



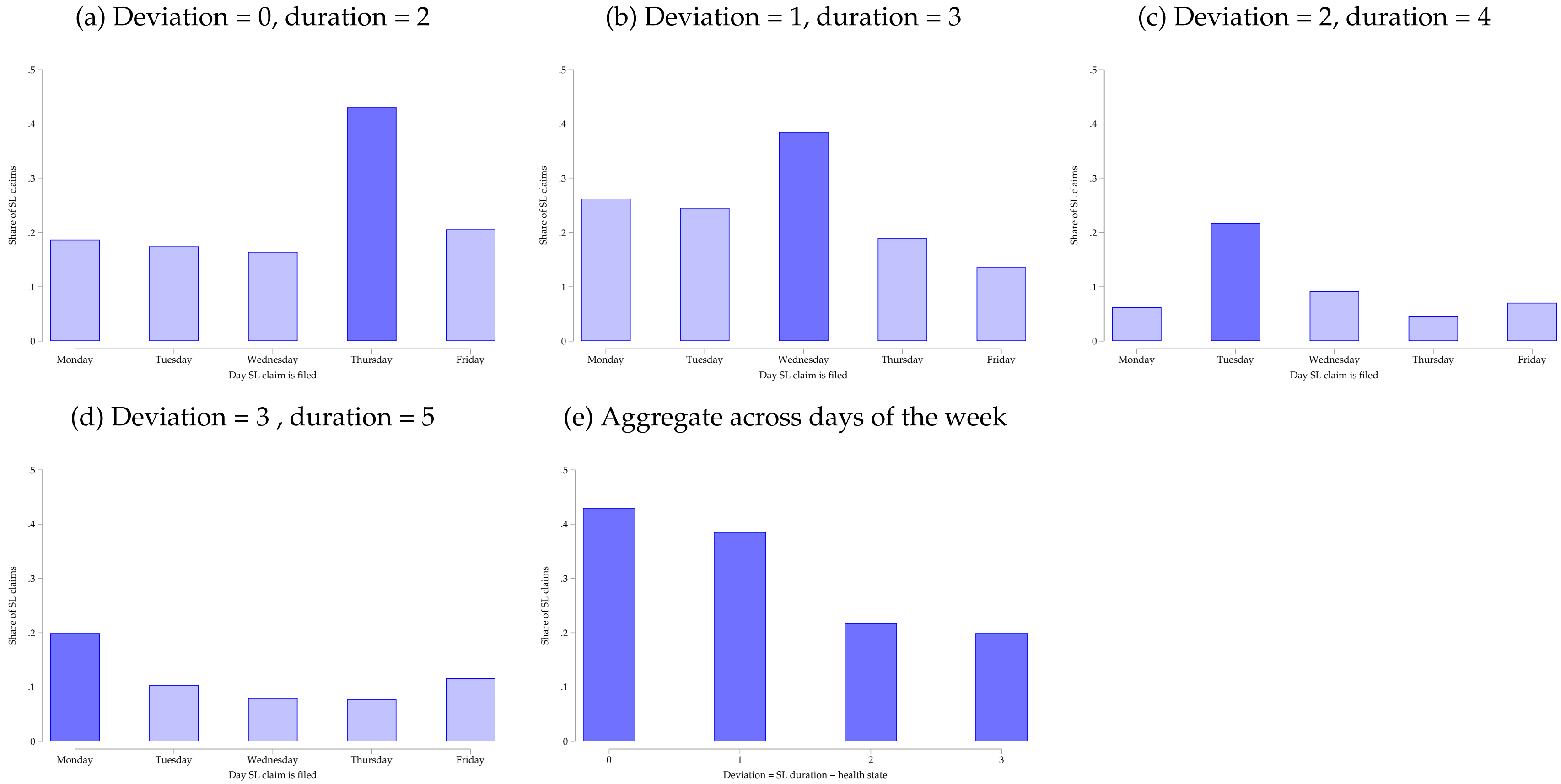
Notes: This figure shows the probability that a worker contracts disease d by by workers' age and occupation for the youngest and oldest group of workers. Diseases are ordered as presented in Table A9. Sample includes male private-sector employees. Blue-collar worker refers to workers who engage in hard manual labor, typically agriculture, manufacturing, construction, mining, or maintenance. White-collar worker refers to workers whose daily work activities do not involve manual labor—e.g., teachers or administrative staff. This figure is referenced in Section IV.B.

Figure A4: Distribution of sick leave claims by duration and day of the week.



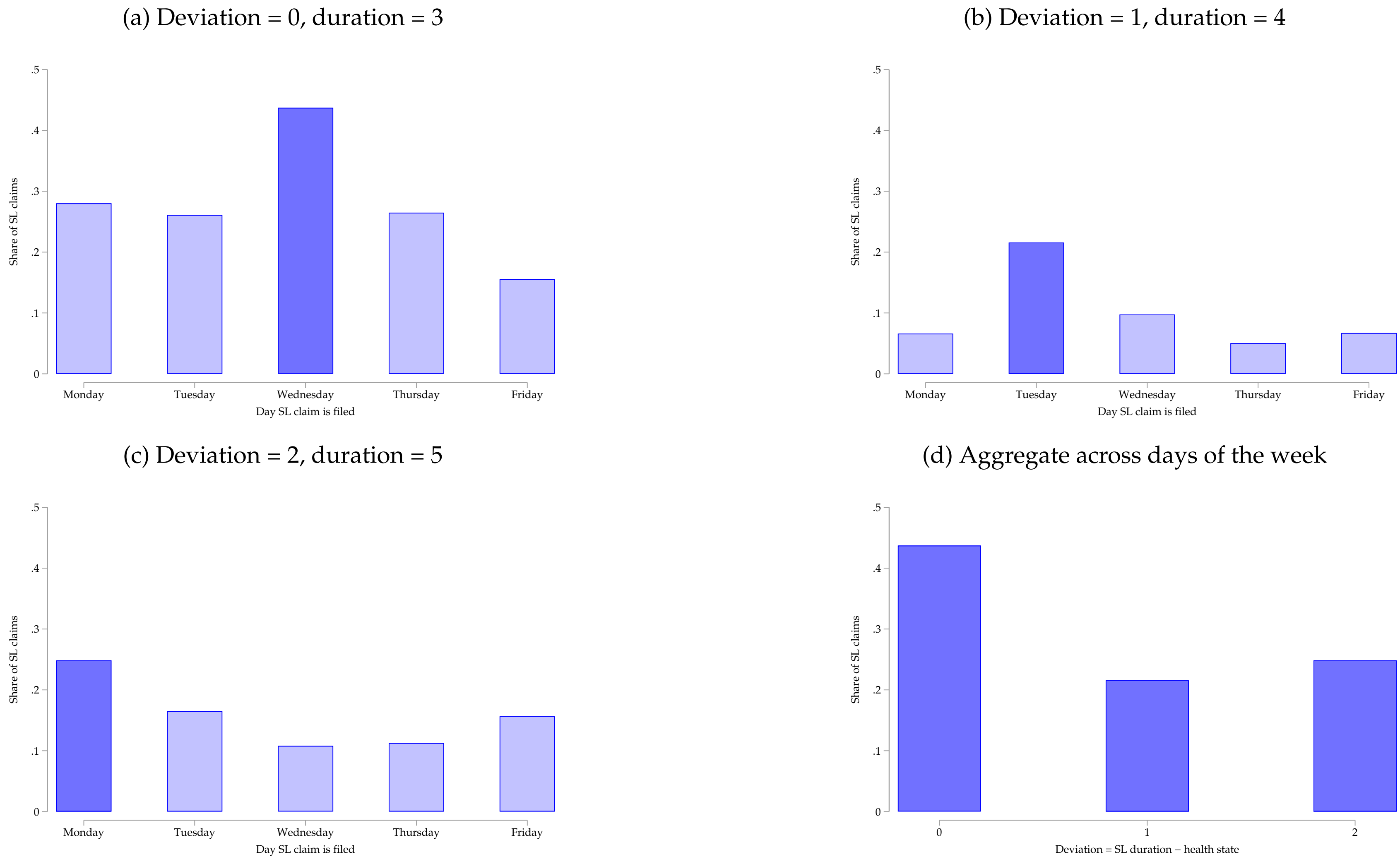
Notes: This figure shows the share of sick leave claims of duration s filed on each day of the week. Each panel aggregates sick leave claims with consecutive duration as stated in the title. This figure is referenced in Section IV.B.

Figure A5: Identification of compliance costs parameter: Sick leave claims by duration and day of the week. Health shock (θ) equals 2-days-long.



Notes: Panels (a) to (d) show the share of sick leave claims with duration s for workers whose main diagnose would implied a health state of 2 days on leave. Panel (e) aggregates the share of sick leave claims across days of the week, including only the weekend-streak combinations, e.g., from panel (a) I only consider the share for Thursday. This figure is referenced in Section IV.B.

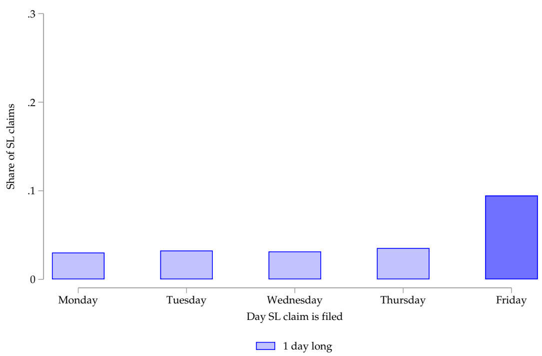
Figure A6: Identification of compliance costs parameter: Sick leave claims by duration and day of the week. Health shock (θ) equals 3-days-long.



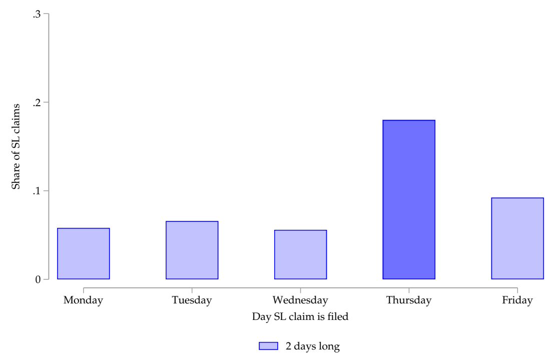
Notes: Panels (a) to (c) show the share of sick leave claims with duration s for workers whose main diagnose would implied a health state of 3 days on leave. Panel (d) aggregates the share of sick leave claims across days of the week, including only the weekend-streak combinations, e.g., from panel (a) I only consider the share for Wednesday. This figure is referenced in Section IV.B.

Figure A7: Days of the week and sick leave claim duration. Conditional to first quarter of the year (Summer quarter).

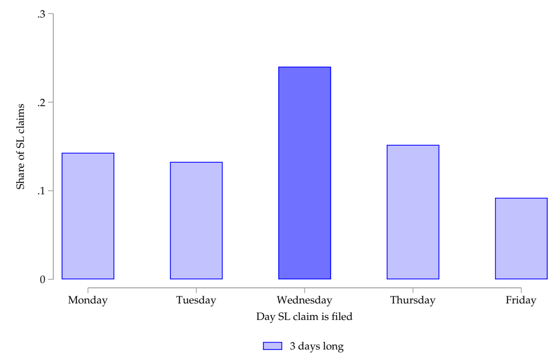
(a) One day long



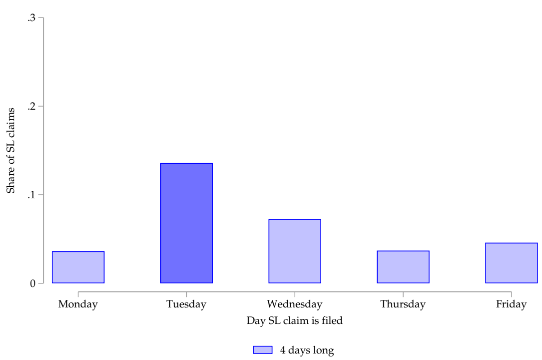
(b) Two days long



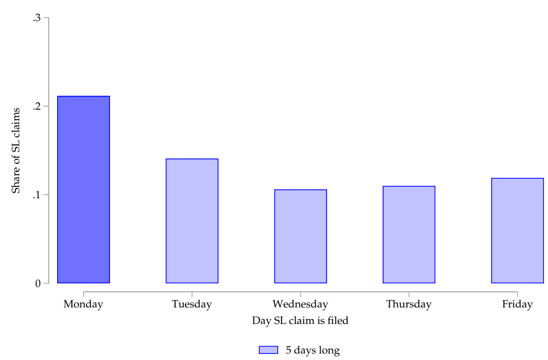
(c) Three days long



(d) Four days long



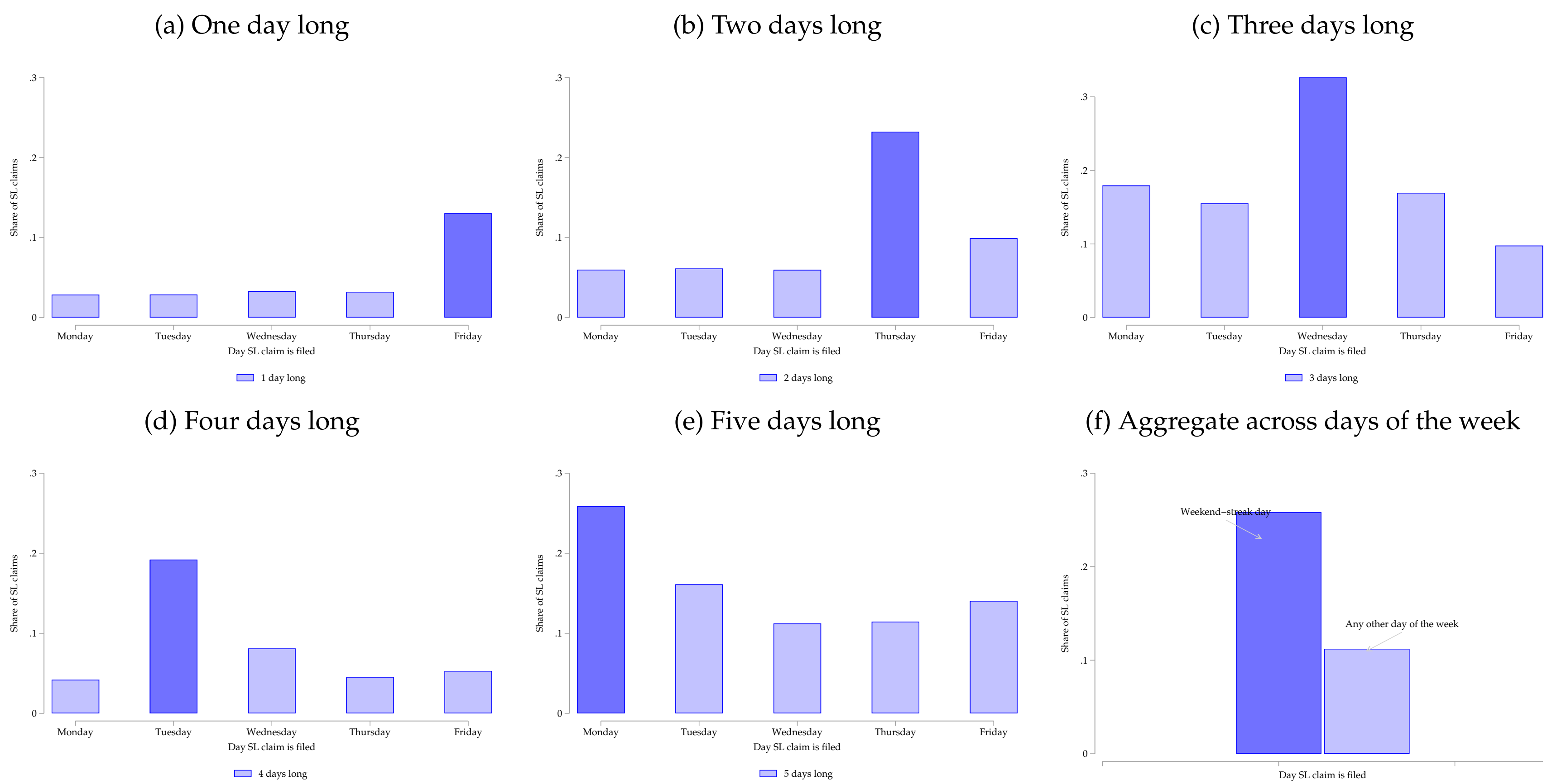
(e) Five days long



(f) Aggregate across days of the week

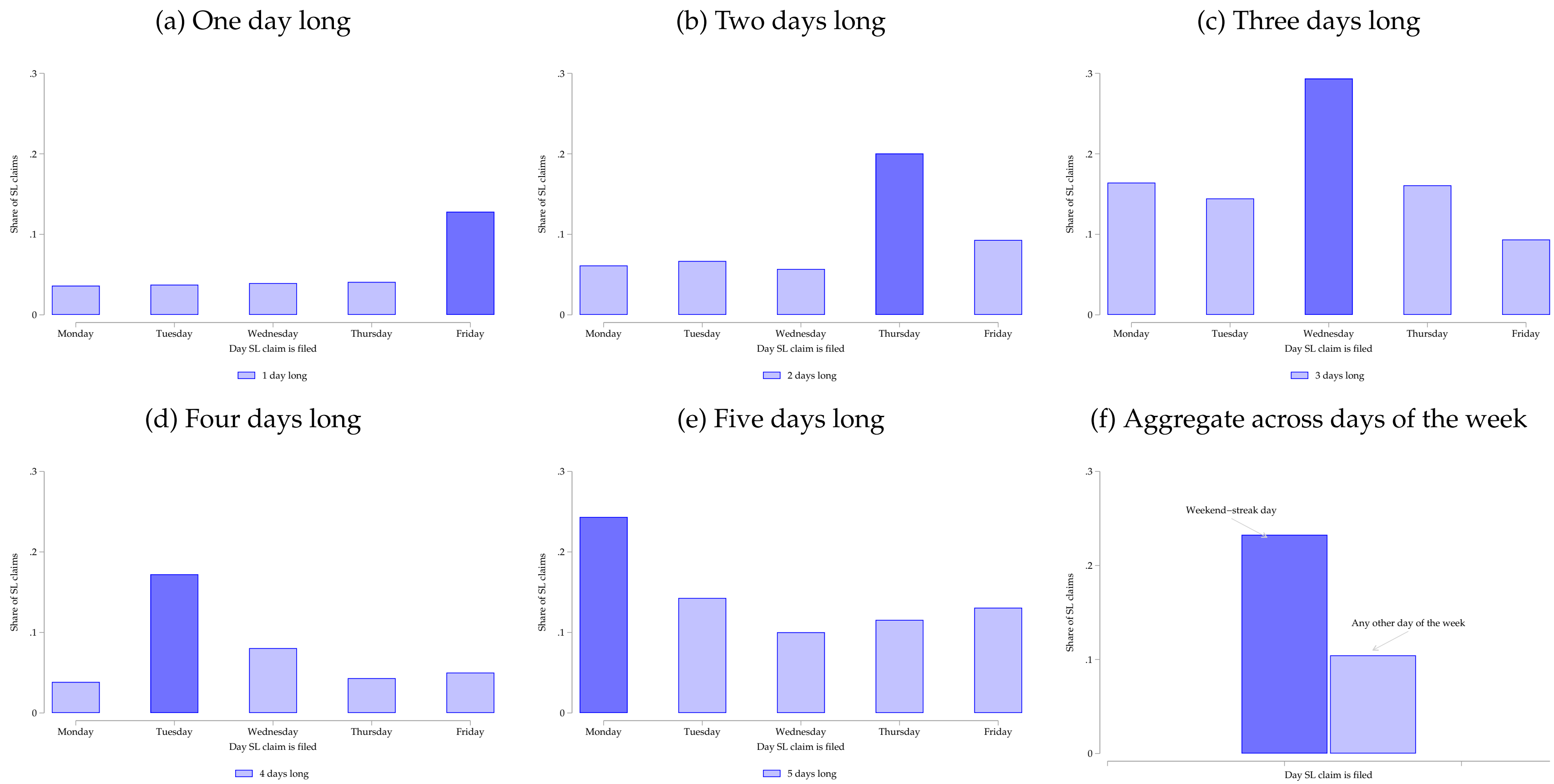
Notes: Panels (a) to (e) show the share of sick leave claims with duration s and the share of seven-days-long sick leave claims filed on each day of the week. Panel (f) aggregates across duration and days of the week: the first bar—labeled "weekend streak"—averages the share of one-to-five-days-long sick leave claims that end of a Friday and are filed any day of the week. For example, one-day-long on Friday, two-days-long on a Thursday, and so on. The second bar—labeled "non-weekend streak"—averages the share one-to-five-days-long sick leave claims filed any other day of the week. For example, two-days-long claims file on Friday. This figure is restricted to sick leave claims filed during the first quarter of the year (Summer quarter in Chile). This figure is referenced in Section V.C.

Figure A8: Days of the week and sick leave claim duration. Conditional to the second quarter of the year (Fall quarter).



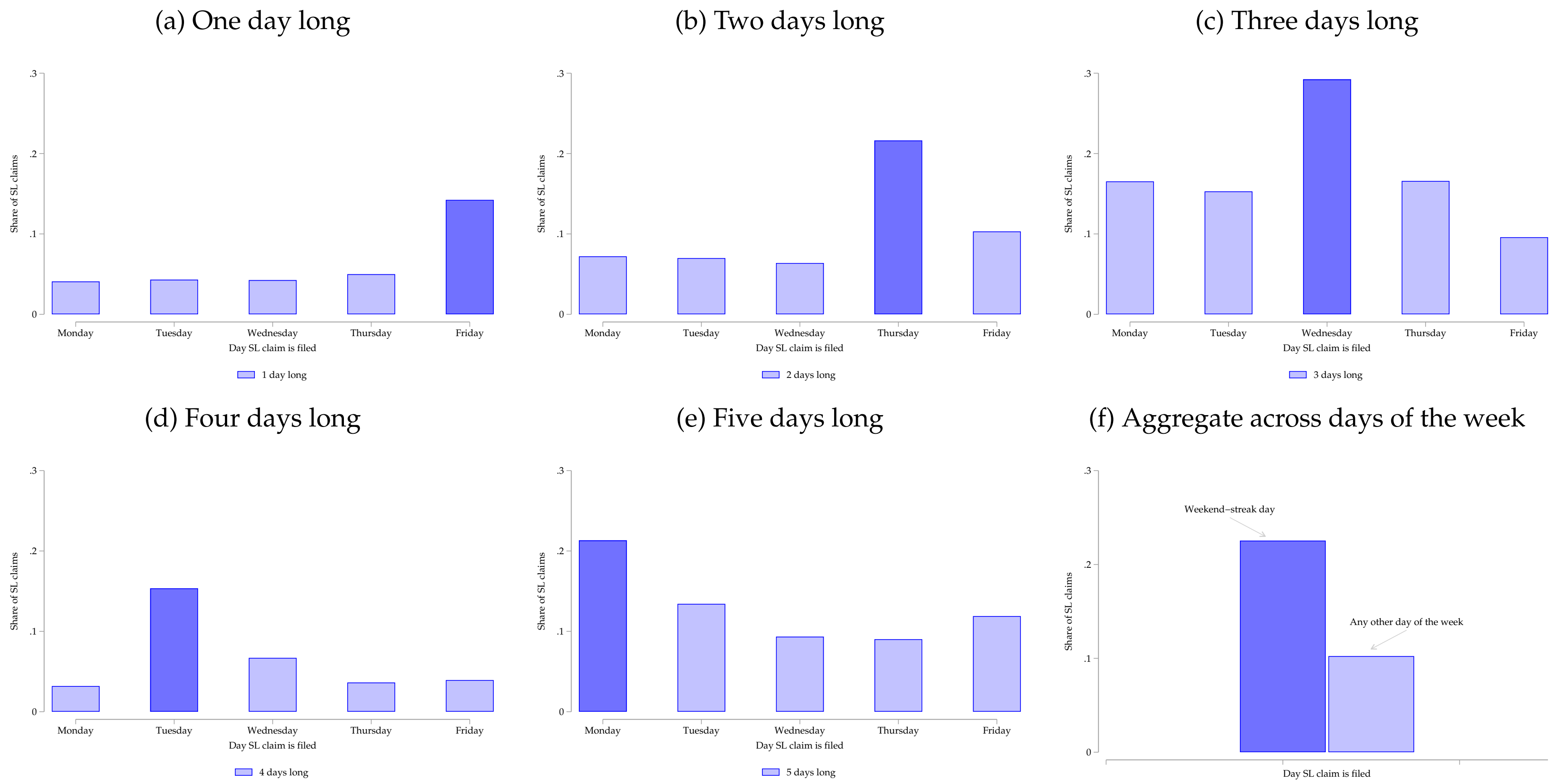
Notes: Panels (a) to (e) show the share of sick leave claims with duration s and the share of seven-days-long sick leave claims filed on each day of the week. Panel (f) aggregates across duration and days of the week: the first bar—labeled “weekend streak”—averages the share of one-to-five-days-long sick leave claims that end of a Friday and are filed any day of the week. For example, one-day-long on Friday, two-days-long on a Thursday, and so on. The second bar—labeled “non-weekend streak”—averages the share one-to-five-days-long sick leave claims filed any other day of the week. For example, two-days-long claims file on Friday. This figure is restricted to sick leave claims filed during the second quarter of the year (Fall quarter in Chile). This figure is referenced in Section V.C.

Figure A9: Days of the week and sick leave claim duration. Conditional to the third quarter of the year (Winter quarter).



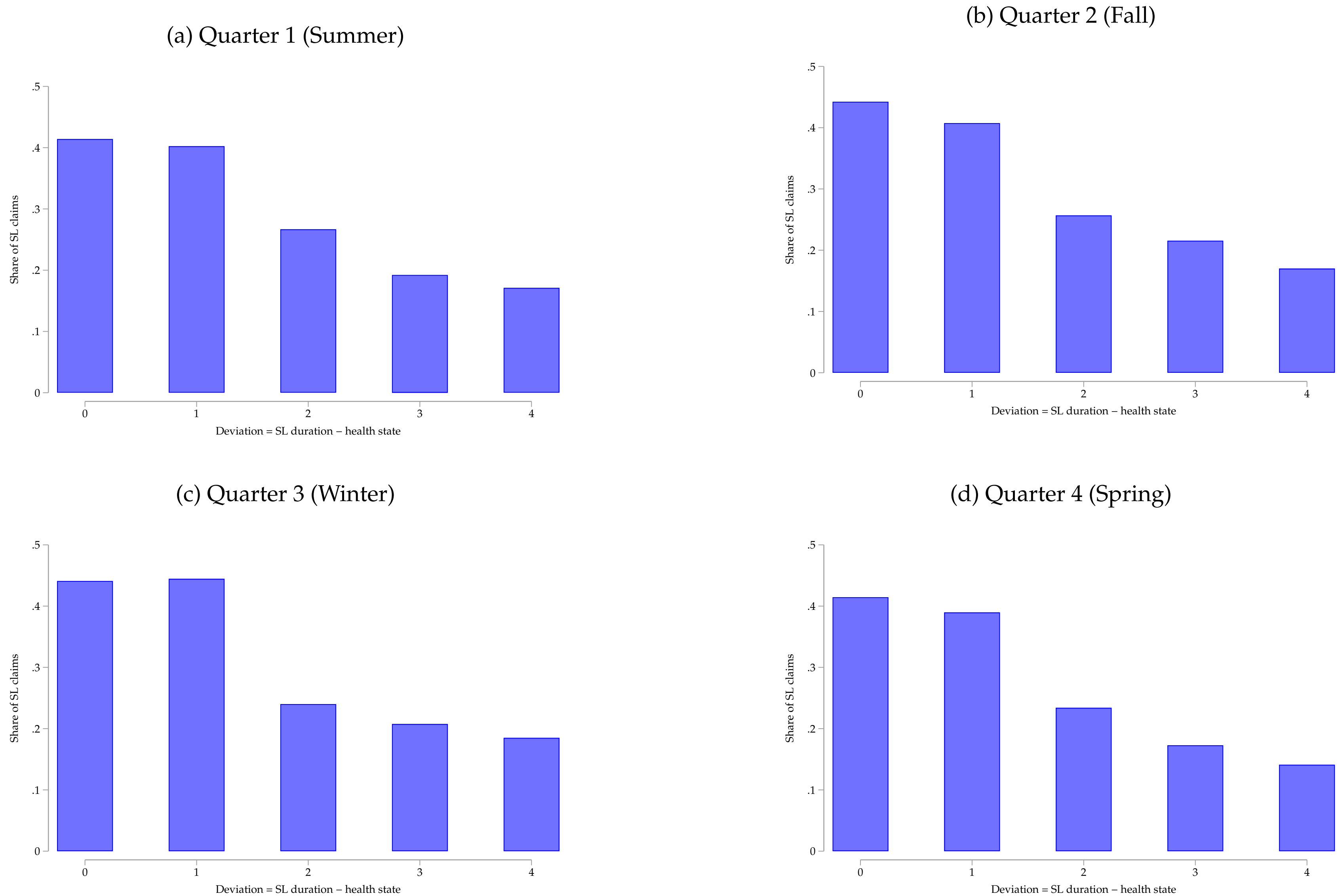
Notes: Panels (a) to (e) show the share of sick leave claims with duration s and the share of seven-days-long sick leave claims filed on each day of the week. Panel (f) aggregates across duration and days of the week: the first bar—labeled "weekend streak"—averages the share of one-to-five-days-long sick leave claims that end of a Friday and are filed any day of the week. For example, one-day-long on Friday, two-days-long on a Thursday, and so on. The second bar—labeled "non-weekend streak"—averages the share one-to-five-days-long sick leave claims filed any other day of the week. For example, two-days-long claims file on Friday. This figure is restricted to sick leave claims filed during the third quarter of the year (Winter quarter in Chile). This figure is referenced in Section V.C.

Figure A10: Days of the week and sick leave claim duration. Conditional to the third quarter of the year (Spring quarter).



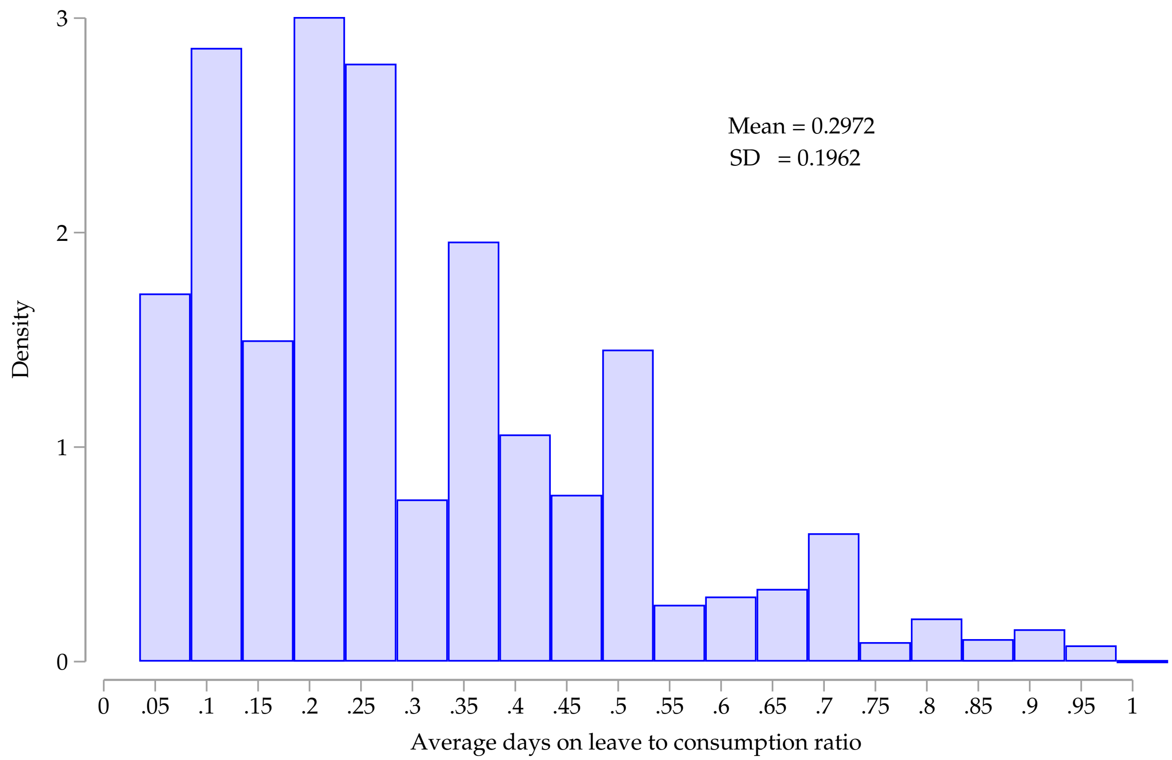
Notes: Panels (a) to (e) show the share of sick leave claims with duration s and the share of seven-days-long sick leave claims filed on each day of the week. Panel (f) aggregates across duration and days of the week: the first bar—labeled "weekend streak"—averages the share of one-to-five-days-long sick leave claims that end of a Friday and are filed any day of the week. For example, one-day-long on Friday, two-days-long on a Thursday, and so on. The second bar—labeled "non-weekend streak"—averages the share one-to-five-days-long sick leave claims filed any other day of the week. This figure is restricted to sick leave claims filed during the fourth quarter of the year (Spring quarter in Chile). This figure is referenced in Section V.C.

Figure A11: Compliance cost function by quarter.



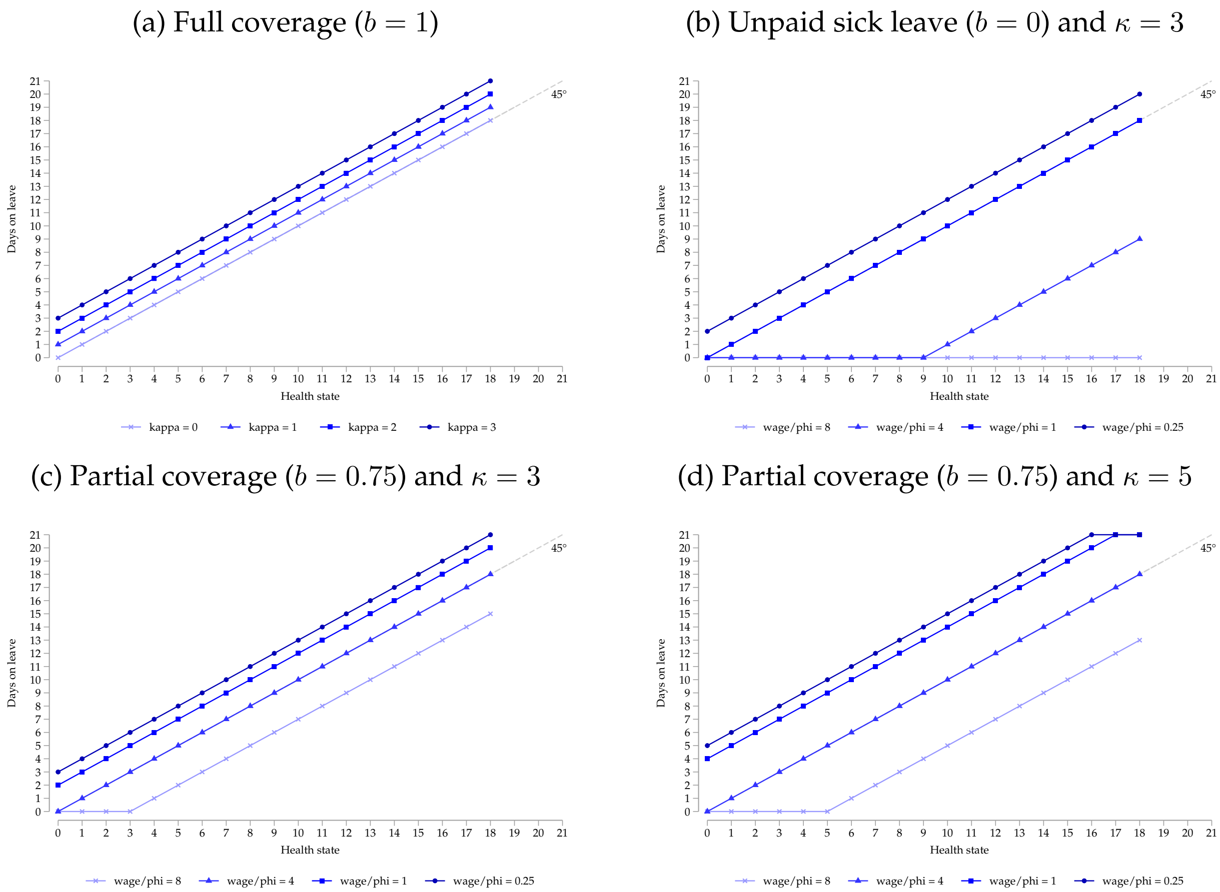
Notes: Panels (a) to (d) show the average share of sick leave claims with deviations between 0 and 4 days for each quarter. Summer quarter goes from Jan to March. The average is computed over sick leave claims with primary diagnosis requiring 1, 2 or 3 days of rest filed on a weekend streak days. This figure is referenced in Section V.C.

Figure A12: Distribution of leisure to consumption ratio from raw data



Notes: This figure shows the distribution of the leisure to consumption ratio LC_i. This figure is referenced in Section IV.B.

Figure A13: Sick Pay Utilization with Linear Contract



Notes: This figure shows the optimal demand of days on leave $s^*(\theta)$ as a function of worker's health status (θ) under the assumption of linear contracts with different levels of coverage. This figure is referenced in Section III.

A.II Additional Tables

Table A1: Summary statistics by healthcare insurance provider

	Government-run insurance (1)	Private insurance (2)	Year(s) (3)
Panel A. Enrollees Characteristics			
Share of enrollees aged			
25 - 34	25.06	31.76	2015-2019*
35 - 44	21.51	28.84	2015-2019*
45 - 54	21.11	20.12	2015-2019*
55 - 64	14.64	11.47	2015-2019*
Share female enrollees	0.44	0.35	2015-2019*
Wages (in USD monthly)			
Average	761.27	1,824.94	2015-2019*
Enrollees w/ wage above median (%)	34.44	86.69	2015-2019*
Metropolitan region (%)	38.04	60.01	2015-2019*
Mining sector (%)	0.50	2.49	2015-2019*
N of enrollees	4,503,474	1,689,240	2015-2019*
Share (%)	72.72	27.28	2015-2019*
Panel B. Sick Leave Claims			
Ratio SL claims to enrollees (%)			
2015	77.66	86.53	2015
2019	98.42	90.66	2019
Approved SL claims (%)	91.94	74.54	2015-2019
Rejected SL claims (%)	5.31	14.76	2015-2019
Ratio days on leave to SL claim	13.09	10.24	2015-2019
Annual cost per enrollee (in USD)	240.69	463.61	2015-2019
Ratio of total annual cost to paid days on leave	24.91	58.90	2015-2019
Annual cost (% of GDP)	0.51	0.37	2015-2020
N of sick leave claims	3,910,482	1,473,540	2015-2019
Share (%)	72.63	27.37	2015-2019

Notes: Panel A presents summary statistics of individuals enrolled in plans offered by each healthcare insurance provider. Only individuals eligible to file a sick leave claim are included in the computations. Panel B shows characteristics of the sick leave claims handled by each insurer. Data come from the Annual Statistics of the Sick Leave System published by the Social Security Administration (SUSESO, 2020; 2019; 2018; 2017; 2016). The reported data are annual counts. Statistics in this table correspond to averages for 2015 - 2019, * indicates that data for 2018 are not available. The median monthly wage is computed from the 2017 CASEN survey, and using this figure I compute the share of workers with monthly salary above the median. GDP data comes from the World Bank national accounts data. SL stands for sick leave. This table is referenced in Section II.A.

Table A2: Sick leave claims and sick leave spells definitions

	(1)	(2)	(3)
Number of sick leave claims	1,483,103	657,125	551,647
Number of sick leave spells	1,030,613	437,418	365,127
N of SL claims in a spell (% of claims)			
One claim	55.43	51.71	51.22
Two claims	16.85	17.19	17.30
Three claims	7.30	7.93	8.00
Four claims	4.49	5.06	5.12
Five claims	3.19	3.63	3.67
Six or more claims	12.75	14.48	14.70
Among sick leave spells with more than 1 claim* (% of claims)			
Diagnoses change within spell			
Yes — 4 digits disease code	30.83	30.01	29.97
Yes — 3 digits disease code	28.67	27.86	27.82
Physician change within spell	31.21	30.87	30.83
Sample: Private sector workers			
Gender	All	Male	Male
Age	18-70	18-70	25-64

Notes: This table presents counts and summary statistics of sick leave claims and sick leave spells. A spell is a group of consecutive claims—these are considered one claim for the computation of sick leave benefits. The first row counts each sick leave claim as one observation and the second row considers the number of sick leave spells. The subsequent rows explore the composition of a spell in terms of number of claims and whether diagnoses and physicians changed withing a spell. * indicates that proportions are computed for spells composed by two to five sick leave claims. This table is referenced in Section II.B and in Section II.C.

Table A3: Average recovery times - Examples from Peruvian Handbook

Workers' characteristics and diagnoses	Correction factor and recovery time
Example 1	
Lumbago with sciatica (M544)	14
43 years old	1.05
Operator/manual worker (blue collar)	1.5
Optimal time	22.05
Example 2	
Common cold (J00)	3
25 years old	0.87
Teacher (white collar)	0.75
Optimal time	2
Example 3	
Infectious gastroenteritis (A09)	2
57 years old	1.3
Office manager (white collar)	0.75
Optimal time	2

Notes: This table presents examples on how to construct the average recovery time based on workers' characteristics and sick leave diagnoses. This table is referenced in Section II.C and Section IV.B.

Table A4: Conditions included in the analysis by ICD-10 group

ICD Group	Description	Included (=1 if yes)	Sick leave claims Number	Share (%)
		(1)	(2)	(3)
A00-B99	Certain infectious and parasitic diseases	1	31,244	8.56
C00-D49	Neoplasms	0	6,515	1.78
D50-D89	Blood and blood-forming organs	0	478	0.13
E00-E89	Nutritional and metabolic diseases	0	3,842	1.05
G00-G99	Nervous system	1	8,758	2.40
H00-H59	Eye and adnexa	1	6,141	1.68
H60-H95	Ear and mastoid process	1	6,246	1.71
I00-I99	Circulatory system	1	15,139	4.15
J00-J99	Respiratory system	1	64,823	17.75
K00-K95	Digestive system	1	25,854	7.08
L00-L99	Skin and subcutaneous tissue	0	8,762	2.40
M00-M99	Musculoskeletal system	1	108,908	29.83
N00-N99	Genitourinary system	1	11,605	3.18
O00-O9A	Pregnancy and childbirth	0	<50	0.01
P00-P96	Certain conditions of the perinatal	0	149	0.04
Q00-Q99	Congenital malformations	0	331	0.09
R00-R99	Abnormal clinical and laboratory findings	1	9,840	2.69
S00-S99	Injuries	1	44,922	12.30
T00-T88	Poisoning and external causes	0	4,385	1.20
U00-U85	Codes for special purposes	0	<50	0.00
V00-Y99	External causes of morbidity	0	3,578	0.98
Z00-Z99	Contact with health services	0	3,577	0.98
Total included			329,312	90.19
Total			365,127	

Notes: This table reports the health conditions included in the analysis, the number of sick leave claims filed in 2017, and what share these represent of the universe of claims. The criteria for excluding the selected health conditions is discussed in detailed in Section Appendix B. This table is referenced in Section II.C.

Table A5: Average recovery time by workers characteristics

		25 - 34 years old		35 - 44 years old		45 - 54 years old		55 - 64 years old	
ICD group	Main diagnoses	Blue c. (1)	White c. (2)	Blue c. (3)	White c. (4)	Blue c. (5)	White c. (6)	Blue c. (7)	White c. (8)
A00-A99	Infectious gastroenteritis	2	1	3	2	3	2	3	2
G00-G99	Migraine and headaches	3	2	4	2	4	2	5	3
G00-G99	Carpal tunnel syndrome	13	8	17	10	17	10	21	13
H00-H59	Conjunctivitis	5	3	7	4	7	4	9	5
H60-H95	Vertigo	4	2	5	3	5	3	6	4
I00-I99	Hypertension	4	3	6	3	6	3	7	4
I00-I99	Myocardial infarction	16	10	21	13	21	13	26	16
J00-J06	Common cold	3	2	4	2	4	2	5	3
J09-J18	Influenza and pneumonia	4	3	5	3	5	3	6	4
J20-J22	Bronchitis	5	3	7	4	7	4	8	5
J23-J99	Other respiratory diseases	8	5	9	6	9	6	11	7
K00-K95	Noninfective gastroenteritis	2	1	2	1	2	1	3	2
K00-K95	Inguinal hernia	6	4	9	5	9	5	11	7
M50-M54	Chronic low back pain	10	6	12	7	12	7	14	8
M50-M54	Lumbago with sciatica	10	6	12	7	12	7	14	8
M60-M79	Tendinitis	8	5	9	6	9	6	10	6
M60-M79	Shoulder lesions	8	5	9	6	9	6	10	6
Other M	Arthritis	9	5	10	6	11	6	12	7
Other M	Knee injuries	12	7	14	8	14	8	16	10
N00-N99	Renal colic	4	3	5	3	5	3	6	4
R00-R99*	Abdominal and pelvic pain	2	1	3	2	3	2	3	2
S00-S99	Injuries (e.g., sprain ankle)	14	8	16	9	16	9	18	11

Notes: This table shows the average recovery time by workers' age and occupation type for 22 disease groups. Blue c. stands for blue collar and white c. stands for white collar. Table A6 indicates what occupations and industries are classified as blue and white collar. This table is referenced in Section II.C and Section IV.B.

Table A6: Workers' occupation, industry and manual work classification

Occupation	Industry	Blue collar (=1 if yes) (1)
Executive, managers	Any	0
Professor, lecturer, teacher	Any	0
Other professional	Any	0
Sales representative	Any	0
Admin staff	Any	0
Factory worker	Any	1
Trabajador de casa particular	Any	1
Technician	Any	1
Unknown	Agriculture	1
Unknown	Natural Resources and mining	1
Unknown	Manufacturing	1
Unknown	Construction	1
Unknown	Utilities	1
Unknown	Retail trade	0
Unknown	Transportation, warehousing and telecommunications	1
Unknown	Service-Providing Industries	0
Unknown	Public administration	0
Unknown	Not specified	n.a.

Notes: This table reports workers' occupation, industry, and whether its combination implies the worker is considered a blue-collar (or manual) worker or not. If information is available on occupation and industry, I use worker's occupation to classified the worker as a blue-collar. If occupation is not available, I use workers' industry information. When neither occupation or industry is available, I drop observations for this worker. "n.a." stands for not applicable. This table is referenced in notes to Table A5 and in Section II.C.

Table A7: Sample construction

	2017
Panel A. Sick leave claims	
Single claims - from clean dataset	2,698,993
Observations without demographic information	22,757
Workers' age not in the interval [18,70]	50,369
Worker is not Chilean	61,740
Worker not enrolled in a public insurance plan	33,560
Observations without income information	8,920
Observations	2,521,647
Panel B. Sick leave spells	
Single spells	1,825,904
Condition on private sector workers	1,030,613
Condition on male workers	437,418
Condition on ages 25-64	365,127
Condition on diagnoses included in analysis	329,312

Notes: Panel A of this table shows the counts of sick leave claims drop due to each sample selection criterion. Panel B shows the counts of sick leave spells—consecutive claims with continuous start and end dates—for each sample selection criterion. A complete list of diagnoses included in the analysis is provided in Table A4. This table is referenced in Section II.C.

Table A8: Summary statistics: Private sector workers who have used SL benefits

	Any	Included conditions	
		All	Up to 30 days
	(1)	(2)	(3)
Age			
Mean	43.39	43.25	42.00
Share of workers aged (%)			
25 - 34 years old	29.21	29.51	33.07
35 - 44 years old	23.71	23.98	25.11
45 - 54 years old	24.70	24.63	23.50
55 - 64 years old	22.38	21.89	18.32
Income (monthly USD)			
Mean	857.97	862.72	870.30
Standard deviation	367.59	368.79	367.30
25th percentile	555.28	559.42	569.20
Median	782.85	788.21	797.88
75th percentile	1,089.33	1,095.32	1,103.05
90th percentile	1,408.30	1,414.46	1,418.81
Region (%)			
Central	46.78	47.32	48.35
Mining intensive	6.77	6.67	6.01
Health - chronic conditions (%)			
Hypertension	16.89	16.71	14.34
Diabetes	8.24	7.79	6.22
Share of workers (%)	100.00	92.59	72.16
Observations	246,017	227,797	177,531

Notes: This table presents summary statistics for all male workers who had used sick leave benefits in the past year based on the conditions and duration of sick leave claims. The sample includes private sector employees age 25 to 64 years old. Income statistics are based on the winsorized distribution where the lowest and highest 5% of the income values are excluded. Sick leave claims of up to 30 days account for 95% of all claims filed in a year. This table is referenced in Section II.C.

Table A9: Probability of filing a SLC for each disease group by workers' characteristics

		25 - 34 years old		35 - 44 years old		45 - 54 years old		55 - 64 years old	
ICD group	Main diagnoses	Blue c. (1)	White c. (2)	Blue c. (3)	White c. (4)	Blue c. (5)	White c. (6)	Blue c. (7)	White c. (8)
A00-A99	Infectious gastroenteritis	12.36	18.26	8.91	13.06	6.51	9.36	4.80	6.57
G00-G99	Migraine and headaches	1.64	2.28	1.57	1.86	1.19	1.64	0.89	1.14
G00-G99	Carpal tunnel syndrome	1.13	1.08	1.31	1.30	1.48	1.31	1.63	1.33
H00-H59	Conjunctivitis	1.57	1.46	1.92	2.40	2.40	2.40	3.03	3.19
H60-H95	Vertigo	1.08	1.24	1.32	1.23	1.85	1.68	2.49	2.64
I00-I99	Hypertension	1.09	0.80	1.88	1.70	2.95	2.85	4.16	3.81
I00-I99	Myocardial infarction	0.12	0.08	0.31	0.34	0.77	0.71	1.55	1.38
J00-J06	Common cold	11.89	17.12	10.64	14.87	8.62	12.79	6.34	10.17
J09-J18	Influenza and pneumonia	3.31	3.92	3.78	4.66	4.68	4.88	5.53	5.58
J20-J22	Bronchitis	5.44	6.44	5.91	7.33	7.18	8.35	8.06	10.80
J23-J99	Other respiratory diseases	0.96	1.00	1.11	1.13	1.30	1.37	1.74	1.71
K00-K95	Noninfective gastroenteritis	5.34	5.99	4.19	4.99	3.95	4.57	3.34	3.93
K00-K95	Inguinal hernia	1.93	1.62	3.02	2.87	3.82	4.02	4.48	4.73
M50-M54	Chronic low back pain	16.85	12.35	16.73	12.78	14.17	11.43	12.02	10.02
M50-M54	Lumbago with sciatica	6.37	4.86	8.11	6.33	7.54	6.60	6.70	5.87
M60-M79	Tendinitis	5.17	3.56	5.99	4.37	6.55	4.98	6.67	4.95
M60-M79	Shoulder lesions	2.47	1.79	3.78	2.49	4.62	3.61	4.43	3.11
Other M	Arthritis	1.91	1.34	2.25	1.79	3.11	2.63	4.38	3.63
Other M	Knee injuries	0.57	0.41	0.70	0.66	1.16	0.99	1.61	1.33
N00-N99	Renal colic	2.42	2.12	3.06	3.23	3.63	3.47	4.29	4.48
R00-R99*	Abdominal and pelvic pain	2.19	2.38	1.99	2.37	2.38	2.72	3.01	2.83
S00-S99	Injuries (e.g., sprain ankle)	14.19	9.89	11.53	8.25	10.13	7.64	8.87	6.82

Notes: This table shows the probability of filling a sick leave claim for each disease (d) by workers' group (b). Each of these probabilities is computed as the ratio of sick leave claims with diagnosis d and all claims from group b, thus columns add up to 100. Main diagnoses indicates the most common condition for a disease group. Blue c. and white c. stand for blue-collar and white-collar occupations respectively. These probabilities are plotted in Figure A3. This table is referenced in Section II.D.

Table A10: Number of business days on leave (s_l)

	Number of days on leave (s_c)							
Day of the week (dow)	1	2	3	4	5	6	7	8
Monday	1	2	3	4	5	5	5	6
Tuesday	1	2	3	4	4	4	5	6
Wednesday	1	2	3	3	3	4	5	6
Thursday	1	2	2	2	3	4	5	6
Friday	1	1	1	2	3	4	5	6

Notes: This table shows the number of business days on leave (s_l) as a function of (total) days on leave (s_c) and day of the week (dow) a sick leave claim is filed. This table is referenced in Section III.A.

Table A11: Identification of weekend-streak utility parameter (q): estimates from raw data

Duration	Day of the week		
	Weekend streak	Non-weekend streak	Difference
	(1)	(2)	(3)
1 day long	0.1219	0.0355	0.0864
2 days long	0.2062	0.0672	0.1391
3 days long	0.2872	0.1482	0.1390
4 days long	0.1640	0.0489	0.1151
5 days long	0.2330	0.1216	0.1114
Simple average	0.2025	0.0843	0.1182
Weighted average	0.2274	0.1041	0.1233

Notes: This table presents the distribution of sick leave claims by duration and day of the week. Weekend streak refers to the day of the week a sick leave claim should start to finish on Friday. For example, when duration is one day, weekend streak refers to Friday, when duration is two days, it refers to Thursday. The non-weekend streak category groups all the other days of the week. The share of sick leave claims of duration s filed on day dow is computed as the ratio between the number of claims with duration s filed on dow and the number of claims of filed on dow with duration between one and fifteen days. Figure 5 presents this table graphically. This table is referenced in Section IV.B.

Appendix B. Distribution of health states

The Peruvian Handbook of Recovery Times specifies an average recovery time for 2,763 unique disease codes at the fourth digit level of the 10th revision of the ICD. This paper focuses on non-mental health conditions; excluding these diagnoses reduces the number of unique diseases to 2,690.[33] Estimating the model with such level of granularity is unfeasible: it would require estimating a probability for each disease and group of observable characteristics (2,690 diagnoses $\times$ 4 age group $\times$ 2 occupation groups = 21,520). For this reason, I group diagnoses in more aggregated categories.

I these categories considering (i) the type of diseases they represent and (ii) how frequently these diagnoses are used in the claims data. Table A4 lists the conditions included in the analysis by their ICD 10th revision group and the share of sick leave claim data with these diagnoses. To compute these shares, I used the sample constructed for the quantitative analysis of this paper. Table A7 provides details on sample construction.

The first criteria for excluding a group of conditions are those not listed in the Peruvian handbook. These are conditions originating in the perinatal period (codes in groups P00-P96) and congenital malformations, deformations, and chromosomal abnormalities (codes in groups Q00-Q99). In fact, 0.15% of the sick leave claim data is reported under these diagnoses. I dropped such observations.

The second criteria for excluding a group of conditions is the nature of the diagnosis which makes very challenging to assign a benchmark recovery time. I exclude poisonings and burns (codes in group T00-T98)—these diagnoses accumulate 1.22% of the sick leave claim data. Additionally, I exclude diseases coded under "special purposes codes" (codes U00-U85), external causes of morbidity (codes V00-Y99), and factor influencing health status and contact with health services (codes Z00-Z99). All these together represent 3.16% of the sick leave claims. These conditions associated with longer recovery times or impairments where full recovery might not be foreseeable, for example, leg amputations and organs transplants. Finally, I exclude conditions with diagnoses C00-D49; these codes are used for neoplasms, which in most cases, are chronic conditions or diseases that would require a longer recovery time. In terms of claims data, these represent 1.78% of the claims. The final sample includes about 86% of all sick leave claims filed by private-sector male workers.

[33] I use codes F01-F99 to define mental health conditions, these are grouped under the chapter "Mental, Behavioral and Neurodevelopmental disorders".