DEPARTMENT OF ECONOMICS
JOHANNES KEPLER UNIVERSITY OF LINZ

PRESENTEEISM WHEN EMPLOYERS ARE UNDER PRESSURE: EVIDENCE FROM A HIGH-STAKES ENVIRONMENT

by

Mario LACKNER
Hendrik SONNABEND

Working Paper No. 2120

December 2021

Johannes Kepler University of Linz
Department of Economics
Altenberger Strasse 69
A-4040 Linz - Auhof, Austria
www.econ.jku.at

Corresponding author: mario.lackner@jku.at

Presenteeism when employers are under pressure: evidence from a high-stakes environment *

Mario Lackner[a,c] and Hendrik Sonnabend[b]

[a]*Department of Economics, Johannes Kepler University Linz (JKU)*
[b]*Department of Economics and Business Administration, University of Hagen*
[c]*Christian Doppler Laboratory "Aging, Health and the Labor Market"*

December 22, 2021

Abstract

This study analyses whether the decision to work while sick can be linked to workload fluctuations. Drawing on data collected from professional soccer, we exploit the dynamics of a season and use additional (national and international) cup games conducted in the second half of a season as a source of exogenous variation. We find robust evidence that players are 6.1 percentage points more likely to return from injuries earlier than expected when their teams are exposed to a high workload. The effect is driven by players who are more important to their teams and those who are less vulnerable to injuries. Finally, we find that presenteeism comes at the cost of an early comeback significantly shortening the time until the next injury by approximately 27 days.

JEL Code: I19, J22, Z2

Keywords: sickness absence; presenteeism; workload variations; soccer

**Contact*: Mario Lackner: Mario.Lackner@jku.at ; Hendrik Sonnabend: Hendrik.Sonnabend@fernuni-hagen.de.
We are grateful for the valuable comments provided by Christian Deutscher, Peter Dawson, Bernd Frick, Rob Simmons, attendees of the ROSES online seminar as well as attendees at the 12^{th} ESEA conference in Bielefeld (online). In addition, we thank Erik Scheil, Houda Ben Said, and Julian Herwig for their help with the data collection.

1. Introduction

Why do employees work even when their health status gives them a reason to stay at home? This phenomenon, often referred to as ‘presenteeism’, has received growing attention in different research areas such as occupational medicine, social psychology, and various fields of economics. There is evidence that, for instance, working conditions and job security, workers’ attitudes, age, and health status, and companies’ absence policies are important drivers of presenteeism (Hirsch et al. 2017, Miraglia & Johns 2016, Arnold 2016, Lohaus & Habermann 2019).

From an employer’s perspective, apart from the context of infectious diseases—‘contagious presenteeism’ has received much attention during the COVID-19 crisis (e.g., Pichler et al. 2020)—, presenteeism acts as a double edged sword. Reduced productivity is better than zero productivity resulting from sickness absence; however, it bears the risk of future health consequences and longer absence times (e.g., Schultz & Edington 2007, Bergström et al. 2009, Hansen & Andersen 2009, Skagen & Collins 2016).

The net utility derived from presenteeism is not necessarily constant over time. This is because workers’ absence may have little consequences in ‘quiet times’ but this could change when the employer is under pressure. One may think of financial auditors in the deadline phase. Consequently, decision-makers might feel an incentive to overvalue the effects of individual absences relative to the potential negative consequences and enforce presenteeism in times of high workloads.

This interplay between presenteeism and employer workload dynamics has not been clearly examined in the existing literature. In this study, we try to fill this gap using data from professional soccer. Specifically, we examine how quasi-random shocks on the number of matches (the workload) affect the recovery time of players (the employees). The idea is that additional games put pressure on teams to field players who were previously unavailable due to well described medical conditions. Depending on factors such as a player’s importance to his team (e.g., his relative productivity), his vulnerability to injuries, and the overall level of absenteeism, players may have an incentive to return from their absence before the scheduled time if their team needs them. In the field of sports, there have been prominent examples of a massive abuse of pain killers and top players participating in decisive matches while being affected by the flu. The ‘playing hurt’ culture in professional sports has received attention in sport science, sport medicine, and sport sociology (e.g. Roderick et al. 2000, Mayer & Thiel 2018, Chen et al. 2019). There is reason to expect that the underlying motives for this behaviour are the same as that in less specific labour market segments. Players can to be loyal to their club and their peers, or they might want to signal their resilience and reliability to the job market. On the employer side, it may be beneficial for teams to take the risk of a secondary injury when the current incentives are sufficiently high.

In general, while firm level data on presenteeism is hardly available and most empirical studies rely on self-reported survey data, our setting allows us to examine the nexus between workload and work-while-sick behaviour using field data in a high-stakes environment. Data on injuries in professional sports have been used before in studies observing issues in the fields of labour economics (Gregory-Smith 2021) and management science (Chan & Fearing 2019). Most closely related to our study, Ngo & Roberts (2021) analyse the relationship between contract status and missed games in the National Basketball Association (NBA). The authors show that for the average player, the likelihood to miss a game decreases toward the end of his contract, whereas the opposite is true for the best players (i.e., players who have been selected for the NBA All-Star Game).

Our empirical strategy is to exploit the dynamics of a standard season in European professional soccer. In the first step, we calculate the average recovery times from the first part of a season where the number of games is fixed and the workload is (to a large extend) predictable. We further analyse how recovery times vary with respect to a team's number of games in the second part of the season where teams are under pressure and the workload is far less predictable. This extra workload is virtually predetermined, as it originates from additional national and/or international cup games, which in turn, result from a team's success in the first part of the season. Our estimates suggest that in the presence of a high workload, players return from their injuries significantly earlier. Specifically, players in the treatment group—depending on the definition and degree of presenteeism—have a higher probability for a reduced injury time compared to the non-treated players by 5 to 7 percentage points (ppts). However, this effect is not equally strong for all types of players. In line with expectations, our analyses indicate that early returns are associated with a high (relative) productivity and a low vulnerability to injuries. Specifics of professional sports such as age and tenure do not seem to drive these results. We also find that presenteeism is associated with future costs: we estimate that the period until the next injury is shortened by 27 days for players exposed to a reduced recovery time. This means that the shorter the healing time, the sooner the next absence due to injury.

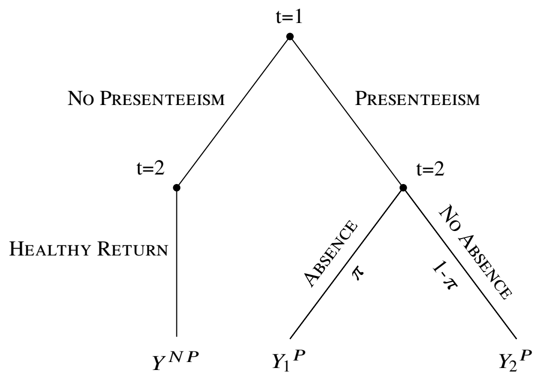
This study structured as follows. In the next section, we outline a simple conceptual framework to motivate our empirical analysis. Section 3 presents the data set and descriptive statistics of the absenteeism of players due to medical conditions. Sections 4 and 5 investigate the nexus between workload and presenteeism within our baseline models and in terms of various kinds of heterogeneity. Section 6 identifies the costs of presenteeism. Robustness checks are provided in Section 7. Section 8 concludes the study.

2. Conceptual framework

To illustrate the key trade-off we want to highlight in this study, we introduce a firm employing workers with different levels of productivity $\alpha_1 \gg \alpha_2 \gg \ldots \alpha_i \ldots > \alpha_{n-1} > \alpha_n$. The difference in abilities decreases in i, implying that the loss in productivity is smaller when worker $n-1$ is absent and worker n is doing her job as compared to a situation where the most productive worker has to be replaced.[1]

Now consider the case where worker i is not in good health (but does not have an infectious disease) in period $t = 1$. Presenteeism means that the worker could continue working in period $t = 1$ at the expense of a lower productivity $\beta\alpha_i$ with $0 < \beta < 1$, and the risk that the disease can develop into a more serious condition that would cause absence in the next period $t = 2$ with probability π_i. Otherwise, without presenteeism, the worker is replaced by a worker with lower productivity $i + 1$ in $t = 1$, and returns in good health in $t = 2$. Figure 1 illustrates the decision problem.

FIGURE 1 — Decision on presenteeism



To account for the fact that the importance of a high productivity may vary between periods, payoffs in period 2 are discounted by a discount factor $0 < \delta < 1$. For instance, the company could be in troubled times in $t = 1$ due to a workload peak or may have to deal with an already high level of overall absentees, whereas a stabilisation is expected in $t = 2$. Payoffs are $Y^{NP} = \alpha_{i+1} + \delta\alpha_i$ and $Y^P = \beta\alpha_i + \delta\left[\pi_i\alpha_{i+1} + (1-\pi_i)\,\alpha_i\right]$. Then, for $\beta > \frac{\alpha_{i+1}}{\alpha_i}$, presenteeism yields higher payoffs than a sick leave *iff*

$$\delta \leq \frac{\beta\alpha_i - \alpha_{i+1}}{\pi_i\left(\alpha_i - \alpha_{i+1}\right)}. \tag{1}$$

[1]We do not explicitly model the way replacements work. For instance, when a worker $i+1$ is doing the job of the absent worker i, this could mean that the workload for $i+1$ increases or that the worker is missing elsewhere. However, the crucial element for our conceptual framework is that there is heterogeneity in 'substitutability'. See Pauly et al. (2002) for a more comprehensive model of absent team members and Skåtun & Skåtun (2004) for an analysis of workload shifting among employees.

Hence, presenteeism is more likely to occur in the presence of (i) high values of β, (ii) low values of δ and π_i, and (iii) the top productivity segment where workers can be less easily replaced because of the greater productivity gap. Since β and π_i depend on both individual predisposition and the disease type, these variables act as determinants of presenteeism. Finally, δ must be sufficiently small to cause presenteeism, all other things being equal. In other words, the firm places a much higher value on productivity in $t = 1$ compared to $t = 2$. Examples of periods of excessive workload and easing thereafter include high occupancy rates in hospitals, all kinds of business deadlines, and the decisive phase of a season in sports.

Note that the mechanisms highlighted in this simple exercise can be easily transferred to a situation where worker i has been absent due to medical conditions in period $t = 0$ and is not fully recovered in $t = 1$. A repeated absence in $t = 2$ then could mean exhaustion or that the disease recurred as the healing process was incomplete. This is exactly the scenario we analyse in our empirical setting.

3. Data and descriptive statistics

We use data from the top soccer leagues in Germany (*Bundesliga*), Spain (*Primera Division*), and Italy (*Serie A*), covering ten seasons from 2010 to 2019. They were collected from *transfermarkt* (`https://www.transfermarkt.de/`), a popular German-based football website.

The focus is on player absences due to medical conditions, such as a pulled hamstring, a traumatic brain injury, and gastroenteritis, among others. For the sake of simplicity, this study uses the term 'injury' to refer to a medical condition that causes a player to be absent from the game.

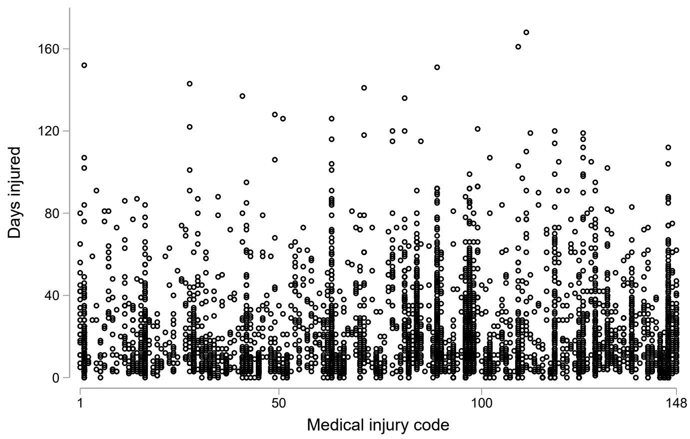
Initially, the raw data encompassed a total of $17,831$ absences, categorised by medical diagnosis and linked with detailed information on the start and end dates as well as player and team characteristics.[2] Since our analysis uses variations in the number of matches conducted in the cup games in the *second* half of a season (January to June), we calculate the average number of days it takes for a player to recover from an injury for each category using the data on the injuries that ended within the *first* part of a season only (July to December). This is important because all absences that end within the first part of a season are not based on the decisions affected by the high workload treatment used in the empirical analysis.

As explained in the next section, we take these expected recovery times as reference points since they are exogenous in the sense that they are not affected by the workload variations under consideration. For the same reason, the data related to the players traded in the second half of a season (i.e., in January when

[2]The recovery time ends when the player is available again and part of the team. Note that this does not necessarily imply that players are fielded in the next match.

the transfer window closes) were excluded. Finally, we are left with 6,479 observations for the period of January to June assigned to 148 different types of injuries with an average duration of 17.67 days of absence and a standard deviation of 18.62.[3] Figure 2 plots the recovery times in days per category.[4]

FIGURE 2 — Observed recovery time in days by medical injury code



Notes: This figure illustrates the absolute injury duration (in days) for all injury categories in our final estimation sample. All injuries occured and ended before July 1. N = 6,479.

We initially group the diagnoses into 21 categories according to the median recovery duration.[5] Note that we do not use the mean, because extreme outliers may bias the results. Figure 3 illustrates the median recovery duration per injury type in ascending order, grouped into 21 categories (1: lowest injury severity group, 21: highest injury severity group). It shows that the dispersion of recovery times is low for the majority of categories (1 to 20), but high for the category of the highest ranks (with medians between about 60 and 180). Since this group appears to be extremely heterogeneous and therefore may bias our results, we excluded it in the analysis conducted below.

4. Empirical strategy and main results

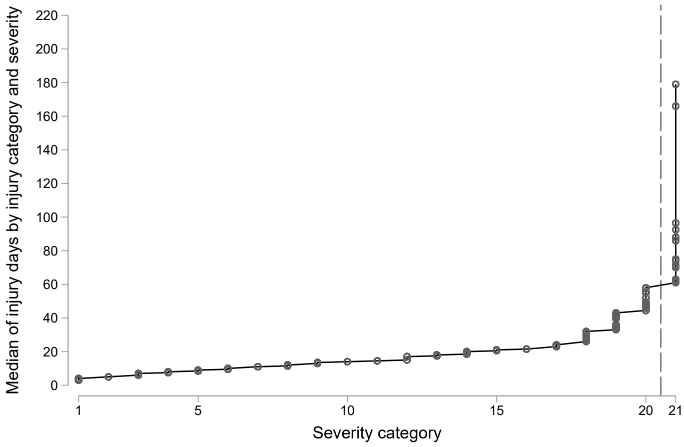
Our empirical strategy is to exploit the dynamics of a standard season in European professional soccer. In soccer, the first half of a season (July to December, 'the reference period'), starting with a preparation period, is characterised by a fixed number of games either in the national league, the first rounds of the national cup(s), and, if qualified, the group stage of European competitions (i.e., *UEFA Europa League* and *UEFA Champions League*). The workload is predictable and teams can adjust their personnel

[3]Please consult the Appendix for a detailed description of the data preparation.

[4]Table A.1 in the Appendix specifies the medical diagnoses.

[5]The number of categories may appear arbitrary, yet it ensures a balanced number of observations across injury types. Nevertheless, as clarified in the next section, we conduct robustness checks with different numbers of categories. Overall, the main results remain unchanged.

FIGURE 3 — Median recovery duration in days per injury type in ascending order and grouped into 20 categories



Notes: The hollow dots indicate the median recovery duration in days per injury type in ascending order, N = 6,479.

decisions. However, in the second half of a season (January to June, 'the relevant period'), the number of matches played varies according to a team's success in the earlier phase of the national and international cup competitions. This is the critical phase of a season where teams are under pressure and the workload is far less predictable.

The basic idea of the research design is to calculate the average recovery times per injury category based on the data obtained for the reference period; we then analyse how recovery times vary with respect to the number of games played by the team in the relevant period. In our setting, presenteeism means that a player returns from the injury of a certain type earlier than the predicted usual recovery duration of the specific injury. An intuitive way to model presenteeism behaviour of player j is to create a binary variable equal to 1 if the recovery time (in days) is at least 1 day shorter than expected for an injury of type i, and 0 otherwise. We call this variable $early\ return_{i,j,k,t}$.

In order to compare teams with high and low workloads during the relevant observation period, we introduce a binary variable defined as

$$high\ load_{k,t} = \begin{cases} 1, & \text{if } load_{k,t} \in Q_4 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

to indicate that the number of national and European cup games played by team k in the relevant period of season t is in the fourth quartile of the continuous variable $load_{k,t}$ (mean 2.68 and st. dev. 3.57). This variable measures the number of matches played after the initial group stages of the UEFA Champions League and the Euro League in the first half of a season. It also includes the number of matches played in the national cup competition after the first three rounds. The fourth quartile is equal to a workload of more than 4 games. The maximum value of $load_{k,t}$ is 13 and the average number of games played by a team exposed to a high workload is 8.

While this definition of a high workload treatment is based on the overall distribution of games in the second half of the seasons, it also has strong theoretical foundations: since the fourth quartile of $load_{k,t}$ equals to five or more games, it follows that for $high\ load_{k,t} = 1$, team k still participates in one or more cup competitions. This means that our measure of workload is, to a large extent, pre-determined with respect to the start and end of all injury times in our outcome sample.[6] In essence, it indicates whether a team experiences *any* extra workload in addition to the mandatory league games during the second half of a season (which applies to 25% of all the team-season combinations in our data). Our estimates should therefore be viewed as a lower bound of the association between a high workload and presenteeism behaviour.

As we demonstrate later in this section, our main findings are robust to alternative cutoff values (3 and 5 matches). Moreover, we check the robustness of our results with a sample restricted to absences starting in the May to July period. This is because we cannot rule out the fact that our estimates may suffer from endogeneity bias in the sense that earlier returns from injuries increase a team's workload due to an increase in performance. It shows that our main results hold and that the effects are even stronger. We explain this by the fact that injuries occuring during the 'crunch time' of a season are highly detrimental to team performance. Thus, the incentive for presenteeism behaviour when the workload is high is greater than that observed earlier in the season.

Table 1 reports the summary statistics of injury characteristics for the high and low workload groups. As expected, the mean duration of injury is almost 4 days shorter for high workload teams compared to low workload teams. In other respects, we do not find relevant differences: the average severity category is basically identical, and the difference in the average starting month (0.12) is statistically significant but economically insignificant.

Table 2 presents our key variables on the team-season level. As per definition, teams with a high workload play additional national and international cup games in the relevant period from January to June. In addition, these teams perform significantly better in the autumn part of the season compared to low workload teams. Finally, there is a difference in the overall number of injuries (2.47, significant at the 10% level), probably due to the higher number of games played. Although the difference is small,

[6]Qualification for the knock-out stage of a European cup competition results from performances in the reference period and is therefore independent of the injuries that occurred during the relevant period.

this could pose a threat to our identification strategy in case the higher number of injuries explains a potential positive association between high workload and presenteeism behaviour. We will address this issue in the heterogeneity analysis presented in Section 5.

TABLE 1 — Summary statistics by treatment status: injury level

	low	**high**	*difference*
days	18.61	14.67	3.94***
	(19.33)	(15.79)	
start month	2.83	2.95	−0.12**
	(1.32)	(1.31)	
end month	3.41	3.43	−0.02
	(1.40)	(1.39)	
severity[a]	9.01	9.71	−0.16
	(5.71)	(5.86)	
N	4,940	1,539	

Notes: All injuries in the second halves of seasons are included (January - July), $N = 6,479$. [a]: *severity* refers to the 20 injury categories. Standard deviations are presented in parentheses. t-tests for the difference in means are presented in the third column, *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

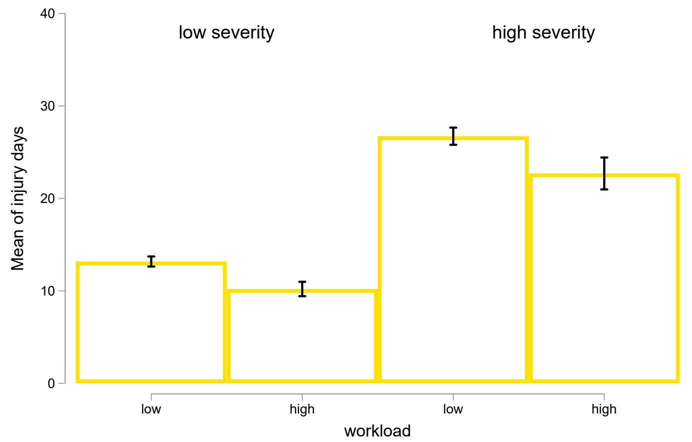
TABLE 2 — Summary statistics by treatment status: team-season level

	low	**high**	*difference*
number of extra games	0.81	8.66	−7.85***
	(1.22)	(2.77)	
rank ratio	0.58	0.28	0.30***
	(0.27)	(0.24)	
injury count	12.47	14.94	−2.47*
	(8.09)	(9.76)	
N	396	103	

Notes: Descriptive statistics of the main variables at the team-season level, $N = 499$. t-tests for the difference in means are presented in the third column, *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The variable $rankratio$ illustrates team performance by dividing the team's current ranking by the worst possible according to the maximum number of teams in the league.

Figure 4 provides an illustration of the association between workload and recovery times. We split the 20 injury severity categories into two main categories: 'low severity' (categories 1 to 10) and 'high severity' (categories 11 to 20). For both workload categories, the average number of days an injured player is absent is lower when the club is exposed to high workload. That is, players of clubs with a high workload are more likely to return earlier than expected.

FIGURE 4 — Average recovery times for high and low workload



Notes: N = 6,479; a high workload is defined by equation (2). Low severity: injuries of categories 1-10, high severity: injuries of categories 11-20.

Next, we document the average treatment effect of having a high workload on the probability of an early return from an injury i of category c by estimating the following equation:

$$early\ return_{i,j,k,t,c} = \beta_0 + \beta_1 high\ workload_{k,t} + \phi' \mathbf{X}_{i,j,k,t,c} + \xi_k + \pi_t + \rho_c + \epsilon_{i,j,k,t,c}\ , \qquad (3)$$

where $\mathbf{X}_{i,j,k,t,c}$ is a vector of injury-, team-, and player specific characteristics including the win ratio and the relative league position of team k after the reference period, the team's total market value, and a dummy indicating the market value quartile of player j.[7] Furthermore, we control for the team-season specific number of injuries in the reference period and the average injury length (in days). To account for the fact that an early return of player j might also be affected by his 'substitutability' within the team, we control for team k's overall squad size as well as the number of injuries occurred (*NoIP*) and the available players (*APP*) assigned to j's position during the time he recovers from injury i.[8] We also add a binary variable that indicates whether or not a player holds the citizenship of the country where the league is based. Player j's tenure with team k is used as a proxy for loyalty.[9][10] In addition, we include the number of matches played by team k in the relevant period below the cutoff value of the $work\ load$

[7]The relative league position is defined as the ranking position divided by the number of teams. The first quartile of the market value distribution serves as the reference category.

[8]The four major positions used in the analysis are goalkeeper, defender, midfielder, and forward.

[9]Galizzi & Boden (2003), for instance, show that the absence of workers with job-related injuries is affected by tenure. Arnold (2016) reports that the number of sickness presenteeism days increases with tenure, whereas individuals show less presenteeism behaviour after a job change within the first year.

[10]Note that our data set does not include information on player contracts. Acknowledging that Ngo & Roberts (2021) find that the contract status has an impact on absenteeism in the NBA (excluding top players), we argue that it should not have an impact on the treatment effect under consideration. Moreover, the fact that tenure does not seem to play a role in the LATE (see Section 5) may support this view.

dummy variable to account for the workload in the reference period. Finally, a binary variable indicates whether or not team k participates in a European competition in season t. Table 3 presents the summary statistics of the main variables of interest.

In addition, we include a set of dummies for the starting month and the category c of the injury, ρ_c. Team fixed effects (ξ_k) account for the unobserved heterogeneity between teams, whereas season fixed effects (π_t) control for the general developments in soccer like injury trends and improvements in medical care. The inclusion of team fixed effects means that we will estimate the high-workload effect based on the variation within teams. Thus, we ensure that our results are not based on unobserved differences between teams, such as financial capabilities.

TABLE 3 — Summary statistics

Variable	**Mean**	**Std. Dev.**	**Min**	**Max**
injury length (days)	17.67	18.62	0	168
tenure (years)	2.93	2.68	0	24
NoIP (inj. players for j's position)	0.63	0.84	0	6
APP (available players for j's position)	8.98	3.18	0	31
European competition (1=yes, 0=no)	0.42	-	0	1
avg. injury length (reference period)	25.86	14.09	3.00	165.50
number of injuries (team-season level, reference period)	17.11	11.25	0	61
team market value (EUR mill.)	196.11	198.37	24.02	1,135.90
player market value (EUR mill.)	8.34	13.58	0.03	180
win percentage (reference period)	0.40	0.19	0	0.95
squad size	35.99	6.73	25	82
nationality of league (1=yes, 0=no)	0.55	-	0	1
workload in reference period	4.97	3.23	1	11

Notes: Descriptive statistics of main variables, $N = 6,479$.

Table 4 presents the results for different specifications of the linear probability model defined in (3). Control variables for the player and the team are added in columns (4) to (6). We estimate that the probability of an early return increases by 11.2 ppts in the most simple and 6.1 ppts in the full model specification if a team's workload in the relevant period is high. Since there exits a considerable number of players with only one observation in our sample, we will refrain from estimating the player fixed effects model and refer to model (6) as our preferred specification.[11]

Table 4 also shows that an early return is more likely to occur when there is a shortage of players to fill in j's position, meaning that it is less easy to replace the absent player. One additional player decreases the probability by 1.1 ppts, whereas another injured player substituting that position is associated with an increase in the probability by 3.5 ppts.

Considering our simple model presented in Section 3, the empirical results confirm the theoretical predictions: presenteeism is more likely to occur when employers place more weight on the present than

[11]In Section 7, we present the results of a model with player fixed effects. Our main results still hold.

future payoffs and when the employees can be less easily replaced. We will elaborate on this implication and further predictions derived from the theoretical model presented in Sections 5 and 6.

TABLE 4 — Effect of treatment on probability to return from injury earlier than the reference mean

	(1)	(2)	(3)	(4)	(5)	(6)
high workload[a]	0.112***	0.109***	0.059***	0.051**	0.056**	0.061**
(1 = yes, 0 = no)	(0.017)	(0.018)	(0.021)	(0.020)	(0.023)	(0.024)
nationality[b]				−0.002	−0.002	−0.004
(1 = yes, 0 = no)				(0.013)	(0.013)	(0.013)
tenure				0.001	0.001	0.001
				(0.003)	(0.003)	(0.003)
player market value				0.004***	0.004***	0.003***
				(0.001)	(0.001)	(0.001)
team value					0.000	0.000
					(0.000)	(0.000)
win percentage					−0.013	0.026
reference period					(0.109)	(0.116)
rank ratio					−0.012	0.003
					(0.065)	(0.068)
roster size					0.000	0.002
					(0.001)	(0.002)
competing in UEFA					0.030	0.025
(1 = yes, 0 = no)					(0.042)	(0.042)
number of workload					−0.007	−0.007
games in reference					(0.007)	(0.007)
number of injuries						0.035***
at players position						(0.009)
available players						−0.011***
in relevant position						(0.004)
avg. injury duration						−0.001
in reference period						(0.000)
number of injuries						0.001
in reference period						(0.001)
add. controls[c]	*no*	*yes*	*yes*	*yes*	*yes*	*yes*
start month FEs	*no*	*no*	*yes*	*yes*	*yes*	*yes*
team FEs	*no*	*no*	*yes*	*yes*	*yes*	*yes*
R^2	0.007	0.033	0.083	0.097	0.356	0.356

Notes: $N = 6,479$. Robust standard errors, clustered on the team-season level, are presented in parentheses, stars indicate significance: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The dependent variable is equal to 1 if the injury is reported to be over before the mean duration in the reference sample, and 0 otherwise. Mean: 0.498, standard deviation: 0.450. [a] Binary variable indicating a high workload for a team-season as defined in equation 2. [b] Binary variable equal to 1 if the injured player holds citizenship of the country where the league is based, 0 otherwise. [c] Additional controls include age, position, injury category as well as season dummies. In addition, we estimate team and injury start-month fixed effects.

Although our $early\ return$ variable presents an intuitive way to measure presenteeism, a concern might be that it is too imprecise. Specifically, it does not account for the widely dispersed distribution of deviations from the median injury duration. However, it makes a difference if a player recovers one day earlier than expected from a first-degree strain or an Achilles tendon rupture. We therefore propose a measure of *relative injury duration* for each injury category c defined by

$$\Delta_{i,c} = \frac{duration_{i,c}}{median\ duration_c} \times 100\,, \tag{4}$$

where $duration_{i,c}$ is the duration (in days) of an injury i of category c suffered in the relevant period (January to June) of season t, and $median\ duration_c$ is the expected length of injuries for that category calculated from the reference period. We prefer using the median over the mean of injury duration, as outliers will certainly affect the precision of our presenteeism measure.

To account for potential non-linear effects, we use interval dummies for the relative injury duration:

$$diff_s = \begin{cases} 1, & \text{if } \Delta_{i,c}, \leq \frac{s}{100} \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

for each $s \in \{5, 10, 15, ..., 90, 95, 100\}$. Thus, for instance, $diff_5 = 1$ implies that the recovery time of injury i was at least 5% shorter than the mean in the reference category c.

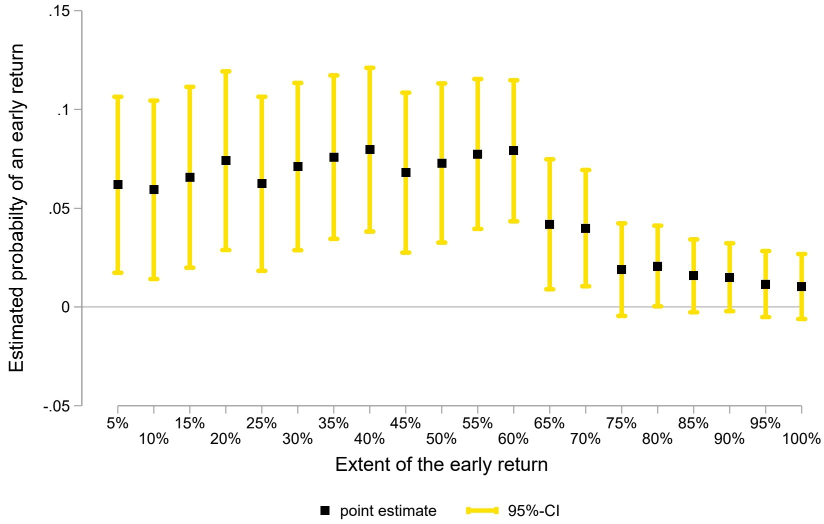
Next, using this vector of dependent variables, we re-estimate model (3) for each interval which yields 20 estimates for β_1. The results are illustrated in Figure 5. We find that a high workload in the relevant period significantly increases (at the 5% level) the probability for an early return for all thresholds between 10% and 80%. The point estimate is approximately 6 ppts for intervals below the 65% threshold, which decreases gradually thereafter. For very high values of Δ which indicates a reduction in the recovery time by 80% or more, $\hat{\beta}_1$ is close to 0 and not statistically significant.

We conclude that the players belonging to teams 'under pressure', who were injured in the second half of a season, have a higher tendency to return earlier than expected. As this relation also holds true for very early returns for which a full recovery seems unlikely, we present our results as evidence for presenteeism.

In the preceding analysis, we used a cutoff value of >4 games (which is the fourth quartile of the overall distribution of extra games) to define a high workload. To test the robustness of this definition, we use alternative cutoffs of >3 and >5. The results are presented in Figure A.2 in the Appendix. Overall, the main effects (as presented in Figure 5) are confirmed qualitatively and quantitatively. We conclude that the exact cutoff at >4 games is not driving our results.

Another issue, mentioned earlier in this section, is that the cutoff criterion for defining a high workload is not fully predetermined, because it includes the cup matches held between February and early April. Consequently, some matches result from prior success in the first two rounds of the playoff stage. To ensure that our main findings are not affected by an endogeneity bias, we check the robustness of our findings with a sample restricted to the injuries that occurred in May or later. Figure 6 suggests that for this sample, the presenteeism effects are even stronger. This is not surprising, because the incentives for

FIGURE 5 — High workload affects the probability of an earlier return from a sick absence due to injury



Notes: Each square represents the point estimate for the binary variable *high workload* and different dependent variables related to the extent of the early return. Full model specification as defined by equation (3) and similar to column (6) in Table 4; N = 6,479.

an early return are highest in the 'crunch time' of a season when it is of foremost importance to succeed in national and international competition. Referring to the conceptual framework, we expect the discount factor δ to be the highest for these periods. Moreover, the sample includes (by definition) shorter and, on average, less severe injuries from which the players recovered before July.

5. Factors that affect the relationship between workload and presenteeism

Our main results presented in the foregoing section indicate that a high workload causes presenteeism behaviour in professional soccer. However, it is reasonable to assume that the nexus between workload and presenteeism is affected by employee (player) and employer (club) characteristics. For instance, our theoretical model predicts that it is more likely that presenteeism occur among the top players and players less vulnerable to injuries. In the following analysis, we focus on these characteristics.

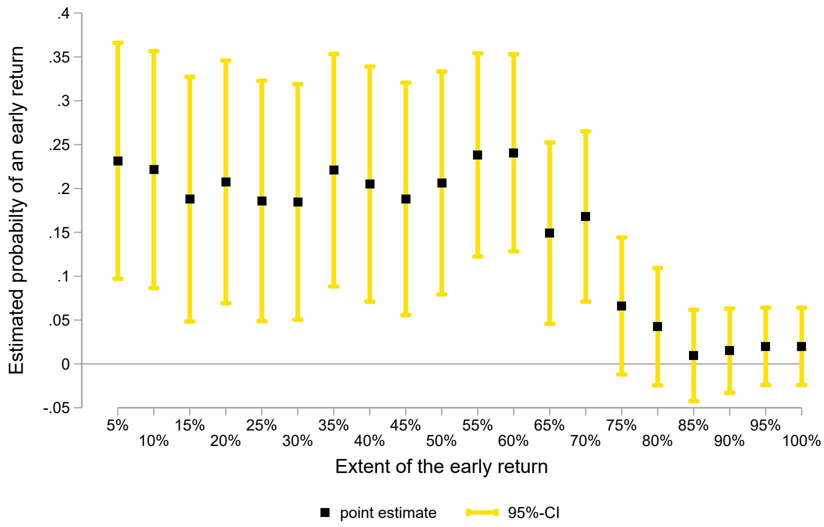
5.1. Individual and team (absolute) ability

We use information on a player's market value in season t as a proxy variable for individual ability.[12] These market values represent the expert estimates of possible transfer fees based on past performances and the market situation. As a general rule of thumb, better players have higher market values.[13] For all

[12]The data on market value were collected from `https://www.transfermarkt.de`.

[13]Note that the *transfermarkt* market values have been used in prior studies such as Krumer & Lechner (2018). Since market values are also affected by the age of a player, which in turn, may not play a role for his actual ability, we conduct the analysis again with age adjusted values. That is, we use the residuals $\mu_{j,k,t}$ derived from estimating the model $market\,value_{j,k,t}$ =

FIGURE 6 — High workload affects the probability of an earlier return from a sick absence due to injury: reduced sample



Notes: Each square represents the point estimate for the binary variable *high workload* and different dependent variables related to the extent of an early return. Full model specification as defined by equation (3) and similar to column (6) in Table 4. Only the injuries that occurred in May or later are included; $N = 753$.

the observations in our sample, the average market value of the observed player at the time of the injury is 8.34 million Euros with a standard deviation of 13.58.

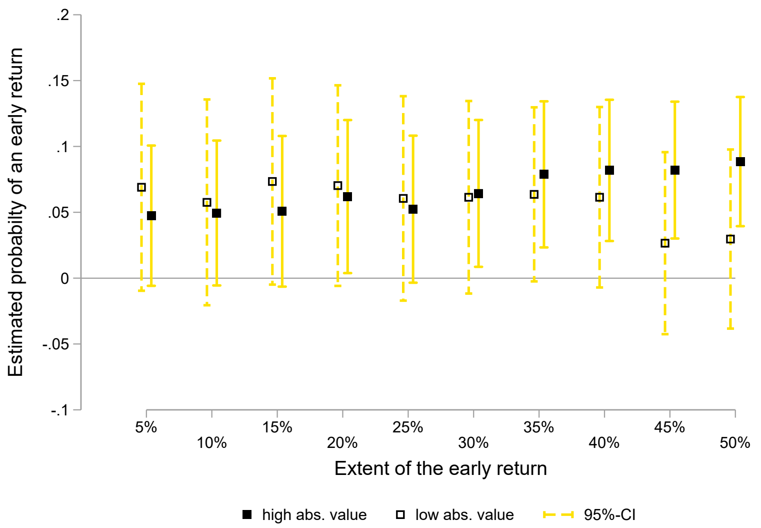
There are two countervailing forces that could shape how the importance or value of a player for a team affects presenteeism behaviour. From a team's perspective, the most able players are also the most important and productive ones who will be missed the most in case of their absence. This speaks in favour of an early return (see Section 2). On the contrary, players with high market values are assets to their teams and require protection. That is, in a worst-case scenario, an inadequate recovery time may result in medium- or long-term negative effects due to chronic ailments. These adverse health effects are a threat not only to a player's availability and performance, but also to possible future transfer revenues.

In the empirical analysis, we use a median split to categorise (absolute) player values as either high or low. Furthermore, following analogy of equation (5), 10 interval dummies are used as dependent variables in our regression model ($s \in \{5, ..., 50\}$), which correspond to a 5% to 50% shorter recovery time compared to the reference group. The results are illustrated in Figure 7. Although we can confirm the significant positive impact of workload on the probability to return earlier from injury, the point estimates for the high and low market value group show a negligible difference. We therefore conclude that the absolute market value is not a driving factor of presenteeism in our setting.

$\alpha_0 + \alpha_1' AGE + \alpha_2' \mathbf{X}_{j,k,t} + \mu_{j,k,t}$ where AGE is a vector of age interval dummies, and $\mathbf{X}_{j,k,t}$ is a vector of player-team characteristics including the position, league, and season. The results remain unchanged and are available upon request.

In addition to the individual productivity of a player, the overall ability of a team may also have an impact on presenteeism. For instance, it might be the case that a high-budget team with a large number of top players has a better ability to replace an injured player. Therefore, a median split is used to categorise the teams into low and high categories of (absolute) market values.[14] Our findings are illustrated in Figure A.3 in the Appendix. It shows that the estimated coefficients are very similar across groups, suggesting that presenteeism does not vary by team market values.

FIGURE 7 — Effect heterogeneity: **absolute** market value of a player



Notes: Each square represents the point estimate for the binary variable *high workload* and different dependent variables related to the extent of the early return. Full model specification as defined by equation (3) and similar to column (6) in Table 4; N = 6,479. Solid (hollow) squares represent the estimates for the group of players with a high (low) absolute market value. Players were categorised using a median split.

5.2. *Relative market values*

The foregoing analyses suggest that the absolute market values—neither for the individual players nor for their teams—significantly impact the workload effect on presenteeism behaviour. Absolute market values, however, might not tell the whole story. The theoretical considerations in Section 2 suggest that it is rather the relative importance of an absent employee relative to her colleagues that causes the employer to demand a return to work while the employee is still recovering.

In our setting, we define the relative ability of player j belonging to team k in season t as

$$relative\ ability_{j,k,t} = \frac{market\ value_{j,k,t}}{team\ market\ value_{k,t}}. \tag{6}$$

Further, the players are divided at the median of $relative\ ability_{j,k,t}$. We then run the same regressions as in the previous section for the players who are above or below the median of relative ability. The

[14]We prefer the median over the mean, because only a few outliers at the top end of the distribution can affect mean market values disproportionately.

results are presented in Figure 8. We find a significant and sizeable workload effect on the probability of an early return for players in the high-ability group but not for those in the low-ability group. For instance, players with a high relative ability are, on average, about 10 to 14 ppts more likely to reduce the expected recovery time by half when their team is under pressure. On the contrary, there is no early return for players from the low-ability group, even when their teams are exposed to a high workload, suggesting that these players can be replaced more easily.

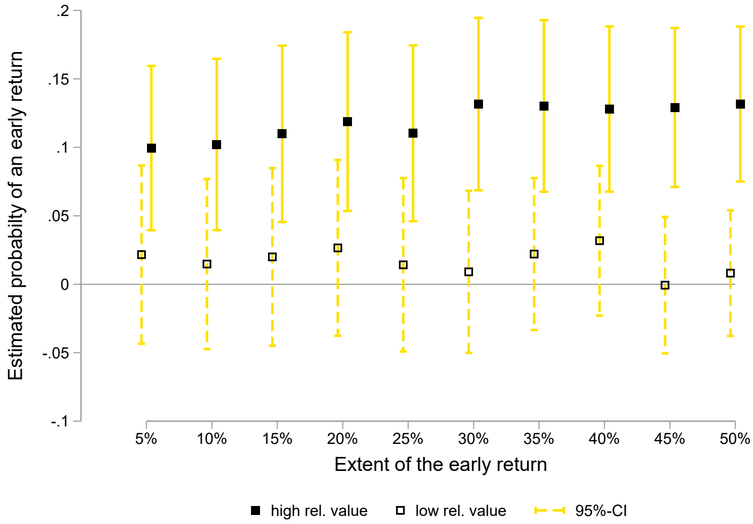
As a further refinement, we follow the idea that professional soccer teams, similar to most conventional enterprises, have employees in key positions. In the case of soccer, these key players are starters and represent the first line of employees. To investigate whether the players with or without such a prominent role exhibit a similar relationship between presenteeism and workload, we define an indicator for key players. In particular, we sort players according to their market values within their teams. Then, we define 'important players' as those players whose ranking position is equal to or below the median ranking position. This median ranking position is 8 for the overall sample. Figure A.4 in the Appendix confirms the prior results: we estimate a significant positive effect of a high workload on the probability to return early from an injury for important players. In contrast, for the group of less important players, the estimated β_1 is not different from zero. The point estimates range from 0.04 to 0.

Similar to other labour markets, the division of labour and specialisation is also present in soccer. Although the degree of specialisation can vary across positions and players (see Kempa 2021, for a recent study on this subject), we account for this issue by introducing team-specific rankings based on market values for the four main player positions. A player is then defined as a *key player* if he is the top goalkeeper or forward player, or if he is among the top four defenders or top five midfielders. Apart from the great importance for their teams, the internal competition and the need to protect their status as regular players may work as additional incentives for these players. Figure 9 indicates that even when we use *key player* as a moderator in our model, the estimates are quite similar to the foregoing specifications.

5.3. Injuries: personal history and the overall injury level

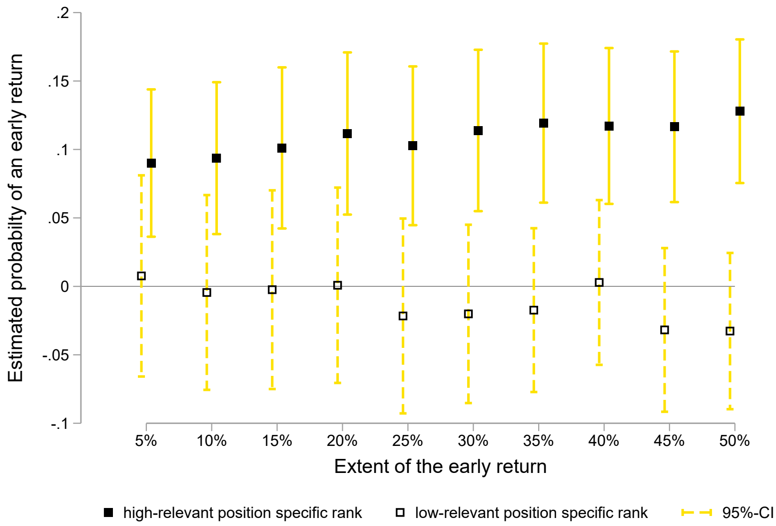
Following the conceptual framework, in this section we test the hypothesis that presenteeism is more likely to occur when the risk of absence in the next period is sufficiently low. We operationalise this idea by using a player's injury history to proxy his 'vulnerability'. Therefore, the sample is split along the median of the player-specific 'number of injuries per number of seasons' ratio. For example, this ratio would be $2/5$ for an individual who had two injuries in five seasons. We then re-estimate model (3) with interval dummies defined by (5).

FIGURE 8 — Effect heterogeneity: relative market value



Notes: Each square represents the point estimate for the binary variable *high workload* and different dependent variables related to the extent of the early return. Full model specification as defined by equation (3) and similar to column (6) in Table 4; N = 6,479. Solid (hollow) squares represent estimates for the group of players with a high (low) relative market value. Players are divided at the median of $relative\ ability_{j,k,t}$ defined by equation (6), which is 0.31.

FIGURE 9 — Effect heterogeneity: key players



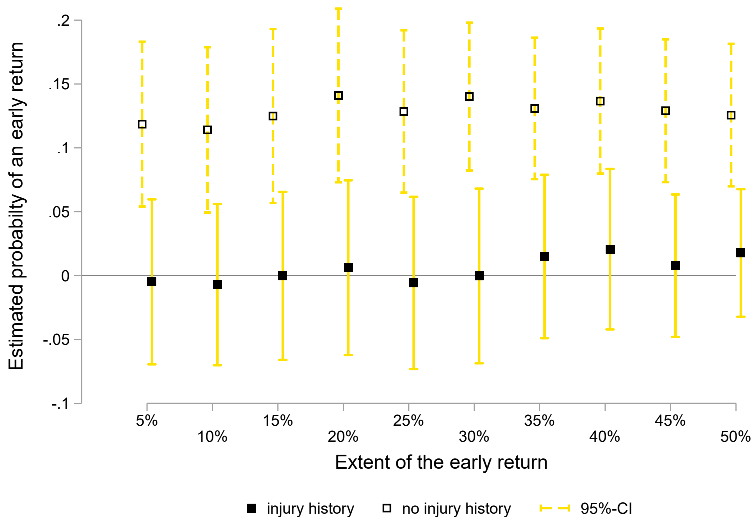
Notes: Each square represents the point estimate for the binary variable *high workload* and different dependent variables related to the extent of the early return. Full model specification as defined by equation (3) and similar to column (6) in Table 4; N = $6,297$. Solid squares represent estimates for the group of *key players*, that is, the top players per position identified by their market values. Hollow squares represent estimates for the group of non-key players.

Figure 10 presents the estimates for β_1. Consistent with our expectations from theory, it shows that the treatment effect of a high workload is restricted to the players who are less vulnerable to injuries. Therefore, players with several prior injury issues do not return to their teams before the median time of healing, even when the team is under pressure. We conclude that for vulnerable players, the benefits of an early return do not outweigh the risks involved.

Figure A.5 in the Appendix complements our prior findings regarding players injured at the same time. While our main results in Table 4 provide evidence that the number of additional injuries at the focused player's position increases the likelihood of an *early return*, the figure demonstrates that the high workload effect is equivalent for teams with a low and high overall number of injuries.

Moreover, studies such as Godøy & Dale-Olsen (2018) demonstrate positive peer effects in absenteeism. In our setting, this would suggest that there might be spillover effects from the teammates who returned early from their absence due to injuries, and this can exert peer pressure on the focused player to do likewise. However, we do not observe such behaviour in our data. That is, (unreported) results indicate that the number of previous early returns in a team per season does not affect subsequent presenteeism behaviour. This not only casts doubts on the role of peer effects in our setting, but also suggests that differences among coaches in handling injured players are not driving our results.

FIGURE 10 — Effect heterogeneity: vulnerability

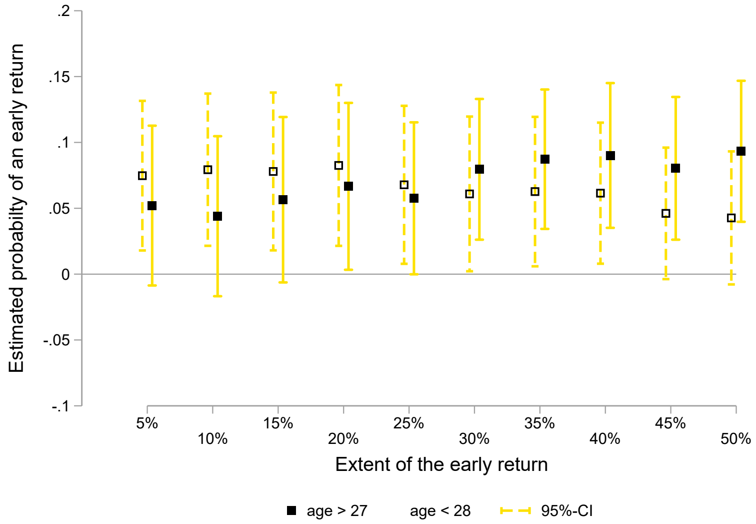


Notes: Each square represents the point estimate for the binary variable *high workload* and different dependent variables related to the extent of the early return. Full model specification as defined by equation (3) and similar to column (6) in Table 4; N = 6,479. Solid (hollow) squares represent estimates for the group of players with a high (low) number of past injuries. Players are divided at the median of the injuries per season ratio, which is 1.

5.4. *Age and tenure*

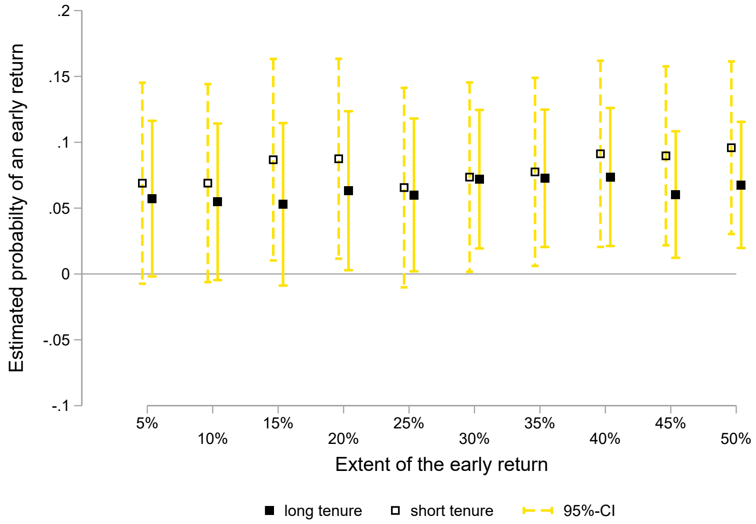
Finally, we present complementary results from the estimates of model (3) with age and tenure as moderators. First, splitting the sample into young and old players (median age: 27) suggests that age alone is not the predictor of an *early return* when the workload is high (Figure 11). As we have seen in the previous section, this might be because it does not capture vulnerability. Second, Figure 12 indicates that the treatment effect is virtually the same for players with long and short tenures. This suggests that we should not overemphasize the importance of loyalty and 'contract compliance' in the present context.

FIGURE 11 — Effect heterogeneity: age



Notes: Each square represents the point estimate for the binary variable *high workload* and different dependent variables related to the extent of the early return. Full model specification as defined by equation (3) and similar to column (6) in Table 4; N = 6,479. Solid (hollow) squares represent estimates for the group of old (young) players. Players were categorised using a median split (median age = 27).

FIGURE 12 — Effect heterogeneity: tenure



Notes: Each square represents the point estimate for the binary variable *high workload* and different dependent variables related to the extent of the early return. Full model specification as defined by equation (3) and similar to column (6) in Table 4; N = 6,479. Solid (hollow) squares represent estimates for the group of long-tenured (short-tenured) players. Players were categorised as *short* with 0 or 1 seasons with the observed team, and categorized as *long* with more than 1year with the team.

6. Cost and consequences of an earlier return

Our main results suggest that there is presenteeism behaviour in professional soccer caused by additional workload. Consistent with theoretical considerations, we find that the effect is driven by those players who are the most important for their teams. In addition to a high workload, we find the number of injuries among co-players at the same position to be another driver of an early return.

Although these findings clearly suggest strong incentives for presenteeism when the teams are under pressure, studies in occupational medicine and sport science document its negative side. Obviously, it is beyond our data to measure the pain associated with an incomplete healing and a possible long-term chronic disability. However, we can estimate the risk of subsequent injuries including re-injuries and new injuries. If presenteeism significantly shortens the time until the next absence, these costs are borne not only by the players themselves (e.g., in the form of reduced earnings and career opportunities), but also by the team that needs to compensate for the absent players.

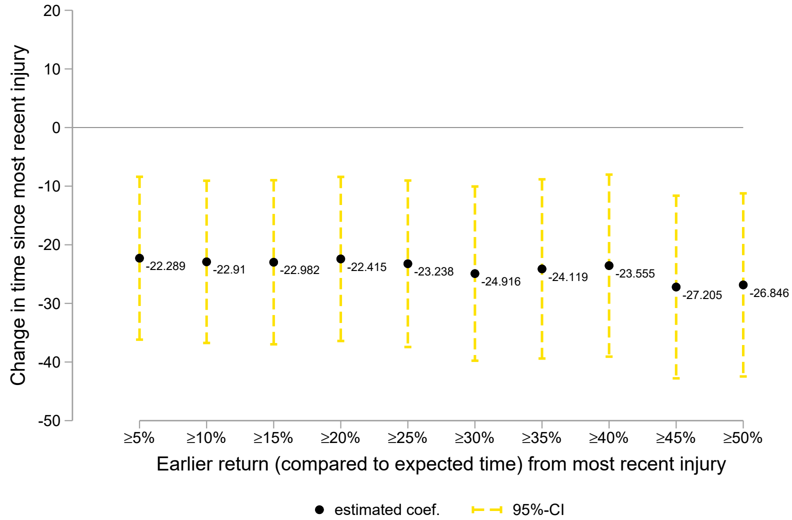
Our empirical approach is to use the full sample (like in Section 4) and link injuries in the relevant period to subsequent injuries. Since our data do not allow us to correctly identify the censoring for players with only one observation, we exclude these players. Hence, the results will indicate the consequences of an early return conditional on re-injuring. The average player in the final sample ($N = 5,032$) suffers 5.07 subsequent injuries in the relevant period (see Section 4). The mean time between events is 188.28 days (standard deviation of 260.12 and a maximum of 3,135) and the mean duration of these injuries is 28.00 days (standard deviation of 45.33).

We estimate a variant of model (3) where the dependent variable is the time spent in good health after a prior injury, that is, the time between the actual and the subsequent injury. Figure 13 illustrates the results. We find that the time spent in good health for players who returned at least 5% earlier than expected is reduced by approximately 27 days. The effect is stable across intervals, suggesting that the negative health consequences are relevant for all grades of presenteeism.

As a sensitivity analysis, Table 5 presents estimates from a model including three binary variables equal to 1 if the recovery time was 5% to 25%, 30% to 50%, or more than 50% shorter than expected, and zero otherwise (reference category: no early return). Results from ordinary least squares (OLS) regressions show that an early return reduces the time until the next absence: by approximately 37 days for a low degree of presenteeism, 23 days for a medium degree of presenteeism, and 21 days for a high degree of presenteeism (column (3), full model specification).

As we use a duration measure as the dependent variable, we also specify a Cox proportional hazard model to check the robustness of our estimates derived from OLS. The results are presented in column (4). We find that an early return increases the risk of subsequent injury. Coding long subsequent injury times (i.e., longer than 365 days) as censored does not change our result qualitatively.

FIGURE 13 — Earlier returns and time before next injury



Notes: Each point represents an estimate from a separate regression, using the number of days between two injuries as the dependent variable. The variable of interest is equal to 1 if the preceding injury time ended at least 5%,10%,15%,...,50% earlier than expected, and 0 if no early return was recorded. The number of observations decreases along the x-axis for each estimate. The full sample of 5,032 injuries is used for the first estimate of a 5% early return. Standard errors are clustered at the player level.

TABLE 5 — Effect of an earlier return from previous injury on time to follow-up injury.

	OLS			*COX model*
	(1)	(2)	(3)	(4)
return 1	−17.817	−38.487***	−37.383***	0.156***
(1 if return ∈ [5%–25%] , 0 otherwise)	(10.868)	(10.873)	(10.977)	(0.046)
return 2	−17.755*	−25.081**	−22.733**	0.080*
(1 if return ∈ [30%–50%] , 0 otherwise)	(9.518)	(10.223)	(10.243)	(0.042)
return 3	−27.832***	−24.849***	−21.261**	0.146***
(1 if return > 50%, 0 otherwise)	(8.006)	(8.719)	(8.709)	(0.040)
add. binary controls	*yes*	*yes*	*yes*	*yes*
team last injury dummies	*yes*	*yes*	*yes*	*yes*
player dummies	*no*	*yes*	*yes*	*no*
end-month previous injury	*no*	*no*	*yes*	*no*
R^2	0.217	0.492	0.496	
N		5,032		5,032

Notes: Robust standard errors clustered at the player level are presented in parentheses; asteriks indicate significance: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The dependent variable in columns (1) to (3) observes to the number of days between the observed injury and the end of the preceding injury. Additional binary controls include season and age dummies as well as dummies indicating the severity category of the preceding injury. Column (4) reports the coefficients derived from a Cox proportional hazard survival model, stratified by the team the player was signed with when the preceding injury ended.

7. Robustness

In our analysis, we group all injuries into 20 categories. This categorization may appear arbitrary, yet it ensures a balanced number of observations across injury types. To check robustness of our main results, we employ an alternative approach, constructing 10, 15, 26, and 30 injury categories. The corresponding results are presented in Figure A.1 in the Appendix. Overall, the main results remain unchanged.

The two most prestigious competitions for national teams, the FIFA World Cup and the UEFA European Championship, are important for players and teams. These tournaments not only shorten the recovery time for players but also are also regarded as career milestones. Miklós-Thal & Ullrich (2016), for instance, show that players adapt their league performances in the run-up of such an event. Hence, it might be the case that the decision for injured players to return to play could be affected. Although we use season fixed effects to account for heterogeneity across years, we want to additionally test whether our main results presented in Figure 5 are driven by international tournaments. For this reason, Figure A.6 in the Appendix illustrates the estimates based on a sample restricted to season without international tournaments. It shows that our main findings can be confirmed, indicating that international tournaments are not the driving force behind our estimates.

We also run an alternative specification to our empirical model defined by equation (3) including player fixed effects. The results are presented in Figure A.7 in the Appendix. Due to the high number of players with only one injury in our data, the effects are estimated with less precision. Our main results, however, are confirmed.

Finally, our definition of $high\ workload$ pertains to national and European cup matches. Since national cups differ in terms of competition formats across leagues, Figure A.8 in the Appendix presents estimates from an alternative definition using European competition games only. The effects are estimated less precisely and are smaller in size. However, as our main results are confirmed, we conclude that the specifics of the national cup competitions are not driving our findings.

8. Conclusion and discussion

This study contributes to the small but growing literature of the analysis of work-while-sick behaviour. Using data collected from professional soccer, we investigate the effect of an increased team workload on the propensity for players to return from absence due to injury or sickness earlier than expected. We propose a novel measure of presenteeism which allows to predict the expected time of absence and to precisely measure the extent of presenteeism. Our estimates suggest that the probability to return earlier than expected from an injury increases by 6.1 ppts when the team is exposed to a high workload.

Furthermore, the setting allows us to study the potential moderators of the nexus between workload and presenteeism behaviour. As predicted by our simple model, we find that the effect is driven by players with a higher relative productivity. These players are more important to their teams and therefore harder to replace. In line with our expectations, we find that the players who are more vulnerable to injuries seem to be protected against an early comeback to competition. The external validity of our findings is supported by the fact that sport specific factors are not the drivers of our results: neither the age nor the tenure of players—two characteristics in professional sports that typically differ from standard work environments—appear to affect presenteeism behaviour.

These observations may suggest that teams balance the pros and cons of an early return in a way that presenteeism occurs when the ratio of benefits to costs is the greatest. Nonetheless, we document that this behaviour has serious consequences. Specifically, it shows that the time in good health between two injury events is reduced by about one month for players who had an early comeback due to a high team workload. We find this effect to be large given that the mean time between two absences, on average, is about six months. First, this suggests that teams may have a general idea of how to deal with the return of injured players in troubled times, but they underestimate the actual risks. Second, it raises the question on why workers in general and soccer players in particular agree to presenteeism when there are substantial health consequences. Possible explanations are related to career concerns (Crichton et al. 2011, Markussen 2012, Ngo & Roberts 2021), reciprocal behaviour (e.g., Charness & Kuhn 2011), direct pressure by the employer, and the fear of losing income due to the low levels of sick payments or even unemployment (e.g., Olsson 2009, Hirsch et al. 2017, Callison & Pesko 2020). However, there might be other and more indirect reasons for a (tacit) agreement on presenteeism. These are work attitudes and social norms (e.g., Godøy & Dale-Olsen 2018), a (perceived) lack of replacement (e.g., Lohaus & Habermann 2019), and peer pressure in the sense that workers want to prevent a workload increase for their colleagues (Skåtun & Skåtun 2004, Barmby et al. 2016).

References

Arnold, D. (2016), 'Determinants of the annual duration of sickness presenteeism: Empirical evidence from European data', *Labour* **30**(2), 198–212.

Barmby, T., Sessions, J. G. & Zangelidis, A. (2016), 'Looking after number two? Competition, cooperation and workplace interaction', *Journal of Economic Behavior & Organization* **131**, 166–182.

Bergström, G., Bodin, L., Hagberg, J., Aronsson, G. & Josephson, M. (2009), 'Sickness presenteeism today, sickness absenteeism tomorrow? A prospective study on sickness presenteeism and future sickness absenteeism', *Journal of Occupational and Environmental Medicine* **51**(6), 629–638.

Callison, K. & Pesko, M. F. (2020), 'The effect of paid sick leave mandates on coverage, work absences, and presenteeism', *Journal of Human Resources* . Online first, `https://jhr.uwpress.org/content/early/2020/05/04/jhr.57.4.1017-9124R2.abstract`.

Chan, T. C. & Fearing, D. (2019), 'Process flexibility in baseball: The value of positional flexibility', *Management Science* **65**(4), 1642–1666.

Charness, G. & Kuhn, P. (2011), Lab labor: What can labor economists learn from the lab?, *in* O. Ashenfelter & D. Card, eds, 'Handbook of Labor Economics', Vol. 4, Elsevier, chapter 3, pp. 229–330.

Chen, Y., Buggy, C. & Kelly, S. (2019), 'Winning at all costs: a review of risk-taking behaviour and sporting injury from an occupational safety and health perspective', *Sports Medicine-Open* **5**(1), 1–21.

Crichton, S., Stillman, S. & Hyslop, D. (2011), 'Returning to work from injury: Longitudinal evidence on employment and earnings', *Industrial and Labor Relations Review* **64**(4), 765–785.

Galizzi, M. & Boden, L. I. (2003), 'The return to work of injured workers: Evidence from matched unemployment insurance and workers' compensation data', *Labour Economics* **10**(3), 311–337.

Godøy, A. & Dale-Olsen, H. (2018), 'Spillovers from gatekeeping–peer effects in absenteeism', *Journal of Public Economics* **167**, 190–204.

Gregory-Smith, I. (2021), 'Wages and labor productivity: Evidence from injuries in the national football league', *Economic Inquiry* **59**(2), 829–847.

Hansen, C. D. & Andersen, J. H. (2009), 'Sick at work—a risk factor for long-term sickness absence at a later date?', *Journal of Epidemiology & Community Health* **63**(5), 397–402.

Hirsch, B., Lechmann, D. S. & Schnabel, C. (2017), 'Coming to work while sick: An economic theory of presenteeism with an application to German data', *Oxford Economic Papers* **69**(4), 1010–1031.

Kempa, K. (2021), 'Task-specific human capital and returns to specialization: evidence from association football', *Oxford Economic Papers* . Online first, `https://doi.org/10.1093/oep/gpab006`.

Krumer, A. & Lechner, M. (2018), 'Midweek effect on soccer performance: Evidence from the German Bundesliga', *Economic Inquiry* **56**(1), 193–207.

Lohaus, D. & Habermann, W. (2019), 'Presenteeism: A review and research directions', *Human Resource Management Review* **29**(1), 43–58.

Markussen, S. (2012), 'The individual cost of sick leave', *Journal of Population Economics* **25**(4), 1287–1306.

Mayer, J. & Thiel, A. (2018), 'Presenteeism in the elite sports workplace: The willingness to compete hurt among German elite handball and track and field athletes', *International Review for the Sociology of Sport* **53**(1), 49–68.

Miklós-Thal, J. & Ullrich, H. (2016), ‘Career prospects and effort incentives: Evidence from professional soccer’, *Management Science* **62**(6), 1645–1667.

Miraglia, M. & Johns, G. (2016), ‘Going to work ill: A meta-analysis of the correlates of presenteeism and a dual-path model.’, *Journal of Occupational Health Psychology* **21**(3), 261.

Ngo, P. T. H. & Roberts, S. (2021), ‘There is no ‘I’ in team: Career concerns, risk-taking incentives, and team outcomes’, *Journal of Economics & Management Strategy* **30**(1), 122–138.

Olsson, M. (2009), ‘Employment protection and sickness absence’, *Labour Economics* **16**(2), 208–214.

Pauly, M. V., Nicholson, S., Xu, J., Polsky, D., Danzon, P. M., Murray, J. F. & Berger, M. L. (2002), ‘A general model of the impact of absenteeism on employers and employees’, *Health Economics* **11**(3), 221–231.

Pichler, S., Wen, K. & Ziebarth, N. R. (2020), ‘COVID-19 emergency sick leave has helped flatten the curve in the United States: Study examines the impact of emergency sick leave on the spread of COVID-19.’, *Health Affairs* **39**(12), 2197–2204.

Roderick, M., Waddington, I. & Parker, G. (2000), ‘Playing hurt: managing injuries in english professional football’, *International Review for the Sociology of Sport* **35**(2), 165–180.

Schultz, A. B. & Edington, D. W. (2007), ‘Employee health and presenteeism: A systematic review’, *Journal of Occupational Rehabilitation* **17**(3), 547–579.

Skagen, K. & Collins, A. M. (2016), ‘The consequences of sickness presenteeism on health and wellbeing over time: A systematic review’, *Social Science & Medicine* **161**, 169–177.

Skåtun, D. R. & Skåtun, J. D. (2004), ‘The impact of turnout on turning up: The complementarity of attendance among co-workers’, *Journal of Economics* **83**(3), 225–242.

A. APPENDIX: Data description, additional tables and figures

Data description

We use all injuries entries available at `https://www.transfermarkt.de/` for the season 2010/11 to 2019/20. Due to the fact that the second half of 2019/20 season was strongly influenced by the COVID-19 pandemic, we discarded observations from the year 2020. This also applies to injuries which start before the pandemic in January 2020, because we expect the players' comeback to be affected by the worldwide shutdown of sports events.

Our data cover the major soccer leagues Germany (*Bundesliga*), Spain (*Primera Division*), and Italy (*Serie A*). We do not use data from the UK and France as the website lists implausibly low injury numbers per year for these leagues.

Furthermore, we also excluded injury codes which are associated with a maximum duration of over 365 days. These very severe injuries are unlikely to involve a potential decision for an early return within the relevant time span of 6 months from January through June. We also dropped injury codes where we only observe less than 10 cases overall to avoid making too severe predictions errors for the expected injury duration. Finally, all injuries categorised as 'unknown injury' as well as 133 injury spells coded as 'Virus infection' were discarded.

After cleaning the data and checking for implausible intervals (like when the end date was before the start date), we end up a sample of $7,108$ observations in the reference periods (autumn season) and $6,479$ observations in the relevant period (spring season). All injury types observed in our final sample are tabulated in Table A.1.

TABLE A.1 — Medical injury codes

Code	Diagnosis	Code	Diagnosis
1	Achilles tendon irritation	75	Knee contusion
2	Abdominal discomfort	76	Knee problems
3	Abdominal muscle strain	77	Laceration wound
4	Adductor problems	78	Lateral ligament injury
5	Angina	79	Leg injury
6	Ankle injury	80	Ligament inflammation in the knee
7	Ankle problems	81	Ligament injury
8	Ankle sprain	82	Ligament stretch
9	Appendix surgery	83	Lumbago
10	Arm injury	84	Lumbar vertebrae problems
11	Back contusion	85	Meniscus injury
12	Back injury	86	Meniscus tear grade 2
13	Back problems	87	Metacarpal fracture
14	Blockage in the back	88	Metatarsal contusion
15	Blow	89	Minor blemish
16	Bronchitis	90	Muscle bundle tear
17	Bruise	91	Muscle contusion
18	Bruise on the ankle	92	Muscle fatigue
19	Calf hardening	93	Muscle hardening
20	Calf injury	94	Muscle injury
21	Calf muscle tear	95	Muscle strain
22	Calf strain	96	Muscle tear
23	Capsule injury	97	Muscular problems
24	Capsule tear	98	Nasal fracture
25	Collarbone fracture	99	Nasal injury
26	Common cold	100	Neck injury
27	Concussion	101	Nose surgery
28	Conservation	102	Outer ligament tear
29	Contracture	103	Outer tape tear
30	Contusion	104	Out-of-band problems
31	Contusion on the knee	105	Overstretching
32	Cruciate ligament strain (grade 1)	106	Patellar tendon irritation
33	Cruciate ligament strain (grades 2 and 3)	107	Pelvic injury
34	Dental surgery	108	Pubic inflammation
35	Distortion on the ankle	109	Pubic irritation
36	Elbow injury	110	Rib contusion
37	Eye injury	111	Rib fracture
38	Facial injury	112	Shin contusion
39	Fatigue fracture	113	Shin injury
40	Fever	114	Shoulder injury
41	Finger fracture	115	Sick
42	Finger injury	116	Sports bar
43	Flesh wound	117	Sprain
44	Flu	118	sprained ankle
45	Flu-like infection	119	Stomach problems
46	Foot contusion	120	Strain
47	Foot injury	121	Strain in the thigh and buttock muscles
48	Fracture	122	Stress reaction of the bone
49	Fracture of the arm	123	Surgery
50	Gastroenteritis/Stomach flu	124	Syndesmosis ligament tear
51	Groin injury	125	Tendinitis
52	Groin problems	126	Tendon irritation
53	Groin strain	127	Thigh injury
54	Hand fracture	128	Thigh muscle tear
55	Hand injury	129	Thigh problems
56	Head injury	130	Thigh strain
57	Heel injury	131	Toe fracture
58	Heel problems	132	Toe injury
59	Hip contusion	133	Tonsillitis
60	Hip problems	134	Torn ankle ligamet (grade 1)
61	Horse kiss	135	Torn ankle ligamet (grade 2)
62	Infection	136	Torn ankle ligamet (grade 3)
63	Inflammation	137	Torn knee ligament
64	Inflammation in the knee	138	Torn ligament
65	Inguinal hernia	139	Torn ligament ankle joint
66	Injury to the abdominal muscles	140	Torn muscle fibre (grade 1)
67	Injury to the ankle	141	Torn muscle fibre (grades 2 and 3
68	Injury to the leg flexor muscle	142	Torn muscle fibre in the adductor area
69	Inner ligament injury	143	Torn tendon
70	Inner ligament strain	144	Training deficit
71	Inner ligament stretching knee	145	Viral disease
72	Inner ligament tear (grade 1)	146	Wound
73	Inner ligament tear 3 (grades 2 and 3)	147	Wrist fracture
74	Inner ligament tear knee	148	Zygomatic fracture

FIGURE A.1 — Robustness check: different numbers of injury categories

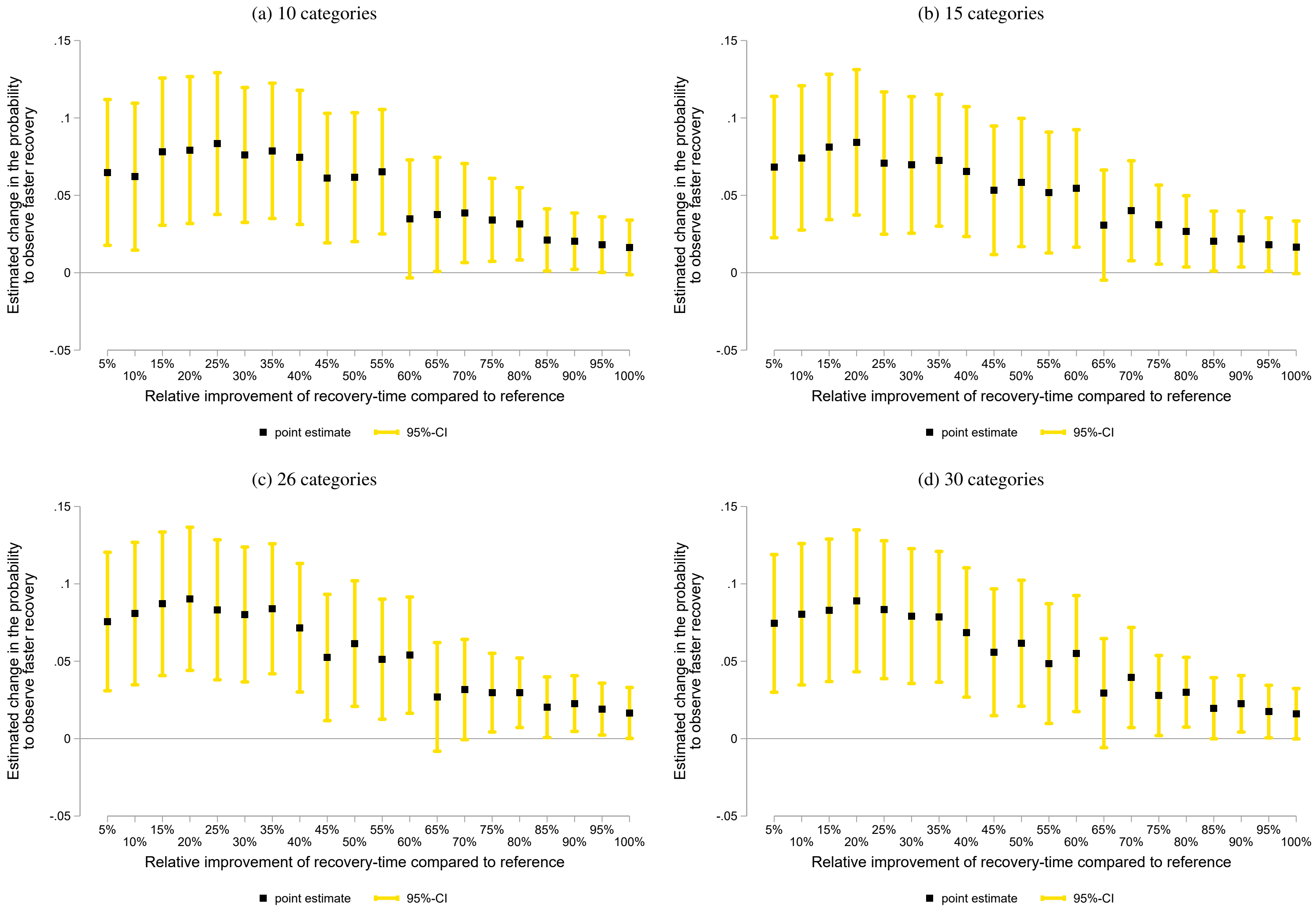
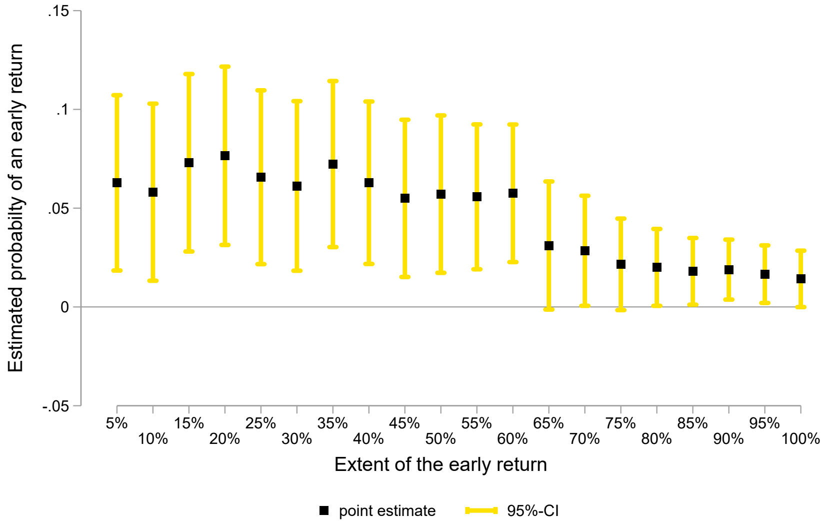
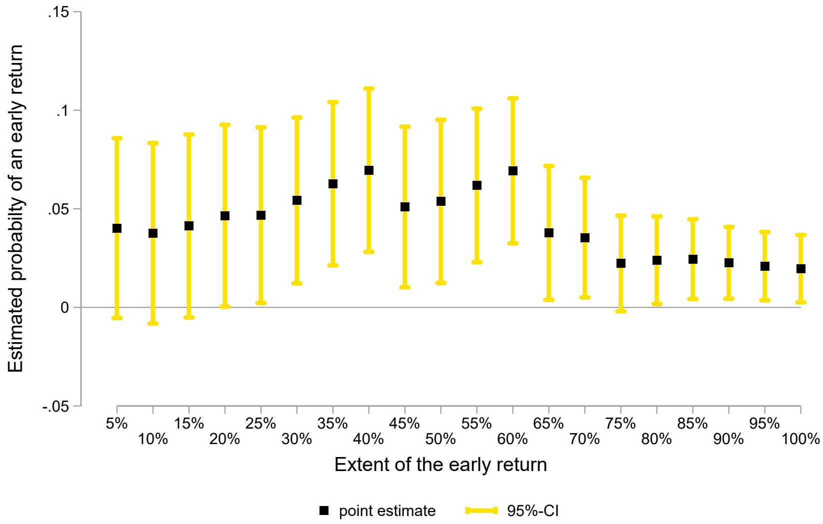


FIGURE A.2 — Robustness check: alternative cutoff values for *high workload*

Panel A. high workload > 3 games



Panel B. high workload > 5 games

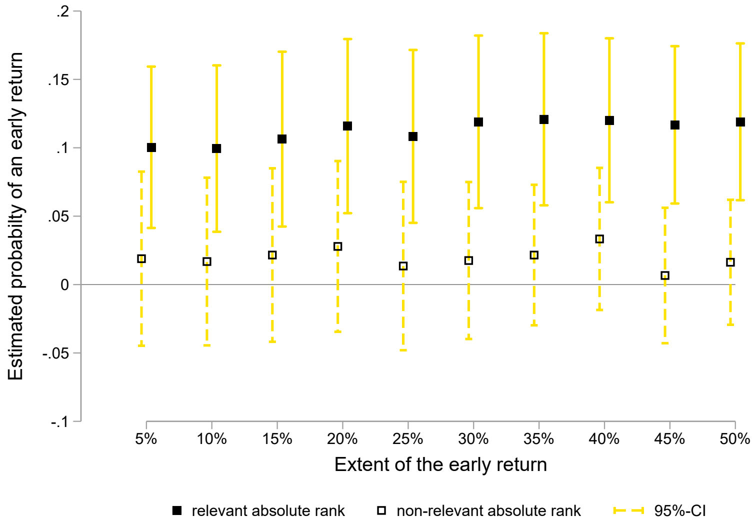


Notes: Each square represents the point estimate for the binary variable *high workload* and different dependent variables related to the extent of the early return. Full model specification as defined by equation (3) and similar to column (6) in Table 4, N = 6,479.

FIGURE A.3 — Effect heterogeneity: high-budget vs. low-budget teams

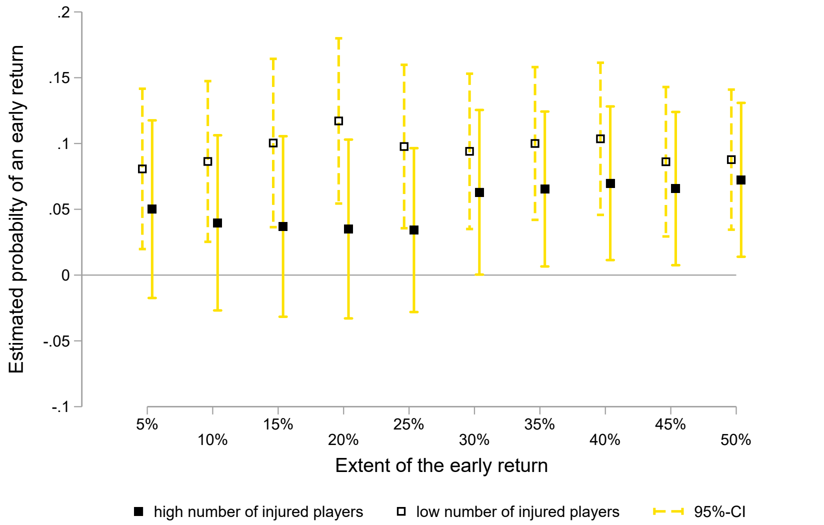
Notes: Each square represents the point estimate for the binary variable *high workload* and different dependent variables related to the extent of the early return. Full model specification as defined by equation (3) and similar to column (6) in Table 4, N = 4,479. Solid (hollow) squares represent estimates for the group of high-budget (low-budget) teams. Teams were categorised using a median split.

FIGURE A.4 — Effect heterogeneity: high-skilled vs. low-skilled players



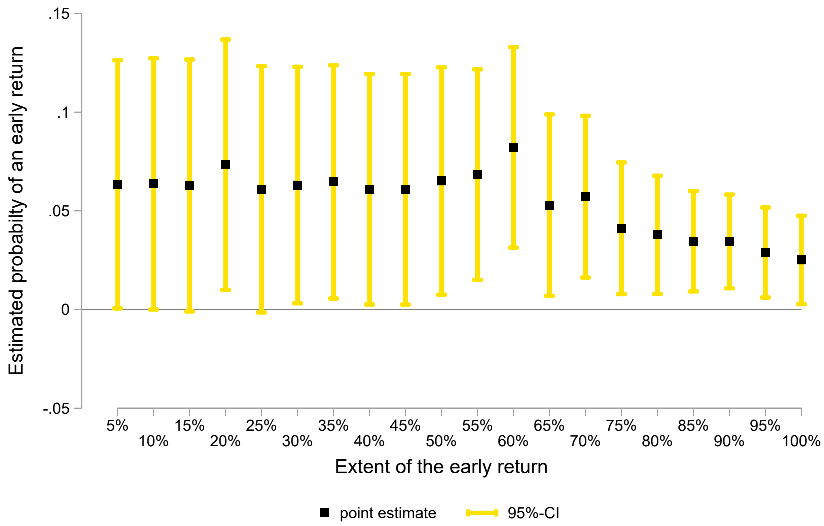
Notes: Each square represents the point estimate for the binary variable *high workload* and different dependent variables related to the extent of the early return. Full model specification as defined by equation (3) and similar to column (6) in Table 4, N = 4,479. Solid (hollow) squares represent estimates for the group of high-skilled (low-skilled) players. Players were categorised according to their market value ranking position within their teams (threshold: 8).

FIGURE A.5 — Effect heterogeneity: team-specific number of injuries



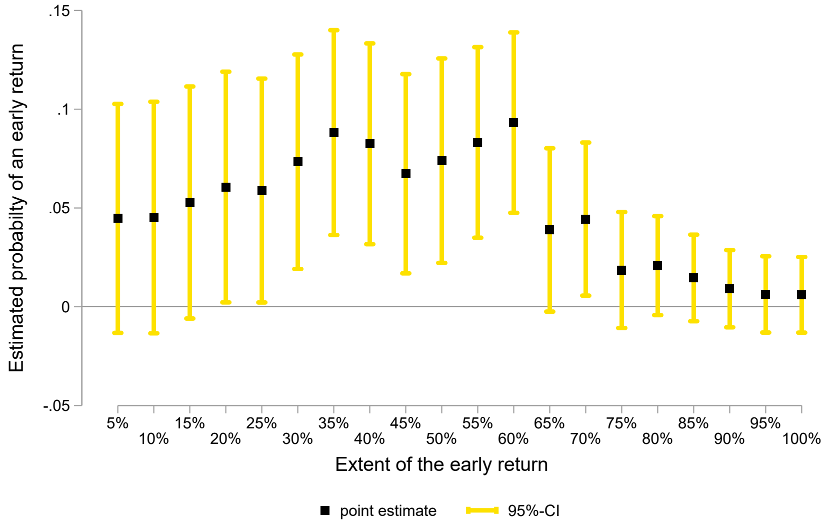
Notes: Each square represents the point estimate for the binary variable *high workload* and different dependent variables related to the extent of the early return. Full model specification as defined by equation (3) and similar to column (6) in Table 4, N = 4,479. Full squares indicate estimates for a high number or parallel injuries on a team, hollow squares for a low number. Solid (hollow) squares represent estimates for the group of teams with a high (low) absolute number of injured players at the time the focused player is absent. Teams were categorised using a median split.

FIGURE A.6 — High Workload affects the probability to return from injury earlier - no seasons with international tournaments



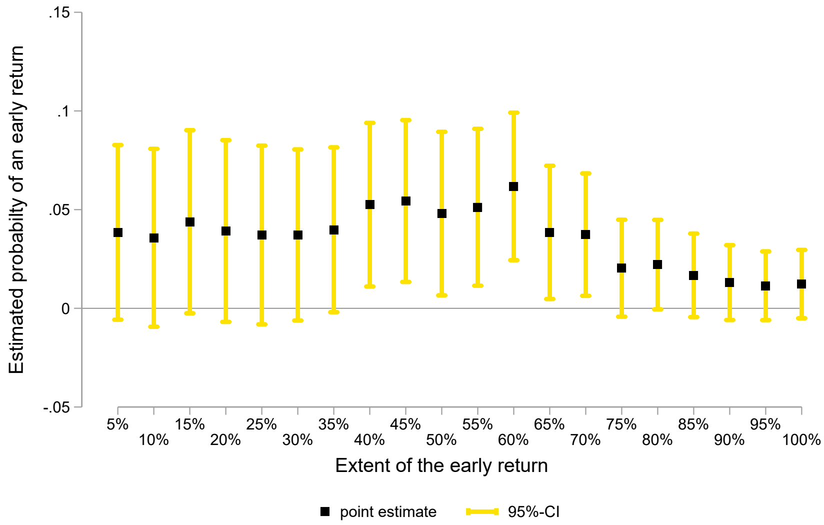
Notes: Each square represents the point estimate for the binary variable *high workload* and different dependent variables related to the extent of the early return. Full model specification as defined by equation (3) and similar to column (6) in Table 4, N = 3,632. Only injuries which started in May or later are included. The sample does not include seasons followed by the FIFA World Championship or the UEFA European Championship.

FIGURE A.7 — High Workload affects the probability to return from injury earlier - player fixed effects.



Notes: Each square represents the point estimate for the binary variable *high workload* and different dependent variables related to the extent of the early return. Full model specification as defined by equation (3) and similar to column (6) in Table 4, N = 5,370, sample is reduced due to dropping singleton observations. Only injuries which started in May or later are included.

FIGURE A.8 — High Workload affects the probability to return from injury earlier - no national cup games included.



Notes: Each square represents the point estimate for the binary variable *high workload* and different dependent variables related to the extent of the early return. Full model specification as defined by equation (3) and similar to column (6) in Table 4, N = 4,479. The variable $high\ workload$ was calculated without national cup competitions.