Safety and Health at Work 15 (2024) 151–157

Contents lists available at ScienceDirect

Safety and Health at Work

journal homepage: www.e-shaw.net

Original article

Comparison of the Association Between Presenteeism and Absenteeism among Replacement Workers and Paid Workers: Cross-sectional Studies and Machine Learning Techniques

Heejoo Park [1], Juho Sim [3], Juyeon Oh [1], Jongmin Lee [2], Chorom Lee [2], Yangwook Kim [1], Byungyoon Yun [3,**], Jin-ha Yoon [3,4,*]

[1] Department of Public Health, Graduate School, Yonsei University, Seoul, Republic of Korea
[2] Department of Occupational Health, Graduate School of Public Health, Yonsei University, Seoul, Republic of Korea
[3] Department of Preventive Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea
[4] The Institute for Occupational Health, Yonsei University College of Medicine, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 26 February 2023
Received in revised form
29 January 2024
Accepted 3 March 2024
Available online 5 March 2024

Keywords:
Absenteeism
Male
Presenteeism
Replacement driver
Replacement worker

ABSTRACT

Background: Replacement drivers represent a significant portion of platform labor in the Republic of Korea, often facing night shifts and the demands of emotional labor. Research on replacement drivers is limited due to their widespread nature. This study examined the levels of presenteeism and absenteeism among replacement drivers in comparison to those of paid male workers in the Republic of Korea.
Methods: This study collected data for replacement drivers and used data from the 6th Korean Working Conditions Survey for paid male workers over the age of 20 years. Propensity score matching was performed to balance the differences between paid workers and replacement drivers. Multivariable logistic regression was used to estimate the adjusted odds ratio (OR) and 95% confidence intervals for presenteeism and absenteeism by replacement drivers. Stratified analysis was conducted for age groups, educational levels, income levels, and working hours. The analysis was adjusted for variables including age, education, income, working hours, working days per week, and working duration.
Results: Among the 1,417 participants, the prevalence of presenteeism and absenteeism among replacement drivers was 53.6% (n = 210) and 51.3% (n = 201), respectively. The association of presenteeism and absenteeism (adjusted OR [95% CI] = 8.42 [6.36–11.16] and 20.80 [95% CI = 14.60–29.62], respectively) with replacement drivers being significant, with a prominent association among the young age group, high educational, and medium income levels.
Conclusion: The results demonstrated that replacement drivers were more significantly associated with presenteeism and absenteeism than paid workers. Further studies are necessary to establish a strategy to decrease the risk factors among replacement drivers.

1. Introduction

A platform worker, also known as a gig worker, connects consumers and workers through an online platform [1]. The characteristics of platform workers are typically low-income, temporary, and on-demand jobs [2,3]. The number of platform workers is increasing, and according to a report released in December 2020, the number of platform workers in the Republic of Korea was estimated to be approximately 1.79 million, accounting for 7.4% of the total workforce [3]. Platform workers are becoming more

Heejoo Park: https://orcid.org/0009-0006-5365-6435; Juho Sim: https://orcid.org/0000-0001-7652-4482; Juyeon Oh: https://orcid.org/0000-0002-9843-4928; Jongmin Lee: https://orcid.org/0000-0002-5265-5864; Chorom Lee: https://orcid.org/0000-0002-5758-767X; Yangwook Kim: https://orcid.org/0000-0002-9162-5350; Byungyoon Yun: https://orcid.org/0000-0001-7055-6424; Jin-ha Yoon: https://orcid.org/0000-0003-4198-2955
* Corresponding author. Department of Preventive Medicine, Yonsei University College of Medicine, Seoul 03722, Republic of Korea.
** Corresponding author. Department of Preventive Medicine, Yonsei University College of Medicine, Seoul 03722, Republic of Korea.
E-mail addresses: yby3721@yuhs.ac (B. Yun), flyinyou@yuhs.ac (J.-h. Yoon).

2093-7911/$ – see front matter
https://doi.org/10.1016/j.shaw.2024.03.001

prevalent worldwide. In the United States, platform workers were estimated to account for 34% of the total workforce, or approximately 55 million individuals, in 2017 [4]. By 2025, it is expected that the majority of the workforce, including platform workers, will engage in the gig economy, contributing nearly $2.7 trillion to global gross domestic product [5].

Replacement drivers, also known as “Daeri drivers”, are one of the most common groups of platform workers in the Republic of Korea [6]. Replacement driving began in 1980 when drunk driving was banned, and a breathalyzer was used to measure the driver’s blood alcohol level [7]. This has continued until the present day. Replacement drivers generally work at night, transporting customers’ vehicles to their destination, and frequently deal with drunk clients [6]. Replacement drivers, unlike Uber and taxis, drive a customer’s vehicle as if it were their own. A qualitative study reported that replacement drivers work in an unsafe environment and engage in emotional labor related to drunk customers [6]. Previous research on the long-term impact of emotional labor indicates that it often leads to both presenteeism and absenteeism. This is particularly relevant for replacement drivers who perform emotionally taxing work during night shifts, raising concerns about simultaneous occurrences of presenteeism and absenteeism in this group [8,9].

Presenteeism refers to the continuation of work when an individual is sick [10], while absenteeism is the absence of employees from their jobs due to illness or injury [11]. Previous research has shown that presenteeism and absenteeism could be affected by working environment characteristics; in particular, presenteeism is associated with health risks and conditions [12,13]. Presenteeism potentially exposes workers to a loss of productivity, accidents, job insecurity, poor health, and depression [10]. According to a study on the potential loss to a company due to presenteeism in conjunction with the sick leave registration data in Finland, the expected annual loss is between 4.6 million and 5.6 million euros [14].

Currently, there is a gap in existing research examining the relationship between presenteeism, absenteeism, and replacement workers, despite its importance. Only the content of social security system updates on presenteeism can be found in previous studies on replacement workers [15]. Thus, this study compares the relationship between presenteeism and absenteeism in replacement workers using a variety of demographic, behavioral, occupational, and social participation characteristics. In addition, the prediction of presenteeism and absenteeism in replacement workers is performed using machine learning algorithms, including deep learning techniques. This study offers a solution to the health problems faced by male replacement workers.

2. Materials and methods

2.1. Study data and population

This study used two different datasets: platform data of replacement drivers collected by the authors and the 6th Korean Working Conditions Survey (KWCS) data for paid workers. The platform data were collected between July 2021 and October 2022, and 457 male replacement drivers were selected from a group of replacement workers. Given the independence of replacement drivers, a snowball sampling method was employed to attract participants, enabling them to recruit additional participants [16]. A community-based participatory research method was applied to promote individual participation, emphasizing equal cooperation between researchers and community members [17].

Basic demographics, socioeconomic status, and job-related and psychological status, including absenteeism and presenteeism, were evaluated. The survey for replacement drivers included the same questionnaires as those in the KWCS. Only male workers in the KWCS and platform data were included in this study owing to insufficient female replacement drivers ($n = 8$).

In compliance with ethical guidelines, informed consent was obtained from all replacement driver participants following institutional review board approval for this study.

Since 2006, the KWCS, which is nationally representative via stratified sampling [18], has been conducted by the Korea Occupational Safety and Health Institutions. It is a nationwide study utilized to acquire basic data for industrial accident prevention plans by investigating various employment and labor conditions that affect occupational safety and health [19,20]. The 6th KWCS examined 50,538 workers in 2020. The participants were economically active people aged 15 years or older. The inclusion criterion for our study regarding KWCS data was male paid workers.

The exclusion criteria in this study were as follows (Fig. S1):

Platform survey data are as follows: 1) Female participants ($n = 8$); 2) missing data ($n = 21$).

KWCS: 1) aged <20 ($n = 70$); 2) female participants ($n = 26{,}795$); 3) people who were not wage workers ($n = 11{,}292$); and 4) missing data ($n = 1{,}568$).

2.2. Outcomes, independent variables, and other covariates

The outcomes of the study were absenteeism and presenteeism. The questions regarding absenteeism were as follows: “Over the past 12 months (or since you started your job), how many days in total were you absent from work due to sick leave or health-related leave?”. Absenteeism was defined as whether individuals experienced it for even just one day or not. The question regarding presenteeism was as follows: “Over the past 12 months (or since you started your job), did you work when you were sick?” This study investigated presenteeism by determining whether individuals experienced it or not. The platform survey participants were “male replacement workers” as an independent variable, while the control group consisted of male paid workers of the KWCS. The adjusted variables were age, education level, income, working hours, working days per week, and working duration. Education level was separated into two categories: higher than graduate university and below high school. Income was classified into 3 categories based on the annual income of 24 million (Korean won [KRW]) and 48 million KRW: “High”, “Middle”, and “Low”. Working hours per week was divided into two categories: “$\leq$40 hours” and “$>$40 hours”. Working days per week were separated into two categories based on five days: “$>$5 days” and “$\leq$5 days”. Working duration was classified into 3 categories based on 3 years and 10 years: “$<$3 years”, “$<$10 years”, and “$\geq$10 years”.

2.3. Statistical analysis

Propensity score (PS) matching with a ratio of 1:3 (caliper width 0.2) was performed to balance the difference between paid workers and replacement drivers while minimizing potential confounding bias. The PS was estimated using logistic regression with variables of age, education, income, working hours, working days per week, and working duration. The standardized mean difference was utilized as an indication for assessing the imbalance between treatment and control group prior to PS matching.

The frequency (%) and mean (standard deviation, SD) of the baseline characteristics were calculated according to the replacement drivers and paid workers. To compare the differences between entire and matching, or replacement drivers and paid workers, the chi-squared test for categorical variables and the *t* test for continuous variables were used. The adjusted odds ratios (ORs) and 95% confidence intervals (CIs) for presenteeism and absenteeism by replacement drivers in the PS-matched population were estimated using univariate and multivariate logistic regression models.

To assess the relationship between presenteeism and absenteeism and replacement drivers in each stratum, stratification analyses by age group, education, income, and working hours were performed. The age group was divided based on the median age of the participants (55 years). Educational level, income level, and working hours were stratified according to previously classified categories.

Machine learning was trained using 70% of the training set and 30% of the test set. Data training was conducted on the training set, and validation was conducted on the test set. Owing to unbalanced data regarding the prevalence of the outcomes, this study conducted data oversampling. Machine learning was utilized for prediction with the variables of age, education, income, working hours, working duration, and working days per week.

Furthermore, to ascertain whether a better prediction model for presenteeism and absenteeism exists, generalized logistic model (GLM), Naive Bayes (NB), recursive partitioning and regression trees (RPAR T), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and generalized boosted models (GBMs) were used as machine learning algorithm systems. Modern neural network layers, activation functions, optimizers, and tools for evaluating, measuring, and debugging deep neural networks are all supported by TensorFlow [21]. The area under the curve (AUC) and balanced accuracy of each machine learning model were calculated in this study.

The occupational characteristics of the control group were classified according to the occupation classification or employment status. First, the occupation classification was classified as "white-collar workers" and "non-white-collar workers" (office workers, managers, and professional technician as white-collar workers and others as non-white-collar workers). The employment status in the control group was classified as "temporary workers" and "regular workers". The adjusted OR (95% CI) was estimated using a logistic regression model to measure the relationship between presenteeism and absenteeism in those subgroups, compared to the replacement driver group.

A p value <0.05 was regarded as statistically significant for all two-sided statistical tests. R (The R Foundation for Statistical Computing, Vienna, Austria, version 4.0.5) was used to conduct all statistical analyses, except for deep learning. The "Caret" package was used to perform machine learning. To develop the deep learning model, Keras/TensorFlow (version 2.9.1) was used via Python (version 3.10.6).

3. Results

The absolute value of the standardized mean difference of each variable was less than 0.1, demonstrating that the matching was acceptable (Table S1). After PS matching, 1,462 individuals remained from 11,339 participants. The mean (SD) age of all participants was 45.06 (13.63), and the proportion of replacement drivers was 26.49% ($n = 428$). The prevalence of presenteeism and absenteeism among replacement drivers was 53.7% ($n = 230$) and 52.6% ($n = 225$), respectively. The prevalence of presenteeism and absenteeism among paid workers was 9.4% ($n =$ 1,025) and 3.7% ($n = 403$), respectively. Replacement drivers had a significantly higher prevalence of experience in presenteeism and absenteeism, high education level, low income level, long working hours, more than 5 working days per week, and working duration was 3 to 10 years (all $p < 0.001$). After PS matching, replacement drivers had a significantly higher prevalence of experience in presenteeism and absenteeism, high education level, medium income level, long working hours, more than 5 working days per week, and working duration was 3 to 10 years (Table 1).

In the adjusted model, replacement drivers had a higher OR of presenteeism (crude OR = 8.41 [95% CI = 6.41–11.03], $p < 0.001$) and a higher OR of absenteeism (crude OR = 23.23 [95% CI = 16.4–33.03], $p < 0.001$) than paid workers. After adjusting for covariates, the relationship was maintained (Presenteeism Model: adjusted OR = 8.77 [95% CI = 6.65–11.58], $p < 0.001$), and replacement drivers had a significant association with absenteeism (Absenteeism Model: adjusted OR = 24.14 [95% CI 16.85–34.58], $p < 0.001$), compared to paid workers (Table 2).

According to the stratification analysis (Table 3), presenteeism has a significantly higher association with replacement drivers in the young age group, high education level, high income level, and long working hours. In contrast, absenteeism has a significantly stronger association with replacement drivers in the old age group, high education level, medium income level, and long working hours.

After adjusting for covariates, in the stratification of regular paid workers according to their occupation classification, replacement drivers (Absenteeism Model: OR = 33.00 [95% CI = 17.20–63.29], $p < 0.001$, Presenteeism Model: OR 8.39 = [95% CI = 5.59–12.59], $p < 0.001$) were significantly associated with absenteeism and presenteeism, compared with non-white-collar workers who were not significantly associated with absenteeism (OR = 1.53 [95% CI = 0.75–3.13], $p = 0.238$) or presenteeism (OR = 0.94 [95% CI = 0.61–1.44], $p = 0.768$) and with white-collar workers. In the stratification of employment status, replacement drivers were significantly associated with absenteeism (OR = 26.61 [95% CI = 17.77–39.83], $p < 0.001$) and presenteeism (OR = 8.69 [95% CI = 6.46–11.69], $p < 0.001$), whereas temporary workers were not significantly associated with absenteeism (OR = 1.52 [95% CI = 0.75–3.06], $p = 0.241$) or presenteeism (OR = 0.96 [95% CI = 0.58–1.58], $p = 0.855$), compared with regular workers (Table 4). Furthermore, when analyzing participants who only had a main job, among the replacement drivers, 273 had a main job, 88 had second jobs, and 22 did not respond. In comparison, 1,079 paid workers had a main job. After adjusting for covariates, replacement drivers were significantly associated with absenteeism (OR = 30.32 [95% CI = 20.49–44.84], $p < 0.001$) and presenteeism (OR = 9.79 [95% CI = 7.17–13.36], $p < 0.001$).

According to the machine learning algorithm of presenteeism and absenteeism (Table S2), the most predictive model for absenteeism was NB (0.923), followed by GBM (0.918) and QDA (0.917). The most predictive model for presenteeism was GLM(0.841) and NB&LDA (0.839), followed by QDA (0.836), based on the AUC values of the models (Fig. S2A, B). According to the deep learning algorithms of absenteeism conducted by TensorFlow, the AUC value of the model was 0.856, with a balanced accuracy of 0.863 (number of nodes: 16, dropout probability: 0.2, learning rate: 0.001, batch size: 32, epochs: 100). The AUC value of the model in presenteeism was 0.791, with a balanced accuracy of 0.850 (number of nodes: 16, dropout probability: 0, learning rate: 0.005, batch size: 32, epochs: 100) (Fig. S2C, D).

4. Discussion

In this study, both presenteeism and absenteeism were found to have a significant relationship in replacement drivers; among them, absenteeism indicated a particularly stronger relationship. A stratified analysis revealed that the association of both presenteeism and absenteeism with replacement drivers was significant,

Table 1
Baseline characteristics of participants by presenteeism and absenteeism

Variable	Entire population			The PS-matched population		
	Paid workers ($n = 10{,}911$)	Replacement drivers ($n = 428$)	p	Paid workers ($n = 1{,}079$)	Replacement drivers ($n = 383$)	p
Age			<0.001			0.067
Mean (SD)	44.8 ± 13.6	56.0 ± 7.9		56.5 ± 9.8	55.6 ± 7.9	
Presenteeism			<0.001			<0.001
No	9,886 (90.6%)	198 (46.3%)		949 (88.0%)	178 (46.5%)	
Yes	1,025 (9.4%)	230 (53.7%)		130 (12.0%)	205 (53.5%)	
Absenteeism			<0.001			<0.001
No	10,508 (96.3%)	203 (47.4%)		1,031 (95.6%)	184 (48.0%)	
Yes	403 (3.7%)	225 (52.6%)		48 (4.4%)	199 (52.0%)	
Education			0.856			0.079
High	6,460 (59.2%)	251 (58.6%)		522 (48.4%)	206 (53.8%)	
Low	4,451 (40.8%)	177 (41.4%)		557 (51.6%)	177 (46.2%)	
Income			<0.001			0.127
High	2,718 (24.9%)	80 (18.7%)		207 (19.2%)	80 (20.9%)	
Medium	6,393 (58.6%)	173 (40.4%)		521 (48.3%)	162 (42.3%)	
Low	1,800 (16.5%)	175 (40.9%)		351 (32.5%)	141 (36.8%)	
Working hour			<0.001			0.574
<40 hours	7,272 (66.6%)	203 (47.4%)		476 (44.1%)	176 (46.0%)	
≥40 hours	3,639 (33.4%)	225 (52.6%)		603 (55.9%)	207 (54.0%)	
Working days for week			<0.001			0.663
>5 days	2,250 (20.6%)	265 (61.9%)		604 (56.0%)	220 (57.4%)	
≤5 days	8,661 (79.4%)	163 (38.1%)		475 (44.0%)	163 (42.6%)	
Working duration			<0.001			0.628
<3 years	3,545 (32.5%)	108 (25.2%)		309 (28.6%)	101 (26.4%)	
<10 years	4,039 (37.0%)	198 (46.3%)		499 (46.2%)	187 (48.8%)	
≥10 years	3,327 (30.5%)	122 (28.5%)		271 (25.1%)	95 (24.8%)	

Values are expressed by n (%) or mean (SD).
SD, standard deviation.

with a prominent association in the young age group, and those with high education and medium income levels. Moreover, after classifying the control group based on occupation or employment status, the extent of the association between presenteeism and absenteeism and replacement drivers was significantly high. Individuals whose main job was replacement driving demonstrated significantly more presenteeism than those whose second jobs were replacement driving. In the machine learning analysis, QDA was the most accurate model for presenteeism and absenteeism.

Replacement workers have the characteristics of temporary work. Replacement drivers among replacement workers have the characteristics of outdoor and night shift work. These characteristics are associated with absenteeism and presenteeism. There is a significant probability of presenteeism in the case of temporary workers because of job instability, such as the conversion to regular workers or maintenance of employment [22,23]. Furthermore, prior research has demonstrated that temporary workers who work shifts have a higher rate of long-term sick leave [24]. Among blue-collar workers, those

Table 2
The ORs (95% CIs) of presenteeism and absenteeism by replacement drivers in logistic regression model

Variable	Values	Presenteeism model*			Absenteeism model*		
		Odds ratios	CI	p	Odds ratios	CI	p
Platform	Paid workers	1.00 (reference)			1.00 (reference)		
	Replacement drivers	8.77	6.65–11.58	<0.001	24.14	16.85–34.58	<0.001
Age	Age	1.00	0.98–1.01	0.656	0.99	0.97–1.01	0.249
Education	High	1.00 (reference)			1.00 (reference)		
	Low	0.93	0.71–1.23	0.621	0.81	0.58–1.14	0.225
Income	High	1.00 (reference)			1.00 (reference)		
	Medium	1.14	0.79–1.64	0.489	1.15	0.73–1.80	0.549
	Low	0.93	0.62–1.39	0.717	1.15	0.71–1.85	0.580
Working hour	<40 hours	1.00 (reference)			1.00 (reference)		
	≥40 hours	1.52	1.07–2.16	0.020	1.57	1.03–2.39	0.035
Working days for week	>5 days	1.00 (reference)			1.00 (reference)		
	≤5 days	0.94	0.66–1.34	0.746	0.95	0.62–1.44	0.805
Working duration	High	1.00 (reference)			1.00 (reference)		
	Medium	1.05	0.75–1.46	0.796	1.77	1.17–2.69	0.007
	Low	1.23	0.84–1.80	0.296	1.29	0.79–2.10	0.304

CI, confidence interval; OR, odds ratio.
* Adjusted by age, education, income, working hour, working days for week, and working duration.

Table 3
Presenteeism and absenteeism according to SES and working hour

	Values	Presenteeism*			Absenteeism*		
		Odds ratio	CI	*p*	Odds ratio	CI	*p*
Age	≤55	10.83	7.15–16.41	<0.001	22.90	13.56–38.65	<0.001
	>55	7.31	4.98–10.74	<0.001	24.66	14.93–40.72	<0.001
Education	High	9.13	6.16–13.55	<0.001	26.90	16.11–44.92	<0.001
	Low	8.43	5.63–12.62	<0.001	20.51	12.27–34.29	<0.001
Income	High	11.48	5.83–22.59	<0.001	23.20	10.03–53.69	<0.001
	Medium	10.42	6.89–15.77	<0.001	31.17	18.02–53.91	<0.001
	Low	5.90	3.63–9.57	<0.001	22.04	11.53–42.15	<0.001
Working hour	<40 hours	7.51	4.89–11.54	<0.001	20.74	11.75–36.62	<0.001
	≥40 hours	10.08	6.95–14.61	<0.001	31.07	19.16–50.39	<0.001

CI, confidence interval.
* Adjusted by age, education, income, working hour, working days for week, and working duration.

engaged in construction or civil engineering with outdoor working characteristics, a trait of replacement drivers, also had higher presenteeism than white-collar workers [25].

Even considering these characteristics, the OR of the replacement driver for presenteeism and absenteeism was higher than that of the regular worker. According to the analysis that employed a stratified control group based on occupation classification or employment status, replacement drivers were more strongly associated with presenteeism and absenteeism than temporary workers or non-white-collar workers, compared to regular or white-collar workers. This implies that in addition to temporary and outdoor characteristics, replacement drivers may have more characteristics, such as the absence of a supervisor, no colleagues, no time pressure, irregularities in work schedule, and income instability, compared to temporary or outdoor shift workers. A previous study indicated that manager's workplace interventions affect presenteeism and absenteeism [22,23]. Some supervisors can help reduce absenteeism and presenteeism in the workplace through healthcare systems and work environment changes [22]. Consequently, replacement drivers with no supervisors are expected to experience a high rate of presenteeism and absenteeism and do not experience time pressure or have coworkers. This may lead to lack of responsibility for work and reduced work efficiency. Furthermore, those who work on a case-by-case basis may be able to take a leave of absence while enduring economic loss and reducing the emotional burden in the case of illness. Therefore, because absence may be determined more quickly for replacement drivers than for paid workers, the extent of the association between absenteeism and replacement drivers was greater than that of presenteeism.

According to previous studies, absenteeism and presenteeism are inversely associated with paid workers' income in Sweden, the United States, Portugal, and Malaysia [26–29], whereas high income levels were significantly associated with presenteeism and absenteeism in this study. Since replacement drivers might work as much as they want and are paid more for each additional job they undertake, presenteeism is thought to be more prevalent among high-income earners. Individuals with high absenteeism are thought to have decided that since they have already profited, they should rest if they feel unwell.

Table 4
The ORs (95% CIs) of presenteeism and absenteeism according to the employment status and the occupation classification in logistic regression model

Variable	Values	The employment status*				The occupation classification*			
		Absenteeism		Presenteeism		Absenteeism		Presenteeism	
		Odds ratio (CIs)	*p*	Odds ratio (CIs)	*p*	Odds ratio (CIs)	*p*	Odds ratio (CIs)	*p*
Job stat	Regular	1.00 (reference)		1.00 (reference)					
	Temporary	1.52 (0.75–3.06)	0.241	0.96 (0.58–1.58)	0.855				
	Replacement drivers	26.61 (17.77–39.83)	<0.001	8.69 (6.46–11.69)	<0.001				
Collar job	White-collar worker					1.00 (reference)		1.00 (reference)	
	Non-white-collar worker					1.53 (0.75–3.13)	0.238	0.94 (0.61–1.44)	0.768
	Replacement drivers					33.00 (17.20–63.29)	<0.001	8.39 (5.59–12.59)	<0.001
Age		0.99 (0.97–1.01)	0.213	1.00 (0.98–1.01)	0.668	0.99 (0.97–1.01)	0.213	1.00 (0.98–1.01)	0.674
Education	High	1.00 (reference)		1.00 (reference)		1.00 (reference)		1.00 (reference)	
	Low	0.79 (0.56–1.11)	0.180	0.94 (0.71–1.24)	0.641	0.77 (0.55–1.09)	0.145	0.94 (0.71–1.26)	0.696
Income	High	1.00 (reference)		1.00 (reference)		1.00 (reference)		1.00 (reference)	
	Medium	1.14 (0.72–1.79)	0.575	1.14 (0.79–1.64)	0.485	1.12 (0.71–1.76)	0.623	1.14 (0.79–1.65)	0.472
	Low	1.11 (0.69–1.80)	0.667	0.93 (0.62–1.40)	0.737	1.11 (0.69–1.80)	0.670	0.94 (0.62–1.40)	0.747
Working hour	<40 hours	1.00 (reference)		1.00 (reference)		1.00 (reference)		1.00 (reference)	
	≥40 hours	1.61 (1.06–2.46)	0.027	1.51 (1.06–2.15)	0.023	1.56 (1.02–2.37)	0.038	1.52 (1.07–2.16)	0.020
Working days for week	>5 days	1.00 (reference)		1.00 (reference)		1.00 (reference)		1.00 (reference)	
	≤5 days	0.95 (0.63–1.46)	0.827	0.94 (0.66–1.34)	0.743	0.96 (0.63–1.46)	0.833	0.94 (0.66–1.34)	0.739
Working duration	High	1.00 (reference)		1.00 (reference)		1.00 (reference)		1.00 (reference)	
	Medium	1.83 (1.20–2.80)	0.005	1.04 (0.74–1.46)	0.823	1.79 (1.18–2.72)	0.006	1.04 (0.74–1.46)	0.808
	Low	1.34 (0.82–2.18)	0.249	1.22 (0.83–1.80)	0.313	1.32 (0.81–2.15)	0.265	1.22 (0.83–1.79)	0.310

CI, confidence interval.
* Adjusted by age, education, income, working hour, working days for week, and working duration.

Predicting absenteeism and presenteeism among replacement drivers is vital to ensure the sustainability of the labor force in the platform work sector, especially in a changing labor market. The fundamental emphasis of platform worker research advances beyond just decreasing prevalence of diseases to improving overall health at work [30]. Despite their irregular work schedules, replacement drivers' absenteeism and presenteeism must be assessed in relation to their working environment. In the machine learning analysis, the AUC of absenteeism was higher than that of presenteeism, which might be due to the high OR value of absenteeism. Thus, absenteeism among replacement drivers is an important characteristic that is required for prediction. In addition, the AUC values of deep learning and machine learning did not differ significantly in this study. In general, the larger the dataset, the better the statistical predictive capacity for pattern recognition [31]. Due to the study's small dataset, it seems that the statistical predictive capacity has not been sufficiently enhanced. This might result in a minor difference between the machine learning and deep learning AUC values. Considering the high AUC value, both absenteeism and presenteeism can be considered as important characteristics representing replacement drivers.

This study has several strengths. First, to the best of our knowledge, this is the first study in Asia to examine the relationship between presenteeism and absenteeism and replacement driver characteristics. Second, owing to the dispersed job characteristics of replacement drivers, the recruitment process presented considerable challenges. However, these challenges were effectively navigated by employing the snowball sampling method, a strategy that proved successful in enlisting participants for the study. The findings of this study highlighted the significant challenges in gathering data from vulnerable populations, underlining their high level of vulnerability and the critical need for research in areas with limited information. Third, this study had the advantage of using a sizable sample from the KWCS dataset as a reference, which is representative of the sociodemographic and job-related variables of the Korean population. However, this study had several limitations. First, as this was a cross-sectional study, we were unable to establish a causal relationship between presenteeism, absenteeism, and replacement driver characteristics. Second, we did not consider several variables that may have acted as unmeasured confounders, including health-related factors, insomnia, and other psychological conditions that might affect presenteeism and absenteeism. Third, it is challenging to recruit a population representative enough for generalization due to the unique job characteristics of replacement drivers. Future research employing random sampling to target replacement drivers specifically is essential for more generalized findings. Further studies on the health outcomes of replacement drivers should utilize a well-designed cohort with detailed job characteristic information. Fourth, memory and recollection biases can affect survey results.

The findings of this study demonstrated that replacement drivers experienced presenteeism and absenteeism to a greater extent than paid workers. This stronger association may be due to temporary work, outdoor work, and absence of supervision, among other replacement driver characteristics. Given the high prevalence of presenteeism and absenteeism among replacement drivers, more comprehensive research and strategy development to control presenteeism and absenteeism of replacement drivers' risk factors is required.

Funding

This research was supported by a grant from the Korea Health Promotion Institute R&D Project, funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HS21C2367).

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgments

This study was approved by our institutional review board and was conducted in accordance with the ethical requirements of the 1975 Declaration of Helsinki and its later amendments (IRB No 4-2022-0509).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.shaw.2024.03.001.

References

[1] Bajwa U, Gastaldo D, Di Ruggiero E, Knorr L. The health of workers in the global gig economy. Global Health 2018;14(1):124. https://doi.org/10.1186/s12992-018-0444-8. PMID: 30563539.
[2] Berg J, Furrer M, Harmon E, Rani U, Silberman MS. Digital labour platforms and the future of work: Towards decent work in the online world; 2018.
[3] Kim S, Kang M, Park J. Digital industrial accidents: a case study of the mental distress of platform workers in South Korea. Int J Soc Welfare 2022;31(3):355–67.
[4] Cao X, Zhang D, Huang L. The impact of COVID-19 pandemic on gig economy labor supply. JNSSoB; 2020.
[5] Allon G, Cohen MC, Sinchaisri WP. The impact of behavioral and economic drivers on gig economy workers. JAaS; 2018.
[6] Jung PK, Won JU, Roh J, Lee JH, Seok H, Lee W, Yoon JH. Workplace Violence Experienced by substitute (Daeri) drivers and its relationship to depression in Korea. J Korean Med Sci 2015 Dec;30(12):1748–53. https://doi.org/10.3346/jkms.2015.30.12.1748. Epub 2015 Nov 30. PMID: 26713049; PMCID: PMC4689818.
[7] Yun J. Seoul's nocturnal urbanism: An emergent night-time economy of substitute driving and fast deliveries. Urban Stud 2022;59(6):1238–54.
[8] Cho SS, Lee DW, Kang MY. The association between shift work and health-related productivity loss due to either sickness absence or reduced performance at work: a cross-sectional study of Korea. Int J Environ Res Public Health 2020;17(22). https://doi.org/10.3390/ijerph17228493. PMID: 33207798.
[9] Sung Won J, June-Hee L, Kyung-Jae L. Assessing the association between emotional labor and presenteeism among nurses in Korea: cross-sectional study using the 4th Korean Working Conditions Survey. Saf Health Work 2020;11(1):103–8. https://doi.org/10.1016/j.shaw.2019.12.002. PMID: JUNG2020103.
[10] Sanderson K, Cocker F. Presenteeism: implications and health risks. Aust Fam Physician 2013;42(4):172–5.
[11] Grinza E, Rycx F. The impact of sickness absenteeism on firm productivity: new evidence from Belgian matched employer–employee panel data. Ind Relat J Econ Soc 2020;59(1):150–94.
[12] Bergström G, Bodin L, Hagberg J, Lindh T, Aronsson G, Josephson M. Does sickness presenteeism have an impact on future general health? Int Arch Occup Environ Health 2009 Nov;82(10):1179–90. https://doi.org/10.1007/s00420-009-0433-6. PMID: 19504117.
[13] Webster JJRpfUNIBEC. Diversity management in the ICT industry: challenges and issues for social dialogue; 2007.
[14] Vänni K, Neupane S, Nygård C-H. An effort to assess the relation between productivity loss costs and presenteeism at work. Int J Occup Saf Ergon 2017;23(1):33–43. https://doi.org/10.1080/10803548.2016.1197578.
[15] Ose S. Precarious work, sickness absence and risk sharing between employers, employees and social insurance: Solveig Ose. Eur J Public Health 2022;32(Suppl. 3). https://doi.org/10.1093/eurpub/ckac129.376.
[16] Julian McFarlane S, Occa A, Peng W, Awonuga O, Morgan SE. Community-based participatory research (CBPR) to enhance participation of racial/ethnic minorities in clinical trials: a 10-year systematic review. Health Commun 2022;37(9):1075–92. https://doi.org/10.1080/10410236.2021.1943978.
[17] Leighton K, Kardong-Edgren S, Schneidereith T, Foisy-Doll C. Using social media and snowball sampling as an alternative recruitment strategy for research. Clin Simul Nurs 2021;55:37–42. https://doi.org/10.1016/j.ecns.2021.03.006.
[18] Lee NR, Kim SW, Joo JH, Lee JH, Lee JH, Lee KJ. Relationship between workplace violence and work-related depression/anxiety, separating the types of perpetrators: a cross-sectional study using data from the fourth and fifth Korean Working Conditions Surveys (KWCS). Ann Occup Environ Med 2022;34:e13. https://doi.org/10.35371/aoem.2022.34.e13. PMID: 35923792.

[19] Shin MG, Kim Y-K, Kim S-Y, Kang DM. Relationship between job training and subjective well-being in accordance with work creativity, task variety, and occupation. Saf Health Work 2020;11(4):466–78. https://doi.org/10.1016/j.shaw.2020.08.006.
[20] Kim YS, Park J, Rhee KY, Kim HM. A comparison between the second Korean Working Conditions Survey (KWCS) and the first KWCS. Saf Health Work 2015;6(2):85–9. https://doi.org/10.1016/j.shaw.2015.02.005. PMID: 26106506.
[21] Kurth T, Smorkalov M, Mendygral P, Sridharan S, Mathuriya A. TensorFlow at Scale: Performance and productivity analysis of distributed training with Horovod, MLSL, and Cray PE ML. Concurrency Comput Pract Ex 2019;31(16):e4989.
[22] Reuter M, Wahrendorf M, Di Tecco C, Probst TM, Ruhle S, Ghezzi V, Barbaranelli C, Iavicoli S, Dragano N. Do temporary workers more often decide to work while sick? Evidence for the link between employment contract and presenteeism in Europe. Int J Environ Res Public Health 2019;16:1868. https://doi.org/10.3390/ijerph16101868.
[23] Kim JY, Lee J, Muntaner C, Kim S-S. Who is working while sick? Nonstandard employment and its association with absenteeism and presenteeism in South Korea. Int Arch Occup Environ Health 2016;89(7):1095–101. https://doi.org/10.1007/s00420-016-1146-2.
[24] Nätti J, Oinas T, Härmä M, Anttila T, Kandolin I. Combined effects of shiftwork and individual working time control on long-term sickness absence: a prospective study of Finnish employees. J Occup Environ Med 2014;56(7):732–8.
[25] Kayaba M, Sasai-Sakuma T, Takaesu Y, Inoue Y. The relationship between insomnia symptoms and work productivity among blue-collar and white-collar Japanese workers engaged in construction/civil engineering work: a cross-sectional study. BMC Public Health 2021;21(1):1244. https://doi.org/10.1186/s12889-021-11273-y.
[26] Wee LH, Yeap LLL, Chan CMH, Wong JE, Jamil NA, Swarna Nantha Y, Siau CS. Anteceding factors predicting absenteeism and presenteeism in urban area in Malaysia. BMC Publ Health 2019;19:1–12.
[27] Taloyan M, Aronsson G, Leineweber C, Magnusson Hanson L, Alexanderson K, Westerlund H. Sickness presenteeism predicts suboptimal self-rated health and sickness absence: a nationally representative study of the Swedish working population. PLoS One 2012;7(9):e44721. https://doi.org/10.1371/journal.pone.0044721. PMID: 22984547.
[28] Susser P, Ziebarth NR. Profiling the U.S. sick leave landscape: presenteeism among females. Health Serv Res 2016;51(6):2305–17. https://doi.org/10.1111/1475-6773.12471. PMID: 26947427.
[29] Martinez LF, Ferreira AI. Sick at work: presenteeism among nurses in a Portuguese public hospital. Stress Health 2012;28(4):297–304. https://doi.org/10.1002/smi.1432. PMID: 22282226.
[30] Framke E, Sørensen JK, Alexanderson K, Farrants K, Kivimäki M, Nyberg ST. Emotional demands at work and risk of long-term sickness absence in 1·5 million employees in Denmark: a prospective cohort study on effect modifiers. Lancet Public Health 2021;6(10):e752–9. https://doi.org/10.1016/S2468-2667(21)00185-7.
[31] Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PLoS One 2019;14(11):e0224365. https://doi.org/10.1371/journal.pone.0224365. PMID: pub.1122390471.