Employment Contracts and Stress: Experimental Evidence

Julia L. Allan[1], Nicole Andelic[2]*, Keith A. Bender[2], Daniel Powell[1 3], Sandro Stoffel[4], Ioannis Theodossiou[2]

[1] Health Psychology, Institute for Applied Health Sciences, University of Aberdeen, Aberdeen, United Kingdom. Email: j.allan@abdn.ac.uk and daniel.powell@abdn.ac.uk

[2] Economics Department, Business School, University of Aberdeen, Aberdeen, United Kingdom. Email: nicole.andelic@abdn.ac.uk, kabender@abdn.ac.uk and theod@abdn.ac.uk

[3] Rowett Institute, University of Aberdeen, Aberdeen, United Kingdom.

[4] University of Basel, Basel, Switzerland. Email: s.stoffel@ucl.ac.uk

Abstract

A growing literature has found a link between performance-related pay (PRP) and poor health, but the causal direction of the relationship is not known. To address this gap, the current paper utilises a crossover experimental design to randomly allocate subjects into a work task paid either by performance or a fixed payment. Stress is measured through self-reporting and salivary cortisol. The study finds that PRP subjects had significantly higher cortisol levels and self-rated stress than those receiving fixed pay, *ceteris paribus*. By circumventing issues of self-report and self-selection, these results provide novel evidence for the detrimental effect PRP may have on health.

JEL: J33, I0, C91

Keywords: performance-related pay, stress, experiment, cortisol

Notes: Authors are listed in alphabetical order. We gratefully acknowledge financial support from the Scottish Economic Society on the pilot study and from the ESRC (Grant ES/R01163X/1). We also extend our thanks to Matthew McGinty, Christine Spencer and participants at the 2020 EALE/SOLE/AASLE World Conference, the 2021 GLO Research Seminars, the Economics Centre at Curtin University, the Economics Department at St Lawrence University and the Applied Health Psychology Research Workshop and Business School at the University of Aberdeen who provided thoughtful comments on an early draft of the current manuscript. Finally, we are grateful to the Editor of the Journal of Economic Behavior & Organization and two anonymous reviewers who improved the quality of the current manuscript. Institutional Review Board (IRB) approval was obtained from the University of Aberdeen, College of Life Sciences & Medicine Ethics Review Board (CERB/2015/5/1198). The study was pre-registered with the Open Science Framework (https://osf.io/sxkb2) prior to data collection.

*Correspondence should be made with Nicole Andelic, Economics Department, Business School, University of Aberdeen, phone: +44(0)1224272709, email: nicole.andelic@abdn.ac.uk.

Declarations of interest: none.

As early as the 18th century in the *Wealth of Nations*, Adam Smith observed that *"Workmen… when they are liberally paid by the piece, are very apt to overwork themselves and to ruin their health and constitution…"*. Although the link between performance-related pay (PRP) and health was well understood over two centuries ago, the use of performance pay in the labour market is still widespread. Indeed, although the rate differs by country and industry as well as how one defines PRP, between 10 to 40% of workers in Europe and the US are paid according to how well they perform in their job (Bryson et al., 2013). If Smith is right, this relatively large portion of the workforce may be at risk of poor health due to their job contracts.

Traditionally, economics has focused on the advantages of performance pay. One of the most famous examples is Lazear (2000) which finds a switch to performance-related pay increased productivity in a windscreen factory by 44% and earnings by 10% on average. Generally, it is associated with higher productivity in both field (Gielen, Kerkhofs, and Van Ours, 2010) and laboratory-based studies (Dohmen and Falk, 2011), higher earnings (Booth and Frank, 1999; Parent, 1999) and increased job satisfaction (Green and Heywood, 2008), just to give a few examples. However, a growing body of literature suggests that there are important negative repercussions of performance pay on the well-being of the workers employed under such conditions. Those paid by performance report higher rates of injuries (Bender, Green, and Heywood, 2012), poorer physical and emotional health (Bender and Theodossiou, 2014; Davis, 2016) and employees of firms using PRP are more likely to use anti-anxiety and anti-depressive medication (Dahl and Pierce, 2019). Data from the UK Health and Safety Executive (HSE) estimate that 26.9 million workdays were lost due to work-related illnesses in 2017/18 (HSE, 2020) with over half of these (12.8 million) due to stress on the job (HSE, 2020). In other words, the social and economic costs of poor health may outweigh potential advantages of PRP (Robertson and Cooper, 2011).

What is unclear in the academic studies linking PRP and health, however, is both the direction of causality and the potential mechanism explaining the relationship. Generally, the previous literature has used observational survey data to investigate the link between PRP and health and therefore struggle to claim that the relationships they find are causal rather than correlational. In addition, while there are a number of possible different underlying mechanisms (covered below), one key mechanism might be the additional stress that the higher productivity and more variable income stream associated with PRP will generate. Consistent exposure to even low-grade stress can cause ill health (DeLongis et al., 1982).

This study, therefore, focuses on the relationship between PRP and stress by highlighting three key features: (i) randomising the assignment of workers to PRP to identify the exogenous effect of PRP on stress, (ii) by employing a crossover design allowing us to control for individual heterogeneity and (iii) examining whether those who are paid by PRP experience higher stress as indicated by both self-report and salivary cortisol (a hormone the body produces when exposed to stress). The findings show that there is a clear causal relationship from PRP on stress as measured by both subjective self-report and by having higher cortisol levels when subjects are in the PRP condition. This relationship is robust to the inclusion of sociodemographic characteristics and controls for individual heterogeneity through fixed effects regressions. Given the established link between repeated exposure to stress and ill health in the medical literature, these results suggest a potentially significant negative attribute of PRP payment systems.

1 Literature review

If working for PRP contracts can lead to poor health, there are a number of potential pathways through which it might do so. One such pathway is that PRP may encourage risk-taking at work and thereby increase the number of work-related accidents (Johansson, Rask, and Stenberg, 2010). For example, Freeman and Kleiner (2005) find that workers' compensation insurance premiums decreased for employees after their pay changed from PRP to salaries, suggesting that accidents decreased as the incentive to work harder and/or faster was removed. Similarly, Saha et al. (2004) find that workers paid by PRP suffered from higher rates of workplace accidents. Consistent with the findings from both industry-specific studies, Bender, Green, and Heywood (2012) analyse a large-scale survey of over 30,000 workers in the EU establishing an association between injuries and piece rates. Furthermore, the same relationship is found in the US, even after controlling for individual and individual-employer fixed effects (Artz and Heywood, 2015).

However, there is also evidence for PRP having a more general effect on health than just increasing the rate of injuries at work. For example, Bender and Theodossiou (2014) find that those who spend longer time in employment contracts paid by PRP are more likely to suffer from poor health, heart problems, stomach problems and poor mental health. A second potential pathway is that PRP makes explicit the opportunity cost of working, specifically, the reduced opportunity to engage in healthy or restorative behaviours. Bender and Theodossiou (2014, p. 838) find that workers paid by PRP work, on average, 1.5 hours more per week than those not paid by PRP. Thus, individuals who are paid by performance are likely to spend more time at work, at the expense of free time which can otherwise be spent on healthy leisure activities, such as exercise or relaxing activities (Mullahy and Robert, 2010). A third explanation for the link between PRP and health is increased engagement in behaviour that is detrimental to one's health. For example, Bender and Theodossiou (2014) report a correlation between PRP and heavy drinking and Artz, Green and Heywood (2020) find that alcohol/drug use is more common in PRP workers than those in standard contracts. Finally, it is possible that poor health is the consequence of PRP contracts causing persistent, increased stress (as suggested by some evidence presented by Bender and Theodossiou, 2014), exacerbated by the increased time pressure and variability in income that is inherent in such payment schemes. Although the human body is well-adapted to and able to recover from brief episodes of stress, a review by Rohleder (2014) on chronic stress finds that long-term stress lead to psychological and physiological damage through higher 'allostatic load', essentially cumulative wear and tear on the body, that eventually compromises the immune system (McEwen, 1998). People who work in PRP contracts for longer and thereby experience more persistent stress over time, may therefore be at higher risk of suffering from poor physical and mental health as a consequence of chronic stress. Although our study looks at PRP as a one-off stress episode, it is likely that repeated exposure to that stressor over time would equate to chronic stress.

Although the above studies suggest reasons that may explain the relationship between PRP and health, it cannot be assumed that PRP is a causal factor. Indeed, Eriksson (2012) argues it is possible that individuals with poor health are more likely to work in jobs which utilise PRP contracts. For example, those with a propensity for worse health may struggle to hold a job with a regular contract, causing them to rely on work in the PRP sector. On the other hand, those who do not perceive stress as a problem because they are relatively risk-loving may self-select PRP jobs which allow them to generate higher earnings. The existing empirical literature on PRP and health outcomes does not allow us to reject any of these theories since there is no explicit control for the endogeneity or sample selection of the choice of PRP-related employment. At best, researchers use statistical methods to try to control for the issues of endogeneity and sample selection using panel data or instrumental variables (see for example Artz, Green and Heywood, 2020; Bender and Theodossiou, 2014). Since

these have well-known potential problems with time-invariance and orthogonality, respectively, one is never sure that the causal mechanism is being identified.

Instead, to disentangle the causal direction of the link between PRP and stress it is necessary to randomly allocate individuals into different types of job contract and compare stress levels between the different groups – a situation that cannot be readily done in real life circumstances of the labour market. Thus, randomised job contracts require an experimental design, but to date, there are few experimental studies that examine the association between PRP and stress. In one of two notable exceptions, Dohmen and Falk (2011) let their subjects solve calculations that were incentivised either by PRP or by standard contracts to study the effect of job contracts on performance. As a secondary topic of interest, they also measure self-reported stress at the end of the experiment. They report that subjects in self-selected PRP contracts are more likely to report higher levels of stress and exhaustion than those in standard contracts. In a similar experimental study, Cadsby, Song and Tapon (2016) find that risk-averse subjects report higher levels of stress in self-selected PRP conditions than the standard contract condition.

Both experiments suggest that PRP contracts generate higher levels of stress in comparison to standard contracts. However, as health is not the main focus of either of the papers, stress is not measured prior to the experiment and so pre-existing group differences cannot be ruled out. In the Dohmen and Falk study, subjects self-select their payment contract which may mitigate some of the stress felt by subjects. Furthermore, in both experiments the fixed payment contract awards the same pay regardless of whether the subject completes some or no calculations, meaning that subjects can earn their participation fee without any effort. Finally, both studies measure stress through self-report.

Another challenge in these studies is the use of subjective self-reported stress. Although self-reported stress is necessary to capture stress appraisals (Shields and Slavich, 2017), it is also subject to biases such as social desirability, confirmation and recall bias (Paulhus and Vazire, 2007). An alternative, often used in psychological and medical research, is to objectively capture the stress response using physiological measures, such as measuring the body's production of the hormone, cortisol.

During stressful episodes, the body typically displays a 'fight-or-flight response' which has previously provided evolutionary advantages for survival (Gleitman, Fridlund, and Reisberg, 2004). The fight-or-flight response is manifested through a range of physiological changes, including changes in heart rate, sweat and digestive processes. One of the most central components in regulating stress is the hypothalamic-pituitary-adrenocortical (HPA) axis, which is activated both as a response to stress and with the purpose to regulate the impact of stress on the body (also referred to as the allostatic process in McEwen, 1998). The end-product of HPA axis activation is cortisol: Cortisol is an anti-inflammatory hormone and chronic exposure of cortisol is generally considered to have a down-regulatory effect on the immune system (Sapolsky, Romero, and Munck, 2000). As it can be measured through non-invasive tests such as saliva samples, cortisol is a commonly used physiological measure in stress research (Nicolson, 2008). Finally, collection of salivary cortisol is easily performed in an experimental setting when the timing of cortisol sampling can be consistently applied to ensure comparability between experimental subjects.

To summarise, while a number of studies have observed a strong and persistent PRP-health link, the nature of the observational data means that they are not able to fully control for selection, endogeneity and individual heterogeneity. It is therefore very difficult to establish a causal relationship between PRP and stress unless an experimental design is implemented. The extant experimental literature has primarily focused on productivity, and until recently there has been an absence of

randomised contract allocation as well as stress measures that are administered both before and after the work task and that are not reliant on only self-report. More recently, Allan, Bender and Theodossiou (2020) undertook a small-scale study in which subjects are randomly sorted into either a PRP contract, in which they are paid by piece rate, or a minimum performance contract, in which they are paid a fixed fee if they have met a set target. In line with the other experiments, the above study includes a self-reported stress measure which fails to find any significant differences between the PRP and the minimum performance contract group. However, to circumvent the issues of self-report measures, a physiological measure of stress, salivary cortisol, is also included. By measuring cortisol before and after the work task, the authors are able to identify the significant impact of pay contracts on acute stress.

The current study aims to build on this experiment by employing a crossover design (where each subject completes both a PRP and a fixed pay task in different sessions, in counter balanced order[1]), using a larger sample, and by measuring cortisol at four separate time points during the experiment. As in Allan, Bender and Theodossiou (2020) study, participants are randomly allocated into a payment contract to control for endogeneity. The most significant and important difference with the Allan, Bender and Theodossiou (2020) study, however, is the utilisation of a counter balanced crossover design, where the same individuals can be compared in the treatment and nontreatment group, netting out individual factors such as mathematical ability (Ratkowsky, Evans, and Alldredge, 1992) and other time invariant factors that may be correlated with stress. A larger sample size has the advantage of increasing power and thereby reducing the chance of type II errors and the ability to control for covariates through the use of regressions during analysis. Finally, although Allan, Bender and Theodossiou (2020) measure cortisol twice, increasing the number of cortisol measurement to four times in the current study allows for more continuous tracking of stress. After implementing these changes, the current study finds that there is an association between PRP and acute, physiological stress. Although previous studies have found an association between PRP and self-reported stress, the current study is the first to do so whilst using within-person tests and controlling for self-report bias, providing evidence for a causal link between performance-related pay and poor health outcomes.

2 Design of the experiment

The current study employs a crossover design, in which subjects are randomly assigned to either a PRP treatment or a minimum performance treatment in their initial session and participate in the alternative treatment a week later. Thus, each individual serves as their own control, allowing for a netting out of individual heterogeneity in the statistical analysis. In line with Dohmen and Falk (2011), subjects are asked to complete mathematical calculations by hand for ten minutes. All subjects are paid a £5 flat show-up fee for participating in the experiment regardless of their performance. In addition to the show-up fee, those in the PRP treatment are paid £0.20 per correct calculation, whereas subjects in the minimum performance (nonPRP) treatment are paid an additional fixed rate of £5 if they are able to complete 10 calculations correctly within the allotted time. All calculations are entered into a computer through the experiment programmed in z-Tree (Fischbacher, 2007).[2] The

[1] Although the current study randomly allocates volunteers to the treatments, workers often self-select into payment contracts which may mitigate their stress levels (though the survey data mentioned above suggests that stress and ill health are not completely mitigated). However, there are instances where PRP is exogenously implemented. For example, both Freeman and Kleiner (2005) and Dahl and Pierce (2019) describe firms that switch from PRP to nonPRP and nonPRP to PRP contracts, respectively.

[2] Screenshots of selected z-Tree screens are given in the attached supplemental material.

experiment requires four cortisol samples from each subject. The samples are obtained by subjects chewing on a synthetic swab for 60 seconds at 4 set time points (before the work task, immediately after the work task, 10 mins after the work task, and 20 mins after the work task) and then placing it in a labelled test tube. All subjects are students at the (university name) and are recruited through the internal study recruitment platform, ORSEE. The study has been reviewed and approved by the (ethics board).

One hundred and fifty-five subjects attended one of 17 sessions that took place during term-time. Out of these, 144 subjects returned for the follow-up session to complete the crossover. To control for the diurnal pattern of cortisol, all experiments took place at 2pm during weekdays. Subjects were asked to refrain from eating or drinking, caffeine, cigarettes or strenuous physical activity for the two hours immediately before the experiment. Upon arriving to the lab, subjects were given information and written consent was taken prior to being randomly assigned to one of the 20 possible lab PCs. Subjects were only informed of the payment of the participation fee and the possibility of earning up to an additional £10 and were not given any advance information about the pay contracts or the crossover design. The crossover session happened a week after the initial session.

The experiment was split into four 10-minute phases (see Figure 1) to allow sufficient time for changes in cortisol to occur (Kirschbaum, Pirke, and Hellhammer, 1993). Phase 1 consisted of 10 minutes of relaxation, during which subjects were provided with colouring-in sheets designed for mindfulness, followed by the first (baseline) cortisol sample and completion of the GHQ-12 (GHQ; Golderberg and Wilson, 1988) asking about general stress, exhaustion, loss of sleep, strain, overcoming difficulties etc.

During phase 2, the subjects completed practice questions and the real calculation task. The practice task consisted of four calculations with unlimited time. As seen in Figure 1, subjects were then randomly assigned by the computer to either the PRP or the minimum performance treatment and informed of their payment contract on the screen. The work task lasted for 10 minutes and consisted of up to 50 calculations. Although the task was based on the task used in the study by Allan, Bender and Theodossiou (2020), the ratio of division calculations to the total calculations was changed to increase the range of performance among subjects. Hence, the task used in the current study consisted of six addition calculations, five subtraction calculations, 20 multiplication calculations and 19 division calculations in a mixed order. Although the type of calculation was fixed across subjects, the values for each calculation were always randomised so that the correct answer could range anywhere between 0-1000 for each individual task (with the exception of addition which could range from 0-2000) and varied across subjects. The number of seconds left for calculations was visible in the top corner of the screen, and subjects were allowed to use scratch paper but not calculators. Once a minimum performance contract subject achieved her or his ten correct answers, a banner appeared on the screen informing them that they had satisfied their contract and that they would receive their fixed payment of £5, but were free to continue solving questions for the remainder of the time if they wished. After the ten minutes, subjects received their results and were then asked to complete a second cortisol sample and stress survey (see Dohmen and Falk, 2011), consisting of the following four items rated on a five-point scale (ranging from “not at all” to “very” or “great”):

- ‘After the task, how stressed do you feel?’
- ‘After the task, how exhausted do you feel?’
- ‘How much effort did you exert solving the mathematical problems in the previous 10 minutes?’
- ‘Did you feel under strain when solving the mathematical problems in the previous 10 minutes?’

Phase 3 and 4 both consisted of 10 minutes of relaxation time during which subjects could continue completion of their colouring-in sheet, followed by a third and fourth cortisol sample, respectively

(see Figure 1). This concluded the initial session of the study, and subjects were paid by the experimenter in a separate room. The follow-up session a week later followed the same procedure with the exception of treatment randomisation. Instead, subjects were manually allocated to the treatment that they had not yet participated in. Coming into the session, subjects did not know into what payment scheme they would be allocated.

All cortisol samples were labelled with unique identifiers and frozen within an hour of completing the experiment. After the final experimental session, all samples were packed and shipped to a commercial laboratory for analysis of cortisol levels.

3.1 Descriptive statistics

Table 1 contains the descriptive statistics of the subjects. The first column includes all subjects who participated in the first session, while the second column includes those who returned for the follow-up crossover session (11 of the 155 did not return) and the third column is based on the crossover sample after removal of cortisol outliers (discussed in the following paragraph). The average subject was more likely to be female, in their first year of studies, younger than 23 years of age, and from arts and social sciences disciplines.

One potential concern is whether there are systematic sample differences between the eleven subjects who did not return for their second session and the final study sample which might generate a sample selection bias into the results. Comparing the 'First Sample' and 'Crossover Sample' columns there are no big differences between the two samples. The only observable difference seems to be that it was males who were more likely not to return, but the rest of the observable characteristics are similar.[3] In line with previous research on cortisol, seven outlier observations with cortisol levels that exceeded four standard deviations from the mean were removed from the data (Nicolson, 2008). Again, except for fewer male subjects, there seems to be little difference for the sample without outliers in observable characteristics. Given this, the results presented below are based on the 137 subject sample after the outliers are removed. Finally, as there are several confounders that may affect cortisol levels, including recent food intake, caffeine, nicotine, time of awakening and medication (Kudielka, Hellhammer, and Wüst, 2009) any subjects who disclosed such activities were coded with a dummy variable to allow us to control for any of these situations in the analysis.

3.2 Subjective stress

Pre-task stress. The average GHQ score for the nonPRP group across the 12 measures is 2.11 (with a standard deviation of 0.48) while for the PRP group it is also 2.13 (0.49). A paired t-test (to account for the fact that each subject undergoes both the PRP and nonPRP treatment) does not show any statistically significant difference between the two groups in the stress level before the task ($p = .182$).

[3] As 3 out of the 11 subjects were nonPRP subjects who did not meet the minimum target of 10 correct answers, to ensure that the non-returning subjects were not more objectively stressed than the returning subjects, we compared cortisol levels from the initial sessions only between those that eventually returned for the second session and those that did not. The t-tests (see Table A.1 in Appendix A) did not find any significant differences at any of the sampling time points, suggesting that those who did not return were not significantly more stressed than the sample used for analysis. Indeed, although the sample is small, the direction of difference suggests that those who did not return had lower cortisol levels and were potentially less stressed than those who did return.

Post-task stress. Out of 137 respondents, 130 completed the post-task survey for both sessions (with the exception of the exhaustion item which had 125 completions). Table 2 reports the average responses by the two groups as well as the paired t-test results. In each case, there is a statistically significant difference in the paired t-test between the PRP and nonPRP group, with subjects in the PRP treatment expressing statistically significantly higher stress, exhaustion, effort or strain. All of these differences remain significant when controlling for a series of standard covariates (available from the authors upon request).

Further mitigating factors. While these t-statistics are suggestive of the increased stress of allocating subjects to the PRP treatment, there are several potential mitigating factors which are useful to investigate. First, there is a clear order effect. Post-task stress was higher in general in the initial session for both the PRP and nonPRP group (although it was significantly higher for the PRP group) and then falls for both groups in the second session when they are exposed to the other treatment. Figure 2 shows the average post-task stress score. Initially, the nonPRP subjects in the first session have an average stress score of 3.01. However, when they return for the second session, when they have the PRP treatment, their stress increases slightly to 3.03 (though the increase is not statistically significant ($t(72)=-0.21$, with a p-value of 0.834). When the first session subjects are allocated to the PRP group, they have an average stress score of 3.49, but this drops significantly when they are given the nonPRP treatment in the second session where stress falls to 2.33, a change that is statistically significant ($t(56)=5.92$, with a p-value<0.001). It is possible that individuals who have experienced PRP are more relaxed in the following nonPRP treatment. In contrast, those who have experienced nonPRP are unfamiliar with the stress of PRP, cancelling out the decrease in stress which might otherwise occur.

In addition to an order effect, task performance is an interesting variable. The literature on PRP and performance suggests that PRP leads to higher productivity and therefore higher performance (Lazear, 2000; Cadsby, Song, and Tapon, 2016). Overall on average, subjects answered 17.42 questions correctly. Out of the 137 subjects, 122 (89%) met the minimum target of 10 correct questions when participating in the nonPRP condition. This is comparable to the PRP condition where 117 out of 137 (85%) of subjects correctly answered a minimum of 10 questions, albeit without a minimum performance threshold to mark this. As expected, performance for the PRP group was higher at 18.89 compared to 16.33 for the nonPRP group, a statistically significant difference ($t(110)=3.64$ with a p-value<0.001)[4]. At the outset, it seems likely that higher performance would be associated with higher levels of stress due to the effort exerted but this may be mitigated by stress having a negative effect on performance. Indeed, there is an overall negative correlation ($r=-0.21$) between performance and reported post-task stress level which reaches statistical significance ($t(215)=-3.09$ with a p-value=0.002). However, this overall correlation hides heterogeneity between the two groups. Figure 3 shows a scatterplot of PRP and nonPRP subjects with a linear correlation line by group added in. Perhaps as expected, there is a strong negative correlation between performance and stress for the nonPRP group ($r=-0.39$, $t(107)=-4.37$ with a p-value<0.001) who risk getting no pay if they do not meet the minimum target. Meanwhile there is no statistically significant correlation with the PRP group ($r=-0.10$, $t(106)=-1.00$ with a p-value=0.320), who do not have a minimum target albeit their performance will dictate the amount they earn. Furthermore, these correlations are statistically different as a

[4] Although lower than in the PRP condition, the performance for nonPRP participants is still notably higher than the target of 10 correct questions. One possible explanation for this is intrinsic motivation; Some participants may be intrinsically motivated to continue solving problems even after achieving the minimum performance target. However, it is not clear how this would affect stress levels. If solving problems is effortful, stress levels may increase. This would bias our results to no effect.

correlation comparison test (Steiger, 1980) gives a χ^2 statistic of 4.63 with 1 degree of freedom, which gives a p-value of 0.031. It seems then that our findings are consistent with the negative effect of stress on performance, although the relationship is not as strong as one might think.

3.3 Cortisol results

Variable treatment. As previously outlined, subjects provided four samples of salivary cortisol as an objective measure of the stress response throughout the experiment: a baseline sample and three post-task samples (post-task, +10 min, +20 min).[5] Higher levels of cortisol as measured in nmol/l indicate higher levels of stress. As cortisol levels naturally vary between individuals, the focus of this analysis is on the change from baseline to post-task cortisol levels within individuals (Kudielka, Hellhammer, and Wüst, 2009). Furthermore, the natural variation of cortisol and individual stress resilience may cause cortisol to peak at different rates. Consequently, three dynamic measures of change were calculated from the cortisol pattern and included in the analysis as a robustness check. These include 'Area Under the Curve with respect to increase' (AUCi – a measure of overall reactivity), peak change in cortisol (the difference between the peak value out of the three post-task samples and the baseline sample) and overall change in cortisol (the difference between the final +20 min and the baseline sample). As these variables measure change in cortisol from baseline, a positive value indicates an increase of cortisol post-task, whereas a negative value indicates a decrease in cortisol post-task. In addition to this, 'AUC with respect to ground' (AUCg - a measure of overall cortisol output) was also calculated.

Simple comparisons. Initially, paired t-tests compared the level of cortisol for each sampling point separately across the two treatments in Figure 4. None of the differences between the groups were statistically significant in these parametric tests nor in two-way within-group ANOVAs (see Supplementary Material for full description of the tests available from the authors).[6] It is also notable that in line with the results in Allan, Bender and Theodossiou (2020), the highest level of cortisol can generally be seen in the baseline sample.

The test of whether PRP affects cortisol comes from analysing the change rather than comparing average levels, however. Therefore, paired t-tests are used on the four outcome variables (Table 3). As expected, there is a significant difference in AUCi as PRP subjects (-12.19) show higher average levels of reactivity than nonPRP subjects (-22.83), t(136)=2.53, with a p-value=0.006. PRP also shows a slight increase in cortisol between peak and baseline (0.07) whereas the nonPRP treatment shows a decrease (-0.30), t(136)=2.33, with a p-value=0.011. Although both treatments show an overall decline of cortisol throughout the experiment (+20 min sample – baseline sample), the PRP group show a smaller decline (-0.75) than the nonPRP group (-1.26), t(136)=2.71, with a p-value=0.004. In contrast, there is no statistically significant difference between PRP (160.91) and nonPRP subjects (160.43) when comparing average AUCg, t(136)=0.27, with a p-value=0.396. These results suggest that although there are no group differences in overall cortisol output (as measured by AUCg), individuals in the PRP treatment show lower rates of cortisol recovery in comparison to when they are experiencing the nonPRP treatment for the other three measures of cortisol change.

[5] The cortisol samples demonstrated appropriate levels of variation as established by the average intra-assay coefficient of 4.33%. This is well within the acceptable cut-off point at <10% (Schultheiss and Stanton, 2009).
[6] Nonparametric tests of differences were also estimated, but none were statistically significant. These results are available from the corresponding author.

Mitigating factors. As with the subjective stress findings, these results suggest that there are differences between the PRP and nonPRP treatments. However, these analyses do not consider external variables which may affect cortisol levels. Indeed, independent t-tests revealed that the group with self-disclosed confounders showed higher levels of cortisol at each sampling point (Table 4). It is clear that confounding variables need to be considered when comparing group differences.

As all subjects took part in two sessions in total but in a randomised order of treatment, it is possible to examine whether order of the treatment matters. Two-way crossover design ANOVAs[7] found that treatment has a significant effect on AUCi, peak change and overall change, but not on AUCg (Table 5). In contrast, AUCg is the only cortisol measurement which is affected by the order of treatment. None of the variables showed a significant carryover effect. Following this, the sample was split by initial session allocation, allowing us to see an order effect although the effect is somewhat different to those found in self-report. As seen in Table 6, subjects who are allocated to the nonPRP treatment in their first session show a significant increase in stress when returning for their second session as measured through AUCi, peak change and overall change, but not AUCg. In contrast, subjects who are allocated to the PRP treatment first are less stressed in their second session, but the effect is never statistically significant (Table 7).

Previous research has found significant associations between PRP and productivity (Dohmen and Falk, 2011). However, despite the significant association between self-reported stress and performance, performance was only weakly related to AUCi (r=-0.02, with a p-value=0.685), AUCg (r=-0.06, with a p-value=0.285), peak change (r=-0.03, with a p-value=0.648) and overall change (r=-0.01, with a p-value=0.925). Unsurprisingly, none of these correlations were statistically significant and remain not significant even when examining the correlations by each treatment separately.

Regressions. The findings in the previous section suggest that there are variables that need to be considered when estimating the effect of performance pay on stress. In addition to these, there are also socio-demographic covariates which may affect stress responses. For this reason, we estimated several regressions to see if the effect of pay contract on stress remained whilst controlling for order effects, confounding variables and socio-demographic covariates (age, gender, year and subject of study), clustering standard deviations at the subject level.[8] In addition, observing all subjects twice allows us to control for individual heterogeneity since each subject serves as their own control. Indeed, this is one of the key reasons for using this method. Therefore, we also estimated fixed effects regressions. Since most of the variables (PRP in the first session, gender, year, discipline and age) are fixed, they drop out of the analysis, leaving the only three variables that can vary – PRP, second session and whether confounders were not present. Table 8 provides the result of the regressions on AUCi, AUCg, peak change or overall change in cortisol respectively.[9] For each cortisol change measure, the first column reports the results estimated by an OLS regression controlling for all covariates, while the

[7] The crossover-design ANOVAs were analysed using the '`pkcross`' command in Stata 15.1, that allows testing for carryover and period effects in crossover studies.

[8] Performance, as measured by the number of correct answers, was considered as a covariate. However, it is likely that it is endogenous in the regression. While increased performance should generate effort and stress, it may be that stress could impact performance. We do not include a performance measure in the regression results below, but including a performance measure in the regressions do not make much of a qualitative difference in the results. These are available from the authors.

[9] Regressions were also estimated for first and second session separately. Although there were differences when estimating the first session, the second sessions showed qualitatively the same findings as the combined sample. For the sake of brevity these results have not been presented in the current paper but can be requested from the corresponding author.

second column contains the estimates a fixed effects regression controlling for individual heterogeneity. The standard errors in all regressions are clustered at the individual level.

The first two columns of Table 8 show the regression models predicting AUCi. In line with our simple group comparisons, PRP is a significant predictor of higher AUCi (9.64), with year "other" and not disclosing a confounder also raising AUCi by 23.50 and 34.16, respectively, compared to their excluded variables. Fourth year is also a significant positive predictor of AUCi at the 0.10 level, compared to being a first year student. The fixed effects regression confirms the significance of PRP and the disclosed confounders whilst controlling for individual heterogeneity. Although the decrease in cortisol reactivity for those with confounders may seem at odds with the significantly higher levels of cortisol for this group, this is most likely because confounding activities, e.g. caffeine, disproportionately inflate the initial sample and cause a steeper decrease in cortisol as the effects of the confounder gradually wears off.

Regression models predicting AUCg can be seen in the next two columns of Table 8. Again, in keeping with the simple comparisons, the OLS regression does not find that being in the PRP contract is a significant predictor (3.81). However, subjects who have confounders and are male have significantly higher levels of overall cortisol output. Other significant predictors are 2nd year and year "other" both of which are significant predictors of less cortisol output compared to first year subjects. Finally, all subjects of study are significant predictors of more cortisol output in comparison to the reference variable. These findings are largely replicated in the fixed effects regression, where the no confounder dummy is the only significant predictor.

Next, the estimated regressions predicting the peak change variable are reported. The OLS regression finds that having the PRP contract (0.33), not disclosing a confounder (1.13), degree in other sciences (0.70), fourth year (0.54) and year "other" (0.86) are significant positive predictors of cortisol change, suggesting an increase of cortisol post-task. Again, PRP contract and no confounder are significant predictors in the fixed effects regression, albeit the no confounder variable is only significant at a 0.10 level.

The final outcome variable is overall change between the +20 min sample and the baseline sample in the final two columns of Table 8. PRP, respondents without confounding variables, 4th year and year "other" are predictors of a positive change between +20 min and baseline. Age group 21-23 generates a negative change, albeit at a 0.10 significance level. As before, the results of the fixed regression reports that PRP and not reporting a confounder are predictors of a positive change in cortisol levels.

In summary, the results show that although there are covariates that impact on cortisol, the effect of PRP remains significant. Comparing the point estimates to the average cortisol level in the sample, the estimated treatment difference is 7.6-10.3% of the average baseline cortisol levels after controlling for covariates depending on the cortisol change measure (e.g. 'peak' change or 'overall change')[10]. To put this into comparison, Badrick, Krischbaum and Kumari (2007) find that the effect of smoking increased cortisol by 13.1%. Although one might expect a larger treatment effect in the initial session when nothing was known about the experiment, the lack of order effects (i.e. that the difference remains significant across both sessions) suggests that being paid by PRP has an effect on stress regardless of whether it is encountered as part of a novel (first) or more familiar (second) task. Furthermore, the similarities of the PRP estimate between the OLS and the fixed effects regressions suggest that the relationship between PRP and stress is robust even when other variables or individual

[10] The magnitude of this difference is comparable to the self-reported treatment difference (17.2% of the average self-reported stress) in the current paper.

time-invariant heterogeneity are omitted. For example, if there was a bias in the OLS regression, we would be able to conclude that individual factors such as ability, or risk preference, may have an impact on the relationship. By estimating a fixed effects regression, we are able to test for this. This does not rule out the influence that time-invariant factors may have on stress in general. Instead, our results suggest that they are unlikely to affect the relationship between PRP and stress.

4 Discussion and future work

The aim of the current study is to examine the link between PRP and stress using an experimental paradigm. The benefit of using this design is to enable the identification of the causal mechanism between PRP and stress since survey data will be subject to sample-selection biases making samples nonrandom and complicating the estimation of the relationship between PRP and stress. The study expands on previous research by using a randomised crossover design and therefore controlling for self-selection and individual differences. In line with previous research by Cadsby, Song, and Tapon (2016) and Dohmen and Falk (2011), the study finds that individuals in a PRP treatment reported significantly higher levels of stress than those in a fixed payment treatment even after controlling for individual effects and implementing a minimum performance target for the nonPRP subjects. While this differs from the findings by Allan, Bender and Theodossiou (2020) - who find a similar trend in subjective stress, but the difference does not reach statistical difference - this likely reflects the relatively low power of the small sample employed by the authors to detect small changes. It is also possible that the changes made to the work task in the current study increased the stress felt by subjects in the present study compared to the earlier study. However, the minimum performance contract was substantially harder to achieve in the current study than in the earlier study. Therefore, any increase in overall stress from the work task is likely to underestimate the difference between the PRP and the minimum performance treatment.

While these findings indicate that there is indeed an effect of pay contract on self-reported stress, the key contribution of the study is the measurement of cortisol, which allows us to capture an additional facet of stress, across different payment contracts. It extends Allan, Bender, and Theodossiou (2020) by measuring it before and three times after the task in ten minute intervals, providing more time for cortisol to peak as a response to the stressor as well as implementing a crossover strategy. By using a crossover strategy, it is possible to control for the natural variation in cortisol response between subjects. For all three measures of cortisol change, there were significant differences in the pattern of change between the two treatments. In line with our findings on subjective stress, these results suggest that being paid by performance decreases the rate of physiological recovery from stress in comparison to fixed pay.

To our knowledge, this is the first study of PRP to use a physiological measure of acute stress to examine the effect of payment contracts on stress. This is arguably a more persuasive piece of evidence for the link between PRP and stress than the subjective measures used to date. However, it does have some limitations that set the stage for future research. Perhaps most importantly, it suffers from common limitations of experimental methods such as having a very limited time frame and low stakes. Since each of these would bias the results towards finding limited effects of PRP on stress, that we find consistent evidence that PRP generates a difference in cortisol levels, suggests a causal effect of PRP on stress. The dynamic aspect of labour markets is also an interesting area of future research and the medical evidence discussed earlier about the negative health effects of constant exposure to even low-grade stress would be evident in PRP jobs, but both are very hard to replicate in an experimental setting. In addition to experimental limitations, job stress can be influenced by a

range of factors which were beyond the scope of the current study. For example, a key part of this paper is to implement randomisation to establish causation, but future research could implement sorting to examine whether it mitigates or exacerbates stress. It is also not clear if intrinsic and extrinsic motivation has different effects on stress depending on the payment contract. If nonPRP subjects are motivated by intrinsic motivation to solve more problems, and thereby experience more stress due to effort exerted, it is likely to narrow the differences between PRP and nonPRP groups. If so, the "true" effect of PRP on stress may be stronger than seen in our sample. This may be a fruitful avenue for future research. Furthermore, there are different kinds of PRP contracts and the current study focused on a task that most closely resembles a piece rate contract. Although piece rate contracts are more common than other types of PRP employment (13%, Bender, Green and Heywood, 2012, p. 575), future research could examine the relationship whilst using a variety of work tasks or PRP schemes. Stress may also be exacerbated by factors such as gender dynamics and job security, and future research could expand the sampling pool to a more heterogeneous workforce to see if the results remain consistent across different groups of workers. Nevertheless, the current study was able to establish an effect of PRP contracts when all other variables are held equal.

Overall, while the productivity incentives of PRP contracts are of great interest to employers, this study suggests that there is a direct causal effect of PRP on both subjective and physiological measures of stress, which may in turn influence the health of workers. This suggests a clear public health dimension to PRP jobs and suggests that either limiting their use or developing policies to mitigate the stress effects of PRP might decrease the public health impacts of stress at work.

5 References

Allan, Julia L., Keith A. Bender, and Ioannis Theodossiou. 2020. “Performance Pay and Low Grade Stress: An Experimental Study.” *Work: A Journal of Prevention, Assessment & Rehabilitation* 67, no 2: 449-457.

Artz, Benjamin, Colin P. Green, and John S. Heywood. 2020. "Does performance pay increase alcohol and drug use?" *Journal of Population Economics*: 1-34. https://doi.org/10.1007/s00148-020-00776-4

Artz, Benjamin, and John S. Heywood. 2015. "Performance pay and workplace injury: panel evidence." *Economica* 82: 1241-1260. https://doi.org/10.1111/ecca.12153

Badrick, Ellena, Clemens Kirschbaum, and Meena Kumari. "The relationship between smoking status and cortisol secretion." *The Journal of Clinical Endocrinology & Metabolism* 92, no. 3 (2007): 819-824. https://doi.org/10.1210/jc.2006-2155

Bender, Keith A., Colin P. Green, and John S. Heywood. 2012. "Piece rates and workplace injury: Does survey evidence support Adam Smith?." *Journal of Population Economics* 25, no. 2: 569-590. https://doi.org/10.2139/ssrn.1632260

Bender, Keith A., and Ioannis Theodossiou. 2014. “The unintended consequences of the rat race: the detrimental effects of performance pay on health.” *Oxford Economic Papers* 66, no 3: 824-847. https://doi.org/10.1093/oep/gpt032

Booth, Alison L., and Jeff Frank. 1999. "Earnings, productivity, and performance-related pay." *Journal of Labor Economics* 17, no. 3: 447-463. https://doi.org/10.1086/209927

Bryson, Alex, Richard B. Freeman, Claudio Lucifora, Michele Pellizzari, and Virginie Pérotin. 2013. "Paying for performance: incentive pay schemes and employees’ financial participation." *Centre for Economic Performance Discussion Paper Series*: 122-123.

Cadsby, C. Bram, Fei Song, and Francis Tapon. 2016. "The Impact of Risk-Aversion and Stress on the Incentive Effect of Performance-Pay.” *Experiments in Organizational Economics* (Research in Experimental Economics) 19: 189-227. https://doi.org/10.1108/s0193-230620160000019007

Dahl, Michael S., and Lamar Pierce. 2020. "Pay-for-performance and employee mental health: Large sample evidence using employee prescription drug usage." *Academy of Management Discoveries* 6, no. 1: 12-38. https://doi.org/10.5465/amd.2018.0007

Davis, Mary E. 2016. "Pay matters: the piece rate and health in the developing world." *Annals of Global Health* 82, no. 5: 858-865. https://doi.org/10.1016/j.aogh.2016.05.005

DeLongis, Anita, James C. Coyne, Gayle Dakof, Susan Folkman, and Richard S. Lazarus. 1982. "Relationship of daily hassles, uplifts, and major life events to health status." *Health Psychology* 1, no. 2: 119. https://doi.org/10.1037/0278-6133.1.2.119

Dohmen, Thomas, and Armin Falk. 2011. "Performance pay and multidimensional sorting: Productivity, preferences, and gender." *American Economic Review* 101, no. 2: 556-90. https://doi.org/10.26481/umaror.2010003

Eriksson, Tor. 2012. "Healthy personnel policies." *International Journal of Manpower* 33, no. 3: 233-245. https://doi.org/10.1108/01437721211234138

Fischbacher, Urs. 2007. “Z-Tree: Zurich toolbox for ready-made economic experiments.” *Experimental Economics* 10, no 2: 171-178. https://doi.org/10.1007/s10683-006-9159-4

Freeman, Richard B., and Morris M. Kleiner. 2005. "The last American shoe manufacturers: Decreasing productivity and increasing profits in the shift from piece rates to continuous flow production." *Industrial Relations: A Journal of Economy and Society* 44, no. 2: 307-330. https://doi.org/10.1111/j.0019-8676.2005.00385.x

Gielen, Anne C., Marcel JM Kerkhofs, and Jan C. Van Ours. 2010. "How performance related pay affects productivity and employment." *Journal of Population Economics* 23, no. 1: 291-301. https://doi.org/10.1007/s00148-009-0252-9

Gleitman, Henry, Alan J. Fridlund, and Daniel Reisberg. 2004. *Psychology*. New York: W. W. Norton & Company.

Golderberg, David, and Paul Wilson. 1988. *User's guide to the General Health Questionnaire*. Windsor, UK: NFER-Nelson.

Green, Colin P., and John S. Heywood. 2008. "Does performance pay increase job satisfaction?." *Economica* 75, no. 300: 710-728. https://doi.org/10.1111/j.1468-0335.2007.00649.x

HSE. 2020. “Table 3. Work-related Illness – Type of Illness.” Accessed August 5. https://www.hse.gov.uk/statistics/lfs/lfsilltyp.xlsx

Johansson, Bo, Kjell Rask, and Magnus Stenberg. 2010. "Piece rates and their effects on health and safety–A literature review." *Applied Ergonomics* 41, no. 4: 607-614. https://doi.org/10.1016/j.apergo.2009.12.020

Kirschbaum, Clemens, Karl-Martin Pirke, and Dirk H. Hellhammer. 1993. "The ‘Trier Social Stress Test’–a tool for investigating psychobiological stress responses in a laboratory setting." *Neuropsychobiology* 28, no. 1-2: 76-81. https://doi.org/10.1159/000119004

Kudielka, Brigitte M., Dirk H. Hellhammer, and Stefan Wüst. 2009. "Why do we respond so differently? Reviewing determinants of human salivary cortisol responses to challenge." *Psychoneuroendocrinology* 34, no. 1: 2-18. https://doi.org/10.1016/j.psyneuen.2008.10.004

Lazear, Edward P. 2000. "Performance pay and productivity." *American Economic Review* 90, no. 5: 1346-1361. https://doi.org/10.1257/aer.90.5.1346

McEwen, Bruce S. 1998. "Stress, adaptation, and disease: Allostasis and allostatic load." *Annals of the New York Academy of Sciences* 840, no. 1: 33-44. https://doi.org/10.1111/j.1749-6632.1998.tb09546.x

Mullahy, John, and Stephanie A. Robert. 2010. "No time to lose: Time constraints and physical activity in the production of health." *Review of Economics of the Household* 8, no. 4: 409-432. https://doi.org/10.1007/s11150-010-9091-4

Nicolson, Nancy. 2008. “Measurement of Cortisol.” In *Handbook of Physiological Research Methods in Health Psychology*, edited by Linda J. Luecken and Linda G. Gallo, 37-62. Thousand Oaks, CA: Sage Publications. https://doi.org/10.4135/9781412976244.n3

Parent, Daniel. 1999. "Methods of pay and earnings: A longitudinal analysis." *ILR Review* 53, no. 1: 71-86. https://doi.org/10.1177/001979399905300104

Paulhus, Delroy L., and Simine Vazire. 2007. “The self-report method.” *Handbook of Research Methods in Personality Psychology* 1: 224-239.

Ratkowsky, David, Richard Alldredge, and Marc A. Evans. 1992. *Cross-over experiments: design, analysis and application*. Vol. 135. Boca Raton, FL: CRC Press. https://doi.org/10.2307/2291237

Robertson, Ivan, and Cary Cooper. 2011. *Well-being: Productivity and happiness at work*. London: Palgrave Macmillan.

Rohleder, Nicolas. 2014. "Stimulation of systemic low-grade inflammation by psychosocial stress." *Psychosomatic Medicine* 76, no. 3: 181-189. https://doi.org/10.1097/psy.0000000000000049

Saha, Asim, Takiar Ramnath, Ramendra Narayan Chaudhuri, and Habibullah Nasrullah Saiyed. 2004. "An accident-risk assessment study of temporary piece rated workers." *Industrial Health* 42, no. 2: 240-245. https://doi.org/10.2486/indhealth.42.240

Sapolsky, Robert M., L. Michael Romero, and Allan U. Munck. 2000. "How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions." *Endocrine Reviews* 21, no. 1: 55-89. https://doi.org/10.1210/edrv.21.1.0389

Schultheiss, Oliver C., and Steven J. Stanton. 2009. "Assessment of salivary hormones." In *Methods in Social Neuroscience,* edited by Eddie Harmon-Jones and Jennifer S. Beer, 17-65. New York: Guilford Press.

Shields, Grant S., and George M. Slavich. 2017. "Lifetime stress exposure and health: A review of contemporary assessment methods and biological mechanisms." *Social and Personality Psychology Compass* 11, no. 8: e12335. https://doi.org/10.1111/spc3.12335

Shirom, Arie, Mina Westman, and Samuel Melamed. 1999. "The effects of pay systems on blue-collar employees' emotional distress: The mediating effects of objective and subjective work monotony." *Human Relations* 52, no. 8: 1077-1097. https://doi.org/10.1177/001872679905200805

Steiger, James H. 1980. "Tests for comparing elements of a correlation matrix." *Psychological Bulletin* 87, no. 2: 245. https://doi.org/10.1037/0033-2909.87.2.245

Figure 1. Experiment timeline

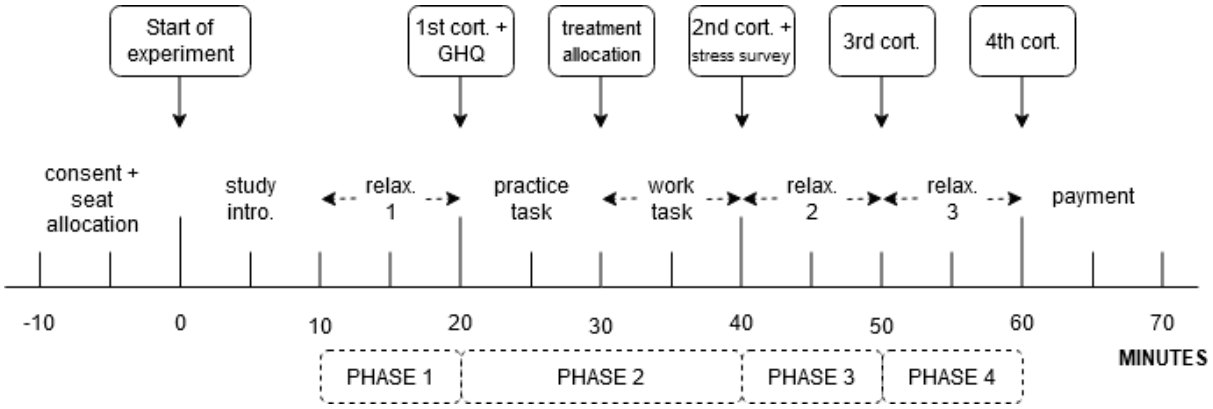

Table 1. Descriptive Statistics

Variable	First Sample	Crossover Sample	Final Sample
Female	52.2%	55.6%	56.0%
Male	47.7	44.4	44.5
1st Year	25.2	27.1	27.7
2nd Year	21.3	21.5	21.2
3rd Year	18.1	16.7	16.1
4th Year	23.9	23.6	23.4
Other Year	11.6	11.1	11.7
Age 18-20	42.6	41.7	41.6
Age 21-23	41.3	41.0	40.1
Age 24-26	6.5	6.9	7.3
Age 27-29	2.6	2.8	2.9
Age 30+	7.1	7.6	8.0
Business Disciplines	16.1	16.0	15.3
Arts and Social Sciences Disciplines	45.2	43.1	43.1
Life Sciences Disciplines	14.2	16.0	15.3
Other Science Disciplines	23.9	24.3	24.1
PRP (Initial Session)	45.8	44.4	43.8
nonPRP (Initial Session)	54.2	55.6	56.2
Number of Observations	155	144	137

Note: "First Sample" contains descriptives of all subjects taking part in the initial session, the "Crossover Sample" describes those who participated in the crossover session and "Final Sample" describes the sample after removing both outliers and non-returners.

Table 2. Post-Task Stress Responses

Measure	PRP Mean (standard deviation)	nonPRP Mean (standard deviation)	Paired t-Test (p-value)
Stress	3.24 (1.06)	2.73 (1.24)	t(129)=3.52 (0.001)
Exhaustion	2.60 (1.10)	2.31 (1.09)	t(124)=2.59 (0.011)
Effort	3.54 (0.90)	3.28 (1.01)	t(129)=2.81 (0.006)
Strain	3.62 (1.13)	3.21 (1.30)	t(129)=3.05 (0.003)

Note: Scores range from 1 to 5 where a higher score indicates higher levels of stress/exhaustion/effort/strain.

Figure 2. Order Effect

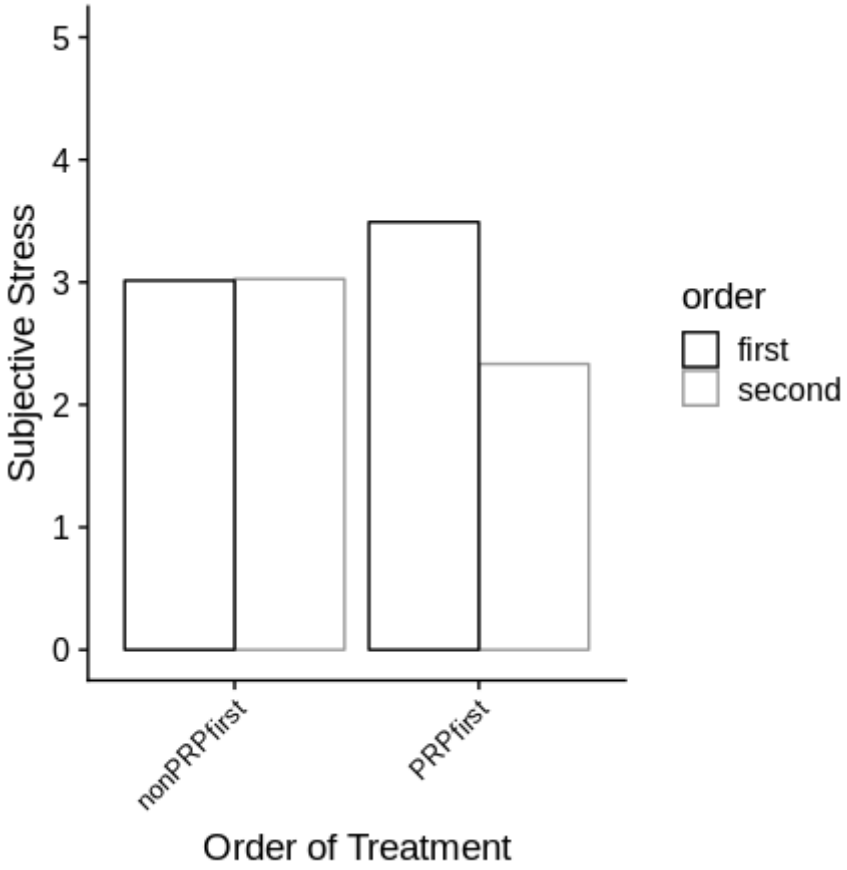

Figure 3. Correlation between Performance and Post-task Stress by Treatment

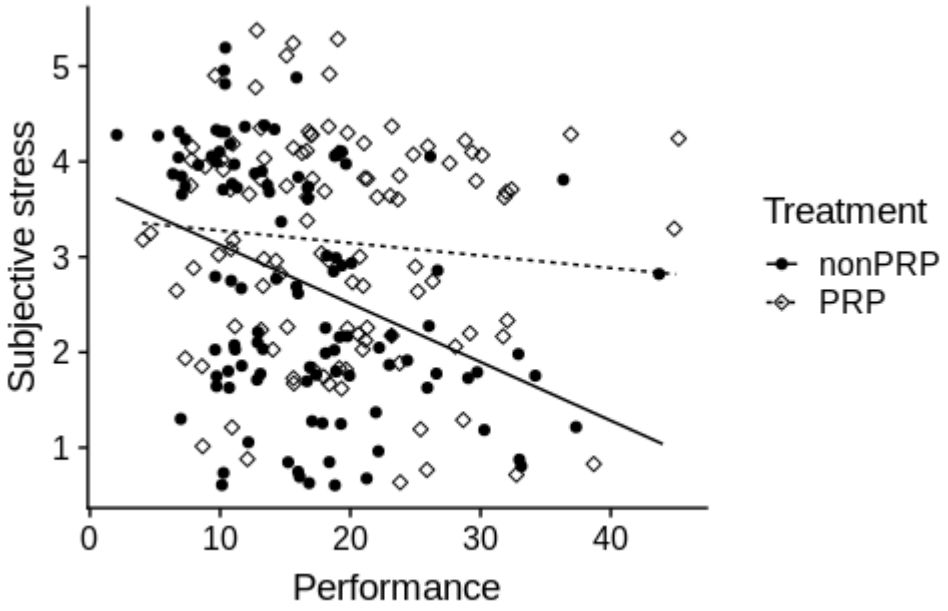

Figure 4. Cortisol Levels

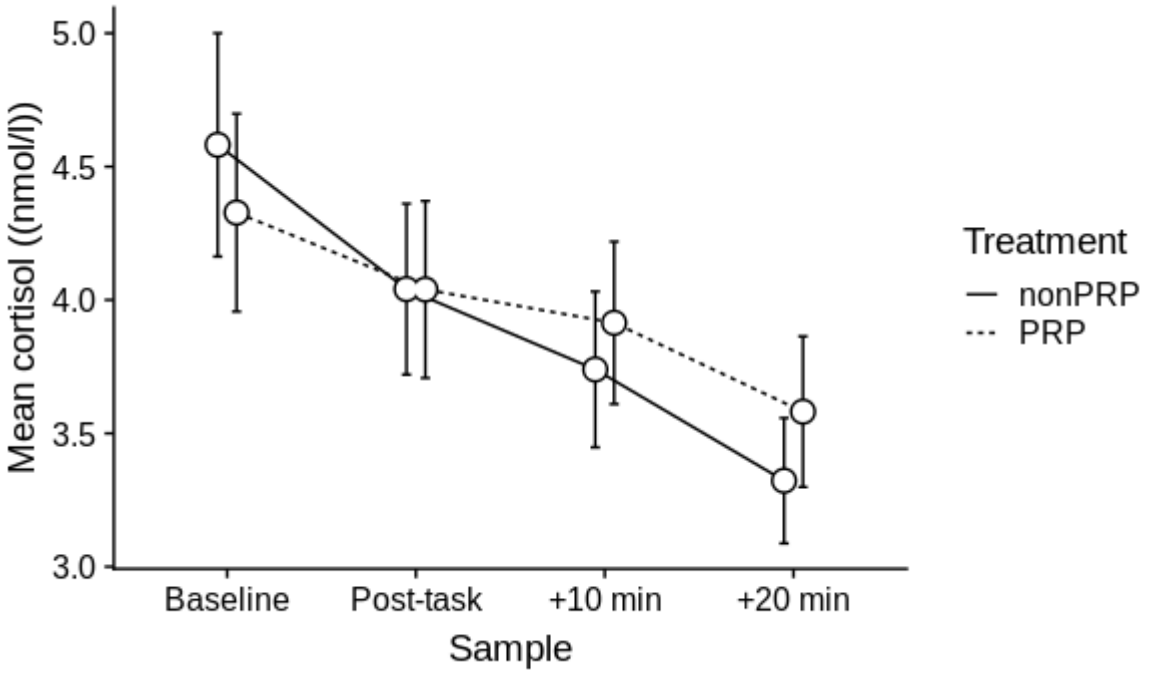

Table 3. Cortisol Outcomes in PRP vs. nonPRP Treatments

Measure	PRP Mean (standard deviation)	nonPRP Mean (standard deviation)	Paired t-Test (p-value)
AUCi	-12.19 (41.04)	-22.83 (41.52)	t(136)=2.53 (0.006)
AUCg	160.91 (72.88)	160.43 (73.58)	t(136)=0.27 (0.396)
Peak Change	0.07 (1.59)	-0.30 (1.43)	t(136)=2.33 (0.011)
Overall Change	-0.75 (1.78)	-1.26 (1.80)	t(136)=2.71 (0.004)

Note: AUCi is a measure of overall reactivity. AUCg is a measure of overall cortisol output. Peak change is the difference between the peak value out of the three post-task samples and the baseline sample. Overall change is the difference between the final +20 min and the baseline sample.

Table 4. Cortisol at each Time-point for Participants with Confounders vs. without Confounders (nmol/L)

Measure	Confounder Mean (standard deviation)	nonConfounder Mean (standard deviation)	Independent t-Test (p-value)
Baseline	6.61 (2.77)	4.27 (2.23)	t(272)=4.63 (<0.001)
Post-task	5.26 (2.03)	3.93 (1.90)	t(272)=3.12 (0.002)
+10 Min	4.58 (1.72)	3.76 (1.77)	t(272)=2.08 (0.038)
+20 Min	3.99 (1.39)	3.41 (1.56)	t(272)=1.70 (0.091)

Note: Participants (n = 22) who disclosed confounding variables such as recent food intake, caffeine, nicotine, waking up or taking medication after 12pm were compared to those who did not disclose any confounding variables, as these activities may affect cortisol levels for up to two hours after the event.

Table 5. The Effect of Treatment and Order on Cortisol Outcomes in a Two-way ANOVA (F)

	Treatment Effect (p-value)	Order Effect (p-value)	Carryover Effect (p-value)
AUCi	6.05 (p=0.015)	0.09 (p=0.760)	0.43 (p=0.512)
AUCg	0.11 (p=0.744)	4.50 (p=0.036)	0.35 (p=0.558)
Peak Change	5.41 (p=0.022)	0.04 (p=0.848)	0.36 (p=0.550)
Overall Change	6.70 (p=0.011)	0.65 (p=0.422)	0.24 (p=0.626)

Note: ANOVA comparing the effects of treatment (PRP vs. nonPRP), order (first vs. second) and treatment*order interaction (carryover effect).

Table 6. Cortisol Outcomes for nonPRP in First Session vs. PRP in Second Session

Measure	nonPRP First Session Mean (standard deviation)	PRP Second Session Mean (standard deviation)	Paired t-Test (p-value)
AUCi	-21.76 (39.12)	-9.97 (36.18)	t(76)=2.31 (0.024)
AUCg	164.89 (73.96)	151.29 (58.77)	t(76)=1.35 (0.180)
Peak Change	-0.23 (1.28)	0.11 (1.45)	t(76)=1.94 (0.056)
Overall Change	-1.27 (1.72)	-0.63 (1.63)	t(76)=2.91 (0.005)

Note: The sample includes only those who were allocated to nonPRP in their first session and PRP in their second session (N = 77).

Table 7. Cortisol Outcomes for PRP in First Session vs. nonPRP in Second Session

Measure	PRP First Session Mean (standard deviation)	nonPRP Second Session Mean (standard deviation)	Paired t-Test (p-value)
AUCi	-15.04 (46.71)	24.21 (44.71)	t(59)=1.30 (0.200)
AUCg	173.26 (86.71)	154.71 (73.32)	t(59)=1.64 (0.107)
Peak Change	0.01 (1.76)	-0.39 (1.61)	t(59)=1.42 (0.161)
Overall Change	-0.90 (1.97)	-1.24 (1.91)	t(59)=1.05 (0.298)

Note: The sample includes only those who were allocated to PRP in their first session and nonPRP in their second session (N = 60).

Table 8. Regression Coefficient Results of the Four Cortisol Outcome Measures

	AUCi		AUCg		Peak Change		Overall Change	
Variable	OLS	Fixed Effects	OLS	Fixed Effects	OLS	Fixed Effects	OLS	Fixed Effects
PRP	9.50 (4.41)	9.97 (4.33)	3.73 (7.67)	3.72 (7.45)	0.34 (0.17)	0.35 (0.17)	0.45 (0.20)	0.46 (0.20)
Second	-0.95 (4.37)	-0.46 (4.18)	-12.93 (7.69)	-11.80 (7.46)	-0.10 (0.17)	-0.09 (0.16)	0.06 (0.19)	0.05 (0.18)
No Confounder	35.01 (11.34)	25.12 (12.75)	-51.23 (18.08)	-60.77 (-17.82)	1.14 (0.35)	0.87 (0.51)	1.50 (0.52)	1.42 (0.67)
Group PRP First	-4.12 (-5.71)		6.35 (9.99)		-0.14 (0.21)		-0.16 (0.24)	
Male	3.22 (6.28)		27.21 (10.29)		0.24 (0.35)		-0.17 (0.27)	
Age 21-23	-9.68 (6.72)		9.69 (13.85)		-0.36 (0.23)		-0.54 (0.30)	
Age 24-26	-3.42 (12.78)		24.07 (22.68)		-0.004 (0.52)		0.09 (0.60)	
Age 27-29	-2.91 (11.51)		-8.86 (34.92)		-0.35 (0.35)		-0.03 (0.63)	
Age 30+	-16.05 (10.81)		11.38 (21.23)		-0.64 (0.45)		-0.48 (0.45)	
2nd Year	3.74 (7.73)		-27.01 (12.59)		0.21 (0.28)		0.31 (0.33)	
3rd Year	4.61 (8.55)		-24.80 (18.02)		0.10 (0.29)		0.32 (0.38)	
4th Year	15.26 (9.00)		-16.86 (17.71)		0.54 (0.32)		0.72 (0.39)	
Other Year	23.78 (10.70)		-44.43 (23.51)		0.87 (0.44)		0.92 (0.46)	

Art and Social Sciences	8.78 (13.25)		56.36 (20.66)		0.51 (0.43)		0.28 (0.56)	
Business	-0.90 (13.41)		51.39 (18.88)		0.23 (0.42)		-0.04 (0.57)	
Life Sciences	2.40 (12.79)		60.37 (20.43)		0.17 (0.42)		-0.21 (0.52)	
Other Science	20.17 (11.60)		27.24 (16.72)		0.76 (0.39)		0.69 (0.48)	
Constant	-64.93 (11.00)		161.20 (18.41)		-1.88 (0.39)		-2.88 (0.46)	

Note: Standard errors in parentheses under coefficient estimate. Reference groups are: females, age 18-20, 1st year of study and unknown subject of study. Number of observations=137. Standard errors in all regressions are clustered at the individual level.

Appendix A

Table A.1. Cortisol Outcomes for Participants That Returned vs. Did Not Return

Measure	Returning Mean (standard deviation)	nonReturning Mean (standard deviation)	Welch t-test (p-value)
Baseline	4.78 (2.68)	4.50 (2.30)	t(10.81)=-0.36 (0.724)
Post-task	4.34 (2.19)	4.12 (2.19)	t(10.56)=-0.33 (0.747)
+10 Min	4.04 (1.94)	3.71 (1.59)	t(11)=-0.62 (0.545)
+20 Min	3.63 (1.68)	3.37 (1.52)	t(10.62)=-0.52 (0.615)

Note: Comparison of raw cortisol values from the initial session between those who returned for the crossover session and those who did not.