RIETI Discussion Paper Series 22-E-046

Educational Gradient in Physiological Risk Factors in a Workplace: A decomposition analysis of biomarkers

AIZAWA, Toshiaki
Hokkaido University

RIETI
Research Institute of Economy, Trade & Industry, IAA
The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 22-E-046
May 2022

Educational Gradient in Physiological Risk Factors in a Workplace: A Decomposition Analysis of Biomarkers*

Toshiaki Aizawa
Graduate School of Economics and Business, Hokkaido University

Abstract

This study explores educational inequality in a workplace in Japan in relation to physiological risk factors. It investigates the difference in the prevalence rates of being overweight, hypertensive, dyslipidemic or diabetic between male employees who have undergone university education and those who have not. Combining the mandatory annual health check-up data and personnel data, we conduct a decomposition analysis to identify the major factors contributing to the inequality, and we measure the extent to which the observed between-group disparity is associated with the differences in observable characteristics. For all four conditions, significant disparities are observed between the groups. Between-group differences pertaining to alcohol consumption, smoking behaviour, job positions, psychological stress and family structure are the major significant contributing factors behind the between-group disparities in health. The results indicate that, along with a universal health-promoting approach, additional efforts to support less-educated employees should be initiated to mitigate the health inequality.

Keywords: health inequality, education, decomposition, physiological risk factors, overweight, hypertension, dyslipidemic, diabetic, health checkup
JEL classification: I12, I14

*This study is conducted as a part of the Project “Productivity Effects of HRM Policies and Management Quality” undertaken at the Research Institute of Economy, Trade and Industry (RIETI). The author is grateful for helpful comments and suggestions by Discussion Paper seminar participants at RIETI.

1 Introduction

The study of the social determinants of health has played an important role in epidemiology and public health for many decades (Marmot & Wilkinson, 2005). A social gradient in health has been defined as a stepwise deterioration in health aligned with decreasing socio-economic status in society (Marmot, 2004), wherein this status is inversely associated with the prevalence and incidence of disorders (Lang et al., 2011; Kaplan & Keil, 1993). Such gradients are often observed not merely in an entire society but also in a smaller community like a workplace. A socio-economic gradient in chronic disease aetiology in a workplace was first recognised in the landmark Whitehall studies of British civil servants (Marmot et al., 1991, 1997; Brunner et al., 1997) which revealed the dramatic differences in health by grade of employment, which could not be fully explained by well-established risk factors such as smoking or drinking. In the Whitehall study, Brunner et al. (1997) and Kumari et al. (2004) found that prevalences of metabolic syndrome and diabetes were associated with lower occupational status, where a progressively greater proportion of the subjects experienced ill health. Socio-economic gradient in the Whitehall study was not limited to the prevalence of chronic diseases. In relation to the difference in mortality between those of higher rank and those of lower rank, only one third of the difference was attributable to the differences in health behaviours (Marmot, 1994).

This study focuses on education as the crucial aspect of socio-economic status and global evidence on the inverse relationship between non-communicable diseases and education has been accumulated in both developed and developing countries (Chatterji et al., 2015; Dalstra et al., 2005; Agardh et al., 2011; Niessen et al., 2018; Mackenbach et al., 2008; Oshio & Kan, 2019; Oshio, 2018). This study aims to examine the contributing factors to education-related inequality. Cutler & Lleras-Muney (2010) review a number of mechanisms through which education affects health, fully acknowledging that they are complementary and none of them alone can entirely account for the relationship between education and health. The most intuitive mechanism would be the one where education improves health through greater earnings, so individuals are provided with the material resources through which they can control their environment through health-promoting lifestyles such as better housing, clothing and food. A second relevant mechanism is that education may alter individual preferences such as the value of the future, discount rate and risk aversion. Education may improve people's outlook on the future and lower time discount rates, thereby encouraging them to invest in promoting their health and avoiding riskier behaviours (Becker &

Mulligan, 1997). The third mechanism is the one where education can provide people with better access to information and improve their critical thinking competence, allowing them to make effective use of new health-related information for their benefit. In this sense, education matters for health beyond fundamental reading and writing skills (Cutler & Lleras-Muney, 2006); it may also enhance the responsiveness to health information. Educated people may be able to take better care of themselves if they manage the information competently and translate their knowledge and intentions into action.

In addition, psychological factors may play roles. In a meritocratic society, education is a pivotal element that determines one's relative position or rank in society or the workplace. As was conspicuous in the Whitehall study, rank by itself might affect health through the psychosocial mechanism (Siegrist & Marmot, 2004; Marmot et al., 1997). A job demand-control hypothesis proposed by Karasek (1979) suggests that a high level of psychological demands combined with a low level of decision authority over task performance and a low level of skill utilisation amplify the risk of a stressful experience, resulting in unpleasant long-term consequences for health. Another psychologically and social mechanism is that educated people tend to have educated colleagues who are more likely to invest in their health and refrain from unhealthy lifestyles (Cutler & Lleras-Muney, 2010). The peer effects in the social network would reinforce the effects of one's own education background. Certainly, not all of these mechanism are directly observable and therefore cannot be unequivocally identified by the observational data, but this study aims to quantify the extent to which the educational inequality in health is attributable to the differences in individual observable characteristics.

This study aims to explore the health disparity across educational backgrounds among full-time male employees of a Japanese chemical manufacturing company. Japan provides a desirable setting for examining the health inequality in a workplace owing to its mandatory annual health check-ups and regular stress check surveys. The Industrial Safety and Health Law mandates that employers must make their employees undergo health check-ups every year. These annual health check-ups are designed to detect potential diseases at an early stage and encourage health-enhancing behaviours through the regular monitoring of people's health. Those who are diagnosed as having risk factors in these check-ups will be offered professional advice from an industrial physician at a later date. The associated costs of the health check-ups are paid by the employers. In addition, a 2016 revision to the Industrial Safety and Health Law requires that companies and organisations with a workforce of 50 or more should carry out an annual stress check to assess the level of psychological burden felt by their employees, with 57 standardised questions regarding job stress (Kawakami & Tsutsumi, 2016). The stress check test is conducted online or in writing during business hours. All full-time workers in these companies are provided with identical national health insurance with comprehensive healthcare coverage at the same co-payment rate.

This study explores the dissimilarities in the prevalence of being overweight, hypertensive, dyslipidemic or diabetic between employees with tertiary education and those without it. These health conditions are associated with a high risk of premature death and they are influenced by socio-demographic factors and health-related behaviours. Decomposition analysis is employed to quantify the degree to which the observed disparities are explained by the differences in observable characteristics between the two education groups. Specifically, this study decomposes the difference found between the two groups into the part which is associated with the differences in observable characteristics, such as socio-demographic, physiological and behavioural risk factors, and the remaining part which is not explained by them. Identifying what factors are strongly associated with the educational health gradient is of paramount importance as a first step in designing a policy to mitigate health inequality.

One of the key features of this study is the use of clinical data to identify individual health risks. This study exploits the anthropometric data and the blood-based biomarkers measured during the health check-ups as they objectively reflect an individual's health and are clinically proven to be reliable (Singh & Newman, 2011). Using objectively measured health data is more beneficial than self-reported health data. Chronic conditions are often asymptomatic in their early stages and not all people are aware of them until they become serious. If education is positively related to the propensity to be aware of having a disease, the prevalence rate based on self-reports or past diagnosis among less educated people would be underestimated due to reporting bias. Consequently, the estimated educational gradient in self-reported illness would be subject to downward bias, resulting in limited validity.

2 Method

2.1 Data

The main data we utilise in this study is the annual health check-up data among full-time male employees of a Japanese chemical manufacturing company with multiple business establishments across the country.

The company conducts research and development (R&D) and produce chemical products. This study combines the health check-up data with the annual stress check and personnel and payroll data for the period 2016-2019. The health check-up records have detailed anthropometric information and clinical data from the blood samples. This study concentrates on male employees aged 35 and over for whom data relating to a comprehensive range of health behaviours, such as smoking and dietary habits, was collected by means of a questionnaire during the check-up. The total sample size is 7,428.

2.1.1 Health outcomes

Four physiological risk factors relating to chronic conditions are examined in this study. The first is the condition of being overweight. Individuals are considered to be overweight if their body mass index (BMI) is 25 or over. A high BMI increases the risk of cardiovascular diseases, stroke, diabetes and musculoskeletal disorders (Ng et al., 2014). The second condition is hypertension, one of the most well-known causes of life-threatening complications such as heart attacks, strokes, kidney failure and premature mortality (World Health Organization, 2013). By using measurements of systolic blood pressure and diastolic blood pressure, this study calculates the means of the three measurements. Individuals are diagnosed with hypertension if their mean systolic blood pressure is 140 $mmHg$ or over, their mean diastolic blood pressure is 90 $mmHg$ or over, or they are taking anti-hypertensive drugs.

The third health condition is dyslipidemia, which elevates the risk of cardiovascular diseases. From the collected blood samples, total cholesterol, high-density lipoprotein (HDL) and low-density lipoprotein (LDL) cholesterols and triglycerides are measured during the health check-ups. Individuals are diagnosed with dyslipidemia if their total cholesterol is 260 mg/dL or over, their HDL cholesterol is 35 mg/dL or lower, their LDL cholesterol is 180 mg/dL or over, their triglycerides are 300 mg/dL or over, or they are taking anti-hyperlipidemic drugs. The final risk factor is diabetes. Glycosylated haemoglobin (HbA1c) and fasting glucose are biomarkers for diagnosing diabetes. Individuals are diagnosed with diabetes if their HbA1c is 6.5% or over, their glucose is 125 mg/dL or over, or they are taking drugs for diabetes.

2.1.2 Educational backgrounds

This study classifies employees into two groups based on their formal schooling history: a higher education group with a university-level education and a lower education group without it. While employees who graduated from technical colleges are placed in the higher education group, those who graduated from

Table 1: Determinants of health

	Category	Variable
1	Age	Age 41-50 years , Age over 50 years
2	Family	Married, Any child
3	Income	Income quartiles
4	Physical activity	Walking habit, Exercise habit
5	Dietary habits	Eating fast, Eat night, Snacking, Skip breakfast
6	Alcohol consumption	Drinking every day, High GGT
7	Smoking status	Current smoker
8	Working hours	Overtime† 10-20 hours, Overtime 21-30 hours, Overtime over 30 hours
9	Job position	Exective, Director, Chief, Senior staff
10	Job category	Sales, R&D, Production
11	Work environment	High workload, High work demand, Low work control, Low support from supervisors, Low support from colleagues
12	Psychological distress	Not lively, Angry, Anxious, Depressed

Note: † Overtime hours are calculated by the month.

vocational schools and two-year colleges are not.[1] The sample sizes for postgraduates, undergraduates and the others are 1,872, 4,440 and 1,116 respectively.

2.1.3 Determinants of health

This study considers socio-demographic, behavioural, psychosocial and psychological factors as determinants of health. The choice of determinants is based on previous studies in social epidemiology, public health and health sciences (Ross & Wu, 1995; Swinburn et al., 2004; Akseer et al., 2020). We consider the following 12 categories of variables as listed in Table 1: (1) age, (2) family, (3) income (4) physical activity, (5) dietary habits, (6) alcohol consumption, (7) smoking status, (8) working hours, (9) job position, (10) job category, (11) work environment and (12) psychological distress. The first two categories relate to socio-demographic factors. Age comprises two age dummy variables; namely, aged between 41 and 50, and aged over 50. The age group between 35 and 40 is a reference category. The second category, family, considers marital status and whether or not an individual has a child. The third category, income, considers the material resources. As the annual total income data is available only in 2019, we estimated the income level in the other fiscal years and defined income quartile dummies for each fiscal year.[2]

Categories (4)-(8) are behavioural health determinants. Physical activity considers whether an individual has a daily habit of walking for more than one hour a day and whether they are in the habit of doing exercise to sweat lightly for over 30 minutes at a time at least twice weekly for over a year. The fifth

[1] Multiple sensitivity analyses are implemented by excluding the technical college graduates from the higher education group, but the results are stable and no substantial differences are detected.

[2] Details on estimating income function are provided in the Appendix.

category, dietary habits, takes into consideration whether an individual feels as if he eats meals quicker than others, whether he eats dinner two hours before bedtime more than three times a week, whether he eats snacks or drinks sweet beverages between meals and whether he skips breakfast more than three times a week. These eating behaviours are significant precursors of having higher body mass (Nishitani et al., 2009).

Alcohol consumption considers the frequency of drinking alcohol, such as sake, beer, wine, whisky or brandy, and the level of gamma-glutamyl transpeptidase (GGT) in an individual's system. GGT is an enzyme that catalyses the transfer of a glutamyl residue, which is often referred to in managing alcohol consumption. A high GGT level is over 100 U/L. Smoking status considers how long an individual has been smoking as well as the number of cigarettes smoked; they will be regarded as a heavy smoker if they have smoked a total of over 100 cigarettes or they have smoked for over 6 months and have been smoking over the past month. The working hours refer to how long an individual works beyond normal office hours. This category consists of three overtime work dummy variables: between 10 and 20 hours per month, between 21 and 30 hours, and over 30 hours. Overtime of less than 10 hours per month is a reference category.

The next three categories, (9)-(11), are related to psychosocial factors. Job position measures the relative position in the hierarchy of the workplace. The highest position is executive, followed by director, chief, senior staff and junior staff. Executives, directors and chiefs are looked upon as managerial workers who supervise and evaluate their subordinates. The tenth category relates to the types of jobs performed. Each employee is classified into one of the following job categories: sales, production, professional and clerk. Being a clerk is a reference category herein. The eleventh category, psychosocial work environment, considers the occupational stress associated with the task characteristics that an individual performs. Whether or not an individual experiences occupational stress is determined by the score of the stress check test under the guidelines developed by the Ministry of Health, Labour and Welfare (Ministry of Health, Labour and Welfare, 2016). High job amounts indicate high perceived job workload and high job demands refer to the difficulty of this workload. Low job control is the perceived lower level of freedom an individual has to decide how they work. A lack of support from supervisors or colleagues indicates the insufficient availability of social workplace support. The negative effects of these occupational stresses on health are reported in high-income countries (Marmot et al., 1997; Bosma et al., 1997; Stansfeld et al., 1998). For example, Toker et al. (2012) found that social support in the workplace is a protective factor contributing to a reduction in the risk of diabetes, and Oshio (2021) observed the negative association between job dissatisfaction and health in Japan.

The final category, psychological distress, considers four distress types: non-liveliness, anger, anxiety and depression. Individuals are tested to see whether they have such psychological distress from the test scores of Stress Check Survey (Ministry of Health, Labour and Welfare, 2016). These negative emotions are reported to be associated with worse health (Adler et al., 1994; Gallo & Matthews, 2003).

2.2 Non-linear Oaxaca-Blinder decomposition

This study employs the non-linear Fairlie decomposition method through which the difference in health between the lower and higher education groups is decomposed into two parts; the first part is the one which is linked with the difference in the distribution of health-related behaviours (the explained part) and the remaining part is the one which is not so explained (residual) (Fairlie, 1999, 2005a). The Fairlie decomposition is an extension of the Oaxaca-Blinder decomposition (Oaxaca, 1973; Blinder, 1973) to the non-linear model. First, $g = \{L, H\}$ indicates the group to which individuals belong. In this study, $g = L$ denotes the lower education group with observation size N^L, while $g = H$ denotes the higher education group with observation size N^H. Y represents a dichotomous health risk variable and $X = \{X_1, ..., X_k\}$ is a vector of individual observable characteristics. As Y is a binary variable, $\bar{Y}^g = \frac{1}{N^g} \sum_{i \in g}^{N^g} Y_i$ denotes the proportion of workers in group g who are diagnosed as having a high health risk. The Fairlie decomposition breaks down the between-group difference in the average predicted probabilities, $\bar{Y}^L - \bar{Y}^H$, as follows:

$$\begin{aligned} \bar{Y}^L - \bar{Y}^H &= \underbrace{\Big\{ \sum_{i \in L}^{N^L} \frac{F(X_i^L \hat{\beta}^P)}{N^L} - \sum_{i \in H}^{N^H} \frac{F(X_i^H \hat{\beta}^P)}{N^H} \Big\}}_{Explained\ part} + \\ & \underbrace{\Big\{ \sum_{i \in L}^{N^L} \frac{F(X_i^L \hat{\beta}^L)}{N^L} - \sum_{i \in L}^{N^L} \frac{F(X_i^L \hat{\beta}^P)}{N^L} + \sum_{i \in H}^{N^H} \frac{F(X_i^H \hat{\beta}^P)}{N^H} - \sum_{i \in H}^{N^H} \frac{F(X_i^H \hat{\beta}^H)}{N^H} \Big\}}_{Residual} \end{aligned} \tag{1}$$

where $F(.)$ denotes a logistic distribution; $\hat{\beta}^H$ and $\hat{\beta}^L$ are the coefficients estimated with the sample in the higher and lower education groups; and $\hat{\beta}^P$ is the coefficient estimated with the pooled sample. In equation (1), $\sum_{i \in g}^{N^g} \frac{F(X_i^g \hat{\beta}^P)}{N^g}$ is an average of counterfactual predicted probability that would be observed if the individual in group g had the coefficient vector estimated from the pooled sample.

The first curly bracket in equation (1) relates to the part of the gap which is explained by the between-group differences in the distribution of X. This part indicates how much the observed difference could be smaller if the between-group difference in the distributions of X were nullified. The second curly bracket represents the remaining part of the observed difference that cannot be explained by the between-group difference in the distribution of X. This term can also be interpreted as the part of the overall difference caused by the between-group differences in the function between X and Y.

Identifying the factors that are strongly associated with the observed difference in health risks is of great consequence to understanding the underlying mechanism that drives the disparity between groups and designing a measure to alleviate health inequality. The first curly bracket in equation (1) can be decomposed further into its contributions made by individual categories and covariates through the simulation. The principle idea of the detailed decomposition proposed by Fairlie (2005b) is to find a one-to-one matching of the observations of each group and sequentially switch each distribution of X.

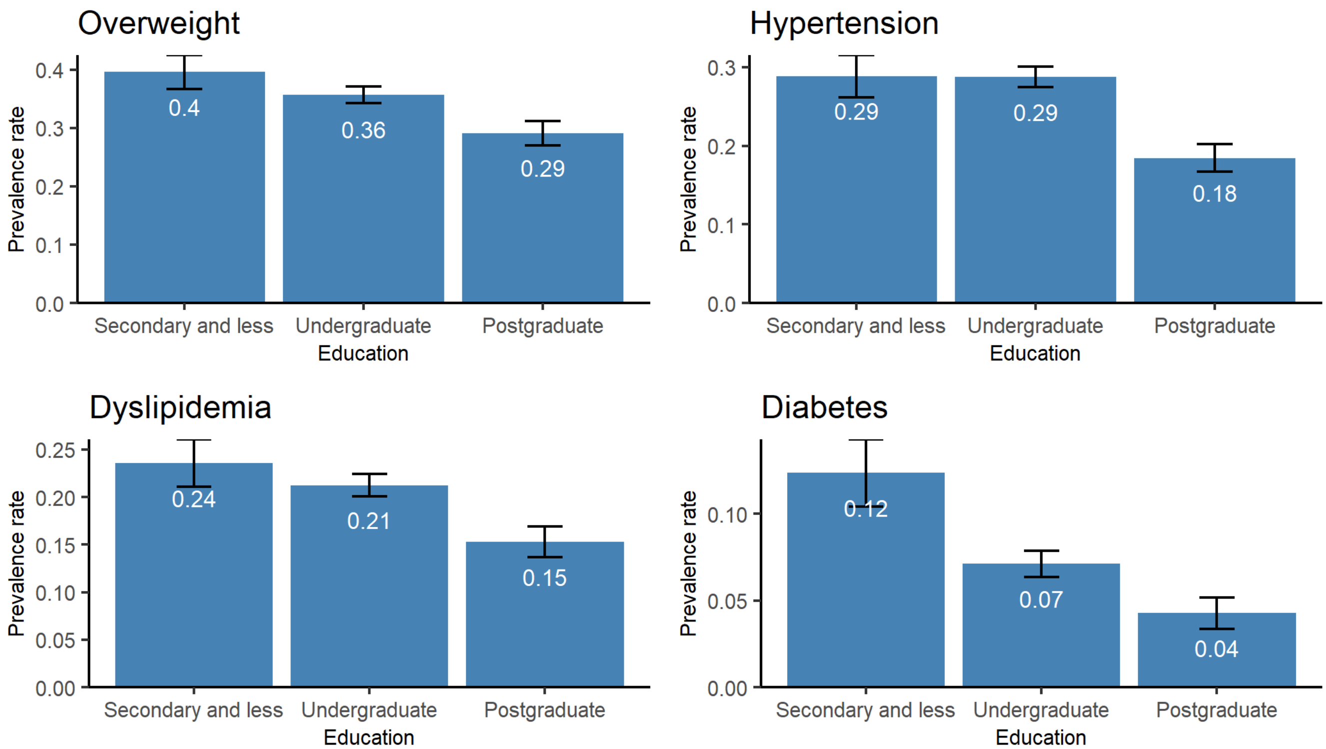
The first step calculates predicted probabilities for each population, namely $\hat{Y}_i^g = F(X_i^g \hat{\beta}^P)$. The second step draws the random samples with size N^L from the observations of the higher education group with a replacement. In the third step, each observation in the lower and higher education groups is separately ranked by the predicted probabilities and matched by their respective rankings. The fourth step calculates the marginal contribution made by each covariate to the explained disparity. The contribution made by each covariate to the disparity is equal to the change in the average predicted probability that is observed when the covariate distribution of the lower education group with that of the higher education group is replaced while the distribution of the other variables is held constant. For example, the marginal contributions made by X_1 and X_2 to the explained part are expressed by

$$\Delta_{X_1} = \sum_{i \in L}^{N^L} \frac{F(\hat{\beta}_0^P + X_{1i}^L \hat{\beta}_1^P + X_{2i}^L \hat{\beta}_2^P +, ..., X_{ki}^L \hat{\beta}_k^P)}{N^L} - \sum_{i \in L}^{N^L} \frac{F(\hat{\beta}_0^P + \tilde{X}_{1i}^H \hat{\beta}_1^P + X_{2i}^L \hat{\beta}_2^P +, ..., X_{ki}^L \hat{\beta}_k^P)}{N^L} \quad (2)$$

$$\Delta_{X_2} = \sum_{i \in L}^{N^L} \frac{F(\hat{\beta}_0^P + \tilde{X}_{1i}^H \hat{\beta}_1^P + X_{2i}^L \hat{\beta}_2^P +, ..., X_{ki}^L \hat{\beta}_k^P)}{N^L} - \sum_{i \in L}^{N^L} \frac{F(\hat{\beta}_0^P + \tilde{X}_{1i}^H \hat{\beta}_1^P + \tilde{X}_{2i}^H \hat{\beta}_2^P +, ..., X_{ki}^L \hat{\beta}_k^P)}{N^L}, \quad (3)$$

where $\tilde{X}_i^H = \{\tilde{X}_{1i}^H, ..., \tilde{X}_{ki}^H\}$ is a vector of covariates of the drawn sample of the higher education group. In a similar way, the marginal contributions can be calculated sequentially for every covariate. Although

Figure 1: Prevalence rates across educational backgrounds



Note: 95% confidence intervals are shown.

the sum of contributions of each covariate equals the explained part, the individual contributions can vary depending on the order of the switching procedure. To deal with this issue, the order of the covariate switching in each replication is randomised. This study implements 5,000 decompositions and approximates average results over all possible orderings.

3 Results

3.1 Descriptive statistics

Figure 1 illustrates the prevalence rates of the respective health risks across the different educational backgrounds. For all health outcomes, prevalence rates differ across education levels (all $p < 0.01$), clearly indicating the striking gradients in relation to being overweight, hypertensive, dyslipidemic and diabetic, where the prevalence rates are the highest among the employees with secondary education or below, and the lowest among the employees with postgraduate education. The decomposition analysis will investigate these health differences between employees with tertiary education and those without it.

Table 2 shows the descriptive statistics. With regard to individual characteristics, compared with the lower education group, a larger proportion of the higher education group are married and have children ($p < 0.01$). The higher education group has a greater income level than the lower education group ($p < 0.01$). While approximately 63% of the employees in the higher education group have income above the median income level, around 20% of them in the lower education group have. In terms of health-related behaviours, 42% of the lower education group walk regularly, while 35% of the higher education group walk regularly. Those in the higher education group are more likely to report that they eat faster than others but they are less likely to eat dinner late or skip breakfast (all $p < 0.01$). Employees in the lower education group are more likely to drink every day ($p < 0.01$) and a larger proportion of them have a high GGT figure ($p < 0.01$) compared with the university graduates. Smoking rates are much higher among the lower education group ($p < 0.01$). While 25% of the higher education group have a smoking habit, 43% of the lower education group have it. Concerning overtime, those in the higher education group work fewer hours ($p < 0.01$). Around a half of non-university graduates work in a production department, whereas most university graduates engage in sales and R&D. University graduates are more likely to be in a higher position in their company ($p < 0.01$). Specifically, around 60% of university graduates are in a managerial position (chief, director or executive) compared with only around 14% of non-university graduates. Employees in the higher education group perceive that their workloads and demands are high ($p < 0.01$). On the other hand, they are less likely to feel that they have low work control and support from supervisors and colleagues ($p < 0.01$). The higher education group are less likely to suffer psychological distress ($p < 0.01$).

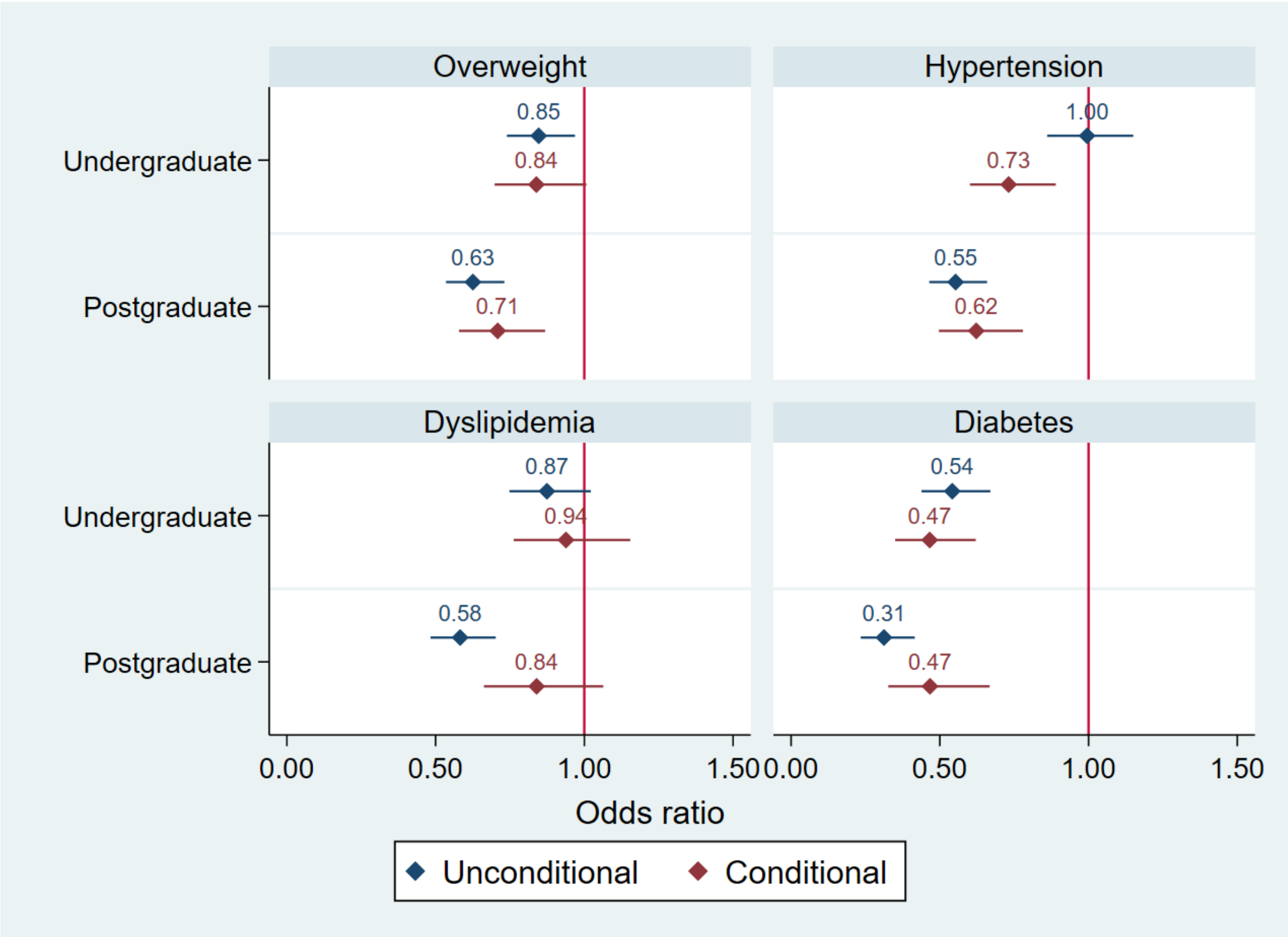
3.2 Regression analysis

Before conducting the decomposition analysis, it is worthwhile exploring the association between health outcomes and health determinants from the regression results. Figure 2 highlights the unconditional and conditional odds ratios for the four risk factors across educational backgrounds, where secondary education is a benchmark. Before controlling for observable characteristics, unconditional odds ratios for being overweight, dyslipidemic and diabetic are significantly smaller than 1 for undergraduate education (all $p < 0.05$). In terms of postgraduate education, the unconditional odds ratios are even smaller for all health outcomes ($p < 0.05$). The inclusion of individual observed characteristics and fiscal year fixed effects attenuated these relationships to varying degrees, suggesting that education-related disparities are partially explained by the behavioural health determinants. After observed characteristics are controlled

Table 2: Descriptive statistics

	All	Lower education	Higher education
	mean	mean	mean
Overweight	0.35	0.40	0.34
Hypertension	0.26	0.29	0.26
Dyslipidemia	0.20	0.24	0.19
Diabetes	0.07	0.12	0.06
Age 40-50	0.47	0.50	0.46
Age over 50	0.43	0.41	0.43
Married	0.87	0.83	0.88
Child	0.72	0.68	0.72
Income quartile 1st (lowest)	0.17	0.39	0.13
Income quartile 2nd	0.27	0.41	0.25
Income quartile 3rd	0.28	0.15	0.31
Income quartile 4th (highest)	0.28	0.05	0.32
Exercise	0.25	0.26	0.25
Walking	0.36	0.42	0.35
Eat fast	0.40	0.36	0.41
Eat night	0.47	0.51	0.46
Snacking	0.14	0.15	0.14
Skip breakfast	0.23	0.27	0.22
Drink everyday	0.38	0.44	0.37
High GGT	0.33	0.37	0.33
Smoke	0.28	0.43	0.25
Overtime 10-20h	0.12	0.21	0.11
Overtime 20-30h	0.15	0.21	0.13
Overtime over 30h	0.09	0.15	0.08
Junior staff	0.03	0.13	0.01
Senior staff	0.45	0.73	0.40
Chief	0.28	0.11	0.31
Director	0.18	0.01	0.21
Executive	0.07	0.01	0.08
Production	0.11	0.47	0.05
Sales	0.21	0.03	0.25
R&D	0.39	0.30	0.40
Clerical work	0.29	0.20	0.30
High workload	0.36	0.27	0.38
High work demand	0.33	0.23	0.34
Low work control	0.14	0.17	0.13
Low support from supervisors	0.26	0.36	0.25
Low support from colleagues	0.35	0.45	0.33
Not lively	0.21	0.27	0.20
Angry	0.27	0.35	0.26
Anxious	0.21	0.23	0.21
Depressed	0.21	0.29	0.19
Observations	7428	1116	6312

Figure 2: Odds ratios for the risk factors across educational backgrounds



Note: Secondary education is the benchmark category. For both unconditional and conditional odds ratios, year fixed effects are controlled for.

for, the significantly smaller odds ratios remained for hypertension and diabetes. For hypertension, the conditional odds ratios are 0.73 (95% CI: 0.60-0.89) for individuals with an undergraduate degree and 0.62 (95% CI: 0.50-0.78) for those with a postgraduate degree. For diabetes, they are 0.47 (95% CI: 0.35-0.62) and 0.47 (95% CI: 0.33-0.67) respectively. Table 3 shows the results of the logistic regression, where the odds ratios are shown and year fixed effects are controlled for. First, age is a strong predictor for all four outcomes ($p < 0.01$). Having a child is negatively correlated with being overweight, hypertensive and diabetic ($p < 0.01$). Higher income levels are negatively associated with health risks but statistical significance of these associations are not confirmed. Doing regular exercise has a negative link with being overweight ($p < 0.01$) and diabetic ($p < 0.05$). Eating quickly is also positively and significantly associated with all conditions ($p < 0.01$). With respect to alcohol consumption, a high GGT level is significantly correlated with the prevalence of all outcomes ($p < 0.01$). Smoking is positively related with dyslipidemia

($p < 0.05$) and diabetes ($p < 0.01$) and negatively with hypertension ($p < 0.05$). Overall, the directions of the relationship between health and these behavioural factors are largely consistent with the previous studies (Swinburn et al., 2004; Akseer et al., 2020).

Workers who do overtime for 20 hours a month or more are less likely to be dyslipidemic and diabetic than those who do so for fewer than 10 hours a month ($p < 0.05$). Regarding job categories, where clerical work is a benchmark category, employees involved in the production process are less likely to be hypertensive and diabetic (both $p < 0.01$). Engaging in R&D is negatively associated with hypertension, dyslipidemia and diabetes (all $p < 0.01$). In relation to the psychosocial work environments, low work control is positively associated with hypertension and dyslipidemia (both $p < 0.10$). On the other hand, high work demands and low support from colleagues are negatively correlated with hypertension ($p < 0.05$ and $p < 0.01$ respectively). Finally, with respect to psychological distress, anger shows positive associations with being overweight ($p < 0.01$), suffering from hypertension ($p < 0.05$) and diabetes ($p < 0.10$).

3.3 Decomposition

Table 4 shows the main results, where the differences in prevalence rates between the lower and higher education groups are decomposed into their contributing factors. Detailed decomposition results are available in the supplementary document. First, apropos weight, 39.6% of the lower education group are overweight compared with 33.7% of the higher education group. The observed difference is 5.9 percentage points (pp) ($p < 0.01$). Among the 12 categories of demographic and behavioural determinants, alcohol consumption shows the largest significant contribution, implying that the between-group difference in the proportion of heavy drinkers with high GGT figure is strongly associated with the between-group difference in the overweight rate. This category accounts for 0.4 pp, corresponding to around 6.8% of the observed difference. The size of this contribution is numerically interpretable; it suggests that in a hypothetical situation where the between-group difference in alcohol consumption was removed altogether, other things being equal, the between-group difference in the overweight rate would shrink by 0.4 pp. Job position indicates a large contribution, but we do not find its statistical significance.

Some categories exhibit negative contributions, signifying that the differences in these characteristics are counterbalancing the between-group health difference. In the case of dietary habits, the significant

Table 3: Regression analysis (odds ratios)

	(1) Overweight	(2) Hypertension	(3) Dyslipidemia	(4) Diabetes
main				
Undergraduate	0.839*	0.731***	0.938	0.466***
	[0.698,1.007]	[0.601,0.890]	[0.763,1.154]	[0.350,0.621]
Postgraduate	0.709***	0.622***	0.839	0.467***
	[0.578,0.868]	[0.497,0.780]	[0.663,1.063]	[0.327,0.667]
Age 40-50	1.553***	3.069***	1.692***	4.526***
	[1.269,1.900]	[2.259,4.170]	[1.296,2.208]	[2.242,9.135]
Age over 50	1.479***	7.177***	2.476***	10.24***
	[1.189,1.839]	[5.233,9.842]	[1.872,3.275]	[5.042,20.82]
Married	0.798**	0.823*	0.859	0.993
	[0.664,0.959]	[0.675,1.004]	[0.697,1.060]	[0.731,1.350]
Child	0.696***	0.780***	0.867*	0.641***
	[0.609,0.796]	[0.674,0.902]	[0.741,1.013]	[0.508,0.809]
Income quartile 2nd	0.916	0.876	1.085	0.911
	[0.726,1.155]	[0.665,1.154]	[0.823,1.431]	[0.593,1.401]
Income quartile 3rd	0.929	0.725	0.835	0.694
	[0.650,1.328]	[0.482,1.090]	[0.545,1.279]	[0.354,1.361]
Income quartile 4th (highest)	1.142	0.705	0.679	0.772
	[0.729,1.789]	[0.429,1.160]	[0.399,1.156]	[0.344,1.736]
Exercise	0.801***	1.011	0.908	0.750**
	[0.708,0.907]	[0.886,1.154]	[0.786,1.049]	[0.592,0.950]
Walking	1.004	0.952	0.974	0.888
	[0.900,1.121]	[0.845,1.072]	[0.857,1.106]	[0.726,1.087]
Eat fast	2.041***	1.463***	1.276***	1.664***
	[1.842,2.262]	[1.307,1.637]	[1.132,1.438]	[1.380,2.006]
Eat night	1.130**	1.059	1.125*	1.155
	[1.016,1.256]	[0.943,1.189]	[0.994,1.273]	[0.952,1.402]
Snacking	1.128	1.053	1.173*	1.543***
	[0.976,1.305]	[0.893,1.240]	[0.991,1.388]	[1.202,1.981]
Skip breakfast	0.919	1.010	1.058	0.737**
	[0.811,1.042]	[0.881,1.159]	[0.917,1.221]	[0.582,0.934]
Drink everyday	0.658***	1.107*	0.769***	0.701***
	[0.589,0.734]	[0.985,1.243]	[0.677,0.873]	[0.574,0.856]
High GGT	2.503***	2.098***	2.374***	2.268***
	[2.247,2.788]	[1.870,2.353]	[2.101,2.682]	[1.875,2.743]
Smoke	1.082	0.867**	1.179**	1.406***
	[0.964,1.214]	[0.764,0.984]	[1.033,1.345]	[1.151,1.717]
Overtime 10-20h	1.038	1.173	0.836	0.911
	[0.810,1.330]	[0.880,1.565]	[0.627,1.116]	[0.588,1.413]
Overtime 20-30h	0.960	1.033	0.682**	0.580**
	[0.735,1.255]	[0.756,1.411]	[0.498,0.935]	[0.354,0.950]
Overtime over 30h	0.875	1.067	0.666**	0.548*
	[0.635,1.207]	[0.733,1.553]	[0.455,0.974]	[0.295,1.020]
Senior staff	1.068	1.098	1.049	1.772*
	[0.765,1.490]	[0.746,1.615]	[0.712,1.545]	[0.964,3.258]
Chief	0.881	1.471	0.925	1.331
	[0.518,1.499]	[0.800,2.702]	[0.499,1.715]	[0.508,3.486]
Director	0.748	1.922*	1.131	1.505
	[0.410,1.366]	[0.977,3.780]	[0.561,2.279]	[0.514,4.410]
Executive	0.911	1.559	1.291	1.698
	[0.489,1.699]	[0.778,3.125]	[0.628,2.656]	[0.563,5.120]
Production	1.015	0.676***	0.988	0.637**
	[0.829,1.242]	[0.542,0.843]	[0.787,1.240]	[0.451,0.901]
Sales	1.246***	1.042	0.945	1.080
	[1.064,1.458]	[0.881,1.233]	[0.788,1.133]	[0.817,1.427]
R&D	1.084	0.799***	0.811***	0.720***
	[0.947,1.241]	[0.692,0.922]	[0.694,0.947]	[0.566,0.918]
High workload	0.974	0.981	1.014	0.963
	[0.860,1.103]	[0.855,1.126]	[0.876,1.174]	[0.759,1.221]
High work demand	1.074	0.872**	0.952	0.852
	[0.951,1.214]	[0.761,0.999]	[0.825,1.099]	[0.674,1.076]
Low work control	0.928	1.166*	1.171*	1.111
	[0.792,1.088]	[0.982,1.385]	[0.979,1.401]	[0.844,1.463]
Low support from supervisors	1.127*	0.913	1.040	1.057
	[0.987,1.287]	[0.788,1.058]	[0.892,1.214]	[0.835,1.337]
Low support from colleagues	0.901*	0.809***	0.986	1.226*
	[0.799,1.017]	[0.709,0.925]	[0.857,1.135]	[0.989,1.520]
Not lively	0.905	0.975	0.970	1.156
	[0.788,1.039]	[0.839,1.134]	[0.827,1.138]	[0.912,1.466]
Angry	1.266***	1.191**	1.026	1.231*
	[1.114,1.438]	[1.035,1.371]	[0.883,1.192]	[0.981,1.546]
Anxious	0.956	0.919	0.930	0.968
	[0.820,1.115]	[0.773,1.091]	[0.776,1.115]	[0.728,1.288]
Depressed	0.988	1.156	1.079	1.052
	[0.844,1.157]	[0.971,1.376]	[0.898,1.297]	[0.792,1.398]
Observations	7428	7428	7428	7428
Year fixed effects	Yes	Yes	Yes	Yes

Exponentiated coefficients; 95% confidence intervals in brackets
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Decomposition of the educational differences in physiological risks

	Overweight		**Hypertension**		**Dyslipidemia**		**Diabetes**	
	Estimates	SEs	Estimates	SEs	Estimates	SEs	Estimates	SEs
Observed difference								
Lower group	0.396***	0.015	0.289***	0.014	0.236***	0.013	0.124***	0.010
Higher group	0.337***	0.006	0.257***	0.005	0.195***	0.005	0.063***	0.003
Observed difference	0.059***	0.016	0.032**	0.015	0.041***	0.014	0.061***	0.010
Decomposition								
Age	0.003***	0.001	-0.001	0.001	0.000	0.001	0.000	0.002
Family	0.006***	0.001	0.004***	0.001	0.002**	0.001	0.001	0.001
Income	-0.006	0.015	0.024	0.016	0.024	0.015	0.007	0.011
Physical activity	-0.001	0.001	0.000	0.001	0.000	0.001	-0.001***	0.000
Dietary habits	-0.006***	0.001	-0.002**	0.001	0.000	0.001	-0.002**	0.001
Alcohol consumption	0.004***	0.001	0.007***	0.001	0.003***	0.001	-0.001	0.001
Smoking status	0.004**	0.002	-0.003	0.002	0.005***	0.002	0.005***	0.001
Working hours	-0.002	0.006	0.003	0.006	-0.010**	0.005	-0.004	0.004
Job position	0.030	0.020	-0.026	0.020	0.000	0.018	0.013	0.013
Job category	-0.003	0.007	-0.016***	0.006	0.006	0.006	-0.002	0.004
Work environment	-0.001	0.002	-0.001	0.002	0.002	0.002	0.006***	0.002
Psychological distress	0.003	0.002	0.004**	0.002	0.001	0.002	0.002**	0.001
Residuals	0.029***	0.012	0.038***	0.011	0.009	0.010	0.037***	0.007

Note: $N^L = 1,116$ and $N^H = 6,312$. Year fixed effects are controlled for. SEs=Standard errors.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

negative contribution is ascribed to the observation that, compared with the lower education group, those in the higher education group are more likely to eat quickly, a key driver of higher body mass (Nishitani et al., 2009). It follows that as far as dietary habits are concerned, the lower education group have preferable habits in relation to the possibility of being overweight. The sum of contributions made by the 12 categories is 3.0 pp, corresponding to 50.1% of the observed difference. The remaining 2.9 pp can be construed as the difference due to the variance in unobservable factors. A typical example of them is the difference in accumulated socio-economic circumstances. Another often-argued unobservable factor involves the time preference and/or the degree of self-control; educated people may have lower time discount rates and/or more self-disciplined, which encourages them to invest more in their education and health and to stay healthy (Fuchs, 2004, 2008).

Next, for hypertension, 28.9% of the lower education group are hypertensive compared with 25.7% of the higher education group. The between-group difference, 3.2 pp, is statistically significant ($p < 0.05$). Among the 12 categories, the difference in alcohol consumption shows the largest significant contribution, explaining around 21.8% of the observed difference ($p < 0.01$). Its large contribution is explicated by the observation that employees in the lower education group are more likely to drink alcohol every day and have a high level of GGT compared with those in the higher education group. The second largest significant contributor is the difference in psychological distress by 0.4 pp ($p < 0.05$). The detailed decomposition results in the supplementary document evince that a higher proportion of the lower education group feeling anger is the main contributor in this category ($p < 0.05$). Indeed, income indicates a large contribution to the observed difference but it exhibits a large standard errors, resulting in an insignificant contribution. Although the differences in these behavioural factors significantly and positively explain the observed between-group disparities, the total contributions made by the observable characteristics included in the model are almost nil. Most parts of the difference explained by the differences in alcohol consumption, psychological distress and family structure are offset by the negative contributions made by the differences in the other categories such as job category and dietary habits. A large negative contribution made by job category is explained by the observation that a larger proportion of employees with the lower education group work in a production department, which is negatively associated with hypertension. On the other hand, a smaller proportion of employees working in a R&D section among the lower education group, which would help to partially mitigate this negative contribution as working in a R&D department also has a negative association with hypertention. However, its effect due to the difference in the proportion of R&D is not large enough to fully offset the negative contribution that job category makes.

For dyslipidemia, the prevalence rate among the lower education group is 23.6%, whilst among the higher education group it is 19.5%. The between-group difference is 4.1 pp ($p < 0.01$). The major significant contributors to the between-group disparity are the differences in smoking status (by 0.5 pp), alcohol consumption (by 0.3 pp) and family structure (by 0.2 pp). They account for approximately 12.2%, 7.3% and 4.9% of the observed difference in the condition of hyperlipidemia respectively. Difference in income shows a large contribution but it indicates a high standard error and no statistical significance is confirmed. Adding all the contributions made by the difference in observed characteristics, the total explained part amounts to 3.2 pp, equivalent to around 78.0% of the observed difference ($p < 0.01$). This relatively large contribution implies that the observed education-related disparity is strongly associated with the differences in observed characteristics.

The prevalence rate of diabetes is 12.4% among the lower education group and 6.3% among the higher education group. Thus, the between-group difference is 6.1 pp ($p < 0.01$). Work environments account for the between-group difference (by 0.6 pp), explaining 9.8% of the observed difference ($p < 0.01$). The detailed decomposition results in the supplementary document reveal that low support from colleagues makes a relatively larger contribution. The descriptive statistics and regression results support this finding as more employees in the lower education group perceive insufficient support from colleagues, which is associated with a higher prevalence of diabetes. The second largest significant contributor is smoking behaviour with a contribution size of 0.5 pp, corresponding to 8.2% of the observed difference ($p < 0.01$). Its large contribution is attributable to the higher smoking rates among the lower education group. The total contribution size made by the observable characteristics is 2.4 pp, corresponding to 39.3% of the observed difference.

3.4 Further analysis of workers with undergraduate and postgraduate degrees

Educational gradients in health as shown in Figures 1 and 2 evince that health disparities exist even among those who have completed tertiary education. This subsection explores the health gaps between those who completed postgraduate education (postgraduate group) and those who only undertook undergraduate education (undergraduate group). The sample sizes are 4,440 and 1,872 respectively, and the descriptive statistics are reported in the supplementary document. The decomposition results are shown in Table 5.

For all outcomes, the prevalence rates are significantly lower among the postgraduate group (all $p < 0.01$). First, for being overweight, the observed difference is 6.6 pp ($p < 0.01$). The main contributing factors are the differences in age, job category and dietary habits (all $p < 0.01$). The total contribution made by the observed characteristics is 5.1 pp, amounting to 77.3% of the observed difference. Next, for hypertension, the observed difference is 10.3 pp ($p < 0.01$) and the main contributing factors are age, job category and alcohol consumption. The total contribution size attributable to the observed characteristics is 8.3 pp ($p < 0.01$). Different from the divergence between those with university education and those without, the difference observed between those with undergraduate education and those with postgraduate education is expounded well by the observed characteristics.

For dyslipidemia, the observed difference is 6.0 pp ($p < 0.01$). The three largest contributing categories are age, job category and alcohol consumption. The total contribution made by the 12 categories is 4.9 pp ($p < 0.01$), accounting for 81.7% of the difference. For diabetes, the observed difference is 2.8 pp ($p < 0.01$) and the major contributing factors are age, job category and smoking behaviour. The total contribution is 2.9 pp, which is slightly larger than the observed difference. This suggests that if workers with only an undergraduate education had the health determinants distribution that workers with a postgraduate education have, they could have a lower prevalence rate than their counterparts.

Overall, the comparison between employees with undergraduate education and those with postgraduate education reveals that most parts of the observed differences are related to the difference in age and job category, which is consistent with the descriptive statistics. The mean age of the undergraduate group is 49.0, while that of the postgraduate group is 45.7. In terms of job category, while the majority of the undergraduate group work in sales, more than 70% of workers in the postgraduate group engage in professional works such as R&D. Intriguingly, in contrast to the finding in the previous subsection, the difference in job position does not show significant associations with health inequality. Plausible reasons for this would be that the proportions of job positions are not substantially different between the undergraduate group and the postgraduate group, and the difference in the proportion of workers in a managerial position is not significant. In the case of the comparison between employees with undergraduate education and those with postgraduate education, the contribution size of income is very small and insignificant. One of the possible explanations for this small contribution is that a difference in income level between these two groups is much smaller than that between employees who have undergone university education and employees who have not.

4 Discussion and conclusion

This study has examined the contributing factors to education-related inequality in a workplace and quantified the extent to which the observed inequality is attributable to the differences in individual observable characteristics. For all four conditions, significant disparities are observed between the groups. Between-group differences pertaining to alcohol consumption and smoking status are the strong contributing factors behind the between-group disparities in health. Concerning the difference in the prevalence of being dyslipidemic, more than half of the disparities are explained by the observable factors. Whilst its contribution size is relatively large, income does not indicate a significant contribution. As found in the regression analysis, income and health risk factors are negatively correlated but these correlations are not significant when we control for other observable characteristics such as job position and job types.

The observation that a certain behavioural characteristic is a significant risk factor does not mean that

Table 5: Decomposition of the differences in physiological risks between the undergraduates and the postgraduates

	Overweight		Hypertension		Dyslipidemia		Diabetes	
	Estimates	SEs	Estimates	SEs	Estimates	SEs	Estimates	SEs
Observed difference								
Undergraduate	0.357***	0.007	0.288***	0.007	0.213***	0.006	0.071***	0.004
Postgraduate	0.291***	0.010	0.184***	0.009	0.153***	0.008	0.043***	0.005
Observed difference	0.066***	0.013	0.103***	0.011	0.060***	0.010	0.028***	0.006
Decomposition								
Age	0.019***	0.004	0.044***	0.003	0.019***	0.003	0.011***	0.002
Family	0.002***	0.000	0.002**	0.001	0.001	0.001	0.001	0.001
Income	0.001	0.002	0.001	0.003	0.004	0.004	0.001	0.003
Physical activity	0.001	0.001	0.000	0.000	0.000	0.001	0.000	0.000
Dietary habits	0.004***	0.001	0.003***	0.001	0.001	0.001	0.000	0.001
Alcohol consumption	0.002**	0.001	0.011***	0.001	0.006***	0.001	0.002**	0.001
Smoking status	0.004**	0.002	-0.003***	0.001	0.003***	0.001	0.004***	0.001
Working hours	0.001	0.002	0.002	0.002	-0.001	0.002	-0.002	0.003
Job position	0.002	0.003	-0.002	0.004	-0.002	0.005	0.003	0.004
Job category	0.015***	0.006	0.021***	0.005	0.016***	0.005	0.009***	0.003
Work environment	-0.001	0.001	0.004***	0.001	0.002**	0.001	0.000	0.001
Psychological distress	-0.001***	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Residuals	0.015	0.010	0.020**	0.009	0.011	0.009	-0.001	0.005

Note: Sample sizes of those with an undergraduate education and those with a postgraduate education are 4,440 and 1,872 respectively. Year fixed effects are controlled for. SEs=Standard errors. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

the difference in such factors is a significant contributor to the observed inequality. Regular exercise, for example, is a significant modifiable risk factor in the regression analysis, but it is not a significant contributor to the educational gradient because its difference across educational backgrounds is not significant. In addition, the weaker contribution to the educational gradient in the decomposition analysis does not negate the importance of improving such behavioural factors for employees' overall health in a workplace (Lynch et al., 2006). Furthermore, some behavioural factors show negative contributions to the observed inequality. Note that negative contributions do not imply that such categories are of less importance for health. For example, the difference in dietary habits indicates the negative contribution to the difference in being overweight and hypertensive, but as the regression result suggests, kicking an unwholesome dietary habit is no less important in improving an individual's health condition.

Certainly, a universal approach should be encouraged to manage illness and ameliorate health conditions, and additional all-out efforts should be made to encourage less educated employees to attenuate the health inequality. Reducing alcohol consumption and smoking should be prioritised because they are modifiable behaviours, even in the middle years of an individual's life, with enthusiastic social support. The decomposition analysis shows that the difference in smoking behaviour accounts for around 12.2% and 8.2% of the educational inequality in dyslipidemia and diabetes respectively. These results suggest that lifestyle change after reaching adulthood would be key in preventing health inequality from evolving in middle and old age. Sustained encouragement should be provided to those in the lower education groups so they retain their salubrious lifestyles even after they have joined the workforce.

The limitations of this study must be contemplated. First, the data used is observational data and therefore we acknowledge that this study may only discern associations. The decomposition results would be subject to the potential reverse causation or unobservable confounding factors. Reverse causation is, however, less serious because biomarkers in middle age and early old age are unlikely to impact formal schooling that is typically completed in young adulthood, although returning to school later in life is not entirely infeasible. As reviewed by Cutler & Lleras-Muney (2010), the possible mechanisms through which education affects health and multiple and not all of them are directly observable or quantifiable. Although the results shows that large parts of the observed health inequality is associate with the observable health risk factors, further research would be necessary to examine the remaining parts of the health disparities that are not explained by the included health risk factors.

Second, the observed educational disparity revealed in this study is a snapshot of the evolving inequality. The associations among socio-economic status, lifestyles and health may have been established over time and they would be mutually reinforcing one another. For instance, opportunities to obtain a university education are not necessarily equally given because educational opportunities are often dependent on the parental socio-economic status and the household circumstances in which a person grew up (Okada, 2012). Future studies should take account of socio-economic status across the life course and the accumulation of risk factors in order to fully explore the possible dynamic associations between disease risks and the life-course socio-economic differences.

Third, the income considered in this study is the individual income level, not the household income. If the more likely a wife is to be in the labour market the lower her husband income is, the dispersion of total household income would be smaller. Moreover, the income data available was total annual income before tax. Hence the actual difference in disposable income after tax can be smaller. Certainly using the quartile levels of income could help to mitigate the severity of these limitations, but we should keep in mind that the contribution made by income could be overestimated if a spouse of employees with lower education is more likely to work for the household.

Fourth, our samples comprise only full-time employees working in a large company who are obliged to undergo the annual health check-ups. It is therefore not clear to what extent the present findings can be generalised to employees in other companies, part-time workers, civil servants, employees in small and medium-sized enterprises, and the self-employed. Moreover, better job opportunities are also dependent on one's educational background as better education increases one's chances of being employed as a full-time worker (Okada, 2012). An even larger education-related health disparity might be observed if we consider part-time workers, people with job insecurity and even the unemployed. Exploring the full extent of the educational gradient is an ultimate and compelling extension of this study. The initial finding of this study would be complemented by considering the different employment statuses across an entire country. We will leave this as a future research agenda.

References

Adler, N., Boyce, T., Chesney, M., Cohen, S., Folkman, S., Kahn, R., & Syme, S. (1994). Socioeconomic status and health. The challenge of the gradient. *American Psychologist*, *49*(1), 15–24.

Agardh, E., Allebeck, P., Hallqvist, J., Moradi, T., & Sidorchuk, A. (2011). Type 2 diabetes incidence and socio-economic position: A systematic review and meta-analysis. *International Journal of Epidemiology*, *40*(3), 804–818.

Akseer, N., Mehta, S., Wigle, J., Chera, R., Brickman, Z. J., Al-Gashm, S., Sorichetti, B., Vandermorris, A., Hipgrave, D. B., Schwalbe, N., & Bhutta, Z. A. (2020). Non-communicable diseases among adolescents: Current status, determinants, interventions and policies. *BMC Public Health*, *20*(1908), 1–20.

Becker, G. S. & Mulligan, C. B. (1997). The Endogenous Determination of Time Preference. *The Quarterly Journal of Economics*, *112*(3), 729–758.

Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources*, *8*(4), 436–455.

Bosma, H., Marmot, M. G., Hemingway, H., Nicholson, A. C., Brunner, E., & Stansfeld, S. A. (1997). Low job control and risk of coronary heart disease in whitehall II (prospective cohort) study. *BMJ*, *314*(558), 558–565.

Brunner, E. J., Marmot, M. G., Nanchahal, K., Shipley, M. J., Stansfeld, S. A., Juneja, M., & Alberti, K. G. M. M. (1997). Social inequality in coronary risk: Central obesity and the metabolic syndrome. Evidence from the Whitehall II study. *Diabetologia*, *40*(11), 1341–1349.

Chatterji, P., Joo, H., & Lahiri, K. (2015). Examining the education gradient in chronic illness. *Education Economics*, *23*(6), 735–750.

Cutler, D. M. & Lleras-Muney, A. (2006). Education and health: Evaluating theories and evidence. Working Paper 12352, National Bureau of Economic Research.

Cutler, D. M. & Lleras-Muney, A. (2010). Understanding Differences in Health Behaviors by Education. *Journal of Health Economics*, *29*(1), 1 – 28.

Dalstra, J., Kunst, A., Borrell, C., Breeze, E., Cambois, E., Costa, G., Geurts, J., Lahelma, E., Van Oyen, H., Rasmussen, N., Regidor, E., Spadea, T., & Mackenbach, J. (2005). Socioeconomic differences in the prevalence of common chronic diseases: An overview of eight European countries. *International Journal of Epidemiology*, *34*(2), 316–326.

Fairlie, R. W. (1999). The Absence of the African? American Owned Business: An Analysis of the Dynamics of Self-Employment. *Journal of Labor Economics*, *17*(1), 80–108.

Fairlie, R. W. (2005a). An Extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models. *Journal of Economic & Social Measurement*, *30*(4), 305 – 316.

Fairlie, R. W. (2005b). An extension of the blinder-oaxaca decomposition technique to logit and probit models. *Journal of Economic & Social Measurement*, *30*(4), 305 – 316.

Fuchs, V. R. (2004). Reflections on the socio-economic correlates of health. *Journal of Health Economics*, *23*(4), 653 – 661.

Fuchs, V. R. (2008). Time preference and health: An exploratory study. In *Economic Aspects of Health* (pp. 93–120). University of Chicago Press, Chicago.

Gallo, L. C. & Matthews, K. A. (2003). Understanding the association between socioeconomic status and physical health: Do negative emotions play a role? *Psychological Bulletin*, *129*(1), 10–51.

Kaplan, G. A. & Keil, J. E. (1993). Socioeconomic factors and cardiovascular disease: A review of the literature. *Circulation*, *88*(4), 1973–1998.

Karasek, R. A. (1979). Job Demands, Job Decision Latitude, and Mental Strain: Implications for Job Redesign. *Administrative Science Quarterly*, *24*(2), 285–308.

Kawakami, N. & Tsutsumi, A. (2016). The stress check program: A new national policy for monitoring and screening psychosocial stress in the workplace in japan. *Journal of Occupational Health*, *58*(1), 1–6.

Kumari, M., Head, J., & Marmot, M. (2004). Prospective Study of Social and Other Risk Factors for Incidence of Type 2 Diabetes in the Whitehall II Study. *Archives of Internal Medicine*, *164*(17), 1873–1880.

Lang, T., Lepage, B., Schieber, A.-C., Lamy, S., & Kelly-Irving, M. (2011). Social Determinants of Cardiovascular Diseases. *Public Health Reviews*, *33*(2), 601–622.

Lynch, J., Smith, G. D., Harper, S., & Bainbridge, K. (2006). Explaining the social gradient in coronary heart disease: comparing relative and absolute risk approaches. *Journal of Epidemiology & Community Health*, *60*(5), 436–441.

Mackenbach, J. P., Stirbu, I., Roskam, A. R., Schaap, M. M., Menvielle, G., Leinsalu, M., & Kunst, A. E. (2008). Socioeconomic inequalities in health in 22 European countries. *New England Journal of Medicine*, *358*(23), 2468–2481.

Marmot, M. (2004). *The Status Syndrome: How Social Standing Affects Our Health and Longevity.* London Bloomsbury Publishing plc.

Marmot, M., Bosma, H., Hemingway, H., Brunner, E., & Stansfeld, S. (1997). Contribution of job control and other risk factors to social variations in coronary heart disease incidence. *The Lancet*, *350*(9073), 235–239.

Marmot, M., Stansfeld, S., Patel, C., North, F., Head, J., White, I., Brunner, E., Feeney, A., Marmot, M., & Smith, G. (1991). Health inequalities among British civil servants: the Whitehall II study. *The Lancet*, *337*(8754), 1387–1393.

Marmot, M. & Wilkinson, R. (2005). *Social Determinants of Health.* Oxford University Press, Oxford UK.

Marmot, M. G. (1994). Social Differentials in Health within and between Populations. *Daedalus*, *123*(4), 197–216. The MIT Press, MA: USA.

Ministry of Health, Labour and Welfare (2016). Manual for stress check program. Available from: URL: https://www.mhlw.go.jp/bunya/roudoukijun/anzeneisei12/pdf/150803-1.pdf.

Ng et al. (2014). Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: A systematic analysis for the global burden of disease study 2013. *The Lancet*, *384*(9945), 766 – 781.

Niessen, L. W., Mohan, D., Akuoku, J. K., Mirelman, A. J., Ahmed, S., Koehlmoos, T. P., Trujillo, A., Khan, J., & Peters, D. H. (2018). Tackling socioeconomic inequalities and non-communicable diseases in low-income and middle-income countries under the Sustainable Development agenda. *The Lancet*, *391*(10134), 2036–2046.

Nishitani, N., Sakakibara, H., & Akiyama, I. (2009). Eating behavior related to obesity and job stress in male Japanese workers. *Nutrition*, *25*(1), 45–50.

Oaxaca, R. L. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, *14*(3), 693–709.

Okada, A. (2012). Education Reform and Equal Opportunity in Japan. *Journal of International and Comparative Education*, *1*(2), 116–129.

Oshio, T. (2018). Widening disparities in health between educational levels and their determinants in later life: evidence from a nine-year cohort study. *BMC Public Health*, *18*(278), 1–10.

Oshio, T. (2021). Job dissatisfaction as a predictor of poor health among middle-aged workers: a 14-wave mixed model analysis in Japan. *Scandinavian Journal of Work, Environment & Health*, *47*(8), 591–599.

Oshio, T. & Kan, M. (2019). Educational level as a predictor of the incidences of non-communicable diseases among middle-aged Japanese: a hazards-model analysis. *BMC Public Health*, *19*(852), 1–11.

Ross, C. E. & Wu, C.-l. (1995). The links between education and health. *American Sociological Review*, *60*(5), 719–745.

Siegrist, J. & Marmot, M. (2004). Health inequalities and the psychosocial environment-two scientific challenges. *Social Science & Medicine*, *58*(8), 1463–1473.

Singh, T. & Newman, A. (2011). Inflammatory Markers in Population Studies of Aging. *Ageing Research Riverview*, *10*(3), 319–329.

Stansfeld, S. A., Bosma, H., Hemingway, H., & Marmot, M. G. (1998). Psychosocial work characteristics and social support as predictors of SF-36 health functioning: The Whitehall II study. *Psychosomatic Medicine*, *60*(3), 247–255.

Swinburn, B. A., Caterson, I. D., Seidell, J. C., & James, W. P. T. (2004). Diet, Nutrition and the Prevention of Excess Weight Gain and Obesity. *Public Health Nutrition*, *7*(1A), 123–146.

Toker, S., Shirom, A., Melamed, S., & Armon, G. (2012). Work Characteristics as Predictors of Diabetes Incidence Among Apparently Healthy Employees. *Journal of occupational health psychology*, *17*, 259–67.

World Health Organization (2013). A global brief on hypertension: Silent killer, global public health crisis. Geneva, WHO.

Appendix

Estimation of income function

We regress logarithmic total annual income on individual demographic and socio-economic characteristics, using the payroll information in 2019. The predicted income is used to impute the missing income information. To take into account the fiscal year effects on income, income quartiles are obtained in each fiscal year. The estimated income function is shown in Table 6.

Table 6: Estimated income function

	(1)	
	ln(total pay)	
Male	0.0937***	(0.0108)
Age 40-50	0.0683***	(0.00959)
Age over 50	0.0914***	(0.00994)
Married	0.0243***	(0.00899)
Child	0.0305***	(0.00734)
Overtime 10-20h	0.0958***	(0.0103)
Overtime 20-30h	0.137***	(0.0107)
Overtime over 30h	0.198***	(0.0116)
Senior staff	0.278***	(0.0120)
Chief	0.568***	(0.0144)
Director	0.676***	(0.0154)
Executive	0.838***	(0.0181)
R&D	0.0166	(0.0106)
Sales	0.0288***	(0.00854)
Professional	-0.0000403	(0.00701)
Undergraduate	-0.00594	(0.00912)
Postgraduate	-0.00514	(0.0101)
Constant	15.41***	(0.0133)
Observations	2208	

Standard errors in parenthesis.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Detailed decomposition results

Table 7: Detailed decomposition results between the non-university graduates and the university graduates

	(1) Overweight	(2) Hypertension	(3) Dyslipidemia	(4) Diabetes
Age 40-50	0.00488***	0.00631**	0.00324	0.00228
Age over 50	-0.00188*	-0.00713**	-0.00311	-0.00376
Married	0.00269**	0.00197*	0.00132	-0.0000580
Child	0.00320***	0.00207***	0.000988	0.00119
Income quartile 2nd	-0.00251	-0.00348	0.00186	-0.000800
Income quartile 3rd	0.00233	0.00853	0.00457	0.00438
Income quartile 4th (highest)	-0.00703	0.0163	0.0164	0.00397
Exercise	-0.000639**	0.0000273	-0.000181	-0.000340
Walking	0.0000290	-0.000416	-0.000273	-0.000378
Eat fast	-0.00655***	-0.00303***	-0.00176***	-0.00172*
Eat night	0.00110**	0.000472	0.000801	0.000594
Snacking	0.000216	0.0000423	0.0000823	-0.00000532
Skip breakfast	-0.000700	0.000207	0.000504	-0.00105*
Drink everyday	-0.00444***	0.00112*	-0.00266***	-0.00165**
High GGT	0.00811***	0.00528***	0.00543***	0.00105
Smoke	0.00351*	-0.00304*	0.00478***	0.00524***
Overtime 10-20h	0.000482	0.00252	-0.00302	-0.00107
Overtime 20-30h	-0.000801	0.000000451	-0.00356**	-0.00189
Overtime over 30h	-0.00207	0.000317	-0.00361*	-0.00164
Production	0.00682	-0.0178***	0.000501	-0.00421
Sales	-0.00951***	-0.0000692	0.00209	0.00141
R&D	-0.000607	0.00276***	0.00309***	0.00124*
Senior staff	-0.00207	-0.00281	0.000464	0.00786
Chief	0.0112	-0.00419	0.00423	0.00167
Director	0.0175	-0.0136	-0.00142	0.00309
Executive	0.00421	-0.00221	-0.00227	0.000553
High workload	0.000723	0.000595	-0.000200	0.000688
High work demand	-0.00137	0.00289**	0.00104	0.00193
Low work control	-0.000467	0.00101*	0.00100	0.000247
Low support from supervisors	0.00260*	-0.00159	0.000703	0.000717
Low support from colleagues	-0.00233	-0.00408***	-0.000215	0.00219*
Not lively	-0.00145	-0.000279	-0.000328	0.000654
Angry	0.00472***	0.00255**	0.000393	0.00124
Anxious	-0.000178	-0.0000665	-0.000126	0.0000212
Depressed	-0.000220	0.00202	0.00107	0.000252
Observations	7428	7428	7428	7428
Year fixed effects	Yes	Yes	Yes	Yes

Standard errors are suppressed.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Detailed decomposition results between the undergraduates and the postgraduates

	(1) Overweight	(2) Hypertension	(3) Dyslipidemia	(4) Diabetes
Age 40-50	0.00225	-0.0119**	-0.00198	-0.00952
Age over 50	0.0168***	0.0560***	0.0210***	0.0193
Married	0.000889**	0.000733	0.000448	0.0000528
Child	0.00127***	0.00133**	0.000750	0.000732
Income quartile 2nd	0.0000533	-0.000823	0.00000375	0.000556
Income quartile 3rd	-0.000638	0.00208	0.00524	0.000741
Income quartile 4th (highest)	0.00179	-0.000474	-0.00123	-0.0000345
Exercise	0.000738**	-0.00000544	0.000245	0.000444
Walking	0.000303	0.000122	-0.000498	-0.0000117
Eat fast	0.00443***	0.00273***	0.00155***	0.00104
Eat night	-0.000172	-0.000122	-0.000198	-0.00000671
Snacking	-0.000293	-0.0000516	-0.000573	-0.000909
Skip breakfast	0.000195	0.000673	0.000717	0.0000315
Drink everyday	-0.0102***	0.000265	-0.00435***	-0.00134*
High GGT	0.0126***	0.0110***	0.0103***	0.00338***
Smoke	0.00360**	-0.00270**	0.00334**	0.00392***
Overtime 10-20h	0.000465	0.00183	-0.00114	-0.000676
Overtime 20-30h	-0.000320	0.00135	-0.00168	-0.00156
Overtime over 30h	0.00128	-0.000932	0.00155	0.000443
Production	0.000711	0.00155	-0.00115	0.00100*
Sales	0.0118**	-0.000167	-0.000901	-0.000251
R&D	0.00188	0.0185***	0.0178***	0.00879***
Senior staff	0.00759	0.00203	0.00186	0.00521
Chief	-0.00654	-0.00571	-0.00535	-0.00202
Director	0.00115	0.00194	0.000788	0.00000816
Executive	0.000479	0.0000930	0.000227	-0.000176
High workload	0.0000425	0.000256	-0.0000123	0.0000917
High work demand	-0.000436	0.00155*	0.000361	-0.0000943
Low work control	-0.000277	0.00131	0.00169*	0.000390
Low support from supervisors	-0.000282	0.000415	-0.000273	0.0000850
Low support from colleagues	0.0000454	0.000275	0.000115	-0.0000884
Not lively	-0.000665**	0.00000740	-0.0000123	0.00000522
Angry	-0.0000244	-0.000133	0.0000103	-0.0000171
Anxious	-0.000177	-0.000256	-0.0000979	-0.0000103
Depressed	-0.0000729	0.000155	0.0000168	0.00000593
Observations	6312	6312	6312	6312
Year fixed effects	Yes	Yes	Yes	Yes

Standard errors are suppressed.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Descriptive statistics of the undergraduates and the postgraduates

	All mean	Undergraduate mean	Postgraduate mean
Overweight	0.34	0.36	0.29
Hypertension	0.26	0.29	0.18
Dyslipidemia	0.19	0.21	0.15
Diabetes	0.06	0.07	0.04
Age 40-50	0.46	0.47	0.45
Age over 50	0.43	0.48	0.32
Married	0.88	0.88	0.89
Child	0.72	0.72	0.73
Income quartile 1st (lowest)	0.13	0.12	0.14
Income quartile 2nd	0.25	0.25	0.24
Income quartile 3rd	0.31	0.29	0.34
Income quartile 4th (highest)	0.32	0.33	0.28
Exercise	0.25	0.25	0.25
Walking	0.35	0.34	0.40
Eat fast	0.41	0.42	0.39
Eat night	0.46	0.45	0.47
Snacking	0.14	0.13	0.16
Skip breakfast	0.22	0.24	0.17
Drink everyday	0.37	0.41	0.29
High GGT	0.33	0.35	0.28
Smoke	0.25	0.29	0.17
Overtime 10-20h	0.11	0.11	0.09
Overtime 20-30h	0.13	0.14	0.12
Overtime over 30h	0.08	0.06	0.13
Junior staff	0.01	0.01	0.01
Senior staff	0.40	0.40	0.38
Chief	0.31	0.29	0.36
Director	0.21	0.21	0.19
Executive	0.08	0.08	0.07
Production	0.05	0.04	0.07
Sales	0.25	0.34	0.03
R&D	0.40	0.27	0.71
Clerical work	0.30	0.35	0.18
High workload	0.38	0.37	0.39
High work demand	0.34	0.32	0.39
Low work control	0.13	0.15	0.09
Low support from supervisors	0.25	0.24	0.26
Low support from colleagues	0.33	0.33	0.33
Not lively	0.20	0.20	0.19
Angry	0.26	0.26	0.26
Anxious	0.21	0.22	0.20
Depressed	0.19	0.19	0.19
Observations	6312	4440	1872