The Association between Depression and Subjective Sleep Quality: Truth or Reporting Bias?

Zhiyong Huang[a] and Fabrice Kämpfen[b*]

[a]Southwestern University of Finance and Economics, Chengdu, China
[b]Department of Economics, HEC, University of Lausanne, Switzerland

September 16, 2018

Abstract

Objectives: *To determine the presence and magnitude of reporting heterogeneity in subjective assessment of sleep quality by those who have depressive symptoms.*

Methods: *An online survey using Visual Analogue Scales and anchoring vignettes was implemented to study the comparability of subjective assessments of sleep quality among university students in Switzerland. Kernel densities and multivariate linear regressions exploiting anchoring vignettes were used to examine the association between self-reported sleep quality and depressive symptoms and the potential role of reporting heterogeneity in moderating this relationship.*

Results: *Our analysis shows that reporting heterogeneity plays a marginal role in moderating the negative association between depression and sleep quality, irrespective of the severity of the depressive symptoms of the individuals.*

Conclusions: *Although present in subjective assessments of sleep quality, reporting heterogeneity doesnâĂŹt appear to explain much of the depression-sleep quality gradient. This suggests that comparisons of subjective sleep quality between individuals are reliable, even among those suffering from depression.*

Keywords: *Sleep quality, Depression, Self-assessment, Visual analogue Scale, Reporting heterogeneity*

*Corresponding author. E-mail address: fabrice.kaempfen@unil.ch, Tel.: +41 21 692 33 20. University of Lausanne, Quartier Chamberonne, 1015 Lausanne. Do not quote or circulate without permission of the authors.

1 Introduction

Sleep quality is associated with a wide range of outcomes affecting one's life. Not only is good sleep quality associated with better physical health, but it has also been found that persons with better sleep quality have greater well-being and better psychological functioning[1–9]. Insomnia and other sleep disorders, on the other hand, are strong predictors of low quality of life, depression, anxiety, stress and neuroticism, to name a few[1].

Sleep quality is also particularly important among adolescents and young adults, as it has been found to have a very strong relationship with mental health; problems with sleep length have been associated with poor mental health such as psychotic-like experiences (PLEs)[10], depression and anxiety[11–17], suicidal thoughts and self-harm behaviors[18–23]. Given the increasing prevalence of depression among adolescents and young adults in the recent years[24] and the fact that depression in adolescents and young adults constitutes an important risk factor for depression and bipolar disorder during adulthood[25], having a clear understanding of the true relationship between sleep quality and depression is crucial if one wants to formulate lifestyle guidance and design policies to promote healthy development of young adults and put in place effective prevention of mental problems.

Many of the existing studies on sleep quality have however relied on subjective assessment of sleep quality[11;26–29]. Because subjective sleep quality is easy to administer and requires little respondent burden, it has been widely used in population health surveys including U.S. Health and Retirement Study (HRS), the Survey of Health, Aging and Retirement in Europe (SHARE), the WHO Study on Global Aging and Adult Health (SAGE), the China Health and Retirement Longitudinal Study (CHARLS) and in clinical settings such as in Pittsburgh Sleep Quality Index. However, subjective sleep quality measures, like any other self-reported health measures, may suffer from reporting heterogeneity. Reporting heterogeneity arises when individuals interpret the possible values attached to a scale differently from one another, i.e., when they adopt different reporting scales to rate their own characteristics. In the case of individuals with depression, the subject of our study, one critical concern would be that people with depression will tend to report poor sleep quality, either because they are indeed deprived of sleep as a result of depression or simply because they are inclined to rate everything more negatively, making subjective sleep quality assessments between individuals with and without depressive symptoms hard to compare. For instance, Visual Analogue Scale (VAS)[1] of sleep quality often takes "best quality of sleep I can imagine" as one endpoint and "worst quality of

[1] Definition of visual analogue scale here

sleep I can imagine"as another. Both of these endpoints, sometime called anchors in the literature, can be interpreted very differently by individuals with different depression level. And because respondents use these anchors to evaluate their sleep quality, the variations in the interpretations of these anchors across groups of respondents could potentially mean that they use different reporting scales, making comparisons between self-reported measures of sleep quality across these groups difficult to interpret. Other factors than depression, such as sex and income, can also bias self-reported sleep quality if respondents interpret the scale of the sleep quality measure differently from others.

Because it is important to understand whether the association between a subjective measure, such as sleep quality, and its important risk predictors –depression in our case– still holds after controlling for reporting differences among people, researchers have designed techniques to account for reporting heterogeneity in self-assessed measures. One example of such techniques is anchoring vignettes.

Anchoring vignettes are short descriptions of hypothetical individuals, which are assessed by respondents alongside their self-assessment of the same domain. For example, an anchoring vignette on sleep quality included in SAGE is profiled as: "[Mark] falls asleep every night within five minutes of going to bed. He sleeps soundly during the whole night and wakes up in the morning feeling well-rested." Respondents are asked to evaluate the sleep quality of this hypothetical person alongside their own sleep quality with a five-point Likert scale, ranging from 'very dissatisfied', 'dissatisfied', 'neither satisfied, nor dissatisfied', 'satisfied' to 'very satisfied'. As anchoring vignettes are predefined and invariant across individuals, their ratings are supposed to only reflect the differences in reporting scales. As long as reporting scales have been identified by these vignettes, the effect of reporting heterogeneity can be purged from the self-assessment, making the adjusted self-assessment better reflect the actual underlying evaluations of individuals.

The objective of this study is to determine the existence of reporting heterogeneity in subjective assessment of sleep quality, paying close attention to reporting heterogeneity among individuals with depressive symptoms. Using data collected from a sample of students in the region of Lausanne in Switzerland whom we asked to complete an online survey[23], we develop five anchoring vignettes adapted from SAGE to correct the reporting heterogeneity in self-assessed quality of sleep. Anchoring vignettes employed in this study cover a wide range of dimensions that elicit the sleep patterns

[2]The survey was conducted using Qualtrics. See https://www.qualtrics.com/ for more information.

[3]Online surveys are particularly well-adapted for this kind of research question[30] and have been used in many other studies to assess reporting heterogeneity in self-reported evaluations[30–32].

of individuals such as the time it takes them to fall asleep, the number of times they wake up during the night, how rested they feel when they wake up in the morning and how easy it is for them to get out of bed in the morning. The five anchoring vignettes that we use in our study are reported in the appendix.

In addition to various sociodemographic information that we collected in our survey, we also asked our respondents to tell us the number of effective hours of sleep they had per night in average over the past 30 days, at what time they usually went to bed, how long it took them to fall asleep and what time they woke up in the morning. On top of that, we asked them to evaluate the difficulty they had with sleep –our measure of sleep quality– along with the five anchoring vignettes using Visual Analogue Scales (VAS)[4], ranging from 0 ('no difficulty') to 100 ('extreme difficulty').

From our sample, individuals suffering from 'mild', 'moderate' and 'severe' depression rated their difficulty with sleep to be 19, 32 and 37 points higher, respectively, compared to those with 'no or minimal' depression symptoms. This association is robust even after controlling for various measures of sleep time and sleep efficiency. In line with the previous literature, we therefore confirm that depression is a strong predictor of poor sleep quality. Although a number of studies have examined the association between depression and sleep quality measured by self-assessment, very few have attempted to assess the existence of reporting heterogeneity and to correct for it. Tareque et al. [33] seems to be the only exception. However, one shortcoming of the study by Tareque et al. [33] is that sleep quality is measured on a Likert scale, making interpretation of the magnitude of reporting heterogeneity difficult. Using self-assessed sleep quality and anchoring vignettes measured on VAS, as this study does, we are able to evaluate the magnitude of reporting heterogeneity in self-assessed sleep quality, which is our main contribution to the literature. Our analysis shows the presence of reporting heterogeneity in subjective assessment of sleep quality by those who suffer from depression. Indeed, individuals suffering from moderate depression have a higher sleep quality than what they report, by about 2.7 points. Nevertheless, with regard to the strong association between sleep quality and depression, this effect is very small. In addition, none of the other sociodemographic characteristics we take into account in our analysis seems to explain reporting heterogeneity when comparing sleep quality between individuals. Therefore, the effect of reporting heterogeneity appears to be only marginal and the strong association between depression and sleep quality cannot be attributed to it.

The rest of the paper is organized as follows. In the next section, we describe the data that we

[4]VAS is a measurement instrument that is used to rate a given characterisic over a continuum of values (a horizontal or vertical line) with "anchors" at the two end-points, representing the best and worst scenarios.

collected, provide some descriptive statistics of our sample and go over the methods we use to derive our results. Section 3 details our analysis and its results. Section 4 explores the significance of our results and concludes the paper.

2 Participants and Methods

2.1 Participants

We recruited our sample of students from various schools of higher education in Canton de Vaud in Switzerland[5][6]. The online questionnaire we used was created on Qualtrics and sent out to students who registered in the University of Lausanne's experiment program. 6,578 emails were sent, out of which 1,938 students completed our survey in the following two weeks (response rate= 29.5%). For their participation, students who completed our online survey were automatically entered into a lottery for which the highest prize was about USD 300[7]. Out of the 1,938 students who completed our online questionnaire, we kept only the observations of individuals who were at most 25 year of age to have a more homogenous sample. After discarding 122 observations that were out of the age range, our final sample consisted in 1,816 observations. To assess the reliability of our questionnaires, we included a question in the end of our online survey that asked how difficult the respondents thought the questionnaire was. 72% of the respondents who completed the survey found it easy to complete the survey and 15% found it very easy. The remaining participants found it difficult and only one person found it very difficult.

2.2 Subjective sleep quality and anchoring vignettes

Respondents were asked to evaluate the difficulties they have had with their sleep patterns, such as having difficulties falling asleep, waking up regularly during the night or too early in the morning in the past 30 days using a VAS, ranging from 0 ('no difficulty') to 100 ('extreme difficulty').

Figure 1 shows the kernel distributions of self-evaluated quality of sleep for both males (dark grey line) and females (light grey line). The distribution of sleep quality appears to be highly bimodal, with the middle inflection point at about 40. The majority of females and males in our sample

[5] Participants were students at the Swiss Federal Institute of Technology in Lausanne (EPFL), the University of Lausanne (UNIL) and Ecole hoteliere de Lausanne (EHL).

[6] The experimental protocol and informed consent were approved by HEC Ethics Committee of the University of Lausanne in February 2017, and all subjects gave informed consent. They were informed that participation was optional and that they could withdraw at anytime.

[7] The distribution of prizes was the following: 3 winners that made USD 300, 10 winners that earned USD 100 and 60 respondents who won USD 20.

evaluated their own sleep quality rather positively (below 50), where the density distribution is the highest at roughly 15 for both males and females. On average, males appear to have a higher quality of sleep (mean of 35) compared to females (mean of 40). There is also a large chunk of our respondents who appear to have difficulty with their sleep, especially females, for whom the density remains relatively high compared to males, even after 75.

[Figure 1 about here.]

The design of our anchoring vignettes on sleep quality relies heavily on existing anchoring vignettes developed by ???. Because ??? does not primarily target student populations, we modify the anchoring vignettes to match the characteristics of the persons described in the vignettes to the characteristics of our student population. We define vignettes with factors including time it takes for individuals to fall asleep, the quality of the sleep, the difficulty the person has to wake up and its consequences as well as whether the person feels rested in the morning. Some of the vignettes share the same factors, but to various intensities. Moreover, the vignettes respondents have to evaluate are sex-specific, matching the sex of the respondent to the sex of the fictitious persons described in the hypothetical scenarios[8]. The five vignettes we use in our study can be found in the Appendix.

In order to determine whether the self-assessed measure of sleep quality suffers from reporting heterogeneity bias, we therefore asked the respondents of our survey to evaluate the sleep quality of persons described in these five anchoring vignettes, using identical VAS also ranging from 0 to 100. The evaluations of the five vignettes are represented in Figure 2. The first plot in the first row is the evaluation of the average of the five vignettes by sex, for which no differences across males' and females' evaluations can be seen. The second plot in the first row represents the evaluation by our respondents of the first vignette. There appears to be unanimity across our respondents: the sleep pattern of the fictitious person described in vignette 1 represents an ideal scenario (most of the density is located at 0). At the complete opposite, the last plot in the second row of figure 2 represents the evaluation by our respondents of vignette 5. Although somewhat more skewed towards the left, most of our sample evaluated scenario 5 as representing the scenario of someone experiencing the most severe difficulty with one's sleep pattern, i.e., lowest quality of sleep. Vignettes 2, 3 and 4 represent mid-range scenarios as can be seen in the other plots of figure 2. It is interesting to note here that the evaluations of vignette 2 appear to be quite discordant, as revealed by the

[8]As documented by Jürges and Winter [34], the sex of the persons described in the vignettes could potentially be a confounder when vignettes are evaluated. Therefore, survey including vignettes usually either assign gender-specific vignettes to respondents [35–38] or control for sex in all estimations [30;31]. We therefore follow the literature and design sex-specific vignettes and control for sex in our analysis.

density which, in addition to displaying bimodality, is high from value 20 to 70. As described below and more in details in the technical appendix, one can use these vignettes evaluations to identify the reporting scale of the individuals. Once these reporting scales have been identified, one can purge out the effect of reporting heterogeneity from self-assessed measures of sleep quality, making the adjusted self-assessment better reflect the actual underlying evaluations of individuals.

[Figure 2 about here.]

2.3 Depressive symptoms

We assess the presence and severity of depression of our respondents using the standard PHQ-9, a 9-question instrument that is a well-validated self-administered questionnaire widely used for assessement of depressive symptoms in clinical practice. Using the PHQ-9 scores of the respondents, we categorize them as having "no or minimum symptoms" if they score between 0 and 9, "mild depression" if their score is between 10 and 14, "moderate depression" if between 15 and 19 and "severe depression" if above 19 points.

2.4 Socio-demographic and health characteristics

We obtained information on our respondents about their age, nationality, relationship status, quality of the relationship with their parents as well as their parents' education level, family income, school performance and number of close friends. To have a rough idea about their health status, we also asked them whether they considered themselves as obese and whether they have already been diagnosed with eating disorder, a disorder that is frequent among teenagers and young adults in Switzerland[39;40] and has been found to be strongly correlated with depression[41].

Table 1 reports the descriptive statistics of our study sample. 887 of our respondents were male (45,8%) and the average age in our sample was about 21 years old. Females were more likely to be Swiss compared to males (55% vs 48%), less likely to report being single (44% vs 58%) and had on average fewer close friends. The bulk of our respondents were from families that were making between USD 7'000.- and 12'000.- per month, which is comparable to the average gross income per month per household in Switzerland in 2014 (about USD 10,079[9]). About 3% of our sample evaluated the relationship with their parents as being bad and more than half of our respondents evaluated their school performance as average. Females were more likely to suffer from 'mild' and

[9]See https://www.bfs.admin.ch in the "Household income and expenditure" section for details.

'moderate' depression (18% vs 13% and 9% vs 3%, respectively) and a larger number of them considered themselves as obese (14% vs 9%) and were diagnosed with eating disorder (11% vs 2%) compared to males. Although not reported in the table, our online survey also collected information on the number of siblings, the school the student was enrolled in and its program (undergraduate or graduate) as well as the level of education of the parents.

[Table 1 about here.]

2.5 Method

Epanechnikov kernel densities of self-assessed sleep quality by categories of depression severity were used to investigate the bivariate relationship between subjective sleep quality and depression. The association of subjective sleep quality and depression was further tested using multivariate linear regressions which adjusted for sociodemographic variables and various measures of objective sleep patterns.

To determine the presence of reporting heterogeneity in subjective sleep quality evaluations, one needs to control for reporting heterogeneity. There are in theory several ways to do so, each of them being based on assumptions that characterize the reporting patterns of the individuals. The first and most straightforward way to control for reporting heterogeneity is to assume that the respondents use the same reporting scale when evaluating their sleep quality as when evaluating the sleep quality of the persons described in the vignettes, an assumption that is referred to as vignette consistency in the literature. If this assumption holds, then one can simply include vignette evaluations as additional control in the linear regression above to control for reporting heterogeneity. This strategy however does not allow us to determine what the factors that determine reporting heterogeneity are. In order to determine the individual characteristics that explain reporting heterogeneity, one can for example run a double=index model. The model allows us to disentangle the true difference in sleep quality across respondents from reporting heterogeneity. Intuitively, a double-index model exploits the fact that we are observing several sleep quality evaluations (self and vignettes) by the same individuals. And because we are assuming that individuals use the same reporting scale when evaluating their own sleep quality and the one of the persons described in the vignettes, one can control for reporting heterogeneity, pretty much as one controls for individual fixed-effect in panel data models where several observations for the same individual are recorded over time. The double-index model allow us in addition to control for reporting heterogeneity to estimate what the characteristics of the individual that drive reporting heterogeneity are. More details on the double-index model are provided in the appendix.

3 Results

3.1 Relationship between depression and sleep quality

We start our analysis by examining the bivariate relationship between self-assessed sleep quality and depression. Figure 3 plots the self-assessed sleep quality of our sample by depression status. As can be seen clearly from the densities, the bimodality of the distribution exists for all four depression categories, although it is somewhat less pronounced for those who are moderately depressed and those who are severely depressed. The relationship between sleep quality and depression is very strong: almost all individuals suffering from depression, whatever the severity, have difficulties with their sleep, with the density shifting towards 100 the more severe the depressive symptoms are.

[Figure 3 about here.]

Results from the multivariate linear regression of self-assessed sleep quality measured from 0 to 100 on our set of covariates are reported in the first column of Table 2. Sex, age, income, relationship status but also perceived obesity and eating disorder appear to have no effect on sleep quality[10]. Individuals suffering from 'mild', 'moderate' and 'severe' depression however have experienced great difficulties with sleep as they have rated their sleep quality 19, 32 and 37 points lower, respectively, compared to persons with 'no or minimal' symptoms. These effects are all significant at the 99.9% confidence.

One may wonder whether the relationship between sleep quality evaluation and depression status could be confounded with sleep time. Indeed, the association could be spurious if it turns out that sleep time or any other variables affecting both sleep quality and depression are not controlled for in the analysis.

[Table 2 about here.]

Column 2, 3, 4, 5 and 6 of Table 2 however show that controlling for various measures of sleep patterns and efficiency does not critically affect the relationship between sleep quality and depression. Controlling first for the reported effective hours of sleep (sleep time) of the respondents (column

[10] Note that we report in the tables only the coefficients associated with sex, age and depression status, as they are the main focus of this study. Details on the coefficient of the other variables are available upon request to the authors.

2), one can see that the effect of depression on sleep quality remains very large and significant. Controlling next for the hours of sleep defined as the difference of the time a person wakes up and the time he or she goes to bed, the effect of depression again remains identical (column 3). In column 4, we control again for hours of sleep as defined in column 3, but this time correcting for the time it takes for the respondent to fall asleep. Again the effect of depression on sleep quality is very robust. The same holds when we control only for the time it takes individuals to fall asleep (column 5). Two things are worth noting here. The first is that this variable captures some of the depression-sleep quality gradient as the coefficients associated to depression decreased by about 30%. Second, the time it takes for individuals to fall asleep explains quite a lot of the sleep quality variation, "independently" of depression status, given the large and statistically significant effect of the variable on sleep quality (about 26 points).

Finally, in column 6, we control for whether it takes more than 20 minutes for an individual to fall asleep and whether the individual sleeps less than 7 hours. In the same regression, we also add an indicator reflecting the sleep efficiency of an individual as defined as taking the value of 1 if the ratio of the number of effective hours of sleep over the difference between the waking up time and the time the individual goes to bed is lower than 0.85. Results show that, including these 3 indicator variables in our statistical specification, although they are individually all positive and highly significant, do not appear to explain the relationship between depression and sleep quality. Sleep patterns and hours of sleep therefore do not seem to explain the effect of depression on sleep quality as the effects of depression on sleep quality still appear to be strong and significant even after controlling for these different pathways. Could the effect of depression on sleep quality be explained, at least partially, by reporting heterogeneity? As explained above, people with depression could report poor sleep quality, either because they are indeed deprived of good sleep as a result of depression or simply because they are apt to rate everything more negatively. Reporting heterogeneity could indeed stem from the fact that the meaning of the two end-points of our VAS, 'no difficulty' and 'extreme difficulty', could be interpreted very differently depending on whether a person suffers from depression or not. We now focus our analysis on reporting heterogeneity and try to answer this question.

3.2 Reporting heterogeneity of sleep quality

We start our analysis by regressing the five vignette evaluations on our set of regressors. If a coefficient turns out to be statistically significant, it would mean that the vignettes are not perceived equally by respondents who have the characteristic associated to that particular coefficient. Table

3 reports the results of the regressions of our 5 vignettes on our usual set of control variables. It is reassuring that very few coefficients are significantly different from 0, suggesting that individuals evaluate vignettes in a rather similar way[11]. One exception stands out however. Individuals suffering from 'moderate' and 'severe' depression appear to evaluate vignette 2 more negatively, with their evaluations being 8.4 and 9.7 point closer to 'extreme difficulty,' respectively, compared to those with 'no or minimal' depressive symptoms, which suggests the presence of reporting heterogeneity in subjective sleep quality in these two subgroups of individuals. The individuals suffering from the same conditions also appear to evaluate vignette 1 differently compared to individuals with no depressive symptoms, although the differences are only weakly significant.

[Table 3 about here.]

To further determine the existence and the magnitude of reporting heterogeneity, we add vignette evaluations directly in the sleep quality regression as control variables. Indeed, assuming that individuals use the same reporting scale to evaluate their sleep quality as the scale they use to evaluate the vignettes[12], reporting heterogeneity can be directly controlled for by adding vignette evaluations as regressors. Table 4 reports these regressions. Column 1 is the results of regressing sleep quality on our set of control variables (same regression as column 1 of Table 2) and column 2 represents the results of the same regression including this time vignettes evaluations as control variables. Although vignettes 1's and 2's evaluations are themselves highly significant, the inclusion of vignettes as control variables does not affect the effect of depression on sleep quality. The conclusion is identical when we add vignettes evaluations (column 4) to the model that also controls for sleep patterns and efficiency (column 3). As can be seen in Table 4, despite the fact that the evaluations of vignettes 1 and 2, are highly significant, which suggests the presence of reporting heterogeneity, the inclusion of the vignettes in our regressions to control for reporting heterogeneity does not alter the deleterious effect of depression on sleep quality. This seems to indicate that the association between sleep quality and depression is unlikely to be severely biased by reporting heterogeneity.

[Table 4 about here.]

Because there nonetheless appears to be some reporting heterogeneity in sleep quality, as evidenced by the statistically significant effect of some of our vignettes, it would be interesting to

[11] Again, details on the coefficient of those variables are available upon request to the authors.

[12] This is commonly known as the vignette consistency assumption in the vignette literature. See the appendix for more details.

understand what factors drive this reporting heterogeneity. Since we observe both self-assessed sleep quality and vignette evaluations, we can run a double-index model which will allows us to estimate the effects of our control variables on both sleep quality free from any reporting heterogeneity and on the reporting heterogeneity itself, hence enabling us to determine the factors explaining reporting heterogeneity.

Results from this model are summarized in Table 5. Odd numbered columns represent the true effect of our control variables, net of any reporting heterogeneity, on sleep quality, and even numbered columns show the effect of our set of regressors on reporting heterogeneity. From column 1, it is clear that the effect of being mildly, moderately and severely depressed on sleep quality remain very strong and are very close to the ones obtained in column 2 of Table 4. The same is true when including sleep patterns and efficiency in the model (column 3). Column 2 and column 4 show that individuals suffering from mild and severe depression do not appear to evaluate the VAS scale differently than those with no depressive symptoms. Respondents with moderate depression however do perceive the VAS scale differently, with their sleep quality score being 2.7 points higher than what it really should be. That being said, note that this effect is very small and that the association between sleep quality and depression remains very strong even after controlling for reporting heterogeneity.

[Table 5 about here.]

4 Discussion and Conclusion

Many studies have analyzed the association between sleep quality and depression. All found that individuals who suffer from depression have lower quality of sleep. However, very few studies have investigated whether that association could be explained, at least partially, by reporting heterogeneity. Based on a sample of students for which we have asked to complete our survey online, we confirm that sleep quality and depression are strongly related. Individuals with mild, moderate and severe depressive symptoms rated their sleep quality to be 19, 32 and 37 points lower, on a 0-100 Visual Analogue Scale, comparing with those who have no or minimal depressive symptoms. This association is robust to the inclusion of several measures of sleeping time and sleep efficiency that could potentially confound this effect. We then ask ourselves whether this association finds its explanation in reporting heterogeneity. Our analysis does indicate the presence of reporting heterogeneity in quality of sleep. We show that those who are moderately depressed tend to be pessimistic about their quality of sleep, rating it 2.7 points lower than what it really is when using the same reporting scale as those who do not suffer from reporting heterogeneity. But given the large association between depression and sleep quality, reporting heterogeneity does not appear to explain much of the depression-sleep quality gradient. Overall, this suggests that comparisons of subjective sleep quality between individuals are reliable, even between individuals suffering from depression and that reporting heterogeneity plays a minor role in moderating the relationship between sleep quality and depression.

Declaration of interest and Funding source

The authors declare that they have no conflict of interest. The authors also confirm that the funding source (HEC Research Fund 2017 - University of Lausanne, Switzerland) has had no involvment in this study whatsoever. The fund has been used exclusively to incentivize the pool of students to participate in the online survey.

5 Appendix

5.1 Vignettes used in this study

Vignette 1

[Firstname][13] falls asleep every night within 5 minutes after having gone to bed. S/He sleeps deeply during the entire night and wakes up in the morning well rested.

Vignette 2

[Firstname] does not have any problem to fall asleep in the evening and does not wake up during the night but s/he has some difficulties to wake up in the morning. S/He uses an alarm clock but falls back asleep after the alarm goes off. S/He is late to school 4 days out of 5.

Vignette 3

[Firstname] easily falls asleep in the evening, but two times per week, s/he wakes up in the middle of the night and can't fall back asleep again.

Vignette 4

[Firstname] wakes up about once every hour during the night. When s/he wakes up during the night, it takes her/him about 15 minutes to fall back asleep again. In the morning, s/he does not feel well rested.

Vignette 5

[Firstname] takes about 2 hours to fall sleep every night. S/He wakes up once or twice per night, feeling panicked and it takes her/him more than an hour to fall back to sleep again.

After reading each vignette, the respondents were asked to answer the following question: *Over the last 30 days, which level of difficulty do you think [Firstname] had with her/his sleep pattern, such as falling asleep with difficulty, waking up often during the night or too early in the morning?*

[13] Note that the firstnames used in the vignettes were sex specific.

5.2 Double-index model

We assume y_{i0}^*, the latent sleep quality of respondent i, can be modeled as a linear function of a set of factors x_i' subject to an error term ϵ_{i0} distributed as $N(0, \sigma^2)$:

$$y_{i0}^* = x_i' b_0 + \epsilon_{i0} \tag{1}$$

When sleep quality y_{i0}^* is self-reported, the measurement can be potentially biased due to reporting heterogeneity, i.e., respondents with different characteristics report value of y_{i0}^* in systematically different ways. By assuming that reporting heterogeneity can take the form of an additive and unobserved individual effect c_i, the reported value of sleep quality of respondent i, y_{i0}, takes the form:

$$y_{i0} = y_{i0}^* + c_i \tag{2}$$

In addition to these assumptions, the use of anchoring vignettes to correct for reporting differences also requires the very common vignette-related assumptions, namely, vignette equivalence and vignette consistency [30;35;42–44]. Vignette equivalence assumes that vignettes are perceived in the same way by all respondents up to an idiosyncratic error term. Under this assumption, the perception of vignette j is:

$$y_{ij}^* = b_j + \epsilon_{ij} \quad \text{for} \quad j \geqslant 1 \tag{3}$$

where b_j is the location of vignette j and ϵ_{ij} is distributed as $N(0, \sigma^2)$. On the other hand, vignette consistency assumes that respondents' evaluations of the anchoring vignettes are subject to the same reporting heterogeneity as their self-reported variable of interest (see (2)), i.e.,:

$$y_{ij} = y_{ij}^* + c_i \quad \text{for} \quad j \geqslant 1 \tag{4}$$

By defining c_i as a linear function of x_i, i.e., $c_i = x_i'\gamma$, we can plug in y_{i0}^* into (2) and y_{ij}^* into (4) to obtain the following system of equations:

$$y_{i0} = x_i' b_0 + x_i'\gamma + \epsilon_{i0} \tag{5}$$

$$y_{ij} = b_j + x_i'\gamma + \epsilon_{ij} \quad \text{for} \quad j \geqslant 1 \tag{6}$$

We can then simultaneously estimate this system of equations with OLS, where b_0 represents the true effect of x on subjective sleep quality and γ represents the effect of x on reporting heterogeneity c.

References

[1] Markku T Hyyppä and Erkki Kronholm. Quality of sleep and chronic illnesses. *Journal of clinical epidemiology*, 42(7):633–638, 1989.

[2] Christoph Augner. Associations of subjective sleep quality with depression score, anxiety, physical symptoms and sleep onset latency in students. *Central European journal of public health*, 19(2):115, 2011.

[3] David A Lombardi, Simon Folkard, Joanna L Willetts, and Gordon S Smith. Daily sleep, weekly working hours, and risk of work-related injury: Us national health interview survey (2004–2008). *Chronobiology international*, 27(5):1013–1030, 2010.

[4] Skai Schwartz, W McDowell Anderson, Stephen R Cole, Joan Cornoni-Huntley, Judith C Hays, and Daniel Blazer. Insomnia and heart disease: a review of epidemiologic studies. *Journal of psychosomatic research*, 47(4):313–333, 1999.

[5] Mélanie LeBlanc, Simon Beaulieu-Bonneau, Chantal Mérette, Josée Savard, Hans Ivers, and Charles M Morin. Psychological and health-related quality of life factors associated with insomnia in a population-based sample. *Journal of psychosomatic research*, 63(2):157–166, 2007.

[6] GJ Gerrit Niet de, BG Bea Tiemens, HHGM Bert Lendemeijer, and GJM Giel Hutschemaekers. Perceived sleep quality of psychiatric patients. *Journal of psychiatric and mental health nursing*, 15(6):465–470, 2008.

[7] Ayhan Algul, M Alpay Ates, Umit Basar Semiz, Cengiz Basoglu, Servet Ebrinc, Omer Gecici, Murat Gülsün, Ejder Kardesoglu, and Mesut Cetin. Evaluation of general psychopathology, subjective sleep quality, and health-related quality of life in patients with obesity. *The International Journal of Psychiatry in Medicine*, 39(3):297–312, 2009.

[8] Andrew G Mayers, Emma AS Grabau, Christine Campbell, and David S Baldwin. Subjective sleep, depression and anxiety: inter-relationships in a non-clinical sample. *Human Psychopharmacology: Clinical and Experimental*, 24(6):495–501, 2009.

[9] Yasuaki Hayashino, Shin Yamazaki, Misa Takegami, Takeo Nakayama, Shigeru Sokejima, and Shunichi Fukuhara. Association between number of comorbid conditions, depression, and sleep quality using the pittsburgh sleep quality index: results from a population-based survey. *Sleep medicine*, 11(4):366–371, 2010.

[10] Norihito Oshima, Atsushi Nishida, Masako Fukushima, Shinji Shimodera, Kiyoto Kasai, Yuji Okazaki, and Tsukasa Sasaki. Psychotic-like experiences (ples) and mental health status in twin and singleton japanese high school students. *Early intervention in psychiatry*, 4(3):206–213, 2010.

[11] Hiroyuki Suzuki, Yoshitaka Kaneita, Yoneatsu Osaki, Masumi Minowa, Hideyuki Kanda, Kenji Suzuki, Kiyoshi Wada, Kenji Hayashi, Takeo Tanihata, and Takashi Ohida. Clarification of the factor structure of the 12-item general health questionnaire among japanese adolescents and associated sleep status. *Psychiatry research*, 188(1):138–146, 2011.

[12] Mark Lawrence Wong, Esther Yuet Ying Lau, Jacky Ho Yin Wan, Shu Fai Cheung, C Harry Hui, and Doris Shui Ying Mok. The interplay between sleep and mood in predicting academic functioning, physical health and psychological health: a longitudinal study. *Journal of psychosomatic research*, 74(4):271–277, 2013.

[13] Nicholas Glozier, Alexandra Martiniuk, George Patton, Rebecca Ivers, Qiang Li, Ian Hickie, Teresa Senserrick, Mark Woodward, Robyn Norton, and Mark Stevenson. Short sleep duration in prevalent and persistent psychological distress in young adults: the drive study. *Sleep*, 33 (9):1139–1145, 2010.

[14] Lauren D Asarnow, Eleanor McGlinchey, and Allison G Harvey. The effects of bedtime and sleep duration on academic and emotional outcomes in a nationally representative sample of adolescents. *Journal of Adolescent Health*, 54(3):350–356, 2014.

[15] Chiara Baglioni, Svetoslava Nanovska, Wolfram Regen, Kai Spiegelhalder, Bernd Feige, Christoph Nissen, Charles F Reynolds III, and Dieter Riemann. Sleep and mental disorders: A meta-analysis of polysomnographic research. *Psychological bulletin*, 142(9):969, 2016.

[16] Hannah G Lund, Brian D Reider, Annie B Whiting, and J Roxanne Prichard. Sleep patterns and predictors of disturbed sleep in a large population of college students. *Journal of adolescent health*, 46(2):124–132, 2010.

[17] Bernd Feige, Chiara Baglioni, Kai Spiegelhalder, Verena Hirscher, Christoph Nissen, and Dieter Riemann. The microstructure of sleep in primary insomnia: an overview and extension. *International Journal of Psychophysiology*, 89(2):171–180, 2013.

[18] Caris T Fitzgerald, Erick Messias, and Daniel J Buysse. Teen sleep and suicidality: results from the youth risk behavior surveys of 2007 and 2009. *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine*, 7(4):351, 2011.

[19] Yu Jin Lee, Seong-Jin Cho, In Hee Cho, and Seog Ju Kim. Insufficient sleep and suicidality in adolescents. *Sleep*, 35(4):455–460, 2012.

[20] Jae Hong Park, Jae-Ho Yoo, and Seong Hwan Kim. Associations between non-restorative sleep, short sleep duration and suicidality: Findings from a representative sample of korean adolescents. *Psychiatry and clinical neurosciences*, 67(1):28–34, 2013.

[21] Cheng-Fang Yen, Bryan H King, and Tze-Chun Tang. The association between short and long nocturnal sleep durations and risky behaviours and the moderating factors in taiwanese adolescents. *Psychiatry research*, 179(1):69–74, 2010.

[22] Misato Matamura, Mamoru Tochigi, Satoshi Usami, Hiromi Yonehara, Masako Fukushima, Atsushi Nishida, Fumiharu Togo, and Tsukasa Sasaki. Associations between sleep habits and mental health status and suicidality in a longitudinal survey of monozygotic twin adolescents. *Journal of sleep research*, 23(3):292–296, 2014.

[23] Stephen P Becker, Melissa R Dvorsky, Alex S Holdaway, and Aaron M Luebbe. Sleep problems and suicidal behaviors in college students. *Journal of psychiatric research*, 99:122–128, 2018.

[24] Ramin Mojtabai, Mark Olfson, and Beth Han. National trends in the prevalence and treatment of depression in adolescents and young adults. *Pediatrics*, page e20161878, 2016.

[25] Eduardo Sabaté. Depression in young people and the elderly. *Priority medicines for Europe and the world, public health approach to innovation*, 6, 2004.

[26] MW Johns, TJ Gay, MD Goodyear, and JP Masterton. Sleep habits of healthy young adults: use of a sleep questionnaire. *British journal of preventive & social medicine*, 25(4):236, 1971.

[27] J Bobes, MP González, J Vallejo, J Sáiz, J Gibert, JL Ayuso, and F Rico. Oviedo sleep questionnaire (osq): A new semistructured interview for sleep disorders. *European Neuropsychopharmacology*, 8:S162, 1998.

[28] Wilse B Webb, Mike Bonnet, and Ginger Blume. A post-sleep inventory. *Perceptual and Motor Skills*, 43(3):987–993, 1976.

[29] Torbjörn Åkerstedt, KEN Hume, David Minors, and JIM Waterhouse. The subjective meaning of good sleep, an intraindividual approach using the karolinska sleep diary. *Perceptual and motor skills*, 79(1):287–296, 1994.

[30] Arie Kapteyn, James P Smith, and Arthur van Soest. Vignettes and Self-Reports of work disability in the united states and the netherlands. *Am. Econ. Rev.*, 97(1):461–473, 2007.

[31] Arthur van Soest, Liam Delaney, Colm Harmon, Arie Kapteyn, and James P Smith. Validating the use of anchoring vignettes for the correction of response scale differences in subjective questions. *J. R. Stat. Soc. Ser. A Stat. Soc.*, 174(3):575–595, 2011.

[32] Raphael Studer. Does it matter how happiness is measured? evidence from a randomized controlled experiment. *Working Paper Series, Working Paper No. 49*, 2011.

[33] Md Ismail Tareque, Nayu Ikeda, Atsushi Koshio, and Toshihiko Hasegawa. Is adjustment for reporting heterogeneity necessary in sleep disorders? results from the japanese world health survey. *BMC psychiatry*, 16:25–25, 2016.

[34] Hendrik Jürges and Joachim Winter. Are anchoring vignettes ratings sensitive to vignette age and sex? *Health Economics*, 22(1):1–13, 2013. ISSN 1099-1050. doi: 10.1002/hec.1806. URL http://dx.doi.org/10.1002/hec.1806.

[35] Gary King, Christopher J L Murray, Joshua A Salomon, and Ajay Tandon. Enhancing the validity and Cross-Cultural comparability of measurement in survey research. *Am. Polit. Sci. Rev.*, 98(01):191–207, 2004.

[36] Joshua A Salomon, Ajay Tandon, and Christopher J L Murray. Comparability of self rated health: cross sectional multi-country survey using anchoring vignettes. *BMJ*, 328(7434):258, 31 January 2004.

[37] Hanna Grol-Prokopczyk, Jeremy Freese, and Robert M. Hauser. Using anchoring vignettes to assess group differences in general self-rated health. *Journal of Health and Social Behavior*, 52(2):246–261, 2011. doi: 10.1177/0022146510396713. URL https://doi.org/10.1177/0022146510396713. PMID: 21673148.

[38] Nicole Au and Paula K. Lorgelly. Anchoring vignettes for health comparisons: an analysis of response consistency. *Quality of Life Research*, 23(6):1721–1731, Aug 2014. ISSN 1573-2649. doi: 10.1007/s11136-013-0615-2.

[39] BAG Im Auftrag des Bundesamtes für Gesundheit. Prävalenz von essstörungen in der schweiz.

[40] Sophie Vust and P-A Michaud. Médecine de l'adolescence: Troubles des conduites alimentaires atypiques. *Revue médicale suisse*, 4(139):40–43, 2008.

[41] CYNTHIA M Bulik. Anxiety, depression and eating disorders. *Eating disorders and obesity: A comprehensive handbook*, 2(1):193–198, 2002.

[42] Teresa Bago d'Uva, Teresa Bago d'Uva, Eddy Van Doorslaer, Maarten Lindeboom, and Owen O'Donnell. Does reporting heterogeneity bias the measurement of health disparities? *Health Econ.*, 17(3):351–375, 2008.

[43] Nicolai Kristensen and Edvard Johansson. New evidence on cross-country differences in job satisfaction using anchoring vignettes. *Labour Econ.*, 15(1):96–117, 2008.

[44] Nigel Rice, Silvana Robone, and Peter C Smith. Vignettes and health systems responsiveness in cross-country comparative analyses. *J. R. Stat. Soc. Ser. A Stat. Soc.*, 175(2):337–369, 2012.

List of Figures

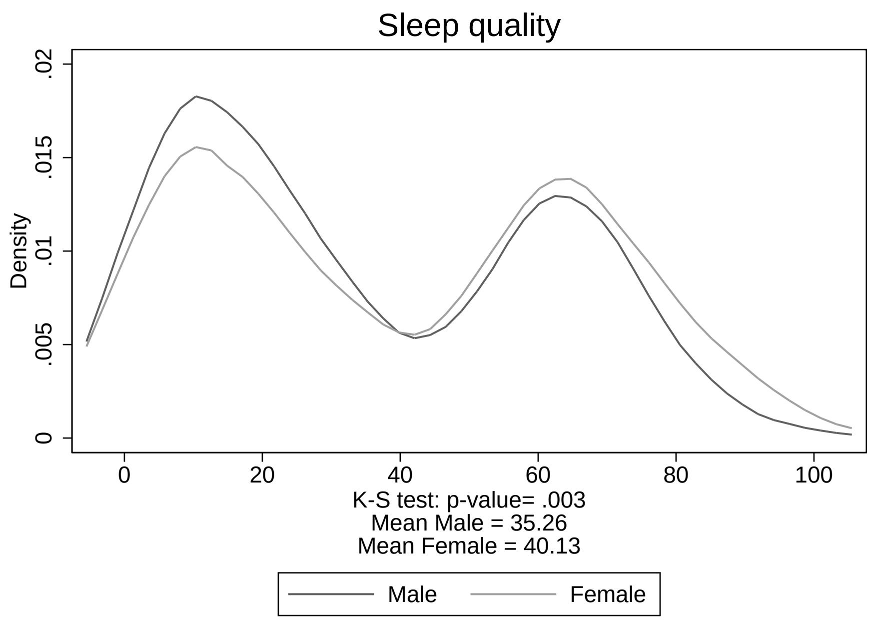

Figure 1: Distribution of the quality of sleep of the respondents by sex. 0 means 'no difficulty' and 100 means 'extreme difficulty' with sleep.

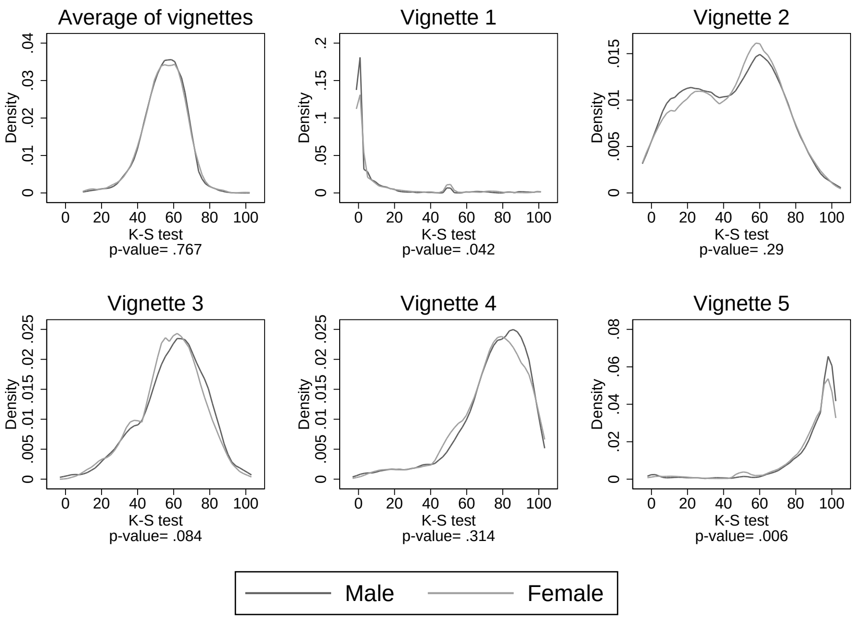

Figure 2: Distribution of the vignettes evaluations by sex. The plot on the top left is the density of the average of the respondents' vignettes evaluations. 0 means 'no difficulty' and 100 means 'extreme difficulty' with sleep.

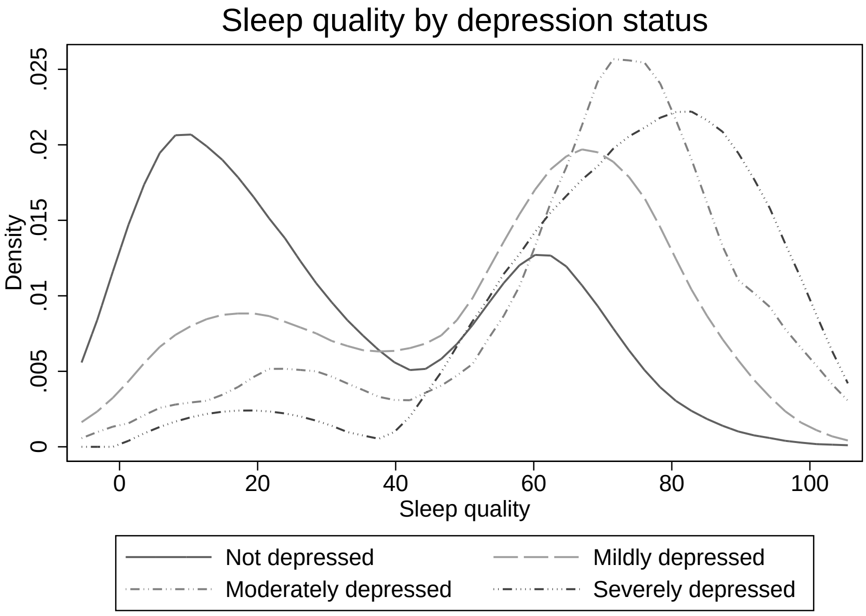

Figure 3: Distribution of the quality of sleep by depression status. 0 means 'no difficulty' and 100 means 'extreme difficulty' with sleep.

List of Tables

Table 1: Descriptive statistics

Variables		Mean or %	
		Male	Female
Age		20.91	21.06
Swiss citizenship		47.69	55.44
Single		57.60	44.45
Number of friends	*None*	2.37	2.48
	One	3.16	5.06
	Two	14.66	16.90
	Three or more	79.82	75.57
Relation with parents	*Good*	71.03	61.57
	Good, but not always	26.16	34.98
	Bad	2.82	3.44
School performance	*Above average*	31.00	23.57
	Average	56.48	62.76
	Below average	12.51	13.67
Depression	*No or minimum symptoms*	82.2	70.5
	Mild	13.4	18.4
	Moderate	3.3	8.6
	Severe	1.1	2.5
Think are obese	*Yes*	9.02	13.56
	No	89.06	83.10
	Don't know	1.92	3.34
Eating disorder	*Yes*	2.25	10.55
	No	95.38	85.58
	Don't know	2.37	3.88
Family income (in USD)	*..., 4000]*	7.55	8.50
	(4000,7000]	18.94	23.79
	(7000,12000]	33.93	33.58
	(12000, ...	29.88	23.14
	Don't know	9.70	10.98
Number of Observations		887	929

Note: Unweighted sample characteristics of the students who have completed our online survey in April 2017. Sample is restricted to respondents we were not older than 25 years of age. Total number of observations: 1,816

Table 2: Linear regressions of Sleep Quality on our set of control variables.

	(1) Sleep Quality	(2) Sleep Quality	(3) Sleep Quality	(4) Sleep Quality	(5) Sleep Quality	(6) Sleep Quality
Sex	1.077 (1.216)	1.444 (1.194)	1.204 (1.227)	1.598 (1.218)	0.835 (1.152)	0.0588 (1.110)
Age	0.192 (0.392)	0.0529 (0.384)	0.192 (0.393)	0.235 (0.390)	0.155 (0.374)	0.128 (0.358)
Mild depression	19.28*** (1.703)	17.76*** (1.684)	19.31*** (1.704)	18.89*** (1.692)	15.37*** (1.670)	13.85*** (1.577)
Moderate depression	31.95*** (2.405)	30.47*** (2.218)	31.97*** (2.387)	30.83*** (2.349)	21.48*** (2.486)	24.39*** (2.266)
Severe depression	36.70*** (3.779)	35.04*** (3.659)	36.95*** (3.737)	36.33*** (3.692)	25.49*** (4.326)	27.71*** (3.466)
Sleep time		-5.212*** (0.594)				
Hours of sleep			-0.697 (0.558)			
Hours of sleep, corrected				-2.902*** (0.541)		
Time to fall asleep (in hrs)					25.95*** (2.715)	
Takes more than 20min to fall asleep						17.67*** (1.163)
Sleep less than 7 hrs						6.671*** (1.423)
Low sleep efficiency						9.302*** (1.480)
Constant	20.62 (10.80)	62.58*** (11.70)	26.60* (11.36)	41.00*** (11.30)	9.162 (10.64)	12.70 (10.09)

Note: Standard errors at the respondent level are in parentheses (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$). We also control for parent's income, education level, nationality, number of siblings, relationship status, relationship with parents, perceived obesity, eating disorders, number of close friends and school performance. These coefficients are not reported in the table but are available upon request.

Table 3: Linear regressions of vignettes evaluations on our set of control variables.

	(1) Vignette 1	(2) Vignette 2	(3) Vignette 3	(4) Vignette 4	(5) Vignette 5
Sex	1.739*	0.236	-1.727**	-0.753	-1.342
	(0.987)	(1.237)	(0.859)	(0.981)	(0.999)
Age	0.218	0.296	-0.287	-0.242	-0.255
	(0.302)	(0.390)	(0.285)	(0.310)	(0.308)
Mild depression	0.274	-0.177	-1.016	-0.614	-1.141
	(1.263)	(1.669)	(1.118)	(1.307)	(1.382)
Moderate depression	5.036*	8.396***	0.105	2.675	-2.627
	(2.730)	(2.622)	(2.164)	(2.182)	(2.552)
Severe depression	-5.913*	9.662**	3.060	-0.705	2.748
	(3.280)	(4.405)	(2.249)	(3.194)	(3.831)
Constant	16.50**	38.96***	71.64***	79.78***	85.49***
	(8.231)	(10.36)	(7.652)	(8.612)	(8.805)

Note: Standard errors at the respondent level are in parentheses (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$). We also control for parent's income, education level, nationality, number of siblings, relationship status, relationship with parents, perceived obesity, eating disorders, number of close friends and school performance. These coefficients are not reported in the table but are available upon request.

Table 4: Linear regressions of Sleep Quality on our set of control variables, including the vignettes evaluations in columns 2 and 4.

	(1) Sleep Quality	(2) Sleep Quality	(3) Sleep Quality	(4) Sleep Quality
Sex	1.077	1.004	0.0588	-0.000311
	(1.216)	(1.209)	(1.110)	(1.101)
Age	0.192	0.173	0.128	0.112
	(0.392)	(0.391)	(0.358)	(0.356)
Mild depression	19.28***	19.35***	13.85***	13.92***
	(1.703)	(1.706)	(1.577)	(1.570)
Moderate depression	31.95***	30.49***	24.39***	22.92***
	(2.405)	(2.396)	(2.266)	(2.205)
Severe depression	36.70***	36.34***	27.71***	27.33***
	(3.779)	(3.717)	(3.466)	(3.326)
Vignette 1		0.0951***		0.101***
		(0.0345)		(0.0316)
Vignette 2		0.0909***		0.0874***
		(0.0264)		(0.0236)
Vignette 3		0.0343		0.0482
		(0.0406)		(0.0355)
Vignette 4		0.0770*		0.0735*
		(0.0441)		(0.0390)
Vignette 5		-0.00214		0.000967
		(0.0437)		(0.0410)
More than 20min to fall asleep			17.67***	17.73***
			(1.163)	(1.149)
Sleep less than 7 hrs			6.671***	6.709***
			(1.423)	(1.406)
Low sleep efficiency			9.302***	9.325***
			(1.480)	(1.479)
Constant	20.62*	7.097	12.70	-1.790
	(10.80)	(11.15)	(10.09)	(10.45)

Note: Standard errors at the respondent level are in parentheses (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$). We also control for parent's income, education level, nationality, number of siblings, relationship status, relationship with parents, perceived obesity, eating disorders, number of close friends and school performance. These coefficients are not reported in the table but are available upon request.

Table 5: Linear regressions of Sleep Quality on our set of control variables, controlling for reporting heterogeneity.

	(1) True effect b_0	(2) Reporting Heterogeneity γ	(3) True effect b_0	(4) Reporting Heterogeneity γ
Sex	1.269	-0.332	0.459	-0.319
	(1.271)	(0.554)	(1.159)	(0.554)
Age	0.231	-0.0496	0.273	-0.0554
	(0.338)	(0.148)	(0.308)	(0.148)
Mild depression	20.55***	-0.478	14.56***	-0.450
	(1.763)	(0.723)	(1.646)	(0.747)
Moderate depression	30.46***	2.707**	21.91***	2.729**
	(2.591)	(1.270)	(2.446)	(1.304)
Severe depression	37.71***	1.794	26.71***	1.761
	(3.724)	(2.175)	(3.361)	(2.189)
More than 20min to fall asleep			18.19***	-0.449
			(1.194)	(0.560)
Sleep less than 7 hrs			6.643***	0.330
			(1.493)	(0.703)
Low sleep efficiency			9.204***	0.217
			(1.609)	(0.757)
Constant	27.79***		18.52**	
	(8.084)		(7.428)	

Note: Standard errors at the respondent level are in parentheses (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$). We also control for parent's income, education level, nationality, number of siblings, relationship status, relationship with parents, perceived obesity, eating disorders, number of close friends and school performance. These coefficients are not reported in the table but are available upon request.