The Effect of Sleep on Academic Performance: A Meta-Analysis

Ruben van den Akker, Maksym Chornyi, Ika Deinum, Rayan Hassaïne, Joost Hoekstra, and Ivo Verhoeven

Introduction

Sleep deprivation is the lack of sleep, or sleeping less than one is supposed to for several nights in a row. Deaconson et al. created a more specific definition, although not universally accepted: "[Sleep deprivation is] defined as the lack of four hours of continuous sleep during the preceding 24 hours" (1). Research has shown that this is not harmless and can have detrimental effects on health and cognitive performance (2). Other research stated that total sleep deprivation showed greater correlation with performance as the duration of deprivation increased (3-7).

Our research will investigate the effects of sleep deprivation on academic performance in university students. The sleep deprivation will be measured with the Pittsburgh Sleep Quality Index (PSQI) instead of the epworth sleepiness scale (ESS), since PSQI is the most cited survey for sleep quality in the literature. Students, specifically, experience the possibility to decide their amount of sleep for themselves for the first time in their lives and are likely to experiment. This is, for instance, supported by evidence from college students where the majority reports that they are suffering from some type of sleep difficulty (8, 9). Our research will try to explain the differences in strength of the relation between sleep deprivation and academic performance by age, department study, and level of study.

There is no sound theoretical justification for a moderator analysis by age, however the assumption is made that older students would have a better indication of time management. This would mean they have a better indication of planning their academics and are less likely to be limited by time constraints.

The department of study and the level of study are based on theory. Multiple studies did research into the relation between academic performance and sleep deprivation in medical students, business and psychology students (10-13). Therefore, it is plausible that the department of study and the level of study can explain some heterogeneity in the relationship between sleep deprivation and academic performance.

Since more and more observational studies have focused on the topic of sleep and academic performance, this review aims to aggregate these results and report summary effect sizes from the synthesized data. The results can be used for future reference and give researchers insight into the possible moderators they can investigate.

Table 1. Frequencies for continent

		North America	Europe	Asia	Africa
Quantity	Participants	1976	36	1981	92
	Studies	8	1	4	1
Quality	Participants	3655	2259	2177	1035
	Studies	7	4	4	3

Methods

Search strategy. Studies were found through a systematic search of eight online scientific databases. Various combinations of the search terms displayed in table 2 were entered in the different databases. Found studies were first screened based on their titles and then based on their abstracts.

Reviewing process. Throughout the reviewing process, all data were collected and updated in a Google Spreadsheet. Reviews were done using a Google Form, in which reviewers filled out if studies met the inclusion criteria and what information about possible moderators was included. The synthesis of included articles (i.e. effect sizes) was done by hand and added to the Spreadsheet.

Inclusion criteria. A study was included if: **1)** It focused on sleep deprivation and academic performance; **2)** Sleep (quality/quantity) was included as an independent variable and GPA as the dependent variable; **3)** Its population consisted of university students; **4)** It was conducted in the past 20 years. *A complete overview of the study search and inclusion process according to PRISMA guidelines can be found in Figure 1.*

Coding. 417 full-text articles were divided over and assessed by all six researchers. Due to time constraints, each researcher individually reviewed 93 articles on average, of which around 30 were control reviews for articles reviewed by the others. Cohen's kappa coefficient was 74% initially. After initial synthesis, a second review round was conducted wherein the methodologist made ad hoc decisions regarding applicability to review format: kappa agreement increased to 98%.

Calculation and analysis of effect sizes. Pearson's r was used for calculating summary effect sizes. If a study reported a different effect size (i.e. mean comparison, Cohen's d, odds ratio, ANOVA, or t-test), these were converted to Pearson's r. To stabilize variance in terms of effect sizes, a Fisher's z transformation was applied. For interpretation, plots, and reporting, all results were back transformed to Pearson's r, however weights were calculated using the inverse z variance.

Variables. Two independent variables were formulated: *sleep quantity*, measured as number of self-reported hours slept in most studies, and *sleep quality*, which was assessed by use of the PSQI scale. The dependent variable in all analyses was *academic performance*, measured in GPA.

Data analysis. All analyses were performed in Rstudio, using R3.5.3. The packages used were *meta*, *metafor*, *metasens*, *readxl*, and *metabias*. Meta-analyses were conducted for sleep quality and sleep quantity separately, using the inverse variance method. Both random and fixed effects models were computed. The random-effects model was the preferred model, considering the large variation in countries and disciplines. The DerSimonian-Laird estimator was used for between-study variance.

Moderator analysis and publication bias. Heterogeneity was assessed with Q-statistics. To explain heterogeneity, moderator analyses were performed using the following moderators: *age*, *department of study*, and *level of study*. Publication bias was assessed through funnel plots and Rosenthal's Fail-Safe N.

Sleep Quantity and GPA

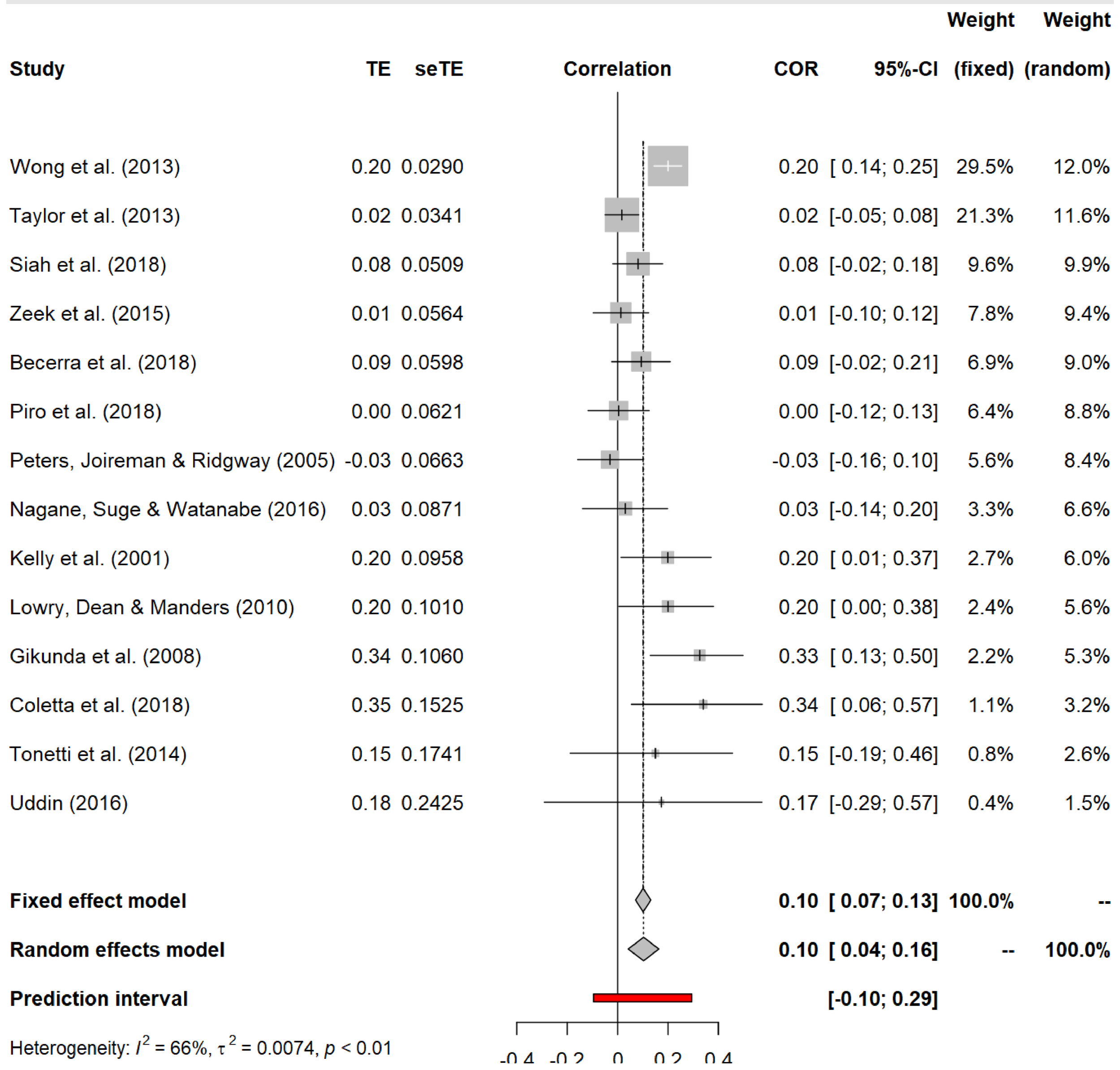

Figure 2. A forest plot for the effect of quantity of sleep on GPA.

For the meta-analysis examining the relationship between sleep duration and GPA, 14 studies fit the inclusion criteria, consisting of an aggregated sample of 4,085 university students (Median_{sample}= 183, [Min, Max: 20; 1,195]). The mean age of the students was 21.0 years and the included studies spanned 17 years (2001-2018). The summary effect (figure 2) indicated that sleep duration and academic performance are positively, but weakly, related, r = 0.10 [95%CI: 0.04; 0.16]. The effect sizes per study do vary substantially, ranging between -0.03 and 0.35. The 95% prediction interval for the dispersion of true effects is between -0.10 and 0.29. This between-study heterogeneity is significant (Q = 38.08, p = .0003). Additionally, the I^2 shows that 66% of all the variation observed is due to between-studies rather than within-study variance.

Subgroup analyses were run to examine the source of the heterogeneity between the studies. Firstly, comparing the summary effect for studies on graduate students (k=2) and studies on undergraduates (k=12); the difference is insignificant ($Q_{between}$=0.55, p = .46). A similar result was obtained for a subgroup analysis on study discipline, where three categories were observed: general (k=10), medical (k=2), and psychology (k=2); the results were again insignificant (Q = 2.29, p = .32). Frequencies are reported in table 3.

Lastly, the publication bias and possible small-study effects were examined. The funnel plot (figure 3a) displays numerous large studies, located at the top, but only a few small studies at the bottom. The plot shows no evidence for the general small-study effect, as the smallest studies are insignificant. This is supported by an insignificant Egger test (p =0.76), implying a lack of bias from publication. Rosenthal's fail-safe N is 167 (p < .0001), implying that 167 additional insignificant studies are required to create an insignificant summary effect. Duval and Tweedie's trim and fill algorithm estimated bias to be 0.02, indicating the true effect to be slightly weaker.

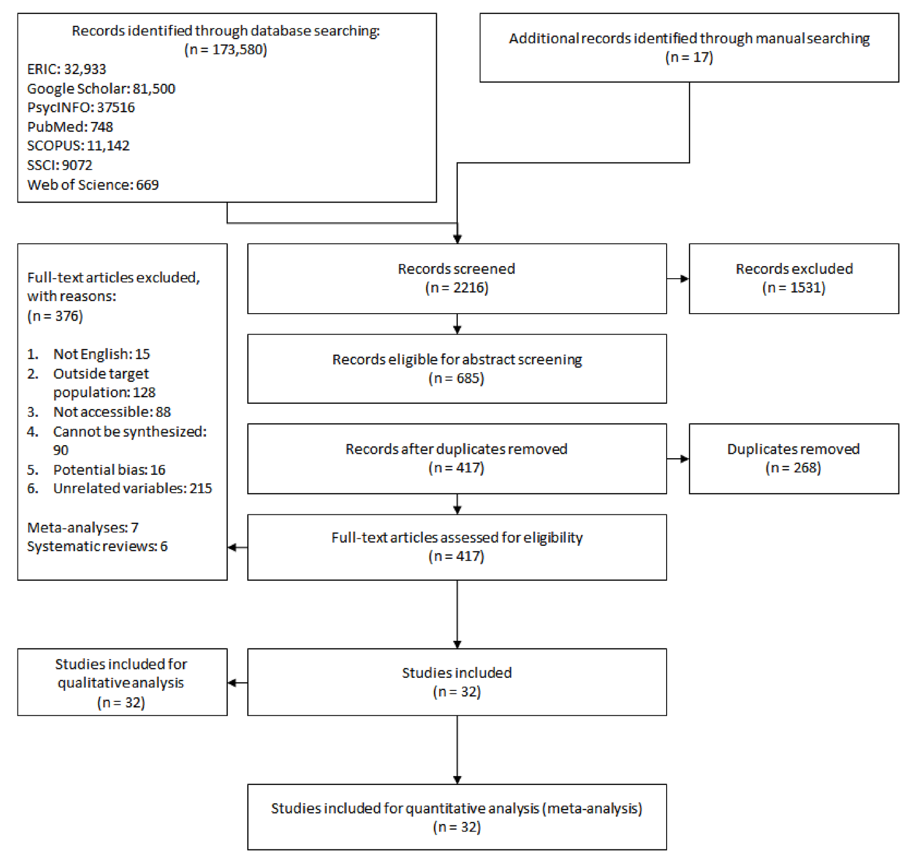

Figure 1. PRISMA flowchart of the selection of studies

Sleep Quality and GPA

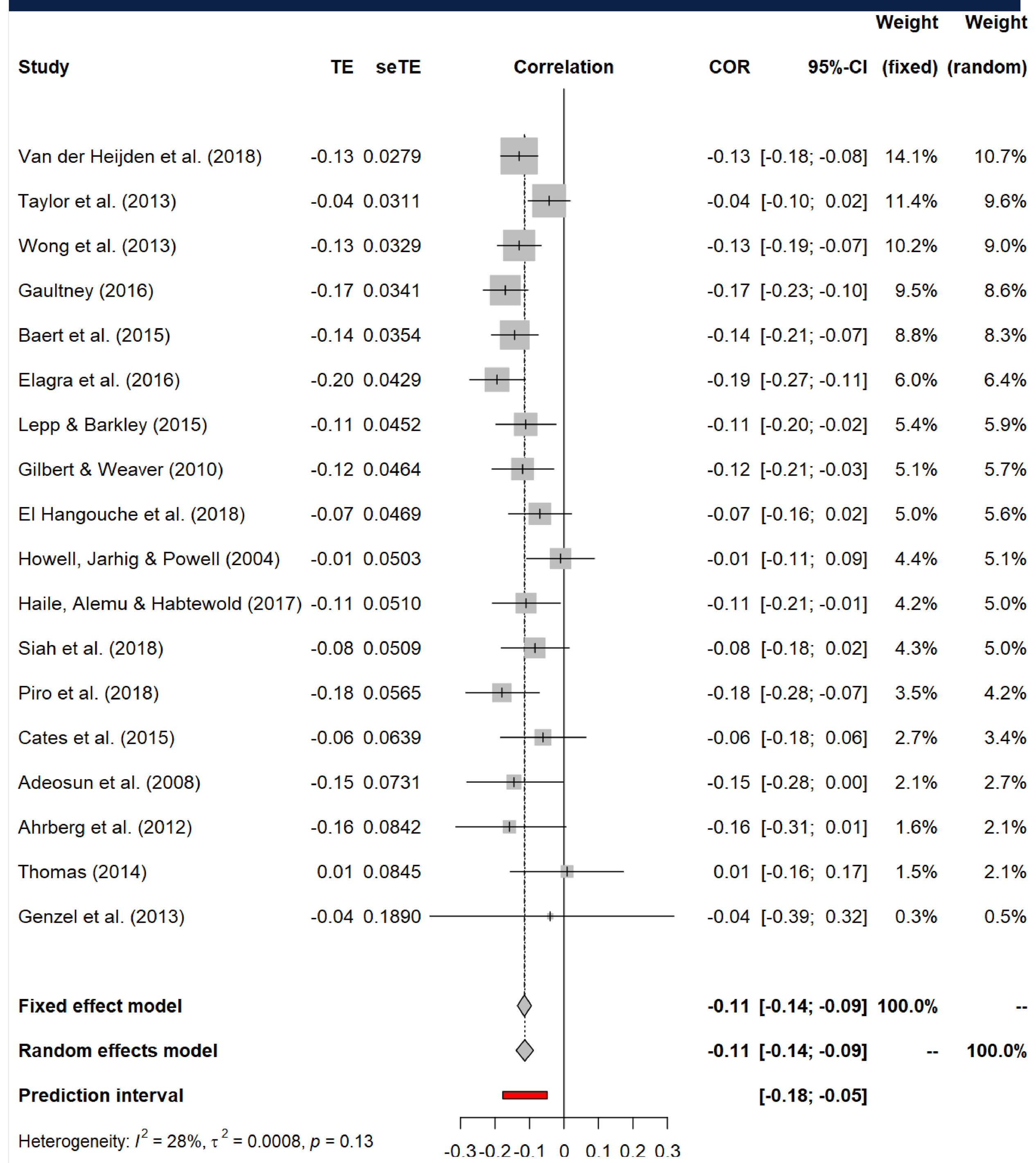

Figure 3. A forest plot for the effect of quality of sleep on GPA.

The second meta-analysis examined the relationship between sleep quality and academic performance, including 18 studies and consisting of a cumulative sample of 9,126 university students (Median_{sample}= 399, [Min, Max: 31; 1,280]). The mean age of the students was 20.4 years and the included studies spanned 10 years (2008 - 2018). The summary effect (figure 3) shows that participants indicating a higher sleep quality, which is denoted by a lower score on the PSQI scale, generally have a higher GPA. The effect size is stronger than in the previous meta-analysis, but remains weak overall, r = -0.11 [95%CI: -0.14; -0.09]. The effect sizes of different studies are grouped closer together than in the previous and the Cochran's Q statistic reflects this (Q = 23.57, p = .17). Furthermore, the I^2 is considerably lower (28%).

Similar subgroup analyses were run. The first analysis run compared the summary effect for studies on graduate students (k=4) and studies on undergraduates (k=14). The heterogeneity between groups was insignificant (Q = 2.44, p =.12). The subgroup analysis on study discipline did find significant results for heterogeneity between groups (Q = 12.48, p = .002). The GPA of participants enrolled in medical studies seemed to be most affected by poor sleep quality (r = -.16), followed by general or not reported (r = -.11), and for psychology students the summary effect turned out to be the smallest (r = -.05). Frequencies are reported in table 3.

The funnel plot (figure 3b) displays many large studies dispersed evenly around the mean effect size. It seems that the analysis is missing some small studies, as almost all studies are located at the top of the plot. Again, no evidence of the small study effect was found as the small studies are mostly insignificant. Rosenthal's fail-safe N is 650 (p < .0001), its insignificance implies that the analysis is not merely an artefact of bias. An application of Duval and Tweedie's trim and fill algorithm resulted in a publication bias of 0.0089, indicating the unbiased effect would have been slightly stronger. This evidence was corroborated by an insignificant (p = .45) but negatively sloped (b = -0.14) Egger bias correction line.

Table 2. Search termst

Population	Intervention		Outcome
students	sleep	sleep disorder	academic performance
university students	sleep deprivation	insomnia	academic probation rate
college students	sleepiness	chronotype	retention rate
undergraduates	sleep health	circadian rhythm	GPA
adolescents	sleep quality	morningness	academic achievement
	sleep duration	eveningness	

Table 3. Frequencies for sub-group moderators

		Undergrad	Grad	Medical	General	Psychology
Quantity	Participants	3748	337	418	3324	343
	Studies	12	2	2	10	2
Quality	Participants	6736	2390	2310	4879	1937
	Studies	14	4	6	8	4

Conclusion

This systematic review provides ample evidence to demonstrate that inadequate sleep has significant negative consequences on various key aspects of student functioning, affecting academic performance. All the included studies contained predominantly healthy students or students with mild sleep problems, except for some cases where the sample was characterized by the prevalence of poor mental health and the common issue of substance abuse. Moreover, there was little evidence of publication bias according to the insignificance of Begg & Egger tests, Rosenthal's N being many factors greater than the found studies and trim and fill summary effect sizes being 0.01 different for both analyses.

Sleep quality (PSQI scale) negatively correlated to GPA, according to the random-effects model the results of which did not differ from the fixed-effects model. According to Ahrberg et al. (2012), students, who are expecting lower grades, are likely to suffer from higher stress, resulting at the same time in worse sleeping quality. The high stress and low sleep quality in turn could negatively influence exam preparation and performance, which again negatively influences stress and sleep quality. The identified similarity of models might be attributable to the use of the standardized scale and the homogenous cut-off boundaries for good and bad sleep quality (i.e. PSQI>5) among the studies. This finding is consistent with the formulated hypothesis, indicating that, indeed, participants with a higher sleep quality are more successful in academic performance.

In addition, sleep duration was positively correlated with GPA, which is, likewise, in agreement with our second hypothesis. Hence, our data support Pilcher and Huffcutt's (1996) statement that sleep deprivation affects cognitive processes. As noted earlier, sleep deprivation was defined by Pilcher and Huffcutt (1996) to be functioning with less than five hours of sleep from the previous night. However, it is important to mention that the effects differed substantially across the studies, that is the increased heterogeneity was observed, compared to the first meta-analysis. A reasonable explanation for this limitation would be the prevalent variability of the definitions of sleep duration, for example, total sleep time, total time in bed, wake time - bed time, etc. Lastly, a small risk of bias was present for the effect sizes, as those could not be calculated for two out of 34 and, therefore, were excluded from the analysis despite meeting the inclusion criteria.

The results of sub-group analysis indicate that students in medical fields are more affected by sleep quality on academic performance than those who study in various departments, ranging from engineering to arts, and psychology. According to Abdulghani et al. (2012), medical students are exposed to a lot more pressure due to academic demands relative to the students from the other disciplines. Hence, the increased pressure on meeting the deadlines often leads to feelings of stress and anxiety which, therefore, restricts the desire to relax during the normal sleeping hours and stimulates the likelihood of them studying overnight.

On the other hand, this was not the case with regards to the analogous analysis for duration of sleep and overall average grade. This limitation may be attributed to the inclusion of a relatively small number of studies in the non-general subgroups, namely two for both medicine and psychology, in comparison to those whose sample consisted of a combination of various departments where the number of subgroups resulted to be 10. Despite these limitations, the overall completeness, the high quality of evidence, and the low risk of bias in the review process support the results of this review.

The conclusions of this review have important clinical implications. Sleep insufficiency and inadequacy have become ubiquitous in western societies, and students are no exception (Dewald et al, 2010). Inadequate sleep may be considered a marker or a prodrome of a wide spectrum of poor functional outcomes. Not unlike other areas of health related outcomes research, these sleep-related negative outcomes may accompany young individuals in their career transition from student to professional life, and may cause more extensive long-term damage. However, in as far as inadequate sleep is pervasive it is also amenable to prevention and treatment, for which at present, research is sorely needed."

Figure 4. The funnel plots for the included studies. Quantity on the left, quality on the right.