University of Zurich[UZH]

MA Seminar Meta-Analysis

How Emotional Exhaustion and Depersonalization are Associated among Nurses

Authors: Yannik Haller | Yang Menz | David Maurenbrecher

Student-ID: 12-918-645 | 18-723-270 | 14-728-000

Contact: `yannik.haller@uzh.ch` | `yang.menz@uzh.ch` | `david.maurenbrecher@uzh.ch`

Prof. Dr. Rainer Winkelmann

Department of Economics

University of Zurich

Zürichbergstrasse 14, CH-8032 Zurich

Date of Submission:

Tuesday 17th November, 2020

Abstract

With the health care systems globally facing a pandemic of unprecedented scale in modern times, occupations like nursing are at increased risk from high levels of stress and burnout. This meta-analysis uses a subscale decomposition of the Maslach Burnout Inventory (MBI), a commonly used three-dimensional measure of burnout, and focuses on two main subscales of the MBI and the associated key drivers. Using a random effects model to meta analytically investigate the correlation between the two focal MBI components "emotional exhaustion" and "depersonalization", we find that they move in tandem. In particular, it is found that the implied average size of the correlation between emotional exhaustion and depersonalization is 0.55. Using a weighted linear model with regions as moderators we find no evidence that this correlation is smaller for more prosperous regions. Considering facility types, however, our analysis provides evidence that the correlation is smaller in critical care facilities compared to both general hospitals and mental health facilities. A multiple-moderator weighted linear regression with continuous socio-demographic study characteristics, provides evidence that the correlation is decreasing in experience, increasing in age and negligibly affected by the gender balance within the sample. These findings provide some important insights into the driving forces of burnout among nurses, which can be used to take preventive measures against burnout and thereby maintain a high quality level of care.

Contents

Glossary

BA Bachelor's degree

CI Confidence Interval

COR Estimated correlation coefficient

DP Depersonalization

EE Emotional Exhaustion

Effect size Effect size is the Fisher-transformed correlation z_r in this paper

EU Central and western Europe

FEM Fixed-effect model

MA Meta analysis, a statistical technique for investigating the combination or interaction of a group of independent studies, for example a series of effect sizes from similar studies conducted at different centers

MBI Maslach Burnout Inventory

MBI-HSS(MP) MBI for medical personnel

MBI-HSS MBI for human services

PA Personal Accomplishment

PCEEDP Pearson correlation between EE and DP

REM Random-effect model

1 Introduction

Among health care workers, nurses historically demonstrate comparatively high rates of burnout and are often the first caregivers patients interact with. There is only limited research that focuses on the correlations between emotional exhaustion (EE), depersonalization (DP) and personal accomplishment (PA), which are the three components of the so called Maslach Burnout Inventory (MBI) – a widely used three-dimensional measure of burnout (Maslach et al., 1997). As we observe that a majority of research frequently refers to the first two perspectives of burnout and their impact on patient care quality (Doulougeri et al. 2016, Tarsis et al. 2007), it is worth to evaluate how they are interrelated regarding burnout in nurses and to investigate potential drivers for this correlation.

The term "burnout" was first used to describe a syndrome of mental weariness specifically observed among human service professionals because they were involved in emotionally demanding contacts with recipients such as clients and patients (Maslach et al., 1997). A brief self-reported questionnaire – the MBI – was developed to assess and quantify burnout. Because the MBI is used so frequently, we find a common measures of burnout across many studies. As this meta-analysis (MA) intends to study the relationship between EE and DP among nurses, the common measures in this case are the obtained scores of these two MBI subscales. The aim of this study is to uncover the drivers for the correlation between EE and DP and provide insights for burnout prevention in nursing.

This paper is organized as follows. Section 2 provides background information on the MBI, its three subscales and their interrelations, particularly the relationship between EE and DP. Section 3 discusses the sample, including the criteria and procedure for paper selection, description of chosen studies, choice of moderators as well as exploratory data analysis. Section 4 describes our hypothesis testing, estimation strategy and moderator analysis strategy. Section 5 presents the meta-analytical result for the PCEEDP among the considered papers as well as the regression results from the analyses of both numerical and categorical moderators. Section 6 concludes with a discussion of implications, methodology and directions for future research.

2 Background

The MBI is a 22 item self-administered questionnaire. It includes three subscales: EE, DP and PA. The answers are statements about the frequency of experiencing the feelings or conditions outlined in the statement (0 = never, 6 = always). Christina Maslach and Susan E. Jackson constructed the original form of the MBI. There are five versions, however, only two are relevant to our context: MBI for human services (MBI-HSS), MBI for medical personnel (MBI-HSS(MP)).

2.1 Three Subscales of the MBI

Though unitary measures are available, empirical evidence suggests that burnouts are multifaceted in nature and thus multiple subscales are appropriate to reliably capture a person's propensity for burnout. The MBI includes three dimensions constituting burnout:

EE, which refers to feelings of being depleted of one's emotional resources, represents the basic individual stress component of Burnout Syndrome and comprises 9 items (e.g. "I feel burned out from work."). Higher scores in the EE-subscale indicate higher levels of burnout.

DP, which refers to negative, cynical, or excessively detached responses to other people at work, captures the stress response towards others and comprises 5 items (e.g. "I don't really care what happens to some of my patients."). Higher scores in the DP-subscale indicate higher levels of burnout.

PA, which refers to feelings of decline in one's competence and productivity and to a lowered sense of efficacy or productivity, captures the stress response towards one-self and comprises the remaining 8 items (e.g. "At my work, I am effective at getting things done."). Lower scores in the PA-subscale indicate higher levels of burnout.

The multidimensional nature of the MBI makes it a more precise and reliable research tool with regards to burnout than single standalone measures such as exhaustion (Maslach 1993, Maslach et al., 1997).

2.2 Importance of the Relationship between EE and DP

Though related from conception the subscales of the MBI are separate. DP being theorized to represent an outward facing coping strategy for EE and PA a self-response to EE (Maslach et al. 1997). Our reasoning behind focusing on the relationship between EE and DP is twofold.

First, within the burnout literature there is no consensus on how the MBI should be translated into a binary burnout/no-burnout diagnosis, the literature seems to rely most heavily on EE and thereafter DP, with PA often being omitted or only supplemental w.r.t. to the other two subscales (Doulougeri et al. 2016). Hence it is reasonable to assume that within the field EE and DP are regarded as the more important components of MBI, making their association the most relevant cross-subscale relationship.

Second, DP is conceptually a response to EE and empirical evidence suggests that EE causes DP more so than DP causing EE (Tarsis et al. 2005). Of the MBI subscales DP has the strongest negative association with quality of care provided, suggesting it is the primary driver of loss of quality experienced by patients due to nurse burnout (Poghosyan et al. 2010, Nantsupawat et al. 2015). Hence, if we are interested in how nurse burnout affects the quality of care the patients experience, then the relationship between EE and its response in DP, which in turn drives negative quality outcomes, is the most relevant interaction within the MBI subscales.

3 Sample Studies

3.1 Searching Procedure

1. We performed general search on literature regarding burnout and MBI in nursing. The academic platforms used were Google Scholar and rechercheportal.ch. Search terms *MBI*, *burnout* and *nurse** were used and most relevant and well-cited papers selected. By reviewing and analysing similarities of the resulting 17 academic papers the precise research question (average magnitude of PCEEDP) and key moderators were defined.

2. A targeted search again on Google Scholar and rechercheportal.ch for studies published since 2000 was performed. In this instance, using combinations *MBI*, *nurse** and key moderater terms, *socio-demographic*, *mean age*, *mean experience*, *share female*, *education* as the search terms.

3. From these searches only studies with sample's consisting of >60% nurses, reporting *PCEEDP*, *sample size*, *sample country*, a *facility type* of either "general hospital", "critical care" or "mental care", and one or more of *share female*, *experience*, *education* were included. The referenced studies of these initial candidate studies were evaluated as well. Finally this yields the 19 academic papers / 26 studies that are included in our analysis.

3.2 Description of Sample Studies

As can be seen in Table 1 our sample includes 26 studies from 19 academic papers[1]. One academic paper accounts for 8 separate studies, namely Poghosyan et al. 2009. All other papers correspond to a single study. Sample sizes range from 60 to 17403 observations, while reported PCEEDP ranges from 0.388 to 0.680.

3.3 Moderator Selection

In addition to the main coefficient of interest, namely the PCEEDP, there are several other variables reported among studies in our sample. Since these variables (referred to as moderators) may have a potentially meaningful influence on the PCEEDP, we intend to reveal their relationship to the PCEEDP by means of a moderator analysis. However, due to the limited number of studies within our sample, many potential moderators (e.g. share of smokers within a sample) are not reported in enough studies to draw reasonable conclusions from an analysis. Nevertheless, we are able to come up with six interesting moderators, which are available in a majority of the sample studies[2]: indicators for the region (i.e. the variable *region*, which is partitioned into the four factors *Central & Western Europe* (EU), *Eastern Europe*, *North America* and *Asia*) and facility type (i.e. the variable *facility_type*, which is partitioned into the three factors *general hospital*, *critical care* and *mental care*) in which the study was conducted, the average age (i.e. the *age_mean* variable) and experience (i.e. the *experience_mean* variable) of the prospects, the share of individuals with a Bachelor's degree or higher (i.e. the

[1] All tables and figures presented in the study at hand are generated using the statistical software R and several packages composed for it (R Core Team, 2020). In particular, we use the package "stargazer" by Hlavac (2018) for Tables 2, 3, 4 and 5, the package "table1" by Rich (2020) for Table 2, the package "kableExtra" by Zhu (2020) for Table 1, the package "ggplot2" by Wickham (2016) for Figures 1 and 3 and the package "meta" by Balduzzi, Rücker and Schwarzer (2019) for Figures 2 and 4.

[2] For some of these moderators we have to apply a disclosure strategy to arrive at the values as we need them. A listing of the concerned moderators together with a detailed description of the strategy for reconstructing the desired value can be found in the subsection on data cleansing and processing.

Study Label	Sample Size	Pearson Correlation (PCEEDP)
Querios et al. (2013)	1157	0.388
Hu et al. (2020)	2014	0.401
Poghosyan et al. (2009) (Russia)	442	0.420
Van Bogaert et al. (2013)	1201	0.461
Iglesias et al. (2010)	80	0.476
Welp et al. (2019)	1148	0.500
Kilfedder et al. (2001)	510	0.500
Kalliath et al. (2000)	197	0.510
Garrosa et al. (2008)	473	0.520
Poghosyan et al. (2009) (Armenia)	398	0.520
Sabbah et al. (2012)	200	0.520
Hamaideh (2011)	181	0.533
Poghosyan et al. (2009) (UK)	9855	0.570
Poghosyan et al. (2009) (NZ)	4799	0.580
Robinson et al. (2003)	286	0.580
Ashtari et al. (2009)	100	0.580
Poghosyan et al. (2009) (Canada)	17403	0.590
Poghosyan et al. (2009) (Germany)	2681	0.590
Poghosyan et al. (2009) (US)	13204	0.600
Poghosyan et al. (2009) (Japan)	5956	0.600
Tselebis et al. (2001)	79	0.600
Paskvan (2011)	569	0.610
Meltzer & Huckabay (2004)	60	0.610
Lee et al. (2013)	1846	0.660
Konstantinou et al. (2018)	78	0.675
Oender & Basim (2008)	248	0.680

Table 1 – Overview of Included Studies.

BA_deg_up_share variable) as well as the share of female subjects (i.e. the *female_share* variable). How the moderator analysis is conducted with these variables and what we expect the outcomes to be is discussed in detail in Section 3.

3.4 Data cleansing & processing

The following assumptions and transformations were performed by variable (if not listed; no transformations or extra assumptions on were necessary):

age_mean: if sample age information was provided in range bins with number of observation per bin, the median of the age range was assumed as the average age of observations within that bin. For open-ended bins, missing the minimum (maximum), a lower-bound (upper-bound) of 18 (65) years was assumed.

BA_deg_up_share: where the education level was reported by categorical bins of years, we assumed individuals with over 2 years of university attained a Bachelor's degree. Similarly if education level was reported using descriptive categories, we assumed that "university level" constituted a Bachelor's degree or higher.

experience_mean: "hospital seniority", "occupational tenure" and "experience" were interpreted as years of experience. If experience information was provided in range bins with number of observation per bin, the median of the experience range was assumed as the average experience of observations within that bin. For bins lacking an upper-bound, average years in said bin was assumed to be 1.1*lower-bound.

facility_type: if the facility type was not expressly stated or was described as a "university hospital" we grouped these with "general hospitals". "Intensive care", "acute care" and "critical care" were grouped together as were "psychiatry" and "mental care".

3.5 Explanatory Data Analysis

From the 19 papers included in this MA we obtain 26 observations as Poghosyan (2009) is comprised of 8 separate samples. Of these, 7 used data collected in Asia, 5 collected in North America, 2 collected in Eastern Europe, and the remaining 12 collected in the EU. Three types of medical facilities are represented in the analysis - general hospitals (17), mental care facilities (5) and critical care facilities (4) - with sample sizes ranging from 60 to 17'403. Means, standard deviations, medians, ranges and number of observations for the continuous measures (PCEEDP, sample size, mean experience, mean age, share female, share BA degree or higher), as well as their correlations can be viewed in Table 2.

The scatter plots, presented in Figure 1, suggest that the negative correlation between *mean experience* and PCEEDP (see Table 2) is significant, so that experience may reduce PCEEDP. By the same logic, the negative correlation between *BA degree or higher* and PCEEDP seems significant, suggesting higher education may decrease PCEEDP. However, for both, *mean age* and *female share*, the observed small positive correlation as well as the scatter plots imply a more ambiguous relationship, if any at all.

Summary Statistics							Table of Correlations					
Statistic	N	Mean	St. Dev.	Min	Median	Max	1	2	3	4	5	6
PCEEDP	26	0.55	0.08	0.39	0.58	0.68	1					
sample size	26	2,506.35	4,425.87	60	491.5	17,403	0.19	1				
mean experience	22	11.51	3.82	2.28	12.14	17.70	-0.4	0.32	1			
mean age	26	35.97	4.97	27.15	35.47	45.47	0.02	0.22	0.73	1		
share female	25	0.84	0.13	0.44	0.87	1.00	0.07	0.4	0.27	0.14	1	
share BA degree or higher	16	0.39	0.26	0.04	0.36	0.84	-0.48	-0.42	-0.31	-0.45	-0.41	1

Table 2 – Table of Summary Statistics: Meta Analysis on PCEEDP among Nurses.

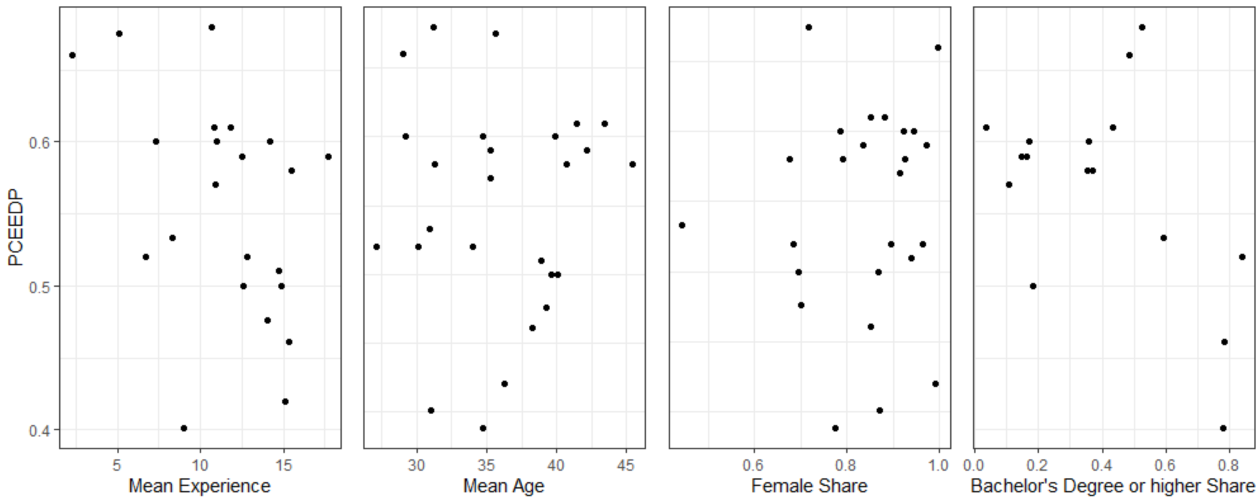

Figure 1 – Scatter Plots: PCEEDP on Continuous Variables.

4 Methodology and Hypotheses

In the first part of this section we shed light on the strategy we follow to estimate the average size of the PCEEDP as well as the tools we use to conduct the analysis. Thereafter, the procedure we apply in the moderator analysis is described in the second part. Furthermore, we denote the hypotheses we aim to investigate for both, the main measure of interest (i.e. the mean size of the PCEEDP) as well as the moderators we analyze.

4.1 Estimation Strategy

Since the sample used in the present analysis consists of $k = 26$ estimates of the PCEEDP (denoted as r_i with $i \in \{1, \ldots, k\}$), we need to ensure that the coefficients we analyze represent comparable measures. Especially the differences in the underlying sample sizes n_i are potentially problematic, as studies with a rather small number of prospects are prone to deliver biased estimators of the population correlation coefficient ρ_i. To account for this issue, we transform the reported correlation coefficient r_i using Equation 4.1, to reveal $G_{(r_i)}$ which serves as an unbiased proxy for ρ_i (DeCoster, 2004). Notice that the correction term, which is added to r to receive $G_{(r)}$, decreases with sample size n. Hence, estimates from studies with small samples are more strongly adjusted than those from studies with large samples, which is exactly what we intend to achieve.

$$G_{(r)} = r + \frac{r(1 - r^2)}{2(n - 3)} \tag{4.1}$$

Thereafter we convert the resulting coefficients by means of Fisher's z transformation for correlations given by Equation 4.2[3].

$$z_r = \frac{1}{2} \ln \left(\frac{1 + r}{1 - r} \right) \tag{4.2}$$

This transformation is necessary because most of the commonly applied models to conduct MAs rely on the assumption that the coefficient under study follows a Gaussian distribution, which usually is not the case for correlation coefficients (DeCoster, 2004). The distribution of a correlation coefficient unequal to zero exhibits a rather skewed shape, whereas the corresponding z_r value is approximately normally distributed (Silver & Dunlap, 1987). In MAs of correlations, it has therefore become a common method to first transform the correlation coefficients into the respective z_{r_i} values, to treat these z_{r_i} values thereafter as the Effect sizes T_i with which the analysis is conducted, and then to transform the results back such that they can again be interpreted as correlation sizes. As this procedure appears to be the most suitable approach in the present setting, we decide to make use of it in the study at hand. The reversed so called z-to-r transformation used thereunto is given by Equation 4.3 (DeCoster, 2004).

[3] Note that the transformations given by 4.2 and 4.3 are applied to $G_{(r)}$ in place of r.

$$r = \frac{e^{2z_r} - 1}{e^{2z_r} + 1} \tag{4.3}$$

Another great advantage of working with z_r is that its normality property allows for a convenient calculation of confidence intervals (CI). One simply has to derive the standard deviation σ_z of z_r using Equation 4.4

$$\sigma_z = \sqrt{\frac{1}{n-3}} \tag{4.4}$$

and then apply σ_z to Equation 4.5 to construct a CI for a desired significance level α.

$$CI_z = [z_r - c^*\sigma_z, z_r + c^*\sigma_z] \tag{4.5}$$

The value c^* stands for to the critical value of the normal distribution, such that the probability of z_r lying within $[-c^*, c^*]$ equals $1 - \alpha$ (DeCoster, 2004). In the present analysis we chose to calculate CIs for $\alpha = 5\%$, which corresponds to a critical value $c^* \approx 1.96$. The resulting boundaries of the CIs for each z_{r_i} value are then converted back into boundaries of CIs for the corresponding $G_{(r_i)}$ using Equation 4.3.

As soon as all k estimates r_i from our study sample are properly converted into the corresponding z_{r_i} values (which are referred to as the Effect sizes T_i for the remainder of this section), we are in a position to start with the actual MA. We apply two different commonly used models: (1) the fixed-effects model (FEM) and (2) the random-effects model (REM) as described by Hedges and Vevea (1998). Although these two models are conceptually rather similar, there is, however, a critical difference in the inference of their results. The FEM assumes that the entirety of the k samples used within the studies under consideration are drawn from a single homogeneous population. Thus, each Effect size T_i is assumed to be an estimate of the same actual Effect size θ_F, such that $T_i = \theta_F + \epsilon_i$ with $\epsilon_i \sim N(0, \sigma_i)$ $\forall i$ (meaning that differences among the T_is therefore solely arise due to within study errors ϵ_i). The main goal then is to reveal the maximum-likelihood estimator $\hat{\theta}_F$ for the true Effect size θ_F. Under the given assumptions $\hat{\theta}_F$ can be obtained by Equation 4.6, which simply is the weighted average of the effect sizes T_i using weights $w_i^{FEM} = \frac{1}{\sigma_i^2}$. The variance $Var(\hat{\theta}_F)$ is then estimated using

$\widehat{Var}(\hat{\theta}_F) = 1/\sum_{i=1}^{k} w_i^{FEM}$ (Schwarzer, Carpenter & Rücker, 2015)[4].

$$\hat{\theta}_F = \frac{\sum\limits_{i=1}^{k} T_i w_i^{FEM}}{\sum\limits_{i=1}^{k} w_i^{FEM}} = \frac{\sum\limits_{i=1}^{k} T_i \frac{1}{\sigma_i^2}}{\sum\limits_{i=1}^{k} \frac{1}{\sigma_i^2}} \tag{4.6}$$

The REM on the other hand takes into account that the k samples might be drawn from a heterogeneous population. It therefore is assumed that each T_i estimates an individual θ_i which itself is normally distributed around the average Effect size of the population (denoted as μ) with variance τ^2 (i.e. the so called between-study variance)[5] (Schwarzer et al., 2015). Formally, this translates into $T_i = \theta_i + \epsilon_i = \mu + \xi_i + \epsilon_i$ with $\epsilon_i \sim N(0, \sigma_i),\ \xi_i \sim N(0, \tau),\ \epsilon_i \perp \xi_i\ \forall i$ (Hedges & Vevea, 1998). The aim of the REM then is to estimate the average Effect size μ of the possibly heterogeneous population investigated by the studies in the study sample. Although the resulting estimates from the FEM and the REM (i.e. $\hat{\theta}_F$ and $\hat{\mu}$ respectively) have different interpretations, the formulas to obtain them look rather similar (the formula for $\hat{\mu}$ is given by Equation 4.7). The only difference is that the underlying assumptions lead to different weighting mechanisms. Namely, the weights in the REM additionally consider the between-study variance τ^2, such that the they are calculated as $w_i^{REM} = \frac{1}{\sigma_i^2+\tau^2}$. The variance $Var(\hat{\mu})$ is then estimated by $\widehat{Var}(\hat{\mu}) = 1/\sum_{i=1}^{k} w_i^{REM}$ (Hedges & Vevea, 1998; Schwarzer et al., 2015).

$$\hat{\mu} = \frac{\sum\limits_{i=1}^{k} T_i w_i^{REM}}{\sum\limits_{i=1}^{k} w_i^{REM}} = \frac{\sum\limits_{i=1}^{k} T_i \frac{1}{\sigma_i^2+\tau^2}}{\sum\limits_{i=1}^{k} \frac{1}{\sigma_i^2+\tau^2}} \tag{4.7}$$

In order to decide which model to choose, we have to assess the plausibility of either model's assumptions in the present context. Namely, we have to evaluate whether the population which our sample studies rely on can be thought of as being homogeneous or not. Schwarzer et al. (2015) describe several methods for this purpose. In the study at hand we decide to

4 Note that the composition of the weights is such that imprecise estimates (i.e. T_is with a large variance σ_i^2) have a relatively low impact on the estimated population effect size θ_F.

5 τ^2 can also be interpreted as a measure for the amount of heterogeneity present among the studies in the study sample. The assumption made in FEMs that each study in the sample investigates a random subset of the same homogeneous population is therefore equivalent to assuming $\tau^2 = 0$, which, if true, would implicate that $w_i^{FEM} = w_i^{REM}\ \forall i$ and therefore $\mu = \theta_F$ (Hedges & Vevea, 1998).

apply the common approach of using the estimated between-study variance $\hat{\tau}^2$ to perform a hypothesis test against the null hypothesis H_0 that $\tau^2 = 0$. Rejecting H_0 then means that there is some perceptible heterogeneity among the sample studies (Hedges & Vevea, 1998). Hence, we rely on the REM if H_0 is rejected, and use the FEM otherwise. To perform the present MA we chose to use the statistical software R (R Core Team, 2020). Furthermore, we largely use functions from an R-package called "meta" by Balduzzi, Rücker and Schwarzer (2019). These functions can be used for a variety of common applications when conducting MA, such as conveniently executing all of the above described calculations to estimate FEMs and REMs or creating useful visualizations (e.g. forest and funnel plots) and nicely arranged output tables to summarize the results.

The main hypothesis of this analysis concerns the correlation between the two MBI subscales EE and DP, with an emphasis on nurses. As explained in Section 2, the risk of a subject suffering a burnout increases with both of these scales. Furthermore, prior literature suggests that EE might be a causative source of DP (Taris et al., 2005). Since among the three MBI dimensions DP has the strongest and most direct negative impact on the quality of care provided, it is of great interest to reveal whether EE appears to be a strong predictor for DP. This insight could be used to take early actions to prevent high levels of DP (and thus to keep the quality of service high) when an increasing level of EE is registered in a facility. Intuitively, it can be reasonably argued that a high EE might be associated with a high DP. We therefore expect to observe a positive PCEEDP, whereas the magnitude of this correlation could be interpreted as the strength of EE as a predictor for DP. Taken together, the main conclusion for the present study can be formalized with the following central hypothesis.

Hypothesis 1: When the MBI-questionnaire is evaluated among nurses in the health sector, EE and DP exhibit a positive correlation (formally: $\text{Corr}(\text{EE,DP}) > 0$).

4.2 Strategy for Moderator Analysis

Moderator analyses are typically carried out by means of weighted linear regressions of the moderators on the coefficient on which the MA is conducted. The weights used in such a regression should correspond to the weights applied in the preferred model of the MA. However, in the special case of MAs on correlation coefficients, it is recommended to run the moderator regressions on the Fisher-transformed correlation z_r instead of directly on the correlation coefficient (Overton, 1998)[6]. Since the present analysis constitutes such a case, we use this approach and therefore treat the z_r values (calculated as described above) as the dependent variable in the subsequent moderator analysis. To get a first impression of the relationship between our continuous moderators and z_r, we create scatter plots for all these variables with the corresponding moderator on the x-axis and z_r on the y-axis. We then start by normalizing each continuous moderator (i.e. *experience_mean*, *age_mean*, *female_share* and *BA_deg_up_share*), such that their distributions are $N(0,1)$[7]. This enables a convenient comparison of the estimated coefficients resulting from the moderator regressions, since all numerical moderators are measured in the same units (namely, their standard deviation). Thereafter, we start by individually regressing each moderator on z_r by means of simple weighted linear regression models. The aim of this step is to identify those moderators which are actually able to explain the variability in z_r. Moderators which appear to have only a weak impact on z_r are then dropped before continuing to a model which incorporates multiple moderators. However, we also intuitively examine our moderators to identify those likely to affect the statistical power of the final multiple regression model, and include or exclude particular moderators when it makes sense to do so. Moreover, we consider excluding moderators who are reported significantly less frequently than others and therefore limit the size of the usable sub-sample for the entire multiple regression model. As soon as all moderator regressions described have been performed, we generate output tables with the regression results using an R package called "stargazer" by Hlavac (2018) and interpret the coefficients obtained.

The discrete moderators we analyze are encoded by the variables *region* and *facility_type*. Since poor infrastructural conditions in a facility (as likely in less developed regions) probably have a noticeable impact on the stress level during work and therefore are likely to increase EE among the nurses working there, we claim that this could also exacerbate how quickly a high EE level transitions to a high level of DP. Hence, we intuitively would expect to obtain a higher PCEEDP in relatively less developed regions of the world. Furthermore, we argue that the cultural and religious background of subjects might also have a meaningful effect on the PCEEDP. However, as these aspects (if not specifically reported in the study sample) cannot easily be generalized for whole segments of the world, we are not in a position to test these considerations. When thinking about the facility type, we claim that stress levels (and therefore also the EE) are likely to be higher in the critical care division than in a general

[6] The reason this is a good idea is because this transformation ensures that the dependent variable used in the regressions is approximately normally distributed.

[7] For the remainder of this subsection, we will therefore always refer to these normalized variables when talking about the continuous moderators.

hospital or in the mental care unit. Moreover, as the patient-fluctuation in critical care facilities is probably significantly higher than in general hospitals or in the mental care unit, we claim that nurses working in critical care facilities are much more likely to be less empathetic towards their patients. This argument would lead to the conclusion that EE and DP both are expected to be higher in critical care facilities, which therefore would hold their correlation approximately unchanged. However, another consideration arising from Garrosa et al. (2008) proposes slightly different conclusions. The study is shows, that nurses' experience with pain and death is negatively associated with both EE and DP, while the effect on DP appears to be relatively stronger. As we suspect that nurses who work in critical care units have more experience with pain and death than nurses working in hospitals or mental health facilities, we consequently expect to observe a comparably lower PCEEDP in critical care facilities. The entirety of these considerations leads us to the following hypotheses on the discrete moderators:

Hypothesis 2: The PCEEDPs obtained in the regions *Asia* and *Eastern Europe* are higher than in the regions *EU* and *North America*.

Hypothesis 3: The PCEEDP obtained in *critical care* facilities is lower than in *general* hospitals and *mental care* facilities.

In the second part of the moderator analysis we investigate how the available continuous moderators (i.e. the variables *experience_mean*, *age_mean*, *BA_deg_up_share* and *female_share*) are associated with the PCEEDP. When thinking about the potential effect of nurses' experience on EE and DP, one might reasonably argue that both are likely to be decreasing in experience. However, we suspect EE to decrease more quickly and to a larger extent than DP. This claim is justified by the assumption, that a nurse's learning curve is directly and inversely associated with movements of EE. Since learning curves are commonly assumed to have a logarithmic shape, we therefore argue that EE is likely to decrease much more strongly than DP at the beginning, whereas both scales are assumed to reach a stable level with a certain amount of nursing experience. Hence, assuming that EE and DP are on a comparable level on a nurse's first day at work, we hypothesise that the PCEEDP is decreasing in experience. Another moderator we aim to investigate is the average age of the staff. Garrosa et al. (2008) estimates the association between a nurse's age and the MBI subscales. It concludes that EE and DP decrease similarly with age, which leads to the conclusion that their correlation should not change by much. The next moderator whose effect on the PCEEDP we discuss is captured by the *BA_deg_up_share* variable. To come up with a first guess of how this relationship could look like, one might reasonably argue that the share of nurses with a Bachelor's degree or higher has a decreasing effect on EE while not affecting the DP. This thought suggests that an increasing *BA_deg_up_share* has a negative effect on the PCEEDP (assuming that EE and DP are at a similar level for non-graduate nurses). The last moderator we intend to analyse is the share of female nurses within a team. Queiros et al. (2013) reports that female nurses appear to have lower levels of DP than males. But what is the likely effect of *female_share* on EE? Assessing this relationship requires some further considerations. In the context of business teams, prior literature vastly emphasises that teams with a well balanced gender mix tend to perform significantly better than less gender-balanced teams (Hoogendoorn, Oosterbeek & Van Praag, 2013). We claim that this finding also applies to nursing teams. Since in almost all of our sample studies the share of female nurses is significantly above 50%, we suspect that decreasing *female_share* would lead to a better performance of the whole team and therefore to less EE. Combining all these considerations suggests that an increasing share of female nurses leads to a decrease in DP while increasing EE. Hence, we expect the PCEEDP to be decreasing in *female_share*. In the following we provide a condense overview of our hypotheses on the numerical moderators.

Hypothesis 4: The PCEEDP is decreasing in *experience_mean*.

Hypothesis 5: The effect of the moderator *age_mean* on the PCEEDP is insignificantly different from 0.

Hypothesis 6: The PCEEDP is decreasing in *BA_deg_up_share*.

Hypothesis 7: The PCEEDP is decreasing in *female_share*.

5 Results

In Subsection 5.1 the results of the MA on the PCEEDP are presented. Subsection 5.2 is dedicated to the results of the moderator analysis and the associated regression models we estimated.

5.1 Meta-Analysis

Figure 2 displays the so called forest plot of our MA. Such plots are widely used to report meta-analytical results, because they allow to present them in a very concise and well-arranged manner. In addition, they provide a brief overview of the sample studies, which facilitates easier understanding of how the final result is determined. In order to improve the legibility even further, we decide to sort the studies in ascending order according to their PCEEDP. For each study, Figure 2 depicts the following information: the sample sizes (contained in the column labeled as "Total"), the estimated population correlation coefficient $G_{(r_i)}$ for the PCEEDP together with the associated 95% CI (denoted in the columns labeled as "COR" and "95%-CI", respectively), a visualization of the correlation coefficient and the 95% CI (depicted in the column labeled as "Correlation") as well as the weights w_i^{FEM} and w_i^{REM} expressed as percentages of the total sum of the study-weights (contained in the columns labeled as "Weight (fixed)" and "Weight (random)", respectively).

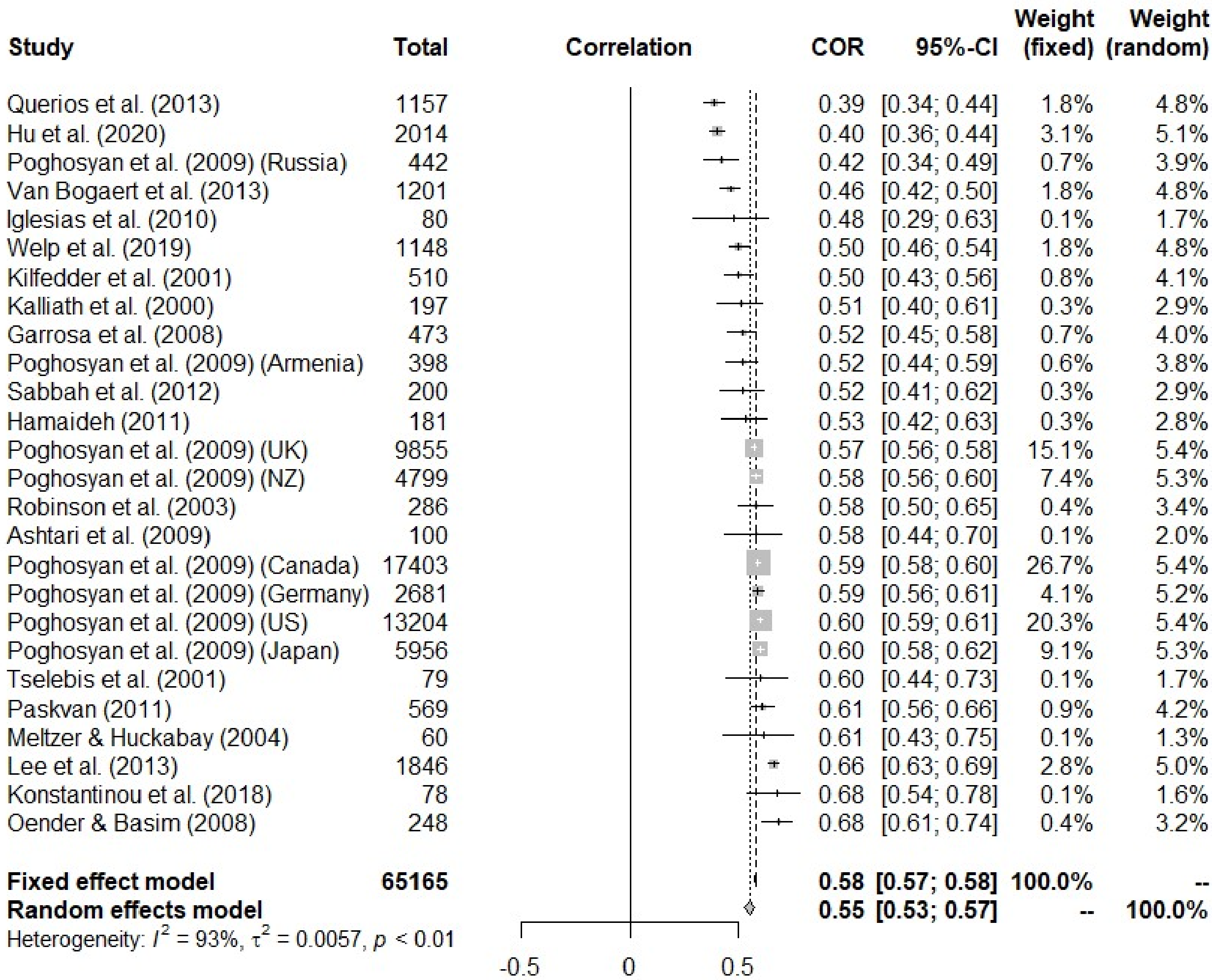

Figure 2 – Forest Plot: Meta Analysis on PCEEDP among Nurses.

The last three rows of Figure 2 present the actual results of the MA. It shows that the sample size of the included studies comprises a total of 65'165 individuals. Furthermore, we observe that the FEM estimates a population Effect size (θ_F) of 0.58 for the PCEEDP with an associated 95% CI ranging from 0.57 to 0.58. The REM on the other hand returns an estimated average PCEEDP of the population (μ) of 0.55 with a 95% CI from 0.53 to 0.57. In addition, these outcomes are listed in the column labeled "Correlation", whereas the black diamond with the corresponding dashed line indicate the results of the FEM and the grey diamond with the corresponding dotted line indicate the results of the REM. The very last line of the forest plot reports estimates for two different measures of the between-study variance (i.e. I^2 and τ^2) as well as the result of a hypothesis test against $\tau^2 = 0$ as described in Section 4[8]. We observe that the estimated τ^2 is 0.0057, which leads to a p-value for the corresponding hypothesis test of < 0.01. Hence, we can reject the null hypothesis and therefore conclude that our study sample indeed exhibits some amount of between-study variance. Consequently, we chose the REM as our preferred model and use w_i^{REM} as weights in the subsequent moderator analysis.

Another commonly used plot in MAs is the so-called funnel plot. This plot visualizes the publication bias as described by Stanley and Doucouliagos (2019). They explain that an MA may be subject to perceptible bias if the used study sample consists solely of studies published in academic journals. The reason is that studies, which deliver insignificant results are more likely to be rejected if submitted for publication, even if the reason for insignificance is solely due to a small sample size (Stanley & Doucouliagos, 2019)[9]. In the present MA, however, the publication bias does not seem to be a major concern, because a majority of studies conducted in this field do not report the PCEEDP as one of the main results. Hence, delivering an insignificant estimate of the PCEEDP is rather unlikely to cause a study to be rejected for publication and investigating a funnel plot is therefore not very meaningful in the present context. Nevertheless, for the sake of completeness we decide to include a funnel plot of our study sample in the appendix (see Figure 4 in the appendix). Since none of our sample studies appears to be inside the triangle originating from 0, we can conclude that each reported estimate for the PCEEDP is significantly different from 0.

5.2 Moderator Analysis

In the first part of the moderator analysis we focus on the weighted linear regression models we applied to our discrete moderators - namely *region* and *facility_type*. As is usual with regression models involving categorical variables, we create an indicator variable for each cat-

[8] A detailed explanation and derivation of the measure I^2 is provided by Schwarzer et al. (2015).

[9] For this reason, publication bias is sometimes referred to as small sample bias.

egory of the respective discrete moderator[10]. The multicollinearity issues that arise when trying to estimate a model that includes all of a discrete moderator's indicator variables along with an intercept are usually prevented by using one of the following methods: drop either the intercept term or one of the indicator variables from the model before estimating it. When the former technique is applied in the present (discrete) moderator regressions, the resulting estimates simply represent the weighted average z_r value in the corresponding category. If, on the other hand, the latter method is used, the category whose indicator variable is dropped serves as the reference group, meaning that that the coefficients of the included indicator variables are interpreted as the difference between the weighted average z_r value of the corresponding category and the weighted average z_r value of the reference group (which is captured by the intercept). In the present analysis we decide to apply both of these methods to our discrete moderators. Table 3 displays the results of the four models we estimated using the moderators *region* and *facility_type* (i.e. Models (1)-(4)). Models (1) and (2) concern the moderator regressions based on *region*, whereas *facility_type* is the focal moderator in Models (3) and (4). The models estimated without an intercept are Models (2) and (4). Note that the *region* category *Asia* is the reference group in Model (1) and the *facility_type* category *critical care* serves as the reference group in Model (3). The results of Model (2) reveal that the weighted average z_r value is 0.601 in *Asia*, 0.693 in *Eastern Europe*, 0.598 in *EU* and 0.665 in *North America*. Hence, we can conclude that in studies like those in our study sample the PCEEDP is on a rather similar level in *Asia* and *EU*, while the one observed in *Eastern Europe* and *North America* tends to be slightly higher, which refutes our Hypothesis 2. However, additionally considering the results from Model (1) suggests that the observed differences are not statistically significantly different from 0 regardless, as the p-values of the estimated coefficients for *Eastern Europe* and *North America* are both above 0.30[11]. We then examine the results of moderator regressions that deal with the different facility types. The estimated coefficients of Model (4) reveal that the weighted average z_r value within our sample is 0.544 among *critical care* facilities, 0.627 among *general* hospitals and 0.636 among *mental care* units. These findings are in line with our conjecture that nurses working in *critical care* facilities exhibit a lower PCEEDP than those working in *general* hospitals or *mental care* units. Furthermore, the results of Model (3) confirm this claim, since the coefficients on the

[10] An indicator variable for a category takes the value 1 when a study belongs to the corresponding category, and is 0 otherwise.

[11] Since the statistical power of the moderator regressions in the present analysis is restricted by the comparatively small sample size of 26 observations, we opt for a rather high p-value threshold of 30% to distinguish between (according to this definition) significant and insignificant results. Furthermore, we refer to an estimated coefficient as being highly significant as soon as its p-value falls below 0.10.

	Dependent variable:			
	Fisher's z			
Model	(1)	(2)	(3)	(4)
region Asia		0.601***		
		p = 0.000		
region Eastern Europe	0.091	0.693***		
	p = 0.348	p = 0.000		
region EU	−0.003	0.598***		
	p = 0.954	p = 0.000		
region North America	0.063	0.665***		
	p = 0.362	p = 0.000		
facility_type critical care				0.544***
				p = 0.000
facility_type general			0.083	0.627***
			p = 0.232	p = 0.000
facility_type mental care			0.092	0.636***
			p = 0.300	p = 0.000
Constant	0.601***		0.544***	
	p = 0.000		p = 0.000	
Observations	26	26	26	26
R^2	0.087	0.972	0.066	0.971
Adjusted R^2	−0.038	0.967	−0.016	0.967
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 3 – Results of the Simple Weighted Linear Regression Models of the Discrete Moderators on z_r.

facility types *general* and *mental care* are not only positive but also statistically significant at a 30% level. We therefore do not reject Hypothesis 3 (according to defined significance threshold) within the sample of studies used.

We then turn to the analysis of our continuous moderators - namely *experience_mean*, *age_mean*, *BA_deg_up_share* and *female_share*. By investigating Table 2 and Figure 1 we can get a first impression of their relationships with the PCEEDP. However, since the moderators are regressed on z_r instead of directly on PCEEDP, we additionally provide scatter plots of these variables versus z_r in Figure 3. Moreover, we add a red line to each of these graphics indicating the unweighted linear regression fit of the corresponding moderator on z_r. Investigating these plots suggests that *experience_mean* and *BA_deg_up_share* exhibit a clear negative relationship with z_r, while for *age_mean* and *female_share* the slope coefficients seem to be rather close to 0. It is important to note, however, that these unweighted regression lines provide just a somewhat naive initial guess of the model results that we are estimating subsequently.

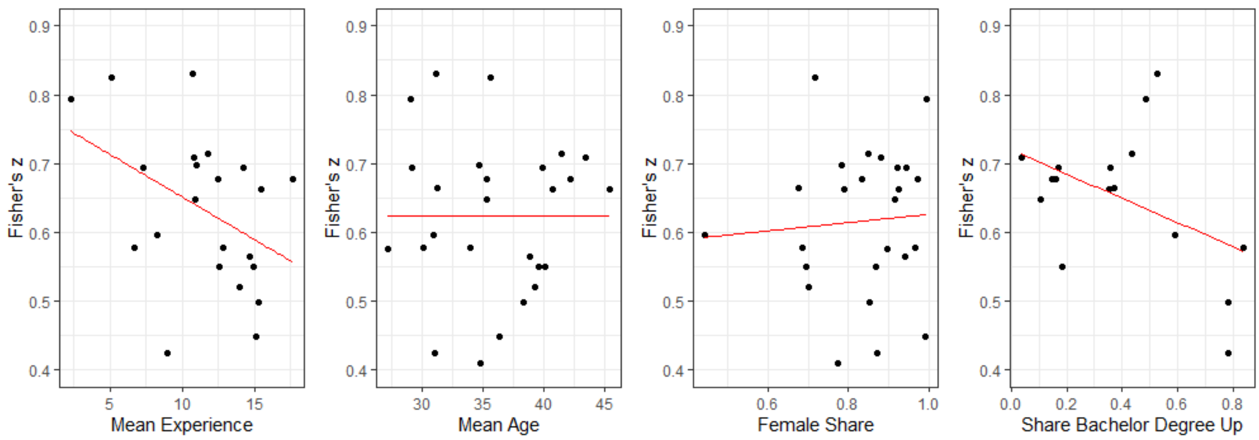

Figure 3 – Scatter Plots including OLS Regression Lines: z_r on Continuous Moderators.

Table 4 summarizes the outcomes of Models (5)-(8). In these four models, the continuous moderators are individually regressed on z_r using simple weighted linear regressions. Since these variables are normalized before we estimate the models, the resulting coefficients are all measured on the same scale and therefore have the following common interpretation: the resulting coefficients capture the shift in z_r in response to an increase by one standard deviation of the corresponding moderator. We can therefore easily identify the relative strength of the moderators as predictors of z_r, since normalization allows a direct comparison of the resulting coefficients. The result of Model (8) suggests that the moderator *BA_deg_up_share* seems to comparatively have the strongest influence on z_r, as the observed coefficient of -0.05 has the largest absolute value among the estimates of the models in Table 4. Moreover, we obtain a p-value of 0.056, which means that estimated coefficient is considered highly significant. Hence, this finding strengthens the credibility of Hypothesis 6 among studies like those considered in

Model	Dependent variable: Fisher's z (5)	(6)	(7)	(8)
experience_mean	−0.037 p = 0.107			
age_mean		0.001 p = 0.957		
female_share			0.023 p = 0.345	
BA_deg_up_share				-0.050^{*} p = 0.056
Constant	0.631^{***} p = 0.000	0.618^{***} p = 0.000	0.607^{***} p = 0.000	0.643^{***} p = 0.000
Observations	22	26	25	16
R^2	0.125	0.0001	0.039	0.237
Adjusted R^2	0.081	−0.042	−0.003	0.182
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 4 – Results of the Simple Weighted Linear Regression Models of the Continuous Moderators on z_r.

this model. However, it is important to note that the generalizability of this result is probably rather limited, since the moderator *BA_deg_up_share* is only reported in 16 studies of our sample. In Model (5) the resulting estimate for the moderator *experience_mean* is reported. We observe that a coefficient of -0.037 is estimated with an associated p-value of 0.107. Thus, *experience_mean* appears to be a slightly weaker predictor for z_r than *BA_deg_up_share*, but still exhibits a significant coefficient. Furthermore, we can see that Model (5) is estimated on a sample of 22 observations. The remaining two models of Table 4 - namely Models (6) and (7) - in contrast to those discussed above, no longer deliver any significant estimates for the focal coefficients. In Model (6) we observe an estimated coefficient of 0.001 with an associated p-value of 0.957 for *age_mean*, whereas in Model (7) an estimated coefficient of 0.023 with an according p-value of 0.345 is reported for *female_share*. In addition, it can be seen that *age_mean* and *female_share* are those continuous moderators which are reported most frequently among the studies in our sample (namely in 26 and 25 studies, respectively).

We then use the above findings to compose a weighted linear regression model containing multiple moderators (which we refer to as the final model). The focal decision is the choice of which moderators to include, which involves a trade-off between the model richness (i.e. number of included moderators) and the statistical power of the model. Hence, there are several aspects to consider. First, some of the moderators are only reported in a subset of our study sample, which limits the sample (and the degrees of freedom) available to estimate the final model when including them. We therefore decide to exclude the least frequently reported moderator *BA_deg_up_share* from the final model. The second issue is that each included variable reduces the degrees of freedom of the model and therefore harms the statistical power of the overall model. Since the inclusion of discrete moderators would imply that all appropriate indicator variables have to be added to the model, we decide to not include them within our final model. Hence, we are left with the following three moderators for the final model: *experience_mean*, *age_mean* and *female_share*. The intersection set of the subsets in which these variables are reported comprises 21 of the 26 studies in our sample. We expect that combining these three moderators within one model delivers more accurate estimates, as biases in the above regressions resulting from omitted variables can be reduced. In Table 2 is shown that *experience_mean* and *age_mean* exhibit a quite strong correlation of 0.73 (which is in line with our intuitive expectation). This gives rise to the claim that the above estimated coefficients for these two moderators are probably suspect to omitted variable bias, which we aim to eliminate by the estimation of the final model.

Table 5 displays the results of Model (9) (i.e. the final weighted linear regression model as described above). It turns out that the estimated coefficients for the included moderators all appear to be statistically significant, whereas those on *experience_mean* and *age_mean* are even highly significant. We observe that the estimated coefficient is -0.103 with p-value 0.004 for *experience_mean*, 0.09 with p-value 0.017 for *age_mean* and 0.03 with p-value 0.155 for *female_share*. These findings suggest that the estimated coefficients for *experience_mean* and *age_mean* from Models (6) and (7) were indeed subject to some bias which is counteracted in Model (9). Moreover can be seen that the adjusted R^2 of Model (9) is noticeably higher than those from all relevant models from above. Hence, these findings strongly suggest rejecting the above stated Hypotheses 5 and 7. In contrast, Hypothesis 4 appears to be supported in the studies used in our MA.

	Dependent variable:
	Fisher's z
Model	(9)
experience_mean	-0.103^{***}
	p = 0.004
age_mean	0.090^{**}
	p = 0.017
female_share	0.030
	p = 0.155
Constant	0.612^{***}
	p = 0.000
Observations	21
R^2	0.429
Adjusted R^2	0.329
Residual Std. Error	0.952 (df = 17)
F Statistic	4.263^{**} (df = 3; 17)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5 – Results of the Multiple Weighted Linear Regression Model of *experience_mean*, *age_mean* and *female_share* on z_r.

6 Conclusion

We started off with the question how EE and DP are associated among nurses. Specifically, we have established and tested seven hypotheses. Our main hypothesis (H1) states that EE and DP exhibit a positive correlation when evaluated among nurses in the health sector. In our analysis we find that the correlation between EE and DP is on average 0.55. Therefore, we do not reject our main hypothesis. Further, we hypothesize that the PCEEDPs obtained in comparatively less developed regions such as Asia and Eastern Europe are larger than in the EU and North America (H2). However, we reject this hypothesis as we find no statistically significant difference between the weighted average PCEEDPs in these regions. Thirdly, we claim that the PCEEDP observed in critical care facilities is lower than in general hospitals and mental care clinics (H3). According to the above defined and justified significance threshold of 30% for the moderator analysis, we find that the regression outcomes support this hypothesis. Fourthly, we put forward the hypotheses that the PCEEDP is decreasing in *experience_mean* (H4) and negligibly related to *age_mean* (H5). Exclusively considering the results of the single-moderator regression models, which separately examine average experience and average age of the staff as predictors for PCEEDP, suggests that neither H4 nor H5 should be rejected. However, consulting the results of the multiple-moderator regression model shown in Table 5 suggests that H5 should instead be rejected as we observe a significantly positive estimate for *age_mean*, while the conclusion that H4 should not be rejected is reinforced. Furthermore, the hypothesis that PCEEDP is decreasing in *female_share* (H7) appears to be rejected by the results of both, the single- and multiple-moderator regression model. Namely, we observe a positive coefficient for *female_share* in both of these models, whereas only in the multiple-moderator regression the coefficient is statistically significant. Lastly, regressing the share of the personnel with an education level of at least a Bachelor's degree on the Fisher-transformed PCEEDP reveals a highly significant and negative coefficient. Hence, this finding suggests that *BA_deg_up_share* is a credible indicator for the variability in PCEEDP and that we therefore should not reject hypothesis 6 (H6), in which we claim that the PCEEDP is decreasing in *BA_deg_up_share*.

Inevitably, this MA has its limitation. We were only able to explore heterogeneity of nursing staff suggested by 19 studies, including average experience, average age, female share, education level, region and facility type as moderators. One could also examine further moderators, such as family support, management style, share of smokers or share of nurses in a stable relationship. However, given the limited number of studies and the concomitant limited degrees of freedom, we were faced with a trade-off between providing a wide scope for examining heterogeneity and maintaining statistical power. A more specific limitation arises due to the small number of critical care or mental care samples included in our MA. Hence, generalizations based on conclusions to H3 on the facility type effects should be avoided. Finally, using 8 samples from Poghosayan (2009) may seem like over-reliance on one paper (and corresponding methodology). However, as we use an REM the weights are more balanced (even for studies with a large sample size n). Furthermore, the studies with the largest n and therefore largest weights are close to the "usual" or average observation, and therefore has a comparatively small leverage, which means that any potentially jeopardizing effect on the final result is rather low, despite their large n.

The main contribution of this paper is to explain how the two most frequently mentioned dimensions of the MBI (i.e. EE and DP) are correlated and to point out important factors that influence the magnitude of this correlation. By doing so, we hope to spark interest for researchers to explore this avenue that opens new possibilities to address burnout from a specific contact point rather than trying to prevent burnout through more holistic approaches. Specifically, the effect of targeted interventions for at-risk Nurses, as predicted by their socio-demographic characteristics, on quality of care, may be an interesting topic for further research.

References

Ashtari, Z., Farhady, Y., & Khodaee, M. R. (2009). Relationship between job burnout and work performance in a sample of Iranian mental health staff. *African Journal Of Psychiatry, 12*(1), 71-74.

Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: a practical tutorial. *Evidence Based Mental Health, 22*(4), 153-160. https://doi.org/10.1136/ebmental-2019-300117

DeCoster, J. (2009) *Meta-Analysis Notes.* Retrieved 30 August 2020, from http://www.stat-help.com.

Doulougeri, K., Georganta, K., & Montgomery, A. (2016). "Diagnosing" burnout among healthcare professionals: can we find consensus?. *Cogent Medicine, 3*(1), 1237605.

Garrosa, E., Moreno-Jiménez, B., Liang, Y., & González, J. (2008). The relationship between socio-demographic variables, job stressors, burnout, and hardy personality in nurses: An exploratory study. *International Journal Of Nursing Studies, 45*(3), 418-427. doi: 10.1016/j.ijnurstu.2006.09.003

Hamaideh, S. (2011). Burnout, Social Support, and Job Satisfaction among Jordanian Mental Health Nurses. *Mental Health Nursing, 32*(4), 234-242. doi: 10.3109/01612840.2010.546494

Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological methods, 3*(4), 486-504.

Hlavac, M. (2018). *stargazer: Well-Formatted Regression and Summary Statistics Tables* [Ebook] (5th ed.). Retrieved 1 November 2020, from https://cran.r-project.org/web/packages/stargazer/index.html.

Hoogendoorn, S., Oosterbeek, H., & Van Praag, M. (2013). The impact of gender diversity on the performance of business teams: Evidence from a field experiment. *Management Science, 59*(7), 1514-1528.

Hu, D., Kong, Y., Li, W., Han, Q., Zhang, X., & Zhu, L. et al. (2020). Frontline nurses' burnout, anxiety, depression, and fear statuses and their associated factors during the COVID-19 outbreak in Wuhan, China: A large-scale cross-sectional study. *Eclinicalmedicine*, 24, 100424. doi: 10.1016/j.eclinm.2020.100424

Kalliath, T., O'Driscoll, M., Gillespie, D., & Bluedorn, A. (2000). A test of the Maslach Burnout Inventory in three samples of healthcare professionals. *Work & Stress, 14*(1), 35-50. doi: 10.1080/026783700417212

Kilfedder, C., Power, K., & Wells, T. (2001). Burnout in psychiatric nursing. *Journal Of Advanced Nursing, 34*(3), 383-396. https://doi.org/10.1046/j.1365-2648.2001.01769.x

Konstantinou, A., Bonotis, K., Sokratous, M., Siokas, V., & Dardiotis, E. (2018). Burnout Evaluation and Potential Predictors in a Greek Cohort of Mental Health Nurses. *Elsevier, 32*(3), 449-456. https://doi.org/10.1016/j.apnu.2018.01.002

Lee, H., Chien, T., & Yen, M. (2012). Examining factor structure of Maslach burnout inventory among nurses in Taiwan. *Journal Of Nursing Management, 21*(4), 648-656. doi: 10.1111/j.1365-2834.2012.01427.x

Lin, F., St John, W., & McVeigh, C. (2009). Burnout among hospital nurses in China. *Journal Of Nursing Management, 17*(3), 294-301. doi: 10.1111/j.1365-2834.2008.00914.x

Lorenz, V., Benatti, M., & Sabino, M. (2010). Burnout and Stress Among Nurses in a University Tertiary Hospital. *Revista Latino-Americana De Enfermagem, 18*(6), 1084-1091. doi: 10.1590/s0104-11692010000600007

Losa Iglesias, M., Vallejo, R., & Fuentes, P. (2010). The relationship between experiential avoidance and burnout syndrome in critical care nurses: A cross-sectional questionnaire survey. *International Journal Of Nursing Studies, 47*(1), 30-37. doi: 10.1016/j.ijnurstu.2009.06.014

Maslach, C., Jackson, S., & Leiter, M. (1997). *The Maslach Burnout Inventory Manual. Evaluating Stress: A Book of Resources, 1*(3), 191-218.

Masterson, L. (2017). *Nurses are burnt out. Here's how hospitals can help.* Healthcare Dive. Retrieved 13 November 2020, from https://www.healthcaredive.com/news/nurses-are-burnt-out-heres-how-hospitals-can-help/442640/.

Meltzer, L., & Huckabay, L. (2004). Critical Care Nurses' Perceptions of Futile Care and Its Effect on Burnout. *American Journal Of Critical Care, 13*(3), 202-208. doi: 10.4037/ajcc2004.13.3.202

Nantsupawat, A., Nantsupawat, R., Kunaviktikul, W., Turale, S., & Poghosyan, L. (2016). Nurse burnout, nurse-reported quality of care, and patient outcomes in Thai hospitals. *Journal of Nursing Scholarship, 48*(1), 83-90.

Önder, Ç., & Basim, N. (2008). Examination of developmental models of occupational burnout using burnout profiles of nurses. *Journal Of Advanced Nursing, 64*(5), 514-523. doi: 10.1111/j.1365-2648.2008.04818.x

Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological methods, 3*(3), 354-379.

Paskvan, M. (2011). *Einfluss von emotionaler Dissonanz auf Burnout* (Doctoral Thesis). Universität Wien.

Poghosyan, L., Aiken, L., & Sloane, D. (2009). Factor structure of the Maslach burnout inventory: An analysis of data from large scale cross-sectional surveys of nurses from eight countries. *International Journal Of Nursing Studies, 46*(7), 894-902. https://doi.org/10.1016/j.ijnurstu.2009.03.004

Poghosyan, L., Clarke, S. P., Finlayson, M., & Aiken, L. H. (2010). Nurse burnout and quality of care: Cross-national investigation in six countries. *Research in nursing & health, 33*(4), 288-298.

Queiros, C., Carlotto, M., Kaiseler, M., Dias, S., & Pereira, A. M. (2013). Predictors of burnout among nurses: An interactionist approach. *Psicothema, 25*(3), 330-335.

Rich, B. (2020). *table1: Tables of Descriptive Statistics in HTML. R package version 1.2.* Retrieved 1 November 2020, from https://CRAN.R-project.org/package=table1

Robinson, J., Clements, K., & Land, C. (2003). Workplace STRESS Among Psychiatric Nurses. *Journal of Psychosocial Nursing And Mental Health Services, 41*(4), 32-41. https://doi.org/10.3928/0279-3695-20030401-12

Sabbah, I., Sabbah, H., Sabbah, S., Akoum, H., & Droubi, N. (2012). Burnout among Lebanese nurses: Psychometric properties of the Maslach Burnout Inventory-Human Services Survey (MBI-HSS). *Health, 4*(9), 644-652. doi: 10.4236/health.2012.49101

Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). *Meta-analysis with R* (Vol. 4784). Cham: Springer.

Silver, N. C., & Dunlap, W. P. (1987). Averaging correlation coefficients: Should Fisher's z transformation be used?. *Journal of Applied Psychology, 72*(1), 146-148.

Stanley, T. D. & H. Doucouliagos (2019). Practical significance, meta-analysis and the credibility of economics, *IZA Discussion Paper.*

Taris, T. W., Le Blanc, P. M., Schaufeli, W. B., & Schreurs, P. J. (2005). Are there causal relationships between the dimensions of the Maslach Burnout Inventory? A review and two longitudinal tests. *Work & Stress, 19*(3), 238-255.

The R Core Team. (2020). *R: A Language and Environment for Statistical Computing* [Ebook] (4th ed.). R Foundation for Statistical Computing. Retrieved 1 November 2020.

Tselebis, A., Moulou, A., & Ilias, I. (2001). Burnout versus depression and sense of coherence: Study of Greek nursing staff. *Nursing And Health Sciences, 3*(2), 69-71.

Van Bogaert, P., Kowalski, C., Weeks, S., Van heusden, D., & Clarke, S. (2013). The relationship between nurse practice environment, nurse work characteristics, burnout and job outcome and quality of nursing care: A cross-sectional survey. *International Journal Of Nursing Studies, 50*(12), 1667-1677. doi: 10.1016/j.ijnurstu.2013.05.010

Wang, S., Liu, Y., & Wang, L. (2013). Nurse burnout: Personal and environmental factors as predictors. *International Journal Of Nursing Practice, 21*(1), 78-86. doi: 10.1111/ijn.12216

Welp, A., Rothen, H., Massarotto, P., & Manser, T. (2019). Teamwork and clinician burnout in Swiss intensive care: the predictive role of workload, and demographic and unit characteristics. *Swiss Medical Weekly, 149*(1112).

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* New York: Springer. Retrieved 1 November 2020, from https://ggplot2.tidyverse.org

Zhu, H. (2020). *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax. R package version 1.3.1.* Retrieved 1 November 2020, from https://CRAN.R-project.org/package=kableExtra

Appendix

Figures

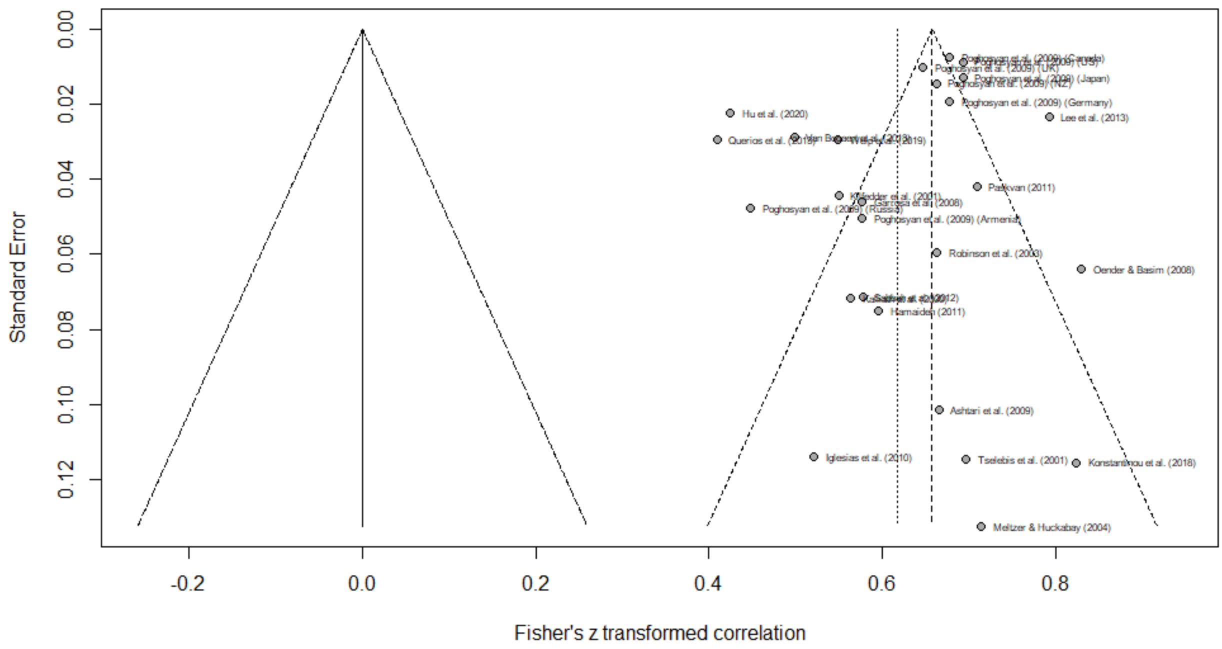

Figure 4 – Funnel Plot: Meta Analysis on PCEEDP among Nurses

R code

```
### Getting started ####
## Pls first download the input file "EE_DP_Corr_data.csv" from https://drive.google.com/file/d/1X00HAx_Q9SCB-kNksyKyyNwUS35_Zfrg/view?usp=sharing

# Set working directory
setwd("~/UZH/20HS/Meta-Analysis (S)/Final Project/R") # change directionary here

# Load packages
# Packages from the system library (do not have to be installed first)
pkg_sys <- c("datasets", "foreign", "MASS", "stats","stats4","nlme","grid")

# Additional packages (need to be installed prior to loading)
pkg <- c("dplyr","tidyr","ggplot2","stargazer","reshape2","readr","haven","dummies",
```

```
           "Hmisc","lmtest","sandwich","doBy","readxl","multiwayvcov","miceadds","car",
           "purrr","knitr","wesanderson","ggvis","lubridate","reporttools",
           "stringr", "data.table", "devtools","tinytex","rmarkdown","matlib",
           "ggiraphExtra", "gridExtra","meta","kableExtra","table1")

# lapply(pkg, install.packages, character.only = FALSE)  # install packages if not already done
invisible(lapply(c(pkg, pkg_sys), library, character.only = TRUE))
rm(list=ls())



### Read in & reshape the data ####
EE_DP_cor <- read.csv("~/Yannik/UZH/20HS/Meta-Analysis (S)/Final Project/R/EE_DP_corr_data.csv", sep = ";") # change directionary here
EE_DP_cor <- as.data.frame(EE_DP_cor)

# Create a column containing an unbiased approximation of the population correlation coefficient G(r)
# See DeCoster Eq. (5.2) p.16
EE_DP_cor$G_cor <- EE_DP_cor$cor + ((EE_DP_cor$cor*(1-(EE_DP_cor$cor^2)))/
                                              (2*(EE_DP_cor$n-3)))

# Take a look at the data and reshape it as desired
str(EE_DP_cor)
EE_DP_cor[EE_DP_cor == "N/A"] <- NA
EE_DP_cor <- as.data.table(EE_DP_cor)
setkey(EE_DP_cor, G_cor)
View(EE_DP_cor)

# Identify the studies with pearson correlation
EE_DP_cor$study_nr <- c(1:nrow(EE_DP_cor))
include <- EE_DP_cor$study_nr[EE_DP_cor$cor_type == "pearson"]
EE_DP_use <- EE_DP_cor[include]
EE_DP_cor$study_nr <- NULL
EE_DP_use$study_nr <- NULL

# Drop unused variables
EE_DP_use$MBI_type                  <- NULL
EE_DP_use$cor_type                  <- NULL
EE_DP_use$country                   <- NULL
EE_DP_use$stable_relationship_share <- NULL

# Clean dataset
str(EE_DP_use)
EE_DP_use$region             <- as.factor(EE_DP_use$region)
EE_DP_use$facility_type      <- as.factor(EE_DP_use$facility_type)
EE_DP_use$experience_mean    <- as.numeric(EE_DP_use$experience_mean)
EE_DP_use$age_mean           <- as.numeric(EE_DP_use$age_mean )
EE_DP_use$female_share       <- as.numeric(EE_DP_use$female_share)
EE_DP_use$BA_deg_up_share    <- as.numeric(EE_DP_use$BA_deg_up_share)
```

```r
str(EE_DP_use)



### Explanatory Data Analysis ####
# Create an overview table (i.e. Table 1)
Overview <- EE_DP_use[, c("study_label","n","cor")]
colnames(Overview) <- c("Study Label","Sample size","Pearson correlation")
View(Overview)

# Create a latex table using kableExtra
Overview_k <- kable(Overview, format = 'latex', digits = 3, booktabs = T,
                    linesep = "", align=c("l","r","c"))
Overview_k <- kable_styling(Overview_k, latex_options = c("striped"))
Overview_k

# Creating table of summary statistics
stargazer(EE_DP_use, out = "./sumstats.html",
          title = "Summary Statistics",
          digits = 2 ,
          omit = "G_cor",
          omit.summary.stat = c("p25","p75"),
          summary.stat = c("n","mean", "sd", "min", "median", "max"),
          covariate.labels = c("PCEEDP","sample size","mean experience", "mean age",
                               "share female","share BA degree or higher"))

# Creating correlation matrix
EE_DP_use_cor <- as.matrix(cbind(EE_DP_use$cor, EE_DP_use$n,
                                 EE_DP_use$experience_mean,
                                 EE_DP_use$age_mean,
                                 EE_DP_use$female_share,
                                 EE_DP_use$BA_deg_up_share))

cormatrix <- round(cor(EE_DP_use_cor, use = "pairwise.complete.obs"), 2)
upper <- cormatrix
upper[upper.tri(cormatrix)] <- ""
upper <- as.data.frame(upper)
colnames(upper) <- c("1", "2","3","4","5","6" )

stargazer(upper, summary=FALSE, rownames=F, out="./upper.html", title = "Table of Correlations")

# Note: the tables "sumstats.html" and "upper.html" are put side-by-side to
# obtain figure "Summary Statistics" (i.e. Table 2)

# Create scatter plots for each continuous moderator
# Mean Experience
expplot <- ggplot(EE_DP_use[!is.na(EE_DP_use$experience_mean),],
                  aes(experience_mean,cor)) +
  geom_point() +
  xlab("Mean Experience") +
```

```
  ylab("PCEEDP") +
  theme_bw() +
  theme(legend.position = "none")


# Mean Age
ageplot <- ggplot(EE_DP_use[!is.na(EE_DP_use$age_mean),],
                  aes(age_mean,cor)) +
  geom_point() +
  xlab("Mean Age") +
  ylab("PCEEDP") +
  theme_bw() +
  theme(legend.position = "none",
        axis.title.y = element_blank(),
        axis.text.y = element_blank(),
        axis.ticks.y = element_blank())


# Female Share
femplot <- ggplot(EE_DP_use[!is.na(EE_DP_use$female_share),],
                  aes(female_share,cor)) +
  geom_point() +
  xlab("Female Share") +
  ylab("PCEEDP") +
  theme_bw() +
  theme(legend.position = "none",
        axis.title.y = element_blank(),
        axis.text.y = element_blank(),
        axis.ticks.y = element_blank())

# Share of individuals with Bachelor degree or higher
BAplot <- ggplot(EE_DP_use[!is.na(EE_DP_use$BA_deg_up_share),],
                 aes(BA_deg_up_share,cor)) +
  geom_point() +
  xlab("Bachelor's Degree or higher Share") +
  ylab("PCEEDP") +
  theme_bw()+
  theme(axis.title.y = element_blank(),
        axis.text.y = element_blank(),
        axis.ticks.y = element_blank())

# Get scatter plots side by side to create Figure 1
grid.newpage()
grid.draw(cbind(ggplotGrob(expplot), ggplotGrob(ageplot), ggplotGrob(femplot),
    ggplotGrob(BAplot), size = "last"))

# Summery table by region to see if there is large differences based on region
    in variables
table1::label(EE_DP_use$cor) <- "PCEEDP"
table1::label(EE_DP_use$experience_mean) <- "Mean Experience"
table1::label(EE_DP_use$age_mean) <- "Mean Age"
table1::label(EE_DP_use$female_share) <- "Share Female"
```

```r
table1::label(EE_DP_use$BA_deg_up_share) <- "Share BA degree or higher"
table1::label(EE_DP_use$n) <- "Sample Size"

rndr <- function(x, name, ...) {
  if (!is.numeric(x)) return(render.categorical.default(x))
  what <- switch(name,
                 cor = "Mean (SD)",
                 n= "Mean (SD)",
                 experience_mean = "Mean (SD)",
                 age_mean = "Mean (SD)",
                 female_share= "Mean (SD)",
                BA_deg_up_share = "Mean (SD)")

  parse.abbrev.render.code(c("", what))(x)
}

table1(~ cor+n+experience_mean+age_mean+female_share+BA_deg_up_share| facility_type, data=EE_DP_use, overall = "All Facilities", render = rndr, output = "html")

table1(~ cor+n+experience_mean+age_mean+female_share+BA_deg_up_share| region, data=EE_DP_use, overall = "All Regions", render = rndr, output = "html")



### Meta analysis ####
# Create the meta object
meta_obj <- metacor(G_cor, n, study_label, EE_DP_use)

# Take a look at the summary
print(summary(meta_obj), digits = 2)

# Create a forest plot (i.e. Figure 2)
forest(meta_obj)

# Create a funnel plot (i.e. Figure 4)
funnel(meta_obj,studlab = T,cex.studlab = 0.5, pos.studlab = 4, ref.triangle = T)



### Moderator analysis ####
# Add a column containing Fisher's z value of G(r)
EE_DP_use$fishers_z <- meta_obj$TE

# Get the weights assigned to each study in our preferred model (i.e. the random effect model)
EE_DP_use$weight <- meta_obj$w.random

## To get a rough idea of the relationship between the continuous moderators of interest and the
## dependent variable we create a scatter-plot for each continuous moderators (
```

```
i.e. Figure 3)

grid.arrange(
# Mean Experience
ggplot(EE_DP_use[!is.na(EE_DP_use$experience_mean),],
       aes(experience_mean, fishers_z)) +
  geom_point() +
  xlab("Mean Experience") +
  ylab("Fisher's z") +
  ylim(0.4,0.9) +
  theme_bw() +
  geom_smooth(method = 'lm', formula = y~x,
              alpha = 0, col = 'red', size = 0.5),

# Mean Age
ggplot(EE_DP_use[!is.na(EE_DP_use$age_mean),],
       aes(age_mean, fishers_z)) +
  geom_point() +
  xlab("Mean Age") +
  ylab("Fisher's z") +
  ylim(0.4,0.9) +
  theme_bw() +
  geom_smooth(method = 'lm', formula = y~x,
              alpha = 0, col = 'red', size = 0.5),

# Female Share
ggplot(EE_DP_use[!is.na(EE_DP_use$female_share),],
       aes(female_share, fishers_z)) +
  geom_point() +
  xlab("Female Share") +
  ylab("Fisher's z") +
  ylim(0.4,0.9) +
  theme_bw() +
  geom_smooth(method = 'lm', formula = y~x,
              alpha = 0, col = 'red', size = 0.5),

# Share of individuals with Bachelor degree or higher
ggplot(EE_DP_use[!is.na(EE_DP_use$BA_deg_up_share),],
       aes(BA_deg_up_share, fishers_z)) +
  geom_point() +
  xlab("Share Bachelor Degree Up") +
  ylab("Fisher's z") +
  ylim(0.4,0.9) +
  theme_bw() +
  geom_smooth(method = 'lm', formula = y~x,
              alpha = 0, col = 'red', size = 0.5),
ncol = 4)

## Note: Analyzing these plots suggests that for all variables a common weighted linear regression
## model seems to be adequate.

```

```r
## Then we prepare our data for the moderator analysis

# Normalize all continuous moderators before running the regressions
EE_DP_use$experience_mean <- as.numeric(scale(EE_DP_use$experience_mean))
EE_DP_use$age_mean        <- as.numeric(scale(EE_DP_use$age_mean))
EE_DP_use$female_share    <- as.numeric(scale(EE_DP_use$female_share))
EE_DP_use$BA_deg_up_share <- as.numeric(scale(EE_DP_use$BA_deg_up_share))


# Run a weighted linear regression for each moderator of interest against the normalized
# Fishers-transformed correlations to identify potentially relevant moderators
mod_reg0 <- lm(fishers_z ~ 0 + region,
               data = EE_DP_use,
               weights = EE_DP_use$weight)

mod_reg1 <- lm(fishers_z ~ 1 + region,
                data = EE_DP_use,
                weights = EE_DP_use$weight)

mod_fac0 <- lm(fishers_z ~ 0 + facility_type,
               data = EE_DP_use,
               weights = EE_DP_use$weight)

mod_fac1 <- lm(fishers_z ~ 1 + facility_type,
               data = EE_DP_use,
               weights = EE_DP_use$weight)

mod_exp <- lm(fishers_z ~ 1 + experience_mean,
               data = EE_DP_use,
               weights = EE_DP_use$weight)

mod_age <- lm(fishers_z ~ 1 + age_mean,
               data = EE_DP_use,
               weights = EE_DP_use$weight)

mod_fem <- lm(fishers_z ~ 1 + female_share,
               data = EE_DP_use,
               weights = EE_DP_use$weight)

mod_BAd <- lm(fishers_z ~ 1 + BA_deg_up_share,
               data = EE_DP_use,
               weights = EE_DP_use$weight)

# Take a look at the results from the simple regression models
# Discrete moderators (i.e. Table 3)
stargazer(mod_reg1,mod_reg0,mod_fac1,mod_fac0,
          type = "latex",title = "Factorial Moderators", report="vc*p",
          omit.stat = c("f","ser"))

# Continuous moderators (i.e. Table 4)
stargazer(mod_exp,mod_age,mod_fem,mod_BAd,
```

```
           type = "latex", title = "Continuous Moderators", report="vc*p",
           omit.stat = c("f","ser"))


## Then we continue by taking the most valuable moderators and combine them in a multiple
## regression model
mod_mult <- lm(fishers_z ~ 1 + experience_mean +
                  age_mean + female_share, data = EE_DP_use,
                  weights = EE_DP_use$weight)

# Take a look at the results from the multiple regression models (i.e. Table 5)
stargazer(mod_mult, type = "latex", report="vc*p")
```