UNIL | Université de Lausanne

FACTORS CONTRIBUTING TO THE ONSET OF BURNOUT AMONG WORKERS: CONSIDERATION OF A SYSTEMATIC REVIEW AND META-ANALYSIS (SRMA)

A Thesis
Submitted in Partial Fulfillment of
the Requirement for the Degree of
Master of Arts in Public Opinion and Survey Methodology

Emna El May
Faculty of social and political sciences

University of Lausanne - Autumn 2020

This thesis was submitted by

Emna El May

Master thesis supervisor: Mr. Matti Langel – University of Neuchâtel

Internship supervisor: Prof. Irina Gusena Canu – Unisanté, Lausanne

Date of Defense: September 1, 2020

Modifications of forms have been made following the submission of this thesis to the Faculty.

Table of content

Abstract

Systematic review with meta-analysis (SRMA) is considered as the best source of evidence on the hierarchy of epidemiological study designs with respect to causal inference. However, this does not mean that meta-analysis is always the most convenient strategy. In some situations, other review synthesis should be preferred.
In the field of burnout, for example, the use of meta-analyses should be avoided because of current burnout's particularities in term of definition and available assessment tools. In the context of a research investigating factors responsible of burnout onset among workers, initiated by the EU-COST Action Omega-Net, I had the opportunity to 1) illustrate when and why meta-analysis is not (and should not be) the preferred review synthesis and 2) test an alternative and more convenient option through combination of vote counting and sign tests. Results of this alternative method are reported and discussed.

1 Introduction

1.1 Systematic review and meta-analysis

1.1.1 History and definitions

Gene Glass defined meta-analysis as "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" (Glass, 1976). Emergence of meta-analysis can be traced back to the 17th century with the increased interest of mathematical approaches allowing to determine how to combine observations made by different astronomers (O'Rourke, 2007). However, it is only in early 20th century, that statisticians began to apply methods to combine results from different clinical studies. Indeed, in 1904, Karl Pearson studied data from different publications in order to assess the efficacy of typhoid vaccination among soldiers serving in various parts of the British Empire (Pearson, 1904). Later, Ronald Fisher, William Cochran and Frank Yates also discussed and encouraged appropriate analysis of several studies in the domain of agriculture (Fisher, 1935; Yates and Cochran, 1938). Statistical methods for meta-analysis began to be further developed in the mid-1980s, by researchers such as Hedges and Olkin (1984).

One prerequisite to meta-analysis is a systematic review. Systematic review provides a summary of studies that tackle a particular question, and which is done following an organized, transparent and replicable protocol. Meta-analysis, in the other hand, refers to the use of *statistical techniques* leading to an overall quantitative summary of the different study-results. Although systematic review and meta-analysis (SRMA) are often commonly used in research, the former remains distinct approach that can be conducted independently from the other. The opposite however is not true: a meta-analysis always requires a systematic review as prerequisite (Last et al., 2001). SRMA have several benefits and offer numerous advantages over individual studies such as the power to solve controversy when individual studies disagree, explain variations and contradictions between studies by investigating potential moderators, enable to answer questions not previously posed in individual studies and strengthens external validity (Matt et al., 2010).

1.1.2 Popularity of SRMA

With the never-ending increasing volume of studies having similar research objectives, the use of methods allowing the synthesis of already available information exploded (Greco et al., 2013). In this context, SRMA is considered as the best source of evidence on the hierarchy of epidemiological study designs with respect to causal inference (Higgins et al.,2019; Crocetti,2016). Indeed, among alternative ways of preforming literature reviews, SRMA, when conducted properly, gives the most reliable and trustworthy synthesis. Stegenga (2011) will even say that meta-analysis is thought by many to be the *platinum* standard of evidence. The reason for high popularity of- and demand for systematic review and meta-analysis can be attributed to the high standards, rigor and strict approach (clear set of rules) it involves.

1.1.3 Theoretical frame and implementation conditions of SRMA

Systematic review starts off with a clearly formulated clinical question based on existing theory/literature. Adequate construction of the research question and bibliographic search is made through PICO strategy. Following the PICO framework, the question clearly defines the **P**opulation (individuals or population of interest, i.e. working population), **I**ntervention, **C**omparators and **O**utcomes (Schardt et al.,2007).

Also, before starting the review, a protocol set out the methods that will be used, the planned review question, the inclusion criteria, the search strategy, data extraction, quality assessment and data synthesis. Moreover, registration (through International prospective register of systematic reviews PROSPERO for example) is done in order to prevent “risk of multiple reviews addressing the same question, reduce publication bias and provide greater transparency when updating systematic reviews” (Greco et al., 2013).

1.1.4 Standards and guidelines on reporting on SRMA

Available textbooks such as the “Systematic reviews: CRD's 16 guidance for undertaking reviews in health care” (Tacconelli, 2010) and the “*Cochrane Handbook*

for Systematic Reviews of Interventions" (Higgins et al.,2019) give detailed precisions on the set of rules that need to be followed in order to perform high quality SRMAs. Alongside, guidelines such as the PRISMA (preferred Reporting Items for Systematic Reviews and Meta-analysis) are made available for researchers to report their results in a transparent and complete way. Main steps and rules necessary for conducting high quality SRMAs derived from above mentioned textbooks and guidelines are described below.

1. Eligibility criteria (inclusion and exclusion criteria) based on each PICO component are specified enough to allow the selection of all the studies relevant for answering the research question.

2. Search of the literature usually based on the PICO elements, are clearly stated and done in various databases in order to ensure that all the relevant studies will be found. Other research strategies for retrieving grey literature (e.g. reports, conference proceedings, thesis, etc.) is also recommended. Finally, in order to reduce the risk of bias (referring to the trend to publish positive results or results going in the same direction), attempts should be performed to include unpublished studies.

3. Selection of the study can be subjective and is required to be done by more than one reviewer. Agreement between reviewers, using kappa statistics, is then reported together with the reason of disagreement.

4. Data extraction is performed by at least two researchers to reduce possible errors and provide accurate and unbiased results. To this purpose, standardized data extraction form is created (including general information, study characteristics, intervention, outcome data, results etc.) and tested before implementation.

5. Data synthesis is done using well established statistical techniques and provides at least: a table describing the included studies, results of quality assessment as well as graphic to depict meta-analysis results called forest plot.

6. Assessment of heterogeneity needs to be done in order to justify use of meta-analysis. Variation of the observed estimates of studies included in meta-analysis is inevitable since no study is large enough for random error to be removed entirely. However, the percentage of variability in prevalence estimates due to heterogeneity rather than sampling error, or chance needs to be assessed. I^2 statistics has to be used to quantify this heterogeneity. "If statistical heterogeneity is observed, then the possible reasons for differences should be explored and a decision made about if and how it is appropriate to combine studies. A systematic review does not always need to include a meta-analysis and, if there are substantial differences between study estimates of effect, particularly if they are in opposing directions, combining results in a meta-analysis can be misleading" (Tacconelli, 2010).

7. Quality of the pooled studies need to be assessed since the conclusions of a meta-analysis strongly depend on the them. The quality of the individual studies should be evaluated with regard to a set of specific rules to assign a quality category, aiming for transparency and reproducibility.

8. Publication bias has to be assessed in order to prevent erroneous conclusion. Funnel plots permit visual detection of publication bias.

When carefully conducted and reported the different above-mentioned steps ensure reliable and trustworthy conclusions which justify the technique's popularity.

However, it is not because meta-analysis is seen as gold standard of evidence and that majority of handbooks prompt its correct and thoughtful use as part of systematic review, that it should be always the preferred method for research synthesis (Koricheva et al., 2013). Indeed, as suggested above, one should recognize and report when meta-analysis of effect estimates is not possible. There exist some scenarios where

the use of other research synthesis should be preferred such as when individual studies show incompletely reported outcome or effect estimate (e.g. no measure of precision, only the direction of the effect is reported,…) or when outcome is treated differently or analyzed using different methods across studies (Higgins et al.,2019).

1.2 Systematic review and meta-analysis in research on burnout

1.2.1 Short history of burnout

The word "burnout" was for the first time used by the American professor Loretta Bradley in 1969, to refer to a state of work-related exhaustion. The first scientific articles describing burnout were published by psychanalyst Herbert J. Freudenberger in 1975 who described psychological demotivation of overinvested employees working with drug addicts in rehab centers[1]. This concept was popularized by American psychologist Christina Maslach and her colleagues in 1976. Since then burnout has received growing attention from both researchers as well as practitioners and still remain subject of first focus in modern societies where the increase of people suffering from burnout keeps growing (Weber & Jaekel-Reinhard ,2000). In January 2020, the "Neue Zürcher Zeitung" newspaper published an article on the drastic increase of sick leave due to burnout among the Swiss working population since 2012. They reported that sick leave related to mental health problems arose by 70% since 2012 and that 60% of them were specifically due to work-related exhaustion or depression. [2]

1.2.2 Lack of harmonized definition and standardized measurement tool

In his famous book "Burnout : The High Cost of Achievement", Herbert Freudenberger decribes burnout as follow : "En tant que psychanalyste et praticien, je me suis rendu

[1] « Syndrome d'épuisement professionnel » in https://fr.wikipedia.org/wiki/Syndrome_d%27%C3%A9puisement_professionnel#cite_note-3 (consulted on July 18th, 2020).

[2] « Les arrêts maladie pour burn-out ont explosé depuis 2012, rapport la NZZ » sur https://www.rts.ch/info/suisse/11006478-les-arrets-maladie-pour-burnout-ont-explose-depuis-2012-rapporte-la-nzz.html (consulted on June 10th, 2020).

compte que les gens sont parfois victimes d'incendie, tout comme les immeubles. Sous la tension produite par la vie dans notre monde complexe, leurs ressources internes en viennent à se consommer comme sous l'action des flammes, ne laissant qu'un vide immense à l'intérieur, même si l'enveloppe externe semble plus ou moins intacte" (Freudenberger & Richerlson, 1981). Despite, a such state of suffering, burnout is still not considered as a *medical condition* in numerous countries[3], including Switzerland. A research conducted by Guseva-Canu et al. (2019) showed that, for only 14 over 37 European countries, burnout syndrome could be acknowledged as an occupational disease. Also, only recently (in May 2019), the World Health Organization (WHO) communicated that burnout is defined in the 11th revision of the International Classification of Diseases (ICD-11) as an "occupational phenomenon". The definition specifies that it is specific to work and that it should not be applied to describe experiences in other areas of life. Its detailed definition goes as follow: "Burn-out is a syndrome conceptualized as resulting from *chronic workplace stress* that has not been successfully managed. It is characterized by three dimensions: feelings of energy depletion or exhaustion; increased mental distance from one's job of negativism/cynicism related to one's job; and reduced professional efficacy". This definition of burnout, as a combination of three subcomponents, actually proposed for the first time by Maslach and colleagues' in the late 1980s. At that time, they developed the well-known and todays-widespread three-subscales-measurement-tool of burnout; the Maslach Burnout Inventory (MBI).

Absence of standardized definition

Although burnout has been described for the first time more than 50 years ago, there still exist no consensus of its definition among researchers and health professionals. Indeed, although the above-mentioned definition remains most prominent in the literature, other definitions have also been proposed and a common definition is still under-research. Among alternative definitions, Kristensen et al. (2005) proposed that fatigue and exhaustion are the core feature of burnout (precising that depersonalization would be a coping strategy, while reduced personal accomplishment would be a

[3] Among the 14 countries were included: Bosnia and Herzegovina, Cyprus, Denmark, Estonia, France, Hungary, Iceland, Latvia, Malta, Netherlands, Portugal, Slovakia, Sweden and Turkey

consequence rather than a defining feature of burnout). Bakker and Demerouti (2007), in the other hand, defined burnout by two core dimensions including affective, physical and cognitive exhaustion and disengagement from work.

Lack of reference method for tis diagnosis

The Maslach Burnout Inventory is currently the most widely used instrument to assess burnout among researchers (Doulougeri et al., 2016). However, depending on the authors, burnout is retrieved as a global score (unique syndrome) or by dimension. Moreover, three different versions (MBI for Human Services workers; MBI for Educator and MBI for General use) with three different recommended cut-offs are available (Maslach et al., 1996). Lack of consensus doesn't only concern the use of a specific tool; author also don't agree on the *type* of tool that should be used. As reported by O'Conner et al. (2018), other *validated* (but also self-reported) tools can be identified in the literature such as the Oldenburg Burnout Inventory (OLBI), the Copenhagen Burnout Inventory (CBI), Pines Burnout Measure, the Psychologists Burnout Inventory and the Organizational Social Context Scale (OSCS).

The lack of consensus definition precludes the development of a standardized measurement tool which inevitably results in high heterogeneity of burnout measurement in both research and practice (Rotenstein et al., 2018).

1.2.3 SMRA conducted on burnout

Many systematic reviews with meta-analysis on burnout have been published to date. Among them, authors investigated topics such as the prevalence of burnout (Rotenstein et al., 2018), interventions to prevent and reduce burnout (West et al., 2016; Ahola et al., 2017), as well as burnout determinants (Lee and Ashforth, 1996; Aronsson et al., 2017). However, the validity of those researches is questionable, especially when a meta-analysis is performed. Indeed, comparison between studies the production of a valid pooled effect is hampered by the lack of consensual definition and the heterogeneity in its assessment (Rotenstein et al., 2018).While this is recognized by Rostenstein (2018) who first planned a systematic review with meta-analysis but finally, judiciously, sticked to the systematic review because "burnout ascertainment methods, as well as statistical heterogeneity, made quantitative pooling

inappropriate". Other authors, such as Aaronson (2017), used meta-analysis anyway and reported pooled estimates.

Hence despite its advantages and increasing popularity, the use of meta-analysis requires making conscious choices in order not to report false conclusions. This is particularly applicable to the field of researches on burnout, where heterogeneity burnout assessment tools can preclude proper meta-analysis. In the state of things, alternative and more appropriate techniques should be preferred.

1.2.4 SR(MA) initiated by EU-COST Action Omega-Net

The European Cooperation in Science and Technology (COST, www.cost.eu) is a European organization for research and innovation networks where researchers can develop their ideas and initiatives. An action COST entitled The Network on the Coordination and Harmonization of European Occupational Cohorts (OMEGA-NET) launched in 2017, has the purpose to optimize the use of occupational, industrial and population cohorts at the European level. OMEGA-NET members of "Task 2.2.1 Psychosocial (Burnout)"[4] leaded by Prof. Irina Guseva Canu launched a study in order to identify *as exhaustively as* possible factors that are the cause of burnout onset among working population.

Research on burnout often focused on one some of predictors (Adriaenssens et al., 2015) with a usual focus on either occupational factors (Lee & Ashforth, 1996; Aronsson et al., 2017) or individual factors (Alarcon et al., 2009). Moreover, those studies usually concerned a specific kind of population (nurses, physicians, teachers, etc.). Also, « Few of the systematic reviews and meta-analyses of burnout in a working context published during the last 15 years have required a longitudinal design; most are based on studies applying a cross-sectional design. There have been many meta-

[4] « WG2 – Harmonization Of Existing Occupational Exposure And Health Outcome Information" in https://omeganetcohorts.eu/working-groups/wg2-harmonisation-occupational-exposur-health-outcome-information/ (consulted July 21st , 2020)

analyses of different occupational groups mostly based on studies applying cross-sectional design ». (Aronsson et al., 2017).

It hence appeared that a systematic overview of the complete panel of possible determinants (including both individual and work-related factors) that prospectively causes burnout in workers, was lacking (Canu et al., 2019). Once addressed accurately, the results of that research would be 1) helpful for the setting of appropriate interventions as well as preventive interventions on factors having significant effect on burnout onset and 2) facilitating decisions regarding individual case of burnout that is determining to justify compensation. Pertaining to this latter, in the above-mentioned study by Canu et al. (2019), we can notice among the few countries which could recognize burnout as an occupational disease, number of persons who were actually compensated by year ranged from 0 to 59.2 only.

To address the question of burnout determinants, meta-analysis was originally planned. However, it was quickly precluded mainly because of variation in predictors and burnout assessment tools as well as the type of effect estimate selected from individual studies. Reasons for precluded meta-analysis, adopted method as well as preliminary results are discussed in next section.

2 Determinants of burnout onset among workers

2.1 Systematic review (as described in PROSPERO) [5]

Articles from 1990 (January) to 2018 (august) with no language limitation were searched in three databases including MEDLINE, PsycINFO and Embase via Ovid.

Titles and abstracts were screened against the eligibility criterion cited below. For studies that are not excluded on the basis of the title/abstract, full text papers was requested and assessed by two reviewers. Any discrepancies were resolved through discussion, and if required, a third reviewer. Studies were screened in both waves (title/abstract and full text articles) using the COVIDENCE systematic reviewing tool.

[5] PROSPERO 2018 CRD42018105901 Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018105901

Approach consistent with the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) was adopted and is depicted in Figure 1.

2.1.1 Eligibility criteria

In order to avoid systemic bias and to guarantee the reproducibility and openness of the study selection strategy, grey literature was not considered. Sources that had less than 50 persons in the exposed group were excluded. However, studies that examined the importance of any risk factor for the onset of burnout among workers were included. Only studies that were relevant for European conditions and which focused on people at work were included. Likewise, samples that examined academic burnout were excluded because they are not the focus of the present study, burnout in the workplace. However, studies that utilized employed students were included in the current study (e.g. Dahlin et al., 2010). In order to control for the time sequence between exposure and the onset of burnout, the choice of study design was limited to longitudinal design and studies that assessed exposure after the burnout onset were excluded. In case where multiple publications describing a single study are identified, the study was included only once and the choice of one of the publications was made as the primary reference under which all the others were listed.
These inclusion criteria yielded a final total of 240 samples from published studies.

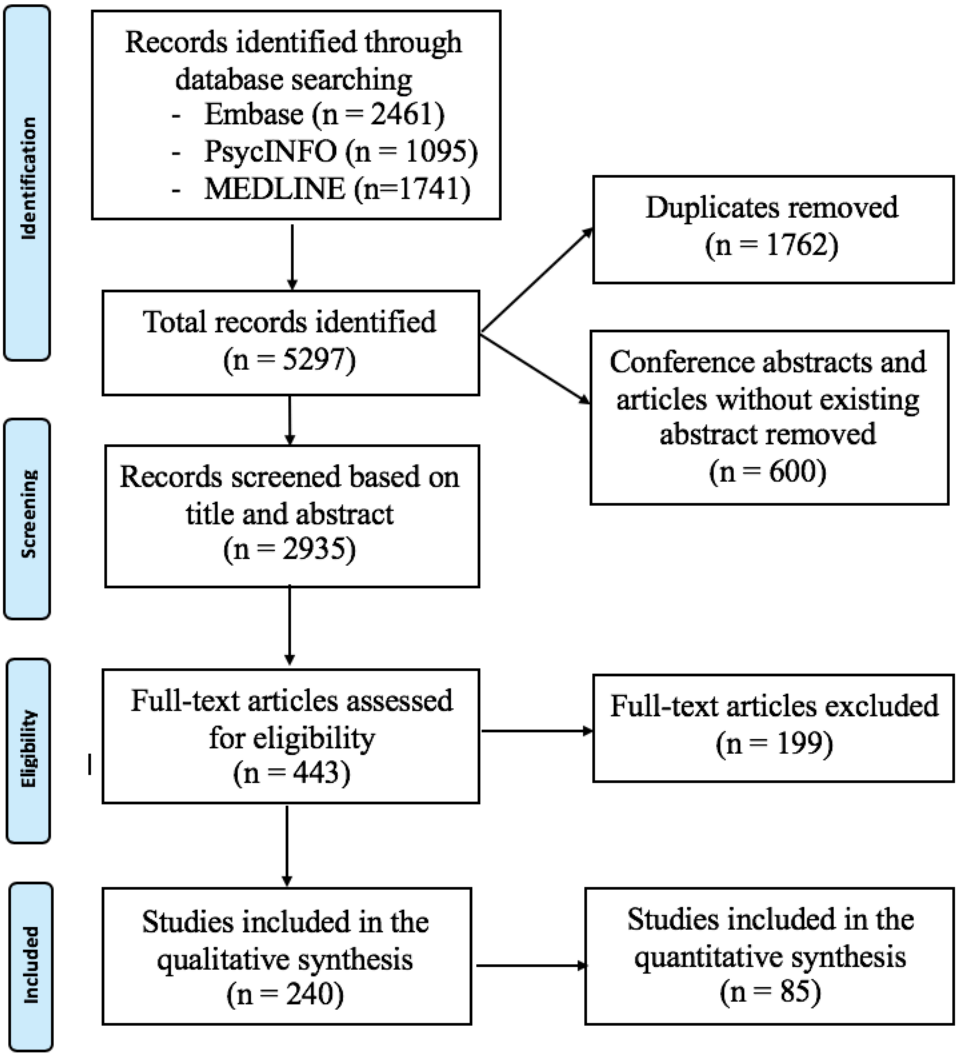


Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) flow diagram. [Retrieved and adapted from task 2.2.1]

2.1.2 Data Extraction

The following data was extracted from each article using a standardized form: Study details including date of study, title, author, country, occupational groups, participants mean age; Research question including hypothesis tested, hypothesis result (confirmed/not confirmed); Methods including study design, predictive variables name and definition, how it is measured/reported; outcome name and definition, outcome type, how it is measured/reported, statistical method used; Results including, actual results for each pair of outcome and predictive variable separately (slope or risk

ratio/equivalent or other) and variability value associated with result (confidence interval, standard error, p-value, other). In case articles assessed burnout and other outcome (e.g. stress) only data concerning burnout was reported. Moreover, data from the studies reporting the results as calculations of association (i.e. risk ratios, rate ratios, odds ratios, and relative risks), using statistical modeling (e.g., multiple logistic regression) was extracted for potential further statistical analysis. When studies contained several models, direct effects from the most adjusted model was extracted. The two mains rationales for that are: 1) past research noticed that the differences between the least adjusted models and the more adjusted ones were pretty small (Aronsson et al., 2017; Van der Molen, 2020) and 2) it better reflects the real relationship between the independent variable and dependent.

2.2 Inclusion criteria into semi-quantitative synthesis

Studies that reported emotional exhaustion dimension score from the Maslach burnout inventory were included. We also considered other measurement tools as long as they reported an emotional exhaustion score. However, studies that only reported an overall burnout score and/or measures other than emotional exhaustion were excluded. The reasons for that are multiple, among them: 1) the combination of subscales scores is not recommended because dimensions are seen as three distinct construct and combining them would reduce the instrument's reliability (Maslach et al., 2001), 2) for many emotional exhaustion is considered as the "core meaning" of burnout (Shirom, 1989; Moore, 2000). Already in 1989, Koeske et al. (1989) showed how emotional exhaustion predict depersonalization and personal accomplishment, the two other components of burnout. Also, Kristensen et al. (2005) proposed that depersonalization is a coping strategy, personal accomplishment a consequence rather than a defining feature of burnout while fatigue and emotional exhaustion are the core feature of it.

In order to be able to assess the *causal* effects, we only included studies that reported linear regression coefficients (from any statistical model as long as it was linear). Only direct effects were considered, effect of adjustment variables and of interactions terms were excluded. Also, at least one of the following variability estimates had to be provided in order to consider the related factors in further analysis: standard error

(ideally), p-value or IC. As suggested in chapter 12.2.1 of the Cochrane Handbook (Higgins et al., 2019), in case studies reported conventional level of significance (e.g. P<0.01), we used the threshold (e.g. 0.01). Only exception was made for the conventional levels of significance of "P<0.05" and "ns", for which we used the p-values of 0.02 and 0.2 respectively in order to be able to distinguish them.
Inclusion criteria into quantitative synthesis yielded a final total of 85 samples from published studies and a total of 282 factors predicting emotional exhaustion.

2.3 Why was meta-analysis precluded?

Meta-analysis is a two-stage process. The first stage concerns the calculation/identification of a *common type* of effect size from the individual studies. An effect size is a measure of the *magnitude* of an observed relationship, treatment effect, or population parameter. In the second stage, the effect sizes from individual studies are weighted before being averaged to produce an overall effect size. "Greater weights are given to the results from studies that provide more information, because they are likely to be closer to the true effect that is being estimated" (Akobeng ,2005).

A standardized regression slope tells us for how many standard deviations the outcome changes when the predictor changes of one standard deviation. A relative regression (beta) coefficient of 0.4, for example means that an increase of one standard deviation on the predictor's scale corresponds to an increase on the outcome scale of 0.4 standard deviations. Hence, the issue of metrics equivalence for both predictors and the outcomes across studies is fundamental for a valid interpretation of the summary estimate. Buteau and Goldberg (2015), showed how, even when considering association of one specific variable such as heart rate variability with ambient air pollution, problem of metrics comparability across studies can arise due to duration of the electrocardiograms which was not uniform, going from 5 minutes to 24 hours as well as conditions in which electrocardiograms are done. Hence, even when considering a same measurement tool, some studies may not be comparable enough to justify the use of meta-analysis. In our case, the problem was even bigger. Indeed, we were confronted to a lack of consensus on many aspects of burnout mentioned

above as well as the huge available possibilities to measure a same construct in psychology. Interpreting the pooled estimate with the assumption that variation of one standard deviation meant the same, *no matter the tool* was a strong assumption we preferred to avoid and replace by an alternative method.

2.4 Data synthesis

First, we regrouped the selected factors into subfamilies in order to synthesize the great amount of information and make each family as homogeneous as possible despite the different ways of measuring one same construct. The created subfamilies were justified with the existent literature and theory as described in section 2.5.
Figure 2 depicts the final content of each family.

Vote counting based on direction and sign test

As an alternative approach to standard meta-analytical methods, we adopted *vote counting* that enabled us to conclude on direction of the studied factors/subfamilies. Vote counting makes one category (protective, detrimental or no effect) the "winner". This technique fitted well to our situation since, as reported by the Cochrane standards (Higgins et al., 2019), "vote counting might be considered in situations when standard meta-analytical methods cannot be applied such as when there is no consistent outcome measure". Vote counting *based on effect direction* was preferred to vote counting *based on statistical significance* as recommended by the updated Cochrane Handbook of systematic reviews of interventions 2018 guidance. Indeed, according to them, it is necessary that the number of studies showing harm be compared with the number showing benefit, regardless of the statistical significance in order to undertake vote counting properly (Higgins et al.,2019). Calculating study results based on statistically significant results have already been criticized (Light & Smith, 1971). One major reason is that it does not take into account the sample size, which affects statistical power.
Sign tests (or binomial tests) was applied to statistically test for the "winner category". This method enabled us to test whether the probability for the (subfamily) effect to be detrimental (or protective) was less than 0.5. A probability of 0.5 being the point where half of the studies would lie on each side of the no-effect line, and where the effect is

hence considered truly ineffective. Sign tests is considered a valid approach when "numeric data are of such different types that they cannot be combined statistically and that studies are so diverse in their populations or other characteristics that a pooled effect size is meaningless, but studies are addressing a questions sufficiently similar that the direction of effects is meaningful" (Coryn & Hobson, 2013). This hence fitted well to our situation.

We further checked whether studies with the highest number of participants showed significant results in line with sign test results. Indeed, as summarized by Kalla Siddharth: « When you have a higher sample size, the likelihood of encountering Type-I and Type-II errors occurring reduces, at least if other parts of your study is carefully constructed and problems avoided. Higher sample size allows the researcher to increase the significance level of the findings, since the confidence of the result are likely to increase with a higher sample size. This is to be expected because larger the sample size, the more accurately it is expected to mirror the behavior of the whole group »[6].

One has to note here that this applies only when "other parts of your study is carefully constructed, and problems avoided". In case it is not, tendency to reject null hypotheses can be *exaggerated* and misleading. Risk of bias assessment will also be conducted to account for this.

Visual display of the data points according to their significance, direction and sample size were done for each subfamily. The data we had enabled us to calculate and plot a t-value for each predictor. Those values were practical since they permitted to simultaneously represent direction and p-values of a given effect estimates. T-values were calculated by dividing the effect estimate by its standard error. If the variability value associated with the beta estimate was a p-value or a confidence interval, formulas were applied to convert them into standard errors. The formula can be found in Appendix C.

[6] Statistical significance and sample size. Retrieved from https://explorable.com/statistical-significance-sample-size

Funnel plots

Despite that meta-analysis was precluded and another method adopted, we still wanted to construct the funnel plots to assess on publication bias. However, since heterogeneity was systematically found and that heterogeneity is a potential source of asymmetry in funnel plots (Sterne et al., 2011), using them to conclude on publication bias was also precluded.

Analyses were performed using R 3.6.2 (R Foundation for Statistical Computing, Vienna, Austria). Example of code used for the created plots as well as binomial test are reported in Appendix C.

2.5 Families of predictors and related theories

In order to perform further analysis, factors were first classified into families and sub-families of predictors on the basis of the predictors' nature and the existent literature in order to create groups that were as homogenous as possible.

2.5.1 Classification overview and methodology

First, based on review of Maslach et al. (2001) on job burnout, we considered 2 main families of factors: situational factors and individual factors. Job characteristics and organizational characteristics were included in the former whereas personality characteristics, work attitudes and non-occupational factors were included in the latter. Moreover, as a category at the intersection between work and personal life, we considered a third main family: work-life interface (Greenhouse & Allan, 2011; Rubio et al., 2015) which refers to factors of personal life that intersect with work factor or vice-versa. Finally, we considered the effect of other job outcomes (than emotional exhaustion) as a fourth main family.

From the four main families, we categorized factors into smaller and smaller families until all predictors of one subfamily 1) referred to a relatively homogenous construct and 2) had the same valence/direction (i.e., 2 subfamilies "maladaptive coping style" and "adaptative coping style" instead of one subfamily "coping style").

Content of those four main families and their subfamilies are depicted in figure 2 (as well as in appendix A in more details). We will now describe the literature we leant on in in order to get to that result.

2.5.2 Situational characteristics

We referred to the Job-demands-control (JDC) model first proposed by Robert Karasek in 1979 and later modified by Johnson & Hall (1988) as the Job-demands- control-support (JDCS) model to consider *job demands*, *decision latitude (control)* and *social support* as subfamilies of situational characteristics.

Job demands refers to requirements/stressors that are set at work including workload, time pressure, physical demands and emotional demands.

Decision latitude (control) is the extent to which an employee has potential control over decisions concerning when, where and how to perform work tasks including job variety (or skill discretion; degree to which job involves a variety of tasks, low repetitiveness, occasions for creativity and opportunities to learn new things) and decision authority (the ability to make its own decisions about how the work is done - i.e. autonomy - as well as influence work team and policies). In regard to control, Rafferty et al. (2001) highlighted how "inconsistent findings may be related to the different ways in which control has been defined and suggest that future research on the JDC model should differentiate between dimensions of control". They indeed showed that emotional exhaustion was more or less associated with higher control depending on how it was defined; authors founds positive association when defined as decision latitude – skill discretion plus decision authority – and autonomy (Landsbergis, 1988 ; de Jonge et al., 1996) , inconsistent association when defined as decision authority (significant in one study but not the other: Taris et al., 1999; de Rijk et al., 1998) and no relation when defined as skill discretion (Taris et al., 1999).

Social support is defined as "the perception and actuality that one is cared for, has assistance available from other people, and most popularly, that one is part of a supportive social network"[7]. Support can be of different type such as emotional support, companionship (sense of belonging) or informational (e.g. advice) and can come from many sources such as family, colleagues, supervisor and/or organizations. According to those models, employees working in situation with high demands, low control and low social support /isolation (specific to JDCS model) experience the lowest well-being ((iso)-strain hypothesis). Conversely, high job control and social support can moderate the negative impact of job demands on well-being (buffer hypothesis).

[7] https://en.wikipedia.org/wiki/Social_support

Alongside, we referred to Job-demand-resources model proposed by Demerouti et al. (2001) as an alternative to the model to the JDC(S) models, to consider job resources as a subfamily of situational characteristics.

Job resources are defined as job-relevant features that help workers achieve work-related goals, lessen job demands, or stimulate personal growth.

According to this model, employee high strain and low well-being is a response to imbalance between demands and resources he or she has to deal with those demands.

Later on, in 2006, Akerboom and Maes proposed the organizational risk factor questionnaire derived from the Tripod accident causation model (Wagenaar, Hudson, & Reason, 1990; Wagenaar, Groeneweg, Hudson, & Reason, 1994) to investigate determinants of psychological well-being from a broader organizational perspective than the JDC(S) model does. These authors considered factors such as "organization (OR)" – which refers to the effectiveness of the organization's structure and processes, and management strategies; communication (CO) – which refers to quality and effectiveness of communications between individuals; and social hindrance – operationalized as disrespectful, unequal and/or unfair treatment, lack of appreciation of extra-effort, etc. Ruehlman and Wolchik (1988) defined social hindrance as "behaviors that are perceived by either the actor or the target person as a) (un-)intentional interference with goal-directed activity or b) (in-) direct expressions of anger or other negative emotions or c) (in-)direct negative evaluations of the target person's character or behavior".

In view of our data and considering those definitions, we included social hindrance, informational climate, lack/conflicting information, enriching leadership and non-collaborative leadership as subfamilies of organizational characteristics.

2.5.3 Individual characteristics

We referred to Maslach et al. (2001) to consider two groups of subfamilies in the individual characteristics: personality characteristics (including intrinsic characteristics, coping strategies and self-evaluation) and job attitude (work-related attitudes including

job involvement and unfavorable job attitude). We also considered non-occupational factors as part of the individual characteristics.

Subfamily "intrinsic characteristics" mainly refers to the Big Five personality traits proposed by Goldberg in 1981 and further developed by Costa & McCrae in the late 80s. The Big Five or "OCEAN" refers to a descriptive model of personality in 5 main traits including openness to experience, conscientiousness, extraversion, agreeableness and emotional stability (inverse of neuroticism). Other intrinsic characteristics including emotional intelligence, optimism, type a behavior and hardiness were also included in this subfamily.

Coping was defined by Lazarus & Folkman in 1984 as "cognitive and behavioral efforts aiming to control, reduce or tolerate demands that threaten or exceed and individual resources". In short, coping refers to efforts to manage stress. In order to fit with the criterion pertaining to classification cited above, coping strategies were divided into two subfamilies, namely, adaptative coping (i.e. strategies that involve confronting problems directly, decrease the actual level of stress and improve long term level of functioning) and maladaptive coping style (i.e. strategies that reduce the symptoms related to stress for a short time, but have no influence on the thing that is causing the stress[8]) and were considered separately.

Self-evaluation included self-esteem – introduced for the first time by William James in 1890, refers to "an individual's sense of value or self-worth, or the extent to which people value, appreciate or like themselves" – and self-efficacy – defined as the belief individuals have in their ability to succeed and their level of competence (Bandura & Adams, 1977). It is often argued that self-esteem and self-efficacy are distinct constructs (Lane et al., 2004). However, a reasonable assumption is that people who have high global self-esteem will predict higher probability of task success (high self-efficacy) and hence they relate both theoretically and empirically (Gardner & Pierce,1998).

In the subfamily "job attitude", we included predictors referring to job involvement (commitment, presenteeism, etc.) and unfavorable job attitudes (absenteeism,

[8] https://study.com/academy/lesson/adaptive-coping-strategies-definition-examples.html

overcommitment, resistance to change). The diversity of predictors also enabled us to make the distinction between intrinsic job involvement (Georges, 1992) and extrinsic job involvement, which we considered as two other subfamilies. The former involves doing something because it's personally rewarding while the latter involves doing something for reward or to avoid punishment.

Finally, non-occupational factors included any determinants related to personal life events.

2.5.4 Work-life interface

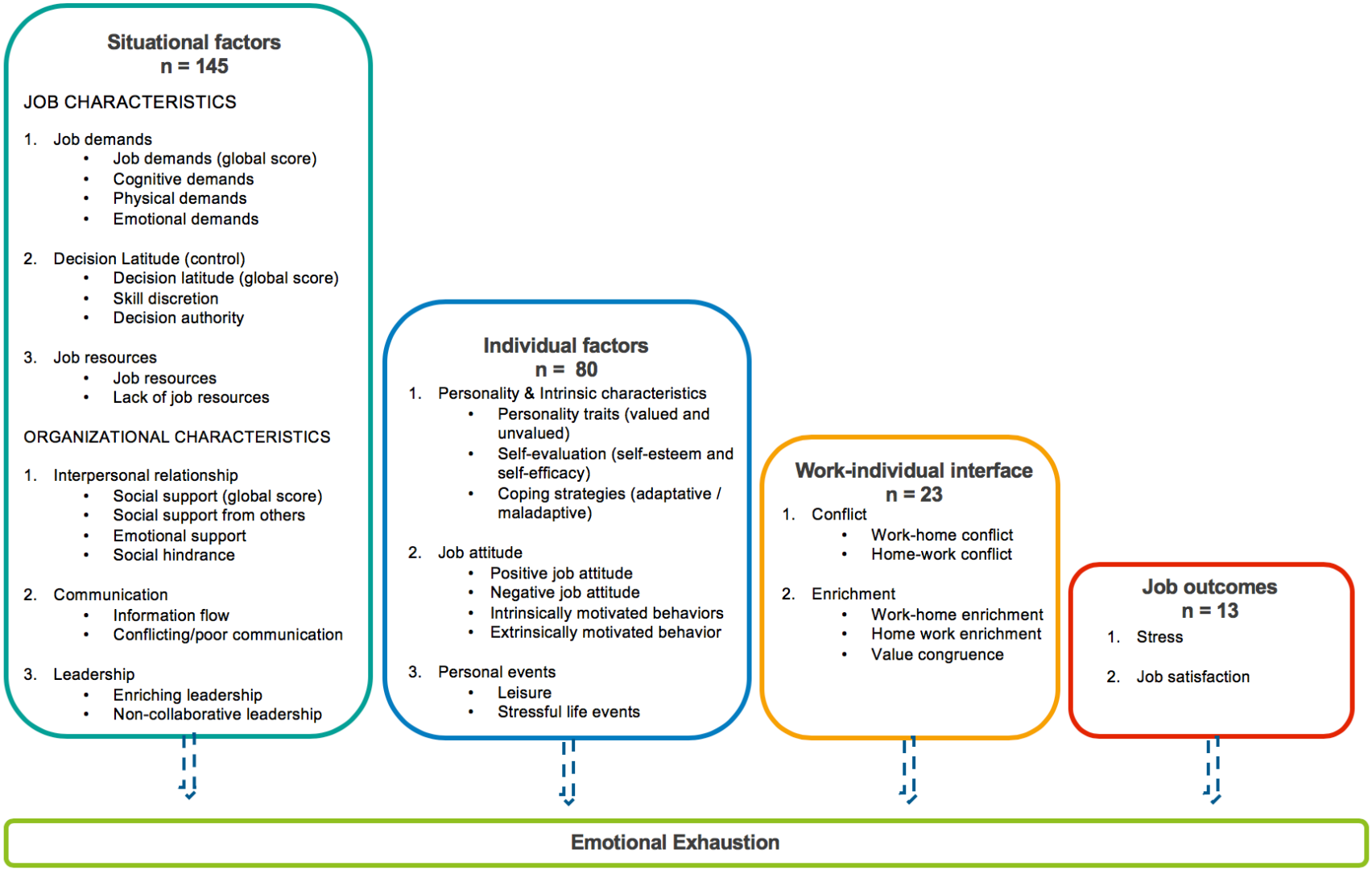
Work-life interface refers to either positive or negative spillover, defined as a process whereby behavior in one domain establishes benefits/resources (positive spillover; enrichment) or detriments (negative spillover; conflict) which then, respectively, improve or impair performance/involvement in the other domain (Greenhaus & Allen, 2011). More precisely*, work - life conflict* occurs when the demands associated with one domain are incompatible with the demands associated with another domain whereas *work-life facilitation* refers to how participation in one role is made better or easier by participation in the other role (Wayne et al., 2004).

Several authors form out dataset investigated the effect of value congruence. *Value congruence* refers to the degree to which an individual's values match the ones found in their work environment. As "value congruence between individuals and the organizations in which they work results in a number of beneficial outcomes, the most common of which are associated with positive attitudes and behaviors" (Molina, 2016), we included all predictors pertaining to value congruence in subfamily work-life enrichment.

2.5.5 Job outcomes

Predictors that related to other job outcomes than emotional exhaustion were included in this group. The two main subfamilies we could form given our data are job satisfaction and stress/insecurity. This classification was inspired Adriaenssens et al. (2015) operationalization of "predictors of stress–health".

Figure 2. Content of factors' families and subfamilies



2.6 Results

2.6.1 Study characteristics

The literature search yielded 5'297 records resulting in 240 studies with 85 studies that met inclusion criteria for the semi-quantitative analysis (Figure 1). From those 85 studies, 261 data points predicting burnout onset were identified.

The majority of data points were *situational factors* (57,5%) with job demands being the most studied subfamily(n=46). Seventy-eight studies (91%) used the Maslash Burnout Inventory (including different versions namely MBI-HSS (Maslach & Jackson, 1981), MBI-GS (Schaufeli et al., 1996), MBI-ES (Maslach & Jackson, 1981), or not specified). From those version, different number of items were used to measure burnout depending on the study. Six samples (7%) used the Oldenburg Burnout Inventory (OLBI) and one sample (Borritz et al., 2005) used the Copenhagen Burnout Inventory (CBI).

2.6.2 Content of subfamilies, vote counting and sign test

Details content of each subfamilies including first authors' name and year, factors' original label, sample size and t-value that were used to construct the plots of this section can be found in appendix A. Vote counting was made easier with the detailed plots described in appendix B. Results of vote counting, and sign tests are reported in Figures 3-15 as well as Table 1.

2.6.2.1 Situational characteristics

2.6.2.2 Job characteristics

Job demands

Majority of the data points (38/47) pertaining to job demands showed detrimental effect (Figure 3). Sign test was in favor of a *detrimental effect* of job demands when it was measured as a global score, quantitative demands and emotional demands. However, sign test was not significant when it was measured as cognitive demands or physical demands. Similarly, data points with highest samples sizes show detrimental and

significant effect and concern quantitative demands, job demands and emotional demands.

One interesting result is effect of cognitive demands for which majority of data points (2/3) indicate a protective effect. Although sign test was not significant, one study with including more than 500 participants, show significant protective effect.

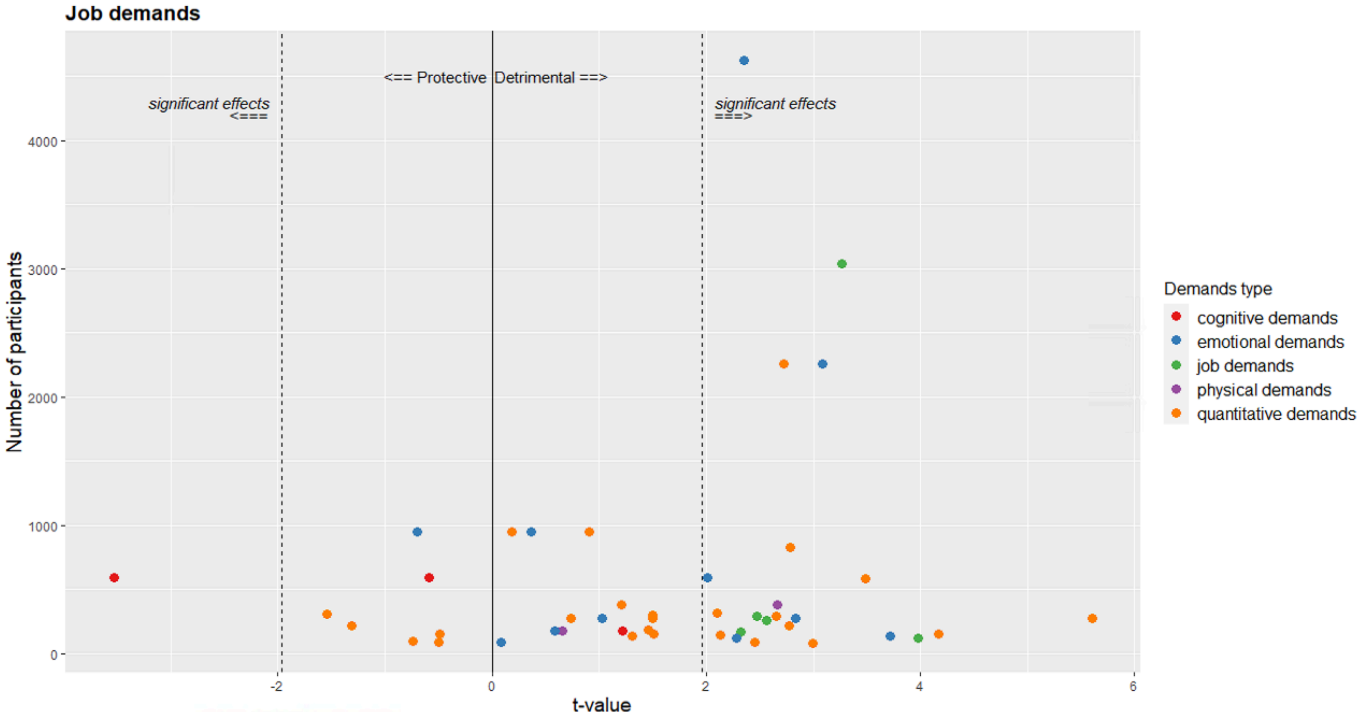


Figure 3. Prospective effects of job demand on burnout. Bold vertical line at 0, represent absence of effect. On the left of that line, data points represent protective effects while on the right data points represent detrimental effects. Dotted lines at ± 1.96 represent limits of the 95% confidence interval. Data points higher than 1.96 or lower than -1.96 are significant effect. Note: data point from Theorell et al. (2013) concerning job demands; sample size = 11'525 participants; t-value = 3.89 showed an extreme number of participants. It was hence removed from the plot for better clarity.

Table 1 vote-counting, sign test and significant effects				
Domain	**Effect count (P/D)**	**Sign test (p-value)**	**Significant effects**	
			Protective	Detrimental
Job demands				
Job demands (overall score)	6 (0/6)	Detrimental (0.0156)	0	5
Quantitative demands	24 (5/19)	Detrimental (0.003)	0	12
Emotional demands	11 (1/10)	Detrimental (0.0058)	0	5
Cognitive demands	3 (2/1)	ns	1	0
Physical demands	2 (0/2)	ns	0	1
Decision Latitude (control)				
Decision latitude (job control)	9(7/2)	Tendency P (0.089)	2	1
Skill discretion (job variety)	5(3/2)	ns	1	0
Decision authority	5(4/1)	ns	1	0
Autonomy	2(1/1)	ns	1	1
Job resources				
Job resources	19 (14/5)	Protective (0.0319)	6	0
Lack of job resources	4 (0/4)	ns	0	3
Interactions at work				
Social support	21(15/6)	Protective (0.0391)	7	1
Good interpersonal relations	6(4/2)	Ns	2	0
Social hindrance	11(2/9)	Detrimental (0.033)	0	5
Communication				
Informational climate	8(6/2)	ns	2	0
Conflicting / poor communication	5(0/5)	Detrimental (0.031)	0	3
Leadership				
Enriching leadership	5(2/3)	ns	1	2
Non collaborative leadership	3(1/2)	ns	0	0
Personality				
Unvalued trait/characteristics	2(0/2)	ns	0	1
Valued trait/characteristics	15(10/5)	ns	4	0
Coping				
Adaptative coping	6(6/0)	Protective (0.0156)	4	0
Maladaptive coping	4(1/3)	ns	0	2
Self-evaluation				
Self-esteem	6(4/2)	ns	2	1
Self-efficacy	9(7/2)	Tendency P (0.08)	4	1

Table 1 continued				
Domain	**Effect count (P/D)**	**Sign test**	**Significant effects**	
			Protective	Detrimental
Job attitude				
Positive job attitude	8(8/0)	Protective (0.004)	5	0
Negative job attitude	9(1/8)	Detrimental (0.019)	0	8
Intrinsically motivated behavior	8(5/3)	ns	2	1
Extrinsically motivated behavior	6(1/5)	ns	0	3
Personal events				
Leisure	5(5/0)	Protective (0.031)	2	0
Stressful life events	5(4/1)	ns	0	3
Work family interface				
Family-work conflict	3(0/3)	ns	0	2
Work-family conflict	13(3/10)	Detrimental (0.0046)	2	8
Family-work enrichment	3(2/1)	ns	2	0
Work-family enrichment	1	-	1	0
Value congruence	3(1/2)	ns	1	0
Job outcomes				
Stress	10(2/8)	Detrimental (0.05)	2	5
Job satisfaction	3(3/0)	ns	3	0

Note: P= protective, D= detrimental, ns = non-significant, tendency = p-value of sign test is almost significant.

Decision Latitude

Majority of the data points pertaining to job control show protective effects (Figure 4). However, 2 data points show significant detrimental effects. Those counterintuitive effects were found by Innstrand et al. (2008) for autonomy and Konze et al. (2017) for job control. Both authors gave a similar explanation which is as follow: job control can be either beneficial or harmful depending on job demands the employee faces (it was noted, respectively, that boundaryless work life and emotional dissonance can emerge along with higher autonomy and job control, which in turn lead to more exhaustion). Sign test was almost significant for decision latitude but not significant for any other dimension of job control. Similarly, study with highest sample size showed protective significant effect of decision latitude.

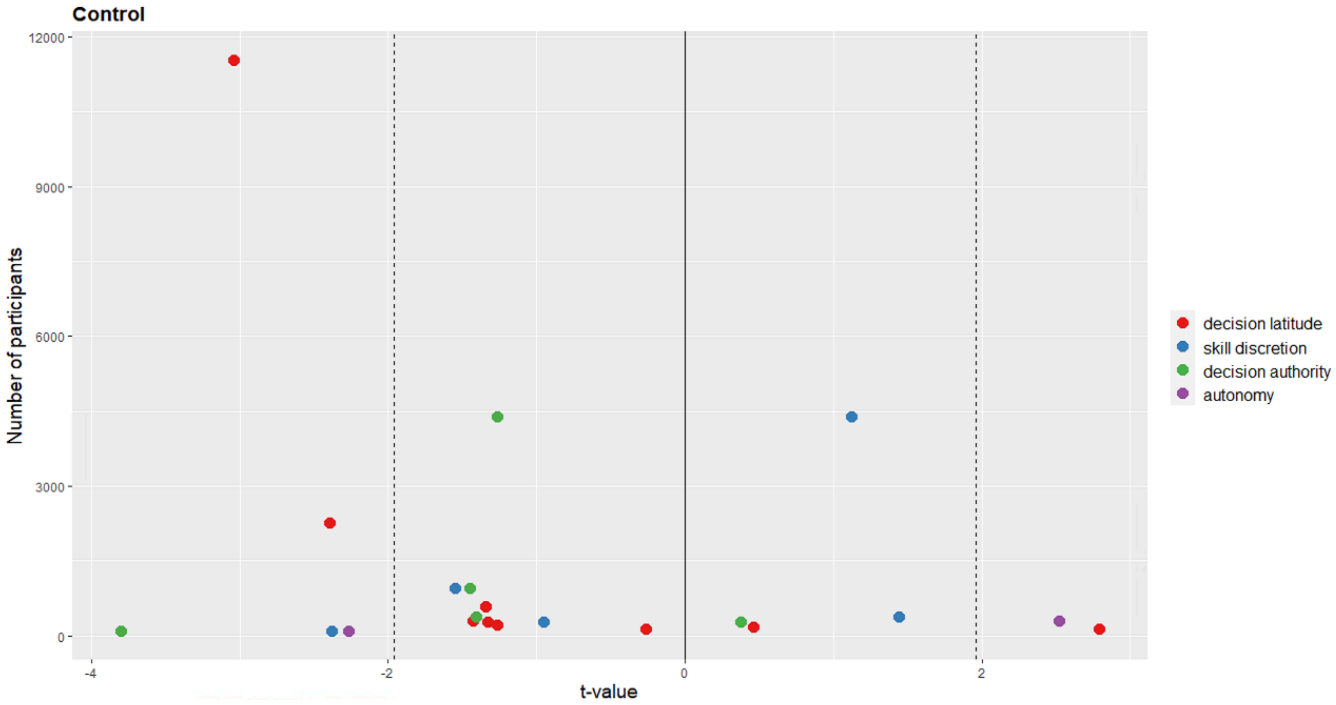


Figure 4. Prospective effects of Control on burnout. Note: two protective points of decision latitude are hidden by others.

Job resources

Of the 19 data points referring to job resources, majority (13/19) showed protective effect (Figure 5). Sign-test was significant (p-value=0.037). Similarly, one study from Theorell et al. (2013) with very high number of participants (n= 11525) showed

protective effect of job resources (more especially cultural activity at work) on burnout onset. Also, all the data points pertaining to *lack* of job resources showed detrimental effects. However, sign test was not significant for this subfamily.

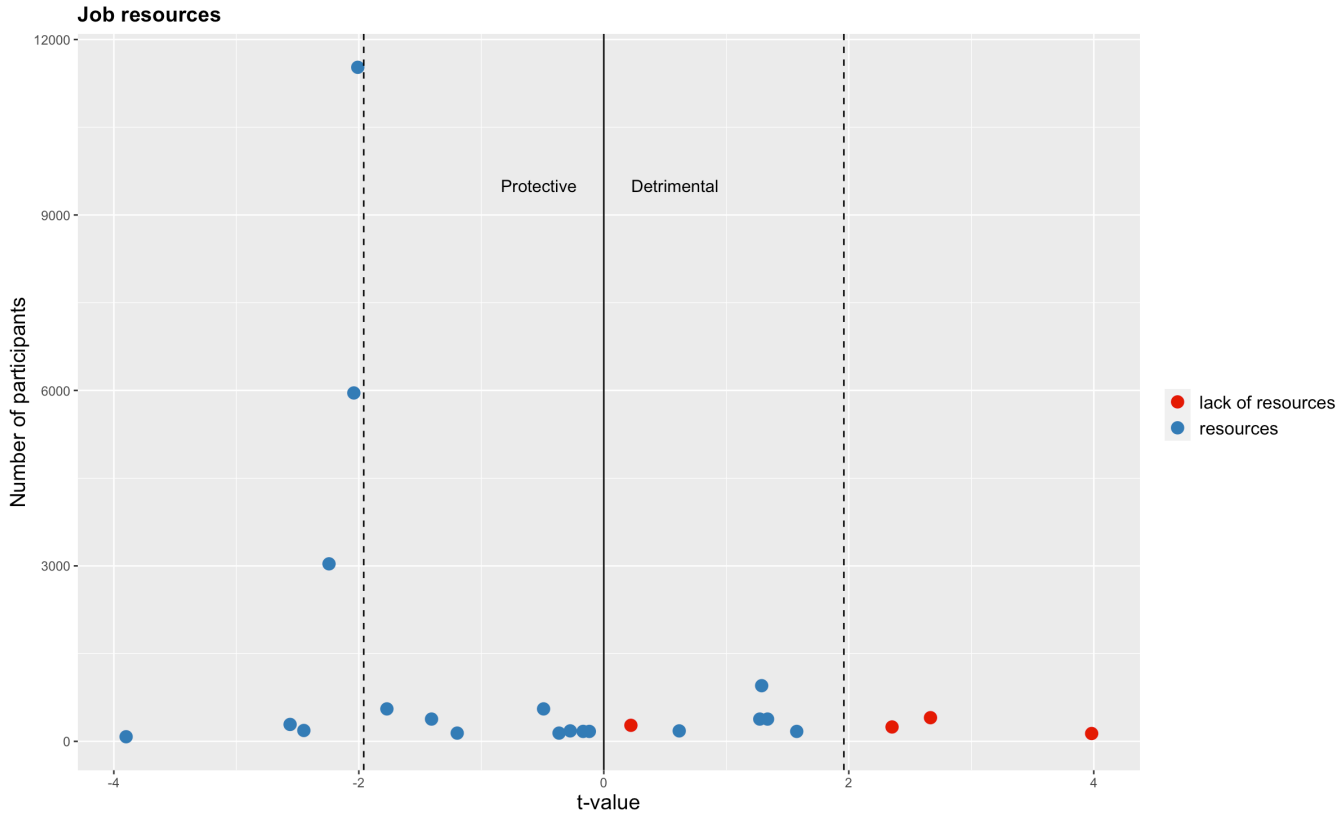


Figure 5. Prospective effects of Job resources on burnout.

2.6.2.3 Organizational characteristics

Interaction at work

Majority (9/11) of data points pertaining to *social hindrance* show detrimental effect (Figure 6). Sign test was significant in direction of a detrimental effect.

Moreover, of the 21 data points pertaining to social support, 6 showed significant protective effect while 1 showed detrimental significant effect. This effect was found by Pisanti et al. (2016) and his explanation was the following: "These results were not in line with our hypothesis. However […] in line with the stress transfer theory (Karasek et al., 1982) less strained people could assimilate the strain of colleagues more stressed. In other words, in situations with strong social bonds, individuals may absorb more feelings of stress from those around them rather than be protected from stress." Sign test of social support was however in favor of a protective effect.

Similarly, data points with higher number of participants concern social hindrance and social support. In line with the above sign test results, these points show a detrimental effect of social hindrance and a protective effect of social support.

However, sign test on data points pertaining to *good interpersonal relations* (i.e.) is not significant although 2 data points showed significant protective effect and that no significant detrimental effect was found.

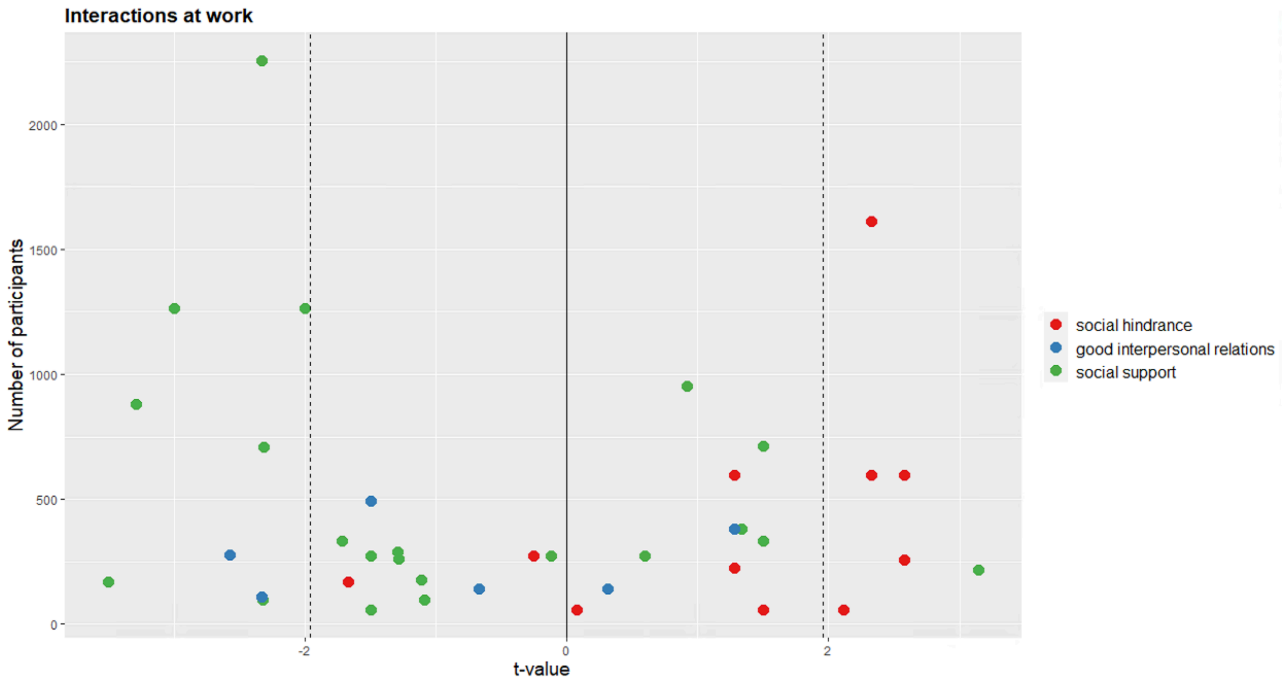


Figure 6. Prospective effects of Interaction at work on burnout.

Communication

Sign-test did not show any significance for informational climate. However, sign test was significant for lack of job resources. Of the 4 data points pertaining to *lack of job resources*, 3 showed significant detrimental effect and no significant protective effect was found.

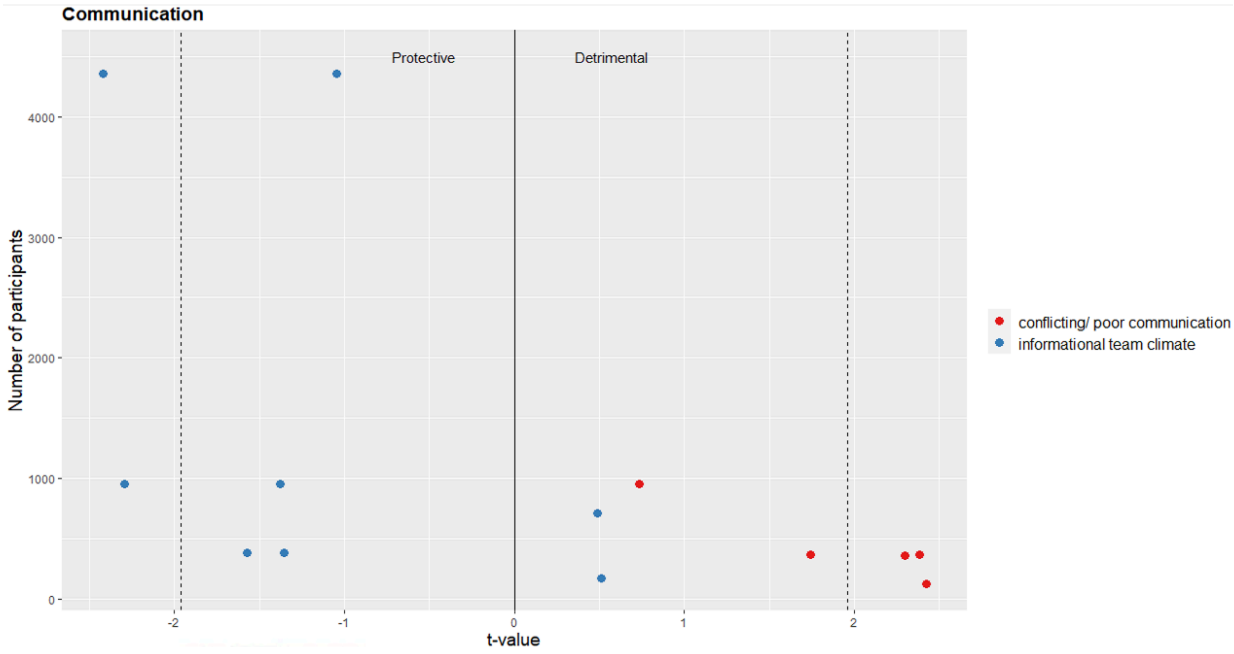


Figure 7. Prospective effects of communication on burnout.

Leadership

We did not find enough evidence for a directed effect of neither enriching leadership nor non-collaborative leadership (p-values =0.5). However, for enriching leadership, we could notice that some data points from studies with high simple size (> 2000 participants) surprisingly show significant detrimental effect Figure 8). Those actually concerned effect of transformational and laissez-faire leadership on leader's own emotional exhaustion (Zwingmann et al., 2016). The authors gave the following interpretation: "whereas previous studies consistently showed health-enhancing effects of transformational leadership by creating resources for followers and themselves (e.g., Wegge et al., 2014), the opposite may also be true: transformational leadership is highly effective in negative sense, that is depleting leader's own resources and increasing leader's emotional exhaustion in the long run".

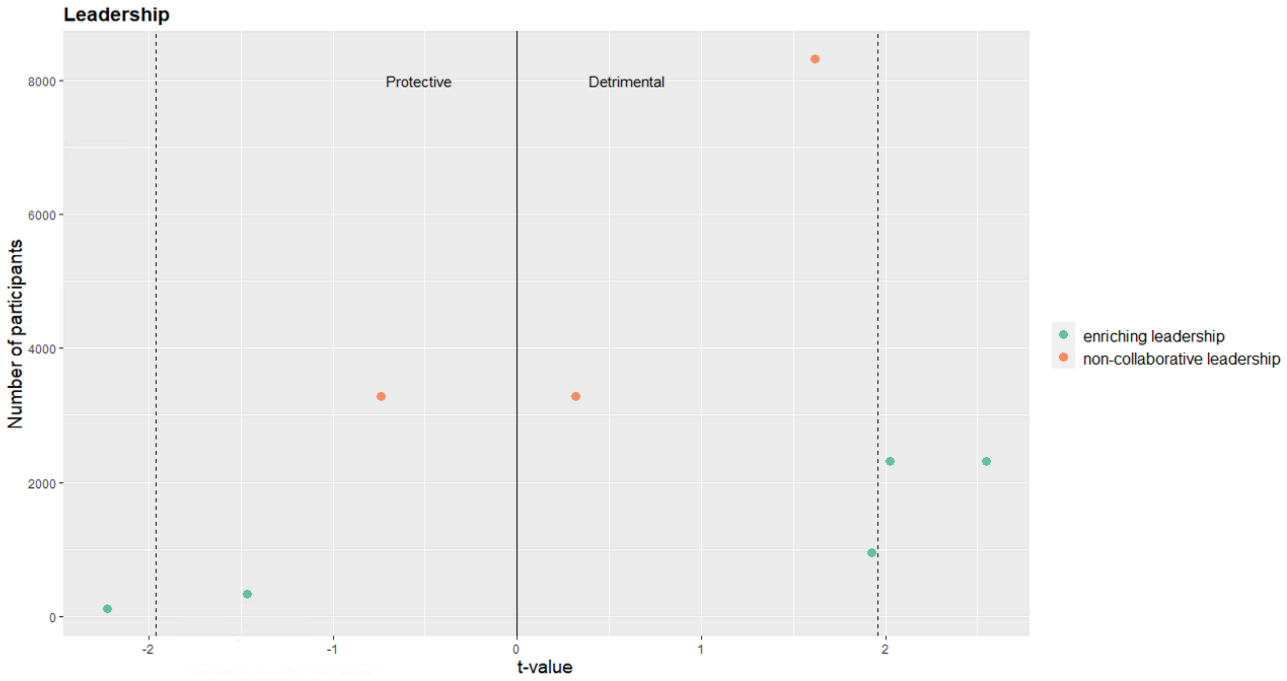


Figure 8. Prospective effects of the kind of leadership on burnout.

2.6.2.4 *Individual characteristics*

(Un)-valued personality traits/characteristics

Seventeen factors were regrouped in subfamily pertaining to intrinsic characteristics including valued [19] and unvalued [6] personality trait/characteristics. We did not find enough evidence to conclude on a directed effect for neither valued personality traits nor unvalued personality traits on burnout onset.

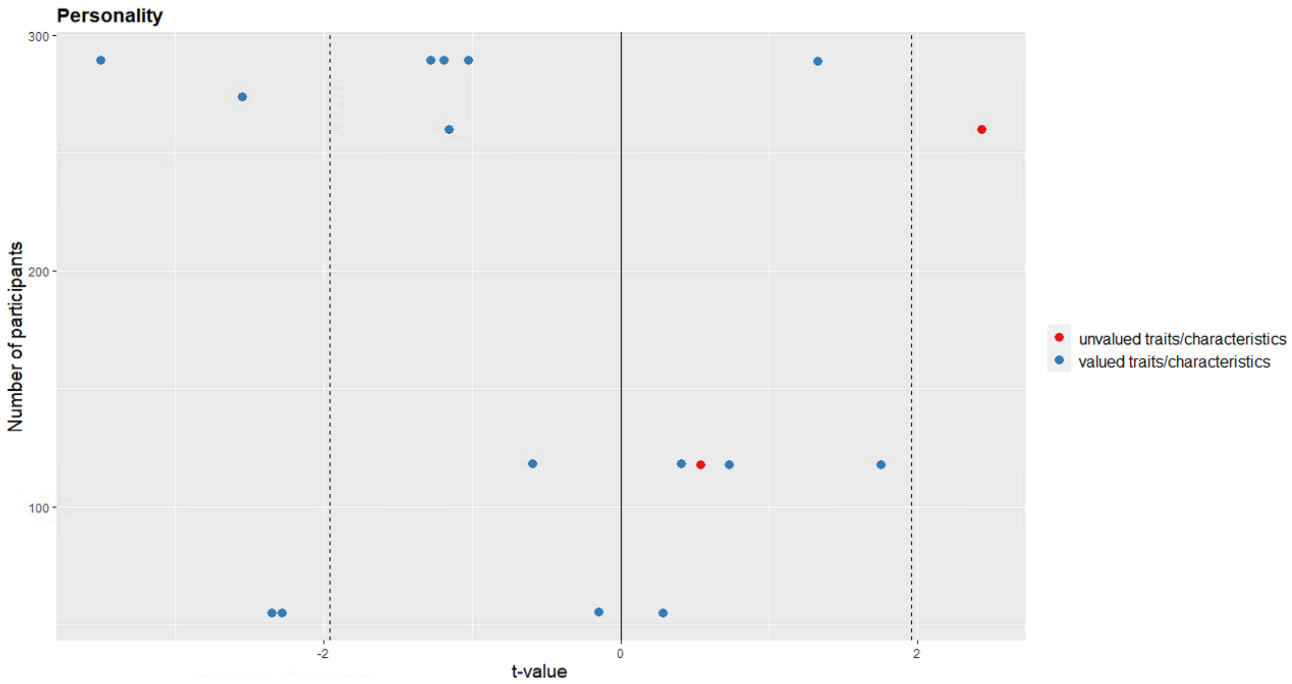


Figure 9. Prospective effects of Personality variables

Coping

Six data points were included in subfamily *adaptative coping* (i.e., problem-oriented coping, emotional support seeking, deep acting) among which all show protective effect. Sign test was significant in this direction (p=0.0156). Similarly, one data point pertaining to adaptative coping, from a study with relatively high number of participants showed protective effect. However, although majority of data points pertaining to *maladaptive cop*ing show detrimental effect, sign test was not significant (p=0.313).

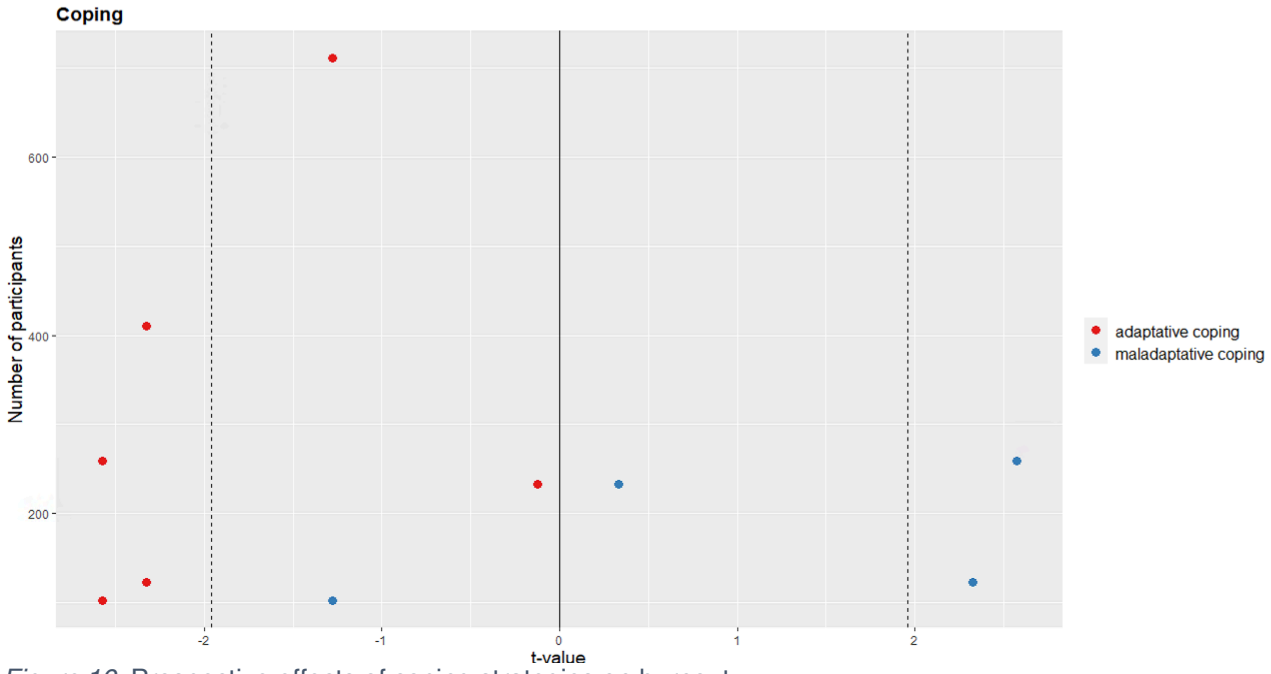


Figure 10. Prospective effects of coping strategies on burnout.

Self-evaluation

Sign test for self-efficacy was almost significant in favor of a protective effect. However, we can notice that 1 data point pertaining to self-efficacy show a significant and detrimental effect. Note that this data points actually concerned factor named "psychological empowerment" investigated by Hochwälder et al. (2008) and which was classified in the current research in subfamily of the self-efficacy. The author gave the following explanation: "a theoretical explanation of this finding, might be if psychological empowerment is not backed with real or structural empowerment, then this state may in long run lead to burnout". Similarly, even though, sign test was not significant for self-esteem, the majority of data points seem favoring a protective effect. One puzzling result though concern the data point from the study including over 3000 participants and which showed significant detrimental effect. This counterintuitive effect was found by Richter et al. (2015) who more specifically investigated effect of performance-based self-esteem and who gave the following interpretation for its positive effect on burnout: "individuals with initial high performance-based self-esteem are said to be more concerned about both their work performance and their accomplishments, which may affect them negatively for instance feeling exhausted".

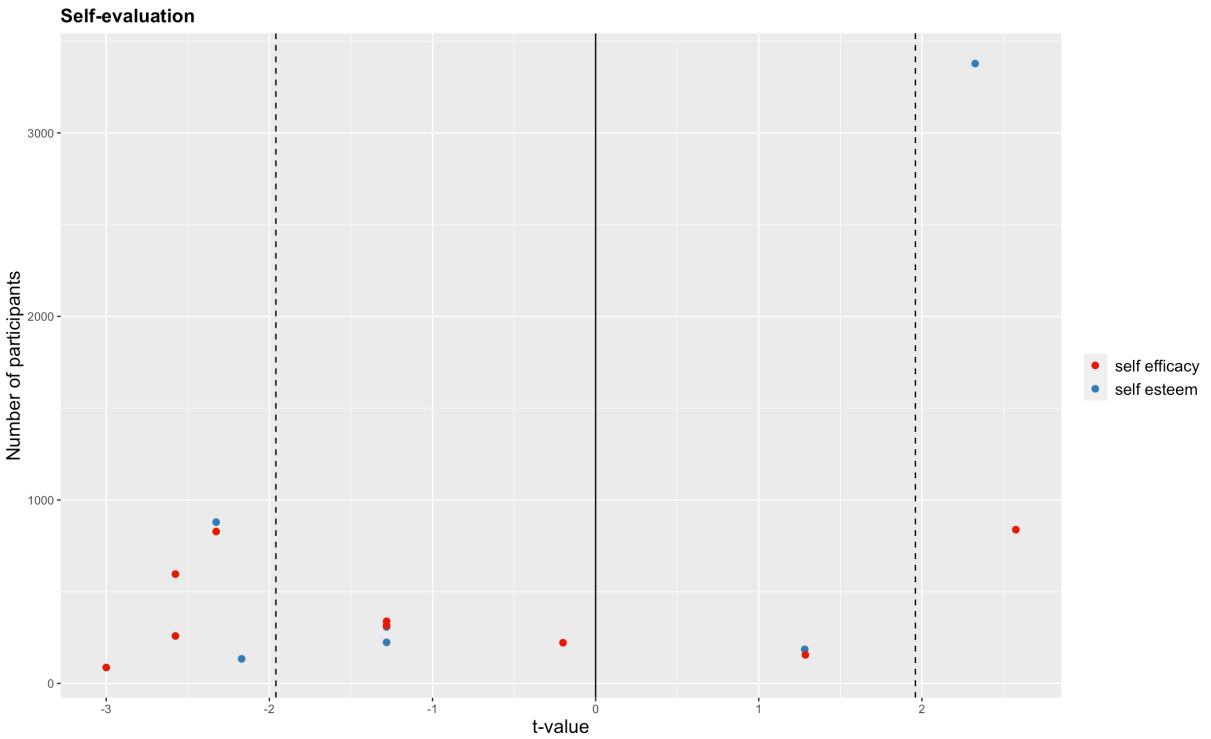


Figure 11. Prospective effects of evaluation of self on burnout.

Job attitude

Significant sign tests were found for both positive job attitude (in favor of a protective effect) and negative job attitude (in favor of a detrimental effect). No evidence such was found for subfamilies related to extrinsically/intrinsically motivated behavior.

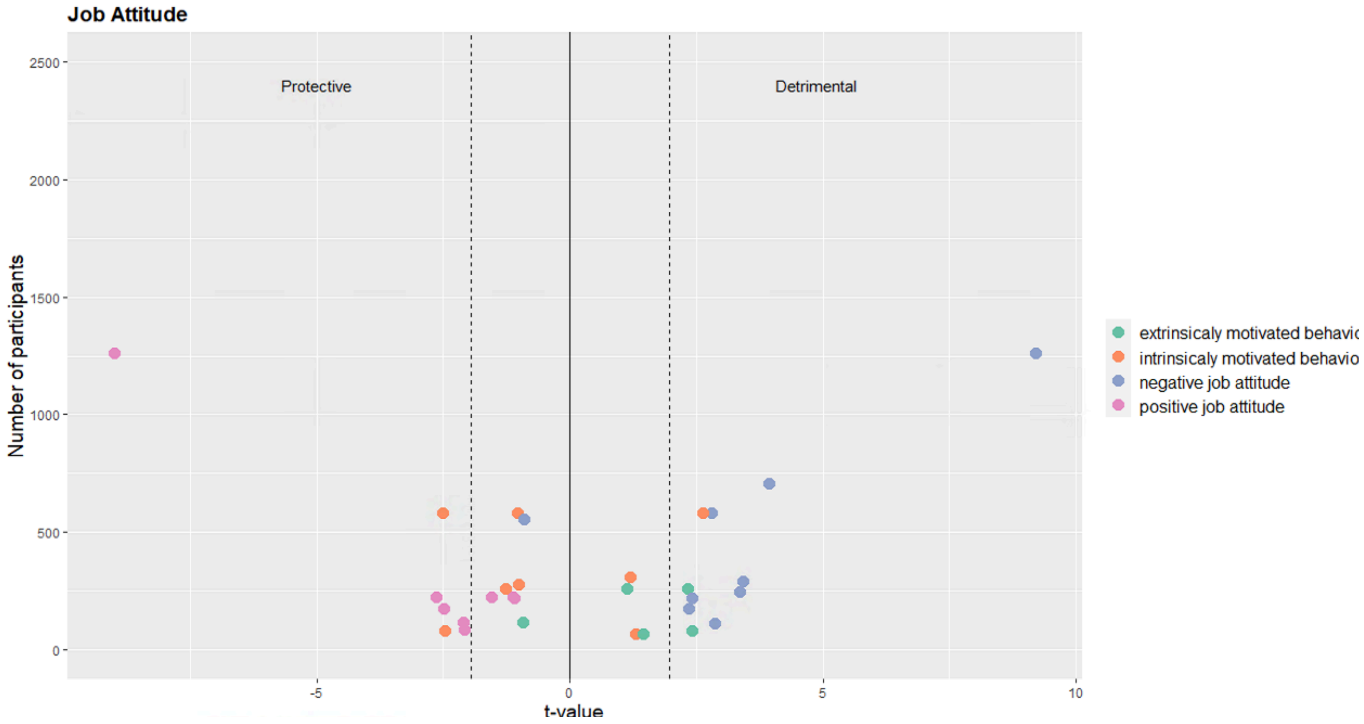


Figure 12. Prospective effects of job attitudes on burnout.

Personal events

All data points are in the expected direction for both leisure (protective) and stressful life events (detrimental). However, sign test is significant only for leisure. We did not find enough evidence for the detrimental effect of stressful life events.

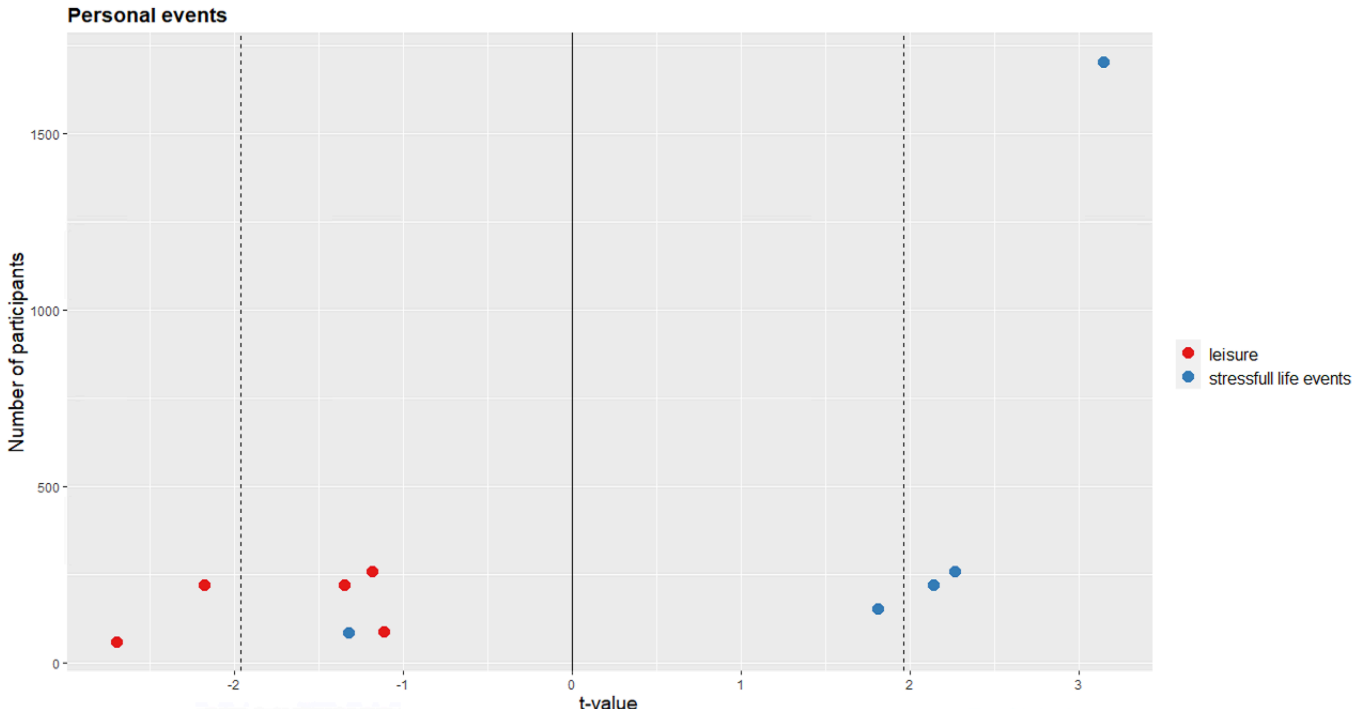


Figure 13. Prospective effects of personal events on burnout.

2.6.2.5 Work life interface

Among the different dimensions of work-life interface, sign test was found significant only for work-family conflict. Although sign test was in favor of a detrimental effect, we can see that 2 data points show a protective and significant effect of work-family conflict on burnout. Those counterintuitive effects were found by Innstrand et al. (2011) and Hertzberg et al. (2016). Innstrand et al. (2011) gave the following explanation "However, the longitudinal findings that suggest an increase in work-to-home conflict is associated with lower levels of exhaustion two years later were more puzzling. Thus, an alternative explanation for these findings is the possible presence of suppression among the variables. A negative or net suppressor effect indicates a variable that has a positive correlation with the dependent variable, but negative beta weights in a regression equation (Tabachnick & Fidell, 2000). Examination of the zero-order correlations supports this notion ». Hertzberg et al. (2016) did not give clear

explanation of this puzzling result. However, we can notice that studies with higher number of participants, systematically found (significant) detrimental effect of work-life conflict, which is more in favor for this conclusion instead of the contrary.

We did not find enough evidence to conclude on direction of the other subfamilies.

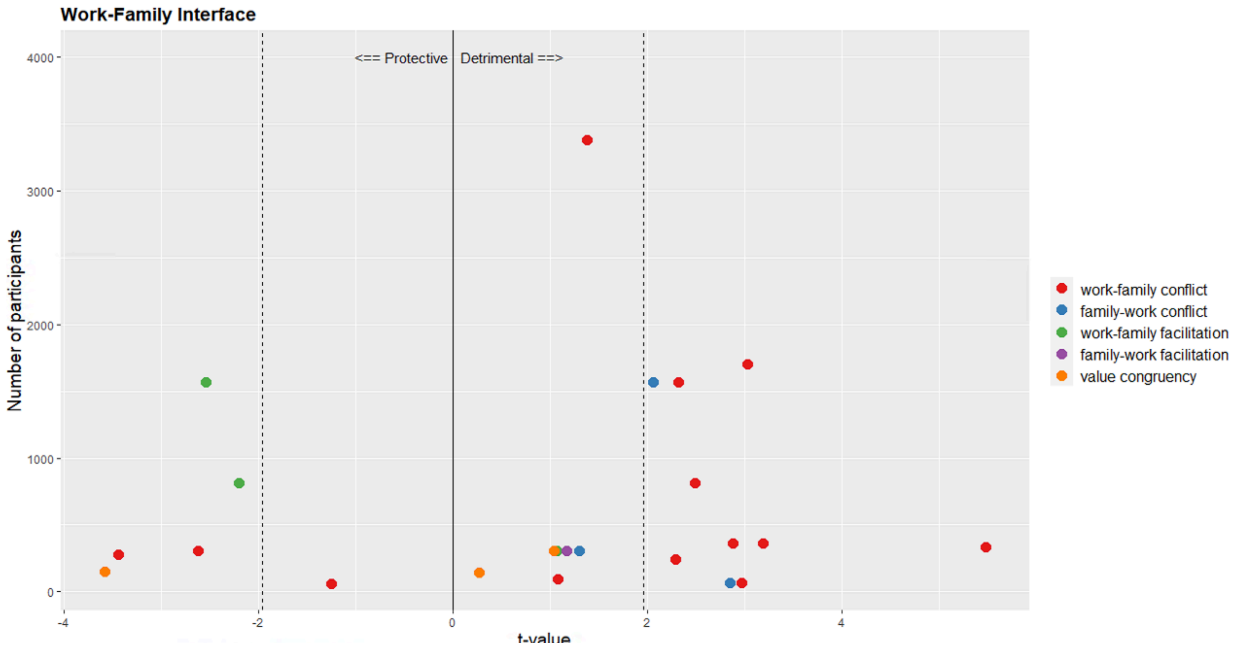


Figure 14. Prospective effects of work-family interface on burnout.

2.6.2.6 *Job outcomes*

Among job outcomes, we found almost significant sign test for stress. No such evidence was found for job satisfaction. Although sign test is in favor of a detrimental effect of stress on burnout, we can see that 1 data point with relatively high number of participants show a protective and significant effect. This effect was found by Taris et al. (2001) who noted that while stressor variables predict increase of burnout cross-sectionally (stress at T1 predict higher burnout at T1 and stress at T2 predict higher burnout at T2), the lagged effect of stressors on burnout (impact of stressors at T1 on burnout at T2) were either not significantly different from 0 or in the reverse direction and interpreted it as such: " this support the reasoning that participant who experience much stress are to some degree successful in mitigating this stress".

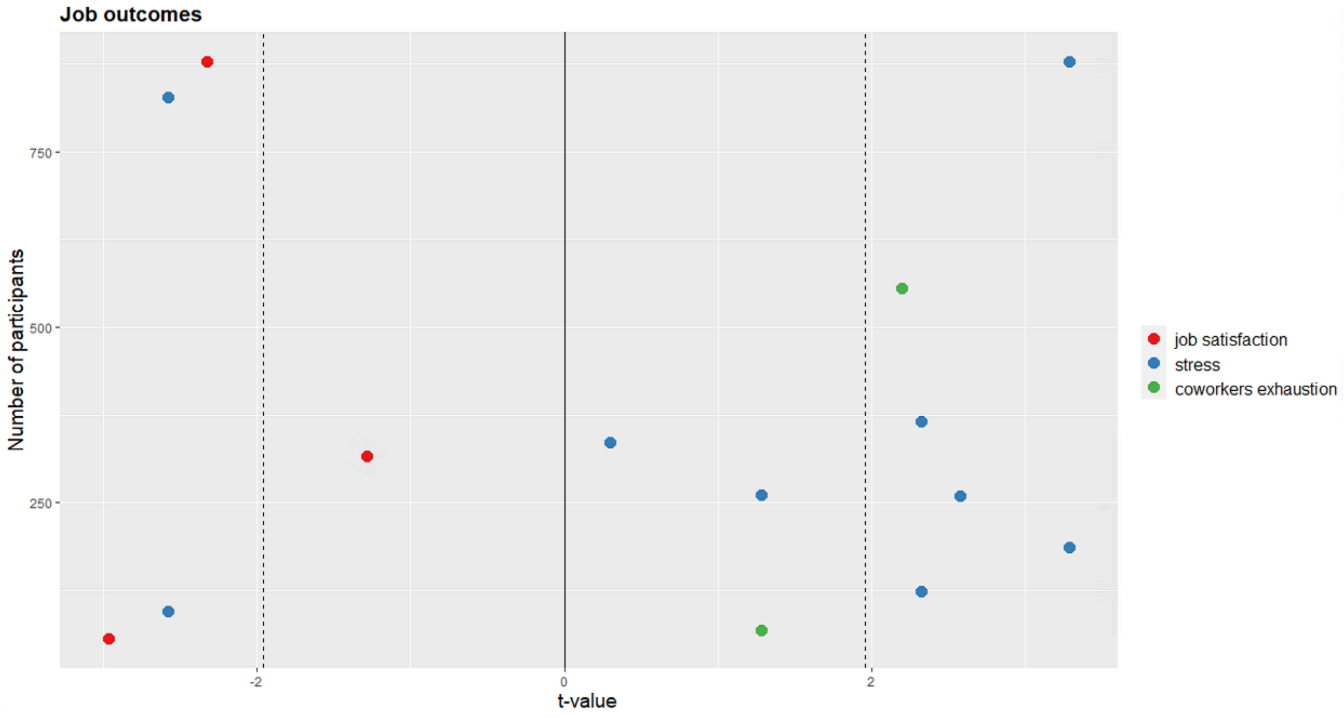


Figure 15. Prospective effects of Job resources on burnout.

2.7 Discussion

2.8 Key (preliminary) findings

It seems that burnout research often focused on one or few types of predictors. A systematic and more exhaustive overview of all kind of factors contributing to burnout onset was lacking. Also, even though many authors performed meta-analysis in similar conditions than ours (i.e., different measurement tools of determinants as well as burnout) and that in practice meta-analysis would have handled our data pretty well (effect estimate and standard error was available for every study), we preferred to adopt vote counting based on effect direction to stick with methodological considerations.

The data included in this review suggest that occupational characterizes are widely investigated. Indeed, as shown in figure 4 majority of factors (almost 58%) concerned this subfamily, letting the other potential causes of burnout proportionally under-investigated. Future research should hence be more “evenly distributed”.
Despite that, this review indicated directed effect in all the four families (considering effects regardless of their associated significance), namely job characteristics, individual characteristics, work-family interface and job outcomes. More precisely:

- For job characteristics, the data from the present study indicated that job demands and especially quantitative demands and emotional demands as well as social hindrance and conflicting/poor communication positively predicted burnout onset, reflecting a detrimental effect. In the other hand, job resources and social support negatively predicted burnout onset, reflecting on their protective effect.
- According to individual characteristics, we found enough evidence to conclude that adaptative coping, positive job attitude and leisure positively predicted burnout onset, reflecting a protective effect. In the other hand, negative job demands positively predicted burnout, reflecting its detrimental effect.
- According to work-family interface, we found enough evidence to conclude that work-family conflict positively predicted burnout onset, reflecting a detrimental effect.

- Similarly, we found enough evidence to conclude that stress positively, positively predicted burnout onset, reflecting a detrimental effect.

Burnout seem hence to be predicted by different kinds of factors and its cause cannot be found in one specific source only. Hence, interventions should also be directed accordingly. The present systematic review indicates that interventions to prevent burnout onset among working population should focus on promotion of informational and, most importantly, supportive working environment through decrease of social hindrance *and* increase of social support. At the individual level, interventions should prepare employees with healthy coping skills (to be able to correctly reduce stress when it appears), promote positive job attitudes and leisure activities. Finally, interventions should be able to detect where and when stress and work-family conflict might occur in order to prevent their detrimental effects.

One should however interpret the above-mentioned results carefully. Indeed, the current review considered *direct effects* in order to get a general overview of how a large panel of factors may directly influence burnout onset *in general* among a working population. In view of results from Pisanti et al. (2016) regarding detrimental effect of social support and stress transfer theory, factors that protects from or increase probability of burnout onset could be different depending on each individual and situation.

2.9 Further steps: quality assessment

As suggested above, however, those results and discussion remain preliminary results. One important future step is to conduct a risk of bias assessment in order to confirm their relevance. Most importantly, the quality of following studies showing reversed significant effects should be investigated: Korunka et al. (2015) showing protective effect of cognitive demand, Innstrand et al. (2008) and Konze et al. (2017) showing detrimental effect of job control, Pisanti et al. (2016) showing detrimental effect of social support and Innstrand et al. (2011) showing protective effect of work-

family conflict. Risk of bias will enable to see whether lack of quality can explain those results, or, if not, confirm where future research might be needed.

Quality assessment on the selected articles is actually currently being conducted by one co-author of the OMEGA-NET team. Results will be added to the above-mentioned results in order to get to a more comprehensive conclusion. We should note here that the purpose won't be to exclude the articles with low quality but rather identify them to discuss the results accordingly. Indeed, Crocetti (2016) recommends to not use results of study quality assessment to exclude studies because it might lead to misleading conclusions. [Usually, moderator analysis is used to test how results are affected by quality. "Doing so, it is possible to derive a comprehensive picture and detect whether lower quality studies found effects that differed systematically from higher quality studies"].

3 Towards a higher-level evidence of burnout etiology

To promote and motivate use of meta-analysis, some would say that this technique handles heterogeneity just fine and that it is precisely the main goal of meta-analysis to explain heterogeneity; actually, the method itself allows to treat the effect as random (in a random effect model versus fixed effect model). Also, meta-regression allows to explore additional amount of the heterogeneity. We agree on that, however, it seems also important to insist on the fact that the participants, methods and settings have still to be reasonably consistent to justify treating the effect as random and get to a summary estimate. In the case where the study methodologies are incompatible, the premise of meta-analysis which is "combining studies of the same factor/treatment in the same population against the same endpoint" is not respected anymore (Conroy, 2015).

In the context of burnout and its predictors, concerns about variation (heterogeneity) across studies – and more specially nonequivalence of measurement tools of both predictors and outcomes– constituted a major obstacle to the feasibility of a meta-analysis. The alternative methods that we used was based on Cochrane handbook

recommendation and seemed more adequate at this stage. Vote counting makes one category (protective, detrimental or no effect) the "winner". Considering sign test based on effect direction enabled us to nuance vote counting based on significant results by 1) considering the total amount of effects and 2) giving some weight to alternative effects. A non-significant sign test could reflect that there was considerable amount of studies going to both directions, or that relatively small number of studies were considered (i.e. sign test of lack of job resources was not significant although 4 studies showed protective effect and 0 was in the opposite direction). Either way, we considered not having enough evidence to consider vote counting result as the "true" effect. Calculating study results based on statistically significant results have already been criticized (Light & Smith, 1971). One major reason is that it does not take into account the sample size, which affects statistical power. Hence, we systematically reported sample size when we talked about significance.

We undoubtfully recognize that this alternative method is far from offering the benefices that a meta-analysis could offer. In absolute, one should rather prefer a meta-analysis as a method of review synthesis; as discussed in section 1, meta-analysis constitute the best source of evidence when performed correctly. However, in order to be addressed properly, a consensus on burnout measurement tool has to be found first. And this cannot be made without a harmonized definition.

In case the consensus is made on the current definition reported by the WHO, measurement tools and cut-offs should be harmonized accordingly. Only then "real" meta-analysis could be considered (as long as enough data is available). In presence of complex databases, where different endpoints are investigated, Crocetti (2016) recommend to adopts an analytic approach by conducting statistical analysis for each outcome in order to avoid mixing "apples and oranges". Also, as reported by Mavridis and Salanti (2013) a more elegant alternative, can be to summarize simultaneously all outcomes of interest through *multivariate meta-analysis*, instead of conducting many separate univariate meta-analyses. As mentioned by Riley (2009), the two main reasons to prefer multivariate meta-analysis over several separate ones are that: 1) overestimated variance of the summary effect size and biased estimates can arise from separate meta-analysis of correlated outcomes and 2) the chance of finding

significant effects are increased with the independent testing of effect on multiple outcomes (that reflect on a same general concept, i.e., 3 dimension of burnout).
However, a harmonization on the current definition might be illusional when considering that this definition actually date of 1976; we might indeed be far from today's reality of things … In line with such doubts, a recent systematic review (that includes the MBI) showed that *only* the Copenhagen Burnout Inventory (CBI) and, to a lesser extent, the Oldenburg Burnout Inventory (OLBI) achieved a satisfactory psychometric validity (Shoman et al., submitted).

More generally, meta-analysis and meta-estimates could be further elaborated through the individual participant data (IDP) approach: "Rather than extracting data from study publication, the original research data are sought directly from the researchers responsible for each study so the data can be reanalyzed and combined in meta-analysis. Instead of coding the study publications, data are extracted directly from study data files which improve data quality, produce more reliable results and reduce risk of bias" (Crocetti, 2016).

4 Conclusion

The present review and internship in which it was embedded enabled me to learn to question and reflect on several issues. Especially, research's result should be analyzed and read commonly with an, even quick, investigation of the methodology and statistical test that were undertaken to get to those results. Publications, even in the most notorious journals, does not make the paper free from bias: an assessment of the quality of 63 meta-analyses published in the famous Psychological Bulletin (considered to represent "state of the arts" research with respect of quality) during the period of 1993 – 2003 showed a non-negligible number of methodological anomalies (College et al.,2009).

"Et si l'on osait une comparaison entre science et magie ? Il était un temps béni où, pour peu que l'on ait quelque sous et si l'on se posait une question, on allait voir l'oracle ou le sorcier. Il avait la réponse. Aujourd'hui, on demanderait d'abord au sorcier ou à

l'oracle quels sont les conflits d'intérêts qu'ils peuvent avoir avec les uns ou les autres. On leur demanderait aussi d'inclure dans leur réponse les liens internet vers d'autres pratiques magiques qui appuieraient leurs conclusions. Est-ce que cela rendra la réponse plus fiable ? Peut-être… Ou peut-être pas."[9]

[9] https://theconversation.com/meta-analyses-de-lart-de-bien-melanger-torchons-et-serviettes-81286

5 References

Adriaenssens, J., De Gucht, V., & Maes, S. (2015). Determinants and prevalence of burnout in emergency nurses: a systematic review of 25 years of research. *International journal of nursing studies*, *52*(2), 649-661.

Ahola, K., Toppinen-Tanner, S., & Seppänen, J. (2017). Interventions to alleviate burnout symptoms and to support return to work among employees with burnout: Systematic review and meta-analysis. *Burnout research*, *4*, 1-11.

Akerboom, S., & Maes, S. (2006). Beyond demand and control: The contribution of organizational risk factors in assessing the psychological well-being of health care employees. *Work & Stress*, *20*(1), 21-36.

Akobeng, A. K. (2005). Understanding systematic reviews and meta-analysis. *Archives of disease in childhood*, *90*(8), 845-848.

Alarcon, G., Eschleman, K. J., & Bowling, N. A. (2009). Relationships between personality variables and burnout: A meta-analysis. *Work & stress*, *23*(3), 244-263.

Aronsson, G., Theorell, T., Grape, T., Hammarström, A., Hogstedt, C., Marteinsdottir, I., ... & Hall, C. (2017). A systematic review including meta-analysis of work environment and burnout symptoms. *BMC public health*, *17*(1), 264.

Bakker, A. B., & Demerouti, E. (2007). The job demands-resources model: State of the art. *Journal of managerial psychology*.

Bandura, A., & Adams, N. E. (1977). Analysis of self-efficacy theory of behavioral change. *Cognitive therapy and research*, *1*(4), 287-310.

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods. 2010;1(2):97-111. doi:10.1002/jrsm.12

Borritz, M., Bültmann, U., Rugulies, R., Christensen, K. B., Villadsen, E., & Kristensen, T. S. (2005). Psychosocial work characteristics as predictors for burnout: findings from 3-year follow up of the PUMA Study. *Journal of occupational and environmental medicine*, *47*(10), 1015-1025.

Bradley, H. B. (1969). Community-based treatment for young adult offenders. *Crime & Delinquency*, *15*(3), 359-370.

Buteau S, Goldberg MS. Methodological issues related to pooling results from panel studies of heart rate variability and its association with ambient air pollution. Environmental Research 2015; 140(0) : 462-5.

Cano-García, F. J., Padilla-Muñoz, E. M., & Carrasco-Ortiz, M. Á. (2005). Personality and contextual variables in teacher burnout. *Personality and Individual differences*, *38*(4), 929-940.

Canu, I. G., Mesot, O., Györkös, C., Mediouni, Z., Mehlum, I. S., & Bugge, M. D. (2019). Burnout syndrome in Europe: towards a harmonized approach in occupational health practice and research. *Industrial health*, 2018-0159.

Conroy, Ronán. (2015). What are the quantitative methods alternative to Meta-Analysis?.

Coryn, C., Hobson, K. (2013) 'Meta-analysis – vote counting, the sign test, power, publication bias, and outliers' [PowerPoint presentation]. Western Michigan University. Available at: https://wmich.edu/sites/default/files/attachments/u58/2015/Lecture_9.pdf (Accessed: 31 July 2020).

Crocetti, E. (2016). Systematic reviews with meta-analysis: Why, when, and how? *Emerging Adulthood*, *4*(1), 3-18.

Dahlin, M., Fjell, J., & Runeson, B. (2010). Factors at medical school and work related to exhaustion among physicians in their first postgraduate year. *Nordic journal of psychiatry*, *64*(6), 402-408.

De Rijk, A. E., Blanc, P. M. L., Schaufeli, W. B., & De Jonge, J. (1998). Active coping and need for control as moderators of the job demand–control model: Effects on burnout. *Journal of Occupational and Organizational Psychology*, *71*(1), 1-18.

Demerouti, E.; Bakker, A.B.; Nachreiner, F.; Schaufeli, W.B. (2001a). "The job demands-resources model of burnout". *Journal of Applied Psychology*. 86: 499–512. doi:10.1037/0021-9010.86.3.499.

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

Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629-634.

Faber, J., & Fonseca, L. M. (2014). How sample size influences research outcomes. *Dental press journal of orthodontics*, *19*(4), 27-29.

Fisher, R. A. (1935). The logic of inductive inference. *Journal of the royal statistical society*, *98*(1), 39-82.

Fısher, R. A. (1935). The design of experiments. *Oliver and Boyd, Edinburgh*.

Freudenberger Herbert, J. (1987). L'Épuisement professionnel: la brûlure interne. *Québec, Editions Gaëtan Morin*.

Freudenberger, H. J., & Richelson, G. (1981). *Burn-out: The high cost of high achievement*. Bantam Books.

Gardner, D. G., & Pierce, J. L. (1998). Self-esteem and self-efficacy within the organizational context: An empirical examination. *Group & Organization Management*, *23*(1), 48-70.

George, J. M. (1992). Extrinsic and intrinsic origins of perceived social loafing in organizations. *Academy of Management Journal*, *35*(1), 191-202.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational researcher*, *5*(10), 3-8.

Greco, T., Zangrillo, A., Biondi-Zoccai, G., & Landoni, G. (2013). Meta-analysis: pitfalls and hints. *Heart, lung and vessels*, *5*(4), 219.

Greenhaus, J. H., & Allen, T. D. (2011). Work–family balance: A review and extension of the literature. In J. C. Quick & L. E. Tetrick (Eds.), *Handbook of occupational health psychology (2nd ed.).* (pp. 165–183). Washington, DC US: American Psychological Association.

Haith-Cooper, M. (2003). An exploration of tutors' experiences of facilitating problem-based learning. Part 1–an educational research methodology combining innovation and philosophical tradition. *Nurse Education Today*, *23*(1), 58-64.

Hedges, L. V., & Olkin, I. (1984). Nonparametric estimators of effect size in meta-analysis. *Psychological Bulletin*, *96*(3), 573.

Hertzberg, T. K., RØ, K. I., VAGLUM, P. J. W., MOUM, T., RØVIK, J. O., Gude, T., ... & Tyssen, R. (2015). Work-home interface stress: an important predictor of emotional exhaustion 15 years into a medical career. *Industrial health*.

Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11):1539-1558. doi:10.1002/sim.1186

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.

Hochwälder, J. (2008). A longitudinal study of the relationship between empowerment and burnout among registered and assistant nurses. *Work*, *30*(4), 343-352.

Innstrand, S. T., Langballe, E. M., & Falkum, E. (2011). The longitudinal effects of individual vulnerability, organisational factors, and work–home interaction on burnout among male church ministers in Norway. *Mental Health, Religion & Culture*, *14*(3), 241-257.

Innstrand, S., Langballe, E., Espnes, G., Falkum, E., & Aasland, O. (2008). Positive and negative work–family interaction and burnout: A longitudinal study of reciprocal relations. *Work & Stress*, *22*(1), 1-15.

James, W. (1890). Principles of psychology. New York: Henry Holt.

Janssen, P. P., Schaufelioe, W. B., & Houkes, I. (1999). Work-related and individual determinants of the three burnout dimensions. *Work & Stress*, *13*(1), 74-86.

Johnson, J. V., & Hall, E. M. (1988). Job strain, work place social support, and cardiovascular disease: a cross-sectional study of a random sample of the Swedish working population. *American journal of public health*, *78*(10), 1336-1342.

Jonge, J. D., Janseen, P. P., & Van Breukelen, G. J. (1996). Testing the demand-control-support model among health-care professionals: A structural equation model. *Work & Stress*, *10*(3), 209-224.

Karasek Jr, R. A. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative science quarterly*, 285-308.

Koeske, G. F., & Koeske, R. D. (1989). Construct validity of the Maslach Burnout Inventory: A critical review and reconceptualization. *The Journal of Applied Behavioral Science*, *25*(2), 131-144.

Konze, A. K., Rivkin, W., & Schmidt, K. H. (2017). Is job control a double-edged sword? A cross-lagged panel study on the interplay of quantitative workload, emotional dissonance, and job control on emotional exhaustion. *International journal of environmental research and public health*, *14*(12), 1608.

Koricheva, J., Gurevitch, J., & Mengersen, K. (Eds.). (2013). *Handbook of meta-analysis in ecology and evolution*. Princeton University Press.

Kristensen, T. S., Borritz, M., Villadsen, E., & Christensen, K. B. (2005). The Copenhagen Burnout Inventory: A new tool for the assessment of burnout. *Work & Stress*, *19*(3), 192-207.

Landsbergis, P. A. (1988). Occupational stress among health care workers: a test of the job demands-control model. *Journal of Organizational behavior*, *9*(3), 217-239.

Lane, J., Lane, A. M., & Kyprianou, A. (2004). Self-efficacy, self-esteem and their impact on academic performance. *Social Behavior and Personality: an international journal*, *32*(3), 247-256.

Langelaan, S., Bakker, A. B., Van Doornen, L. J., & Schaufeli, W. B. (2006). Burnout and work engagement: Do individual differences make a difference?. *Personality and individual differences*, *40*(3), 521-532.

Last, J. M., Harris, S. S., Thuriaux, M. C., & Spasoff, R. A. (2001). *A dictionary of epidemiology*. International Epidemiological Association, Inc.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. Springer publishing company.

Lee, R. T., & Ashforth, B. E. (1996). A meta-analytic examination of the correlates of the three dimensions of job burnout. *Journal of applied Psychology*, *81*(2), 123.

Light, R., & Smith, P. (1971). Accumulating evidence: Procedures for resolving contradictions among different research studies. *Harvard educational review*, *41*(4), 429-471.

Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. Oxford University Press.

Maslach, C. (1976). Burned-out. Human Behavior, 5(9), 16–22

Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of organizational behavior*, *2*(2), 99-113.

Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). Job burnout. *Annual review of psychology*, *52*(1), 397-422.

Maslach, Christina & Leiter, Michael. (2008). Early Predictors of Job Burnout and Engagement. The Journal of applied psychology. 93. 498-512. 10.1037/0021-9010.93.3.498.

Matt, G. E., Brewer, A., & Sklar, M. (2010). External validity.

Mavridis, D., & Salanti, G. (2013). A practical introduction to multivariate meta-analysis. *Statistical methods in medical research*, *22*(2), 133-158.

Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

Molina, A.D. (2016). Value congruence. In Farazmand, E. (Eds.), Global Encyclopedia of Public Administration, Public Policy, and Governance (pp.1–7). Cham, Switzerland: Springer.

Moore, J. E. (2000). One road to turnover: An examination of work exhaustion in technology professionals. *MIS quarterly*, 141-168.

Nieminen, P., Lehtiniemi, H., Vähäkangas, K., Huusko, A., & Rautio, A. (2013). Standardised regression coefficient as an effect size index in summarising findings in epidemiological studies. *Epidemiology, Biostatistics and Public Health*, *10*(4).

O'rourke, K. (2007). An historical perspective on meta-analysis: dealing quantitatively with varying study results. *Journal of the Royal Society of Medicine*, *100*(12), 579-582.

O'Connor, K., Neff, D. M., & Pitman, S. (2018). Burnout in mental health professionals: A systematic review and meta-analysis of prevalence and determinants. *European Psychiatry*, *53*, 74-99.

Pisanti, R., van der Doef, M., Maes, S., Meier, L. L., Lazzari, D., & Violani, C. (2016). How changes in psychosocial job characteristics impact burnout in nurses: a longitudinal analysis. *Frontiers in Psychology*, *7*, 1082.

Rafferty, Y., Friend, R., & Landsbergis, P. A. (2001). The association between job skill discretion, decision authority and burnout. *Work & stress*, *15*(1), 73-85.

Richter, A., Schraml, K., & Leineweber, C. (2015). Work–family conflict, emotional exhaustion and performance-based self-esteem: reciprocal relationships. *International archives of occupational and environmental health*, *88*(1), 103-112.

Riley, R. D. (2009). Multivariate meta-analysis: the effect of ignoring within-study correlation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*(4), 789-811.

Rotenstein, L. S., Torre, M., Ramos, M. A., Rosales, R. C., Guille, C., Sen, S., & Mata, D. A. (2018). Prevalence of burnout among physicians: a systematic review. *Jama*, *320*(11), 1131-1150.

Rubio, C., Osca, A., Recio, P., Urien, B., & Peiró, J. M. (2015). Work-family conflict, self-efficacy, and emotional exhaustion: A test of longitudinal effects. *Revista de Psicología del Trabajo y de las Organizaciones*, *31*(3), 147-154.

Ruehlman, L. S., & Wolchik, S. A. (1988). Personal goals and interpersonal support and hindrance as factors in psychological distress and well-being. *Journal of personality and social psychology*, *55*(2), 293.

Schardt, C., Adams, M. B., Owens, T., Keitz, S., & Fontelo, P. (2007). Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC medical informatics and decision making*, *7*(1), 16.

Schaufeli, W. B., Leiter, M. P., Maslach, C., & Jackson, S. E. (1996). MBI-General Survey (MBI-GS). *Palo Alto, CA: Mindgarden*.

Shirom, A. (1989). Burnout in work organizations.

Simpson, R. J. S., & Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *The British Medical Journal*, 1243-1246.

Stegenga, J. (2011). Is meta-analysis the platinum standard of evidence?. *Studies in history and philosophy of science part C: Studies in history and philosophy of biological and biomedical sciences*, *42*(4), 497-507.

Sterne JAC, Egger M. Funnel plots for detecting bias in meta-analysis. J Clin Epidemiol. 2001;54(10):1046-1055. doi:10.1016/S0895-4356(01)00377-8

Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., ... & Tetzlaff, J. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *Bmj*, *343*.

Tacconelli, E. (2010). Systematic reviews: CRD's guidance for undertaking reviews in health care. *The Lancet Infectious Diseases*, *10*(4), 226.

Taris, T. W., Peeters, M. C., Le Blanc, P. M., Schreurs, P. J., & Schaufeli, W. B. (2001). From inequity to burnout: The role of job stress. *Journal of Occupational Health Psychology*, *6*(4), 303.

Taris, T. W., Schreurs, P. J., & Schaufeli, W. B. (1999). Construct validity of the Maslach Burnout Inventory-General Survey: A two-sample examination of its factor structure and correlates. *Work & Stress*, *13*(3), 223-237.

Theorell, T., Osika, W., Leineweber, C., Hanson, L. L. M., Horwitz, E. B., & Westerlund, H. (2013). Is cultural activity at work related to mental health in employees?. *International archives of occupational and environmental health*, *86*(3), 281-288.

van der Molen, H. F., Nieuwenhuijsen, K., Frings-Dresen, M. H., & de Groene, G. (2020). Work-related psychosocial risk factors for stress-related mental disorders: an updated systematic review and meta-analysis. *BMJ open*, *10*(7), e034849.

van Wely, M. (2014). The good, the bad and the ugly: meta-analyses.

Wagenaar, W. A., Groeneweg, J., Hudson, P. T. W., & Reason, J. T. (1994). Promoting safety in the oil industry. The ergonomics society lecture presented at the ergonomics

society annual conference, Edinburgh, 13-16 April 1993. *Ergonomics*, *37*(12), 1999-2013.

Wagenaar, W. A., Hudson, P. T., & Reason, J. T. (1990). Cognitive failures and accidents. *Applied Cognitive Psychology*, *4*(4), 273-294.

Wayne, J. H., Grzywacz, J. G., Carlson, D. S., & Kacmar, K. M. (2004, May). Defining work-family facilitation: A construct reflecting the positive side of the work-family interface. In *annual meeting of the Society for Industrial and Organizational Psychology, Chicago, IL*.

Weber, A., & Jaekel-Reinhard, A. (2000). Burnout syndrome: a disease of modern societies?. *Occupational medicine*, *50*(7), 512-517.

West, C. P., Dyrbye, L. N., Erwin, P. J., & Shanafelt, T. D. (2016). Interventions to prevent and reduce physician burnout: a systematic review and meta-analysis. *The Lancet*, *388*(10057), 2272-2281.

World health organization. Burn-out an "occupational phenomenon": international classification of diseases what are social determinants of health? 2019. Available https://www.who.int/mental_health/evidence/burnout/en/?mc_cid=c8878361fa&mc_eid=8d4606394a [Accessed 21 july 2020].

Yates, F. and Cochran, W. G. (1938). The analysis of groups of experiments. J. Agric. Sci. 28: 556-580.

Zwingmann, I., Wolf, S., & Richter, P. (2016). Every light has its shadow: a longitudinal study of transformational leadership and leaders' emotional exhaustion. *Journal of Applied Social Psychology*, *46*(1), 19-33.

Appendix A: Datasets used to construct plots by subfamily

Situational characteristics

Job demands

Article	Factor's original name	Included in	Sample size	T-value
Chrisopoulos_2010	cognitive demands	cognitive demands	179	1.250
Korunka_2015	intensified learning demands	cognitive demands	587	-3.289
Kubicek_2015	job complexity	cognitive demands	591	-0.429
Borritz_2005 (CBI)	emotional demands	emotional demands	952	-0.470
Borritz_2005 (CBI)	demands for hiding emotions	emotional demands	952	0.600
Chrisopoulos_2010	emotional demands	emotional demands	179	0.429
Feuerhahn_2013_2	emotional job demands	emotional demands	87	0.118
Hertzberg_2016	emotional demands	emotional demands	274	1.316
Idris_2014_1	emotional demands	emotional demands	117	2.125
Konze_2017	emotional dissonance	emotional demands	139	3.600
Kubicek_2015	emotional rule dissonance	emotional demands	591	1.750
Lorente Prieto_2008	job demands (emotional demands)	emotional demands	274	2.667
Van_de_ven_2013	emotional job demands	emotional demands	4622	2.564
Vegchel_2004	emotional demand	emotional demands	2255	3.333
Adriaenssens_2015	job demands	job demands	170	2.326
Garbarino_2013	job demand	job demands	289	2.326
Hakanen_2008	job demands	job demands	3035	3.286
Hudek-Knezevic_2011_1	job demand (quantitative overload + role conflict)	job demands	118	3.884
Laugaa_2008	job demands (quantitative and qualitative)	job demands	259	2.576
Theorell_2013	psychological demands	job demands	11525	3.891
Chrisopoulos_2010	physical demands	physical demands	179	0.571
Gelsema_2006	physical demands	physical demands	381	2.576
Borritz_2005 (CBI)	quantitative demands	quantitative demands	952	-0.113
Borritz_2005 (CBI)	work pace	quantitative demands	952	0.912
Dahlin_2010 (OLBI)	weekly working hours	quantitative demands	186	1.600
Feuerhahn_2013_2	time pressure	quantitative demands	87	-0.231
Fong_1993	job demands (quantitative)	quantitative demands	84	2.321
Fritz_2006 OLBI	workload after vacation	quantitative demands	221	2.571
Garbarino_2013	Effort	quantitative demands	289	2.581
Gelsema_2006	work and time demands	quantitative demands	381	1.282
Gil-Monte_2008	work overload	quantitative demands	316	2.308
Goddard_2006	work pressure	quantitative demands	79	3.295
Gregory_2015	workload	quantitative demands	153	3.888
Hertzberg_2016	time pressure	quantitative demands	274	1.690
Hertzberg_2016	working hours per week	quantitative demands	274	0.845
Hornung_2013	work overload	quantitative demands	95	-0.543
Huang_2012	job demands (work fast and hard, great deal to do, too little time)	quantitative demands	299	1.297
Instrand_2011_1	workload	quantitative demands	308	-1.277
Jimenez_2017_1	workload	quantitative demands	141	2.073
Konze_2017	quantitative workload (time pressure and work volume)	quantitative demands	139	1.400
Korunka_2015	work intensification demands	quantitative demands	587	3.301
Lorente Prieto_2008	role overload	quantitative demands	274	5.400
Park_2016	hours of direct services	quantitative demands	152	1.640
Pisanti_2016	job demands (work and time pressure)	quantitative demands	217	-1.290
Taris_2010	work overload/time pressure	quantitative demands	828	2.581
Vegchel_2004	quantitative demands	quantitative demands	2255	2.553

Decision Latitude (control)

Article	Factor's original name	Included in	Sample size	T-value
Dubois_2014	autonomy	autonomy	96	-2.3225806
Instrand_2011_1	autonomy	autonomy	308	2.5806452
Leiter_2013	decision authority	decision auhtority	4396	-1.25
Borritz_2004	influence	decision authority	952	-1.5973856
Dubois_2014	informal power	decision authority	96	-3.8905919
Gelsema_2006	decision authority	decision authority	381	-1.2857143
Salanova_2005	Indiscipline managment (defined as the chance of admonishing problematic students)	decision authority	274	0.25
Adriaenssens_2015	decision lattitiude	decision latitude	170	0.3333333
Garbarino_2013	decision lattitude	decision latitude	289	-1.2857143
Huang_2012	job control	decision latitude	299	-1.2888889
Jimenez_2017_1	control	decision latitude	141	-0.2173913
Konze_2017	job control	decision latitude	139	2.8
Kubicek_2014	job control	decision latitude	591	-1.2727273
Pisanti_2015	decision lattitiude	decision latitude	217	-1.2903226
Theorell_2013	decision lattitude	decision latitude	11525	-3
Vegchel_2004	job control	decision latitude	2255	-2.3255814
Borritz_2004	possibilities for development	skill discretion	952	-1.4740061
Dubois_2014	opportunities for stimulating work	skill discretion	96	-2.3300971
Gelsema_2006	skill discretion	skill discretion	381	1.2857143
Leiter_2013	skill discretion	skill discretion	4396	1.25
Salanova_2005	cLass managment (chance of changing type or dynamics of the activities)	skill discretion	274	-1

Job resources

Article	Factor's original name	Included in	Sample size	T-value
Boamah_2017	short-staffing	lack of resources	405	2.5758293
Meier_2015	lack of reward	lack of resources	246	2.3867077
Salanova_2005	technical obstacles	lack of resources	274	0.25
Spence-Laschinger_2008	Effort-reward imbalance	lack of resources	134	3.8905919
Adriaenssens_2015	material resources	resources	170	-0.25
Adriaenssens_2015	personnel resources	resources	170	-0.25
Adriaenssens_2015	reward	resources	170	1.6666667
Borritz_2005	meaning of work	resources	952	1.2772544
Chrisopoulos_2010	cognitive ressources	resources	179	-0.2222222
Chrisopoulos_2010	physical ressources	resources	179	0.5
Dahlin_2010	learning climate	resources	186	-2.326348
Garbarino_2013	reward	resources	289	-2.5758293
Gelsema_2006	material ressources	resources	381	-1.2815516
Gelsema_2006	personnel resources	resources	381	1.2815516
Gelsema_2006	reward	resources	381	1.2815518
Goddard_2006	worplace innovation	resources	79	-3.8905919
Gonzales-Morales_2010	quality of work place facilities	resources	555	-1.6759777
Gonzales-Morales_2010	teacher-students ratio	resources	555	-0.4285714
Hakanen_2008	job ressources	resources	3035	-2.3263479
Jimenez_2017_1	job ressources	resources	141	-0.2211187
Jimenez_2017_1	reward	resources	141	-1.2815516
Kutney-Lee_2013	staffing (number of patients nurses cared for)	resources	5957	-2.1700904
Theorell_2013	cultural activity at work	resources	11525	-2

Interactions at work

Article	Factor's original name	Included in	Sample size	T-value
Chrisopoulos_2010	emotional ressources (emotional support)	emotional support	179	-1.1111111
Feuerhahn_2013	emotional ressources (emotional support)	emotional support	56	-1.5
Van_de_ven_2013	emotional job ressources	emotional support	711	1.5
Fernet_2010	quality of relationships with coworkers	good interpersonal relations	276	-2.5758293
Gelsema_2006	nurse-doctor collaboration	good interpersonal relations	381	1.2815516
Jimenez_2017_1	community	good interpersonal relations	141	0.3120533
Jimenez_2017_1	fairness (captures the extent to which decisions and resources allocation at work are perceived as fair and equitable)	good interpersonal relations	141	-0.6666433
Ramarajan_2008	organizational respect	good interpersonal relations	108	-2.3263479
Welp_2016	interpersonal teamwork -> quality of relationships at work	good interpersonal relations	493	-1.5
Adriaenssens_2015	social harassment	social hindrance	170	-1.6666667
Angelo_2015	organizational demands >> conflict and interpersonal problems	social hindrance	1610	2.3263479
Feuerhahn_2013	classroom disruption	social hindrance	56	0.0769231
Feuerhahn_2013	parents' criticism	social hindrance	56	2.1111111
Feuerhahn_2013	conflicts with colleagues	social hindrance	56	1.5
Fida_2018	work incivilitiy from supervisor	social hindrance	596	1.2815516
Fida_2018	work incivilitiy from collegues	social hindrance	596	2.3263479
Fida_2018	work incivilitiy from physician	social hindrance	596	2.5758293
Lapointe_2013	psychological contract breach (employees' perception that the organization has failed to meet obligations)	social hindrance	224	1.2815516
Laugaa_2008	conflicts and interperosonal problems	social hindrance	259	2.5758293
Salanova_2005	parents/students obstacles	social hindrance	274	-0.25
Adriaenssens_2015	social support	social support	170	-3.5
Birkeland_2018	perceived supervision support	social support	1263	-3
Birkeland_2018	perceived coworkers support	social support	1263	-2
Borritz_2005	interpersonal relations and leadership (social support)	social support	952	0.9223141
Dubois_2014	Ressource loss (supervisor support)	social support	96	-1.0833333
Dubois_2014	support from colleagues (group cohesion)	social support	96	-2.3188406
Fusilier_2005	social support	social support	260	-1.2815516
Garbarino_2013	occuparional stress variables (support)	social support	289	-1.2941176
Gelsema_2006	social support supervisor	social support	381	1.3333333
Gelsema_2006	social support collegues	social support	381	1.3333333
Hertzberg_2016	support from partner	social support	274	-0.1219512
Hertzberg_2016	support from colleagues	social support	274	-1.5
Lizano_2012	supervisory support	social support	335	-1.7142857
Lizano_2012	organizational support	social support	335	1.5
Pisanti_2015	social support	social support	217	3.1428571
Poulin_1993	organisational variables (supervisor support)	social support	879	-3.2857143
Salanova_2005	social facilitators	social support	274	0.6
Turgut_2016	perceived organizational support	social support	709	-2.3125
Vegchel_2004	job ressources (social support)	social support	2255	-2.3333333

Communication & Leadership

Article	Factor's original name	Included in	Sample size	T-value
Borritz_2005	role conflict	conflicting/ poor communication	952	0.905
Lizano_2015	role ambiguity	conflicting/ poor communication	361	2.576
Travis_2016	role ambiguity	conflicting/ poor communication	362	1.625
Travis_2016	role conflict	conflicting/ poor communication	362	2.100
Van_der_Ploeg_2003	poor communication	conflicting/ poor communication	123	2.326
Borritz_2005	quality of leadership	enriching leadership	952	1.726
Gregersen_2014	transformational leadership	enriching leadership	339	-1.282
Idris_2014_1	psychosocial safety climate (management commitment, organzational communication)	enriching leadership	117	-2.326
Zwingmann_2016	transformational leadership (leader encourage, inspire and motivate employees)	enriching leadership	2324	2.326
Zwingmann_2016	laissez-faire leadership ("guided liberty" where employees get all necessary skills to get to their goal)	enriching leadership	2324	2.326
Adriaenssens_2015	work agreements	informational team climate	170	0.250
Borritz_2005	role clarity	informational team climate	952	-1.264
Borritz_2005	predictability	informational team climate	952	-2.047
Gelsema_2006	work agreements	informational team climate	381	-1.282
Gelsema_2006	communication	informational team climate	381	-1.282
Leiter-2013	information flow (amount and quality of information to which respondents had access)	informational team climate	4356	-2.576
Leiter-2013	predictability	informational team climate	4356	-1.282
Turgut_2016	informational team climate	informational team climate	709	0.381
Theorell_2012	non-listening leadership (Does your manager listen to you and pay attention to what you say?)	non-collaborative leadership	3285	-0.529
Theorell_2012	self centered leadership (calculated from three questions (non-participating, asocial and loner)	non-collaborative leadership	3285	0.033
Theorell_2013	non-listening manager	non-collaborative leadership	8315	1.625

Individual characteristics

Personality, Coping & Self evaluation

Article	Factor's original name	Included in	Sample size	T-value
Laugaa_2008	coping centered on the problem	adaptative coping	259	-2.575829
Laungaa_2009	problem oriented coping	adaptative coping	410	-2.326348
Van_de_ven_2013	emotional support seeking	adaptative coping	711	-1.281552
Firoozabadi_2018 (OLBI)	problem solving	adaptative coping	123	-2.326348
Philipp_2010	deep acting (change of the inner emotional state - regulating feelings)	adaptative coping	102	-2.575829
Martinez-inigo_2016	deep acting	adaptative coping	233	-0.125
Laugaa_2008	traditional teaching coping	maladaptative coping	259	2.575829
Firoozabadi_2018 (OLBI)	affective rumination	maladaptative coping	123	2.326348
Philipp_2010	surface acting (superficial expression of an emotion which is not actually felt - regulating expressions)	maladaptative coping	102	-1.281552
Martinez-inigo_2016	surface acting	maladaptative coping	233	0.333333
Feuerhahn_2013_2	self efficacy	self efficacy	87	-3
Taris_2010	professional efficacy	self efficacy	828	-2.326348
Gregersen_2014	occupational self efficacy	self efficacy	339	-1.281552
Fida_2018	relational self-efficacy	self efficacy	596	-2.575829
Gil-Monte_2008	self efficacy	self efficacy	316	-1.281552
Laugaa_2008	self efficacy	self efficacy	259	-2.575829
Park, 2016	general self efficacy	self efficacy	156	1.285714
Pomaki, 2009	goal self efficacy	self efficacy	222	-0.2
Hochwälder_2008	psychological empowerment (meaning, competence , self determination, impact)	self efficacy	838	2.575829
Spence-Laschinger_2008	core self evaluation	self esteem	134	-2.17009
Instrand_2011 OLBI	job performance-based self-esteem	self esteem	308	-1.281552
Poulin_1993	personnal variables (self esteem)	self-esteem	879	-2.326348
Lapointe_2011	OBSE : organizational based self esteem	self-esteem	224	-1.281552
Dahlin_2010	Performance-based self-esteem	self-esteem	186	1.281552
Richter_2015	performance-based self-esteem	self-esteem	3378	2.326348
Fusilier_2005	type A behavior (excessive competitiveness and aggression and a fast-paced life style)	unvalued traits/characteristics	260	2.575829
Hudek-Knezevic_2011	neurotcism	unvalued traits/characteristics	118	0.622
Hudek-Knezevic_2011	extraversion	valued traits/characteristics	118	1.680
Hudek-Knezevic_2011	agreeableness	valued traits/characteristics	118	-0.622
Hudek-Knezevic_2011	conscientiousness	valued traits/characteristics	118	0.633
Hudek-Knezevic_2011	openess to experience	valued traits/characteristics	118	0.516
Fusilier_2005	hardiness (perception of events as opportunities and challenges rather than stressors)	valued traits/characteristics	260	-1.281552
Garbarino_2013	extraversion	valued traits/characteristics	289	-1.282
Garbarino_2013	agreeableness	valued traits/characteristics	289	-1.282
Garbarino_2013	consciouentiousness	valued traits/characteristics	289	-1.282
Garbarino_2013	emotional stability	valued traits/characteristics	289	-3.291
Garbarino_2013	openess	valued traits/characteristics	289	1.282
Lindeman_2017	agreeableness personalit trait	valued traits/characteristics	55	-2.575829
Lindeman_2017	emotional stability personality trait	valued traits/characteristics	55	-0.426148
Lindeman_2017	conscientiousness personality trait	valued traits/characteristics	55	0.568051
Lindeman_2017	emotional intelligence	valued traits/characteristics	55	-2.326348
Lorente Prieto_2008	personal resources (emotional competences)	valued traits/characteristics	274	-2.5

Job attitudes

Article	Factor's original name	Included in	Sample size	T-values
Lapointe_2011	normative commitment	extrinsicaly motivated behavior	260	2.32634789
Lapointe_2011	continuance-sacrifice commitment	extrinsicaly motivated behavior	260	1.28155157
Lapointe_2011	continuance-alternatives commitment	extrinsicaly motivated behavior	260	2.32634788
Hudek-Knezevic_2011	continuance commitment	extrinsicaly motivated behavior	118	-1.0110343
Tonjes_2009	performance-avoidance goal orientation	extrinsicaly motivated behavior	80	2.57582932
Childs_2012	socially prescribed perfectionism	extrinsicaly motivated behavior	69	1.28155157
Lapointe_2011	affective commitment	intrinsicaly motivated behavior	260	-1.2815516
Fernet_2010	self-determined work motivation	intrinsicaly motivated behavior	276	-1.2815516
Petrou, 2015	self-initiated resources seeking	intrinsicaly motivated behavior	580	-0.7891917
Petrou, 2015	self-initiated challenges seeking	intrinsicaly motivated behavior	580	-2.3263479
Petrou, 2015	self-initiated reducing demands	intrinsicaly motivated behavior	580	2.32634787
Instrand_2011 OLBI	goal orentiation	intrinsicaly motivated behavior	308	1.28155159
Tonjes_2009	learning goals orientation	intrinsicaly motivated behavior	80	-2.3263479
Childs_2012	self oriented perfectionnism	intrinsicaly motivated behavior	69	1.28155158
Turgut_2016	resistance to change	negative job attitude	709	3.66666667
Petrou, 2015	sensitivity to change	negative job attitude	580	2.57582934
Garbarino_2013	overcomitment	negative job attitude	289	3.29052675
Birkeland_2018	obsessive passion	negative job attitude	1263	9
Fernet_2014	obsessive passion	negative job attitude	175	2.57582931
Lavigne_2012	obssesive passion	negative job attitude	113	2.57582932
Fritz_2006 (OLBI)	negative work reflection	negative job attitude	221	2.57582929
Gonzales-Morales_2010	absenteeism rate	negative job attitude	555	-1
Lu_2013	sickness presenteeism	negative job attitude	245	3.29052675
Lapointe_2013	organizational commitment	positive job attitude	224	-2.3263479
Lapointe_2013	commitment to the supervisor	positive job attitude	224	-1.2815516
Hudek-Knezevic_2011	affective-normative commitment	positive job attitude	118	-1.9953933
Birkeland_2018	hamonious passion	positive job attitude	1263	-9
Fernet_2014	hamonious passion	positive job attitude	175	-2.5758293
Fritz_2005 OLBI	positif work reflection	positive job attitude	87	-1.959964
Fritz_2006 (OLBI)	positive work reflection	positive job attitude	221	-1.2815516
Pomaki_2009	perception of goal attainability	positive job attitude	222	-1.8

Personal events

Article	Factor's original name	Included in	Sample size	T-value
Fritz_2005	social activity (week end)	leisure	87	-1.281552
Fritz_2006 (OLBI)	relaxation	leisure	221	-1.281552
Fritz_2006 (OLBI)	mastery	leisure	221	-2.326348
Fusilier_2005	physical exerices	leisure	260	-1.281552
Ragsdale_2016	cell phone attachement	leisure	59	-2.575829
Fritz_2005	non work hassles	stressfull life events	87	-1.281552
Fritz_2006 (OLBI)	nonwork hassles	stressfull life events	221	2.326348
Fusilier_2005	stressfull life events	stressfull life events	260	2.326348
Jensen_2017	psychological health complaints	stressfull life events	1702	3.290527
Park, 2016	mental heatlh	stressfull life events	152	1.644737

Work-family interface

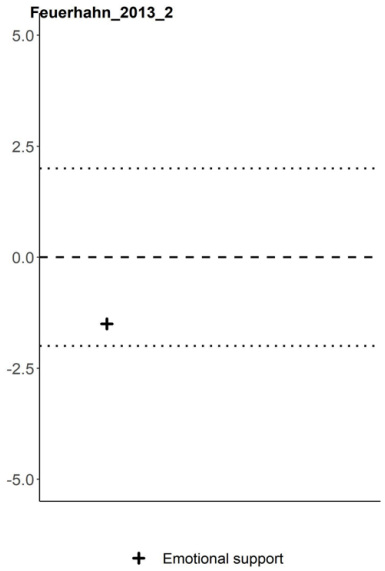
Article	Factor's original name	Included in	Sample size	T-value
Innstrad_2008 (OLBI)	family to work conflict	family-work conflict	1565	2.33
Instrand_2011 OLBI	family work conflict	family-work conflict	308	1.28
Westman_2008	family-work conflict	family-work conflict	66	2.81
Instrand_2011 OLBI	family work facilitation	family-work facilitation	308	1.28
Gregory_2015	values	value congruency	153	-3.89
Instrand_2011 OLBI	value congruency	value congruency	308	1.28
Jimenez_2017_1	values (" my values and the organization's values are alike")	value congruency	141	0.23
Hertzberg_2016	work home interface stress	work-family conflict	274	-3.29
Hornung_2013	work-family conflict	work-family conflict	95	1.28
Innstrad_2008 (OLBI)	work to family conflict (pressure at work hamper functioning at home)	work-family conflict	1565	2.33
Instrand_2011 OLBI	work-home conflict	work-family conflict	308	-2.33
Jensen_2017	work-family conflict	work-family conflict	1702	3.29
Lizano_2012	work family conflict	work-family conflict	335	5.54
Lizano_2015	work family conflict	work-family conflict	361	2.58
Mauno_2015	work family conflict	work-family conflict	814	2.58
Ragsdale_2016	work related cell phone use (during non work time)	work-family conflict	59	-1.28
Richter_2015	work– family conflict	work-family conflict	3378	1.28
Rubio_2015	work-family conflict	work-family conflict	242	2.33
Travis_2016	work-family conflict	work-family conflict	362	3.25
Westman_2008	work-family conflict	work-family conflict	66	2.81
Innstrad_2008 (OLBI)	work to family facilitation	work-family facilitation	1565	-2.33
Instrand_2011 OLBI	work-home facilitation	work-family facilitation	308	1.28
Mauno_2015	work family enrichment	work-family facilitation	814	-2.33

Job outcomes

Article	Factor's original name	Included in	Sample size	T-value
Figueiredo-Ferraz_2012	work satisfaction	job satisfaction	316	-1.28
Lindeman_2017	total positive work experiences	job satisfaction	55	-2.97
Poulin, 1993	satisfaction with clients	job satisfaction	879	-2.33
Dahlin_2010 OLBI	worries about futur endurance/competence	stress	186	3.29
Fusilier_2005	stressfull work events	stress	260	1.28
Hornung_2013	patient demands	stress	95	-2.58
Laugaa_2008	perceived stress	stress	259	2.58
Lizano_2012	job stress	stress	335	0.30
McManus_2002	stress	stress	365	2.33
Poulin, 1993	job stress	stress	879	3.29
Taris_2001	stressors (Students)	stress	828	-2.58
Taris_2001	stressors (colleagues)	stress	828	-2.58
Van_der_Ploeg_2003	job physicial strain	stress	123	2.33

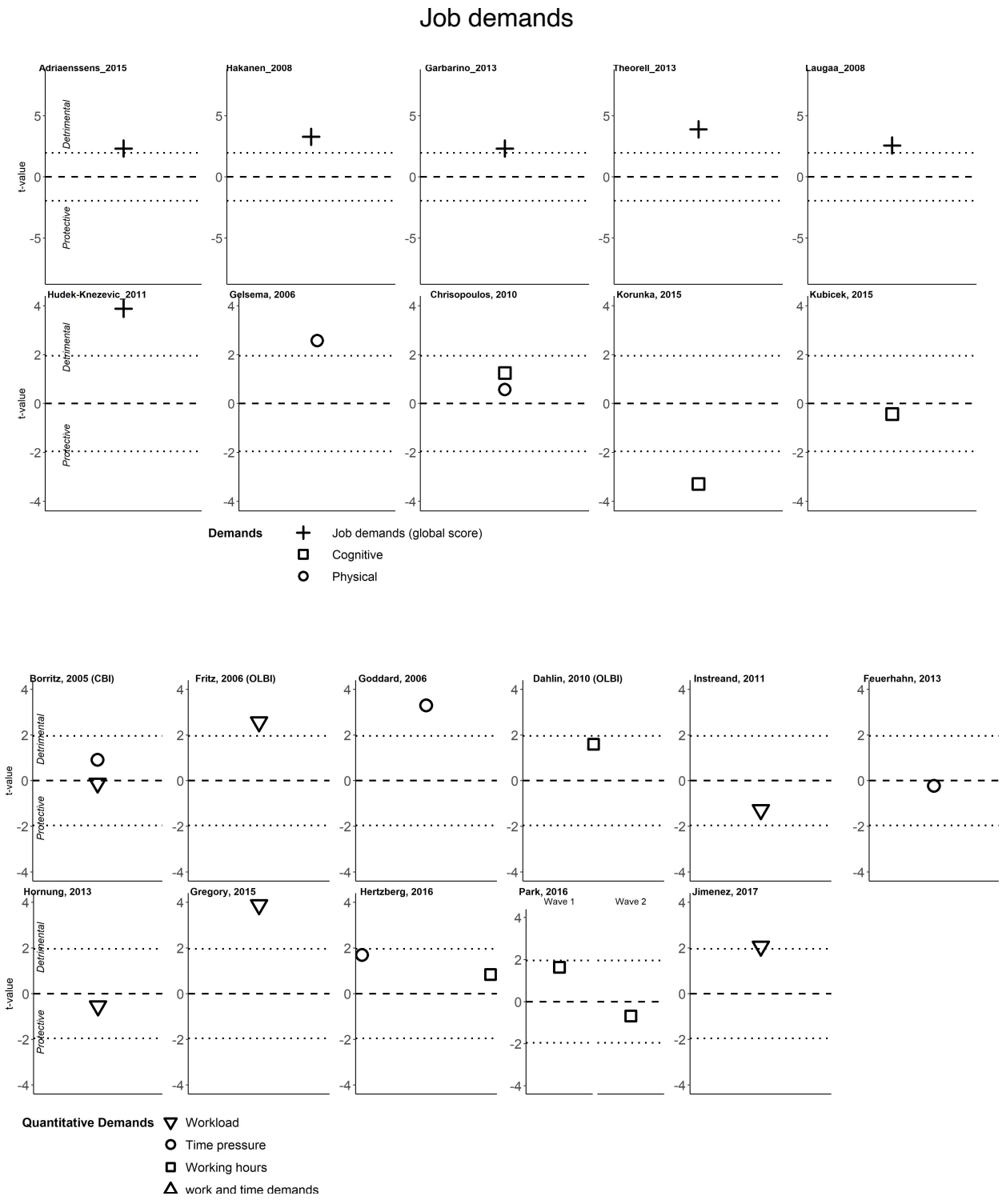
Appendix B: First version detailed plots

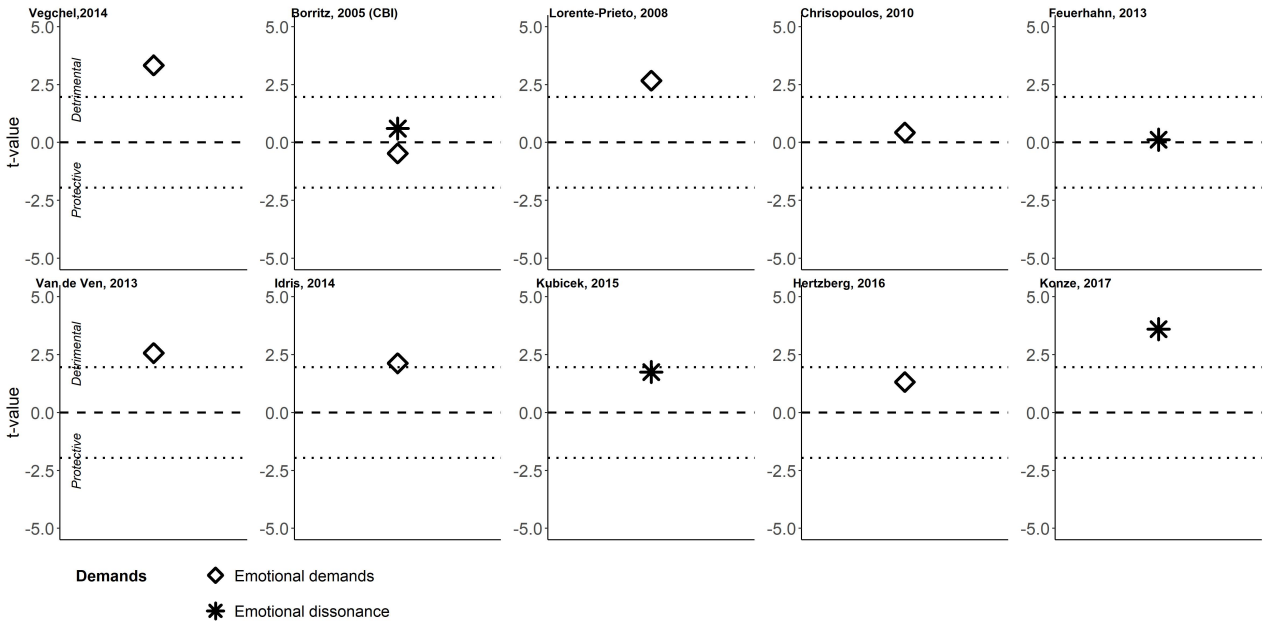
Description of the constructed plots

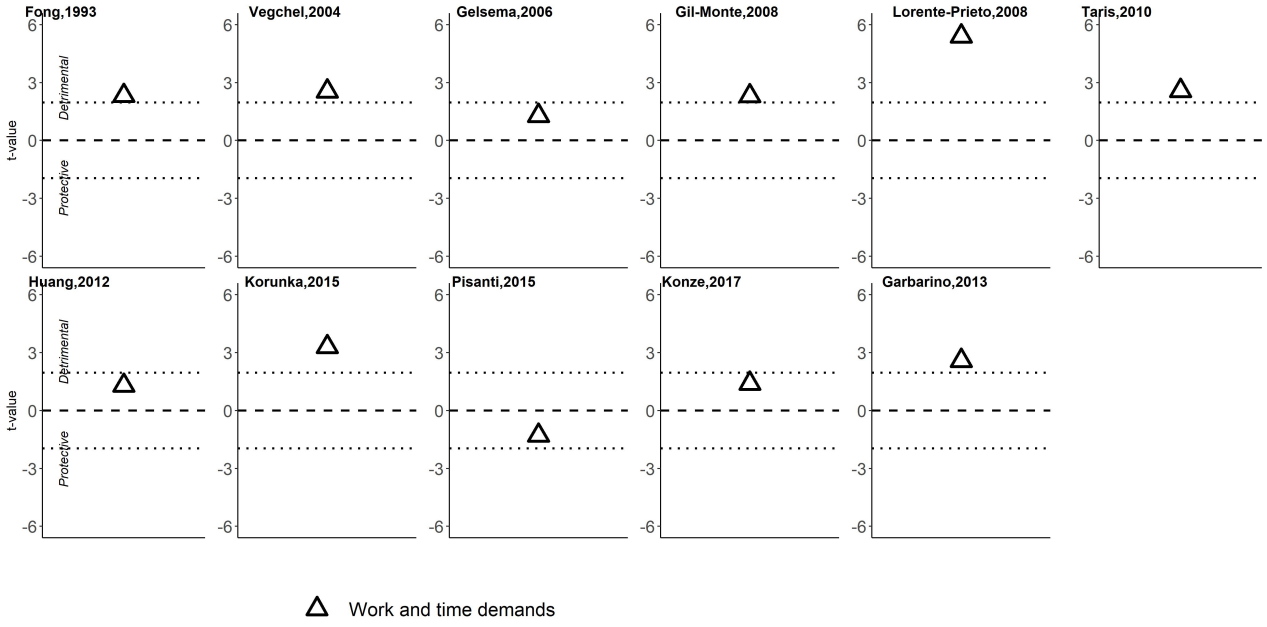


At zero there is no relationship between the predictor and burnout. Values above 0 represent a positive relationship between the predictor and emotional exhaustion which is labelled as a detrimental effect. Values under 0 represent a negative relationship between the predictor and emotional exhaustion which is labelled as a protective effect. Dotted lines at – 1.96 and + 1.96 represent the limits of the interval of 95% confidence interval. Hence, values that are higher than 1.96 or lower than – 1.96 represent significant relationships.

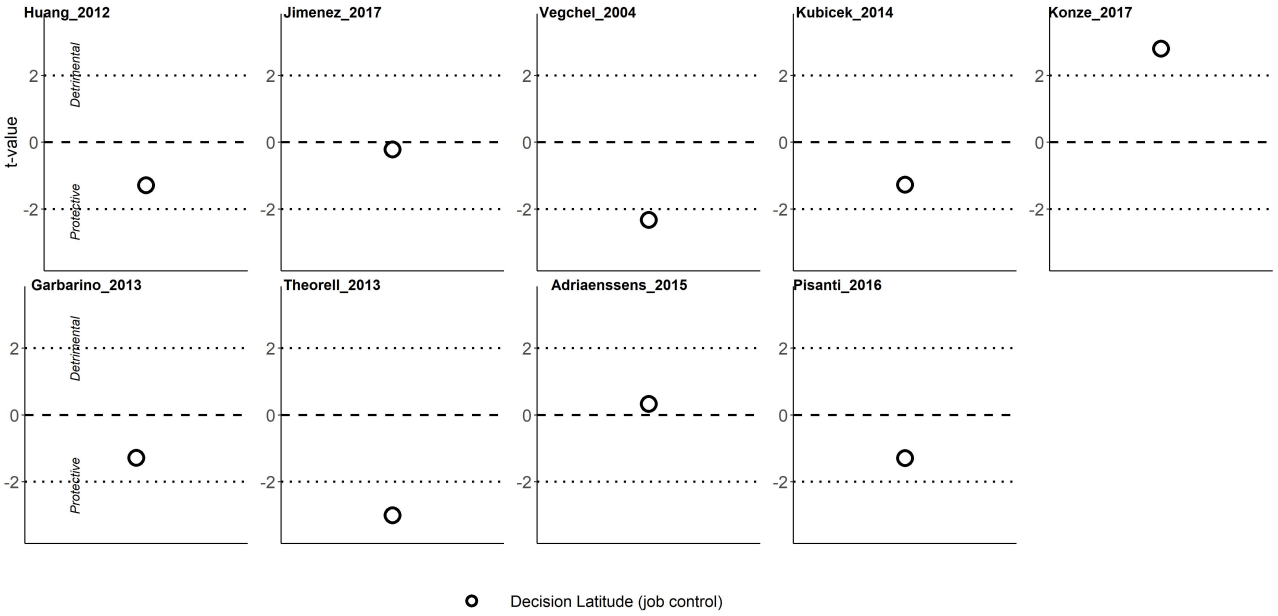
Situational characteristics

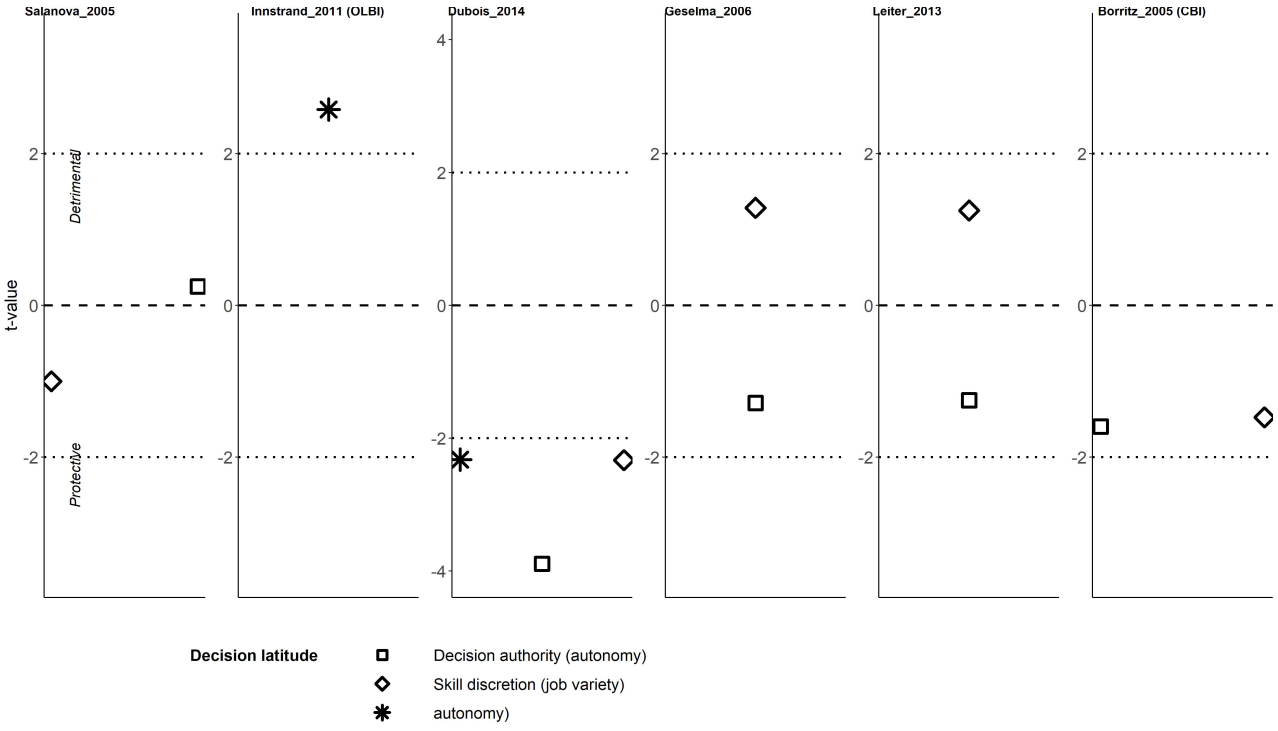



Vegchel,2014
Borritz, 2005 (CBI)
Lorente-Prieto, 2008
Chrisopoulos, 2010
Feuerhahn, 2013
Van de Ven, 2013
Idris, 2014
Kubicek, 2015
Hertzberg, 2016
Konze, 2017
t-value
Detrimental
Protective
5.0
2.5
0.0
-2.5
-5.0
Demands
Emotional demands
Emotional dissonance

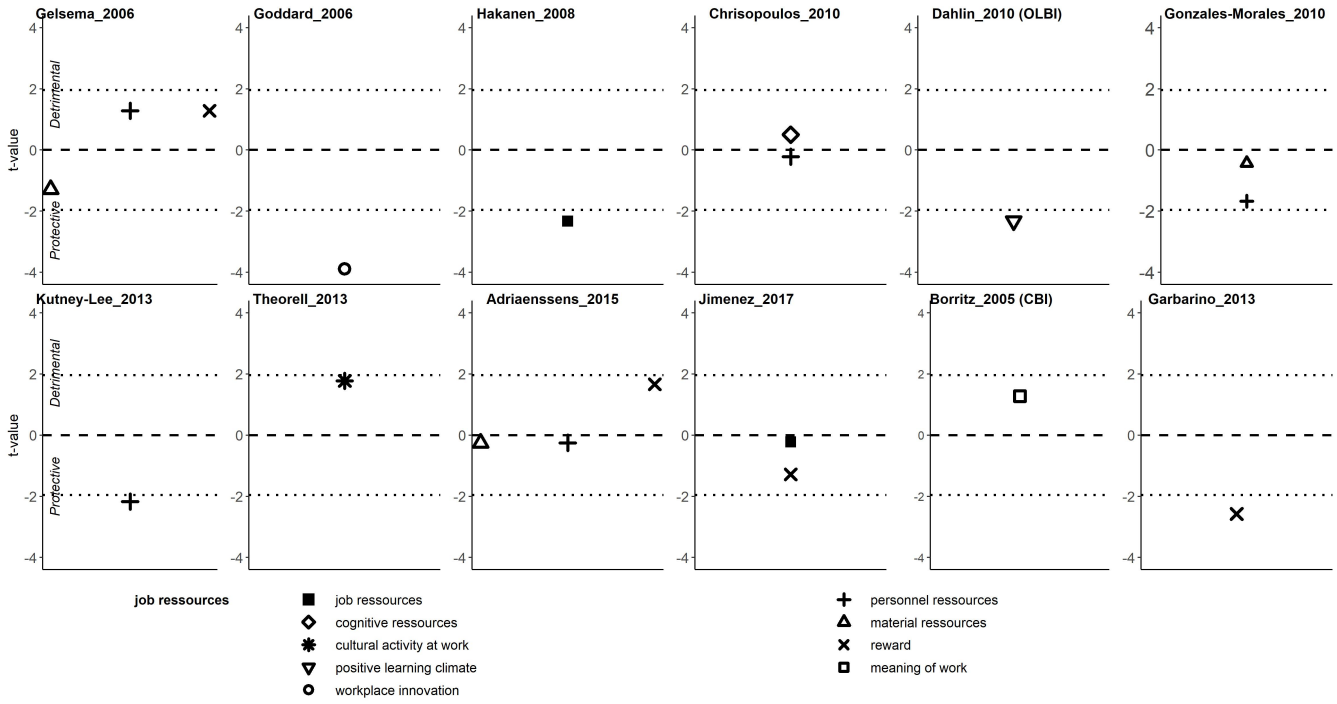

Fong,1993
Vegchel,2004
Gelsema,2006
Gil-Monte,2008
Lorente-Prieto,2008
Taris,2010
Huang,2012
Korunka,2015
Pisanti,2015
Konze,2017
Garbarino,2013
t-value
Detrimental
Protective
6
3
0
-3
-6
Work and time demands

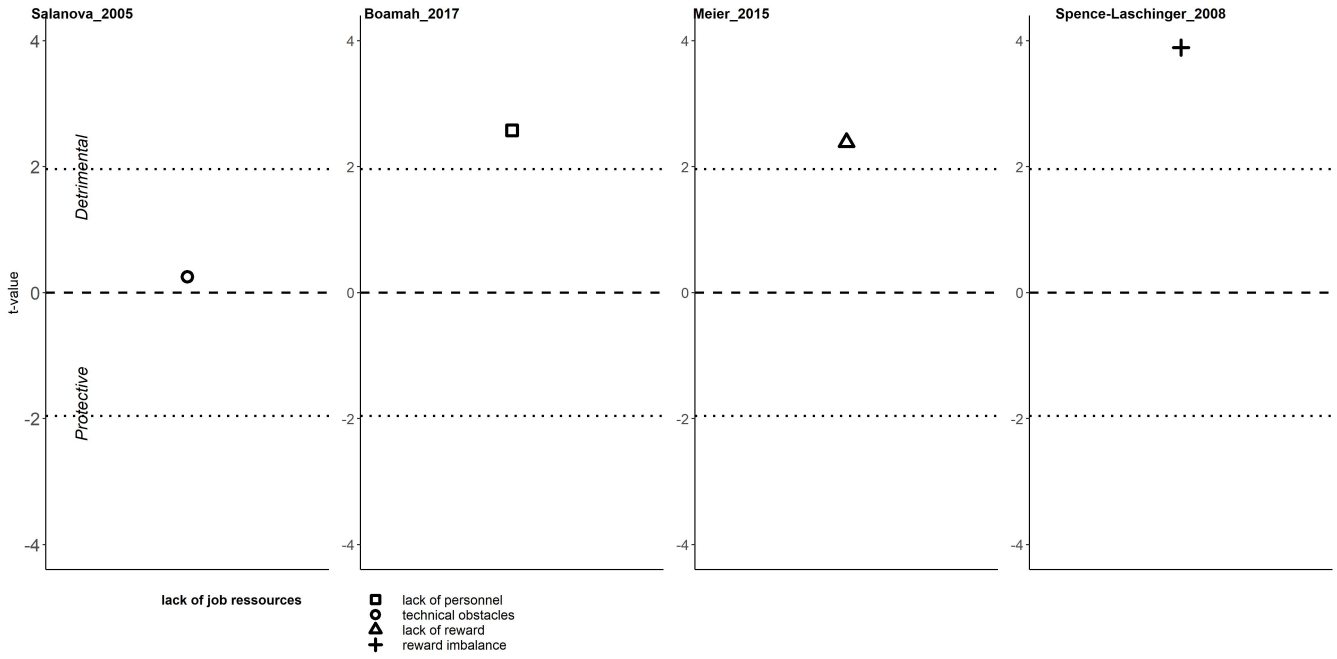
Decision Latitude (Control)



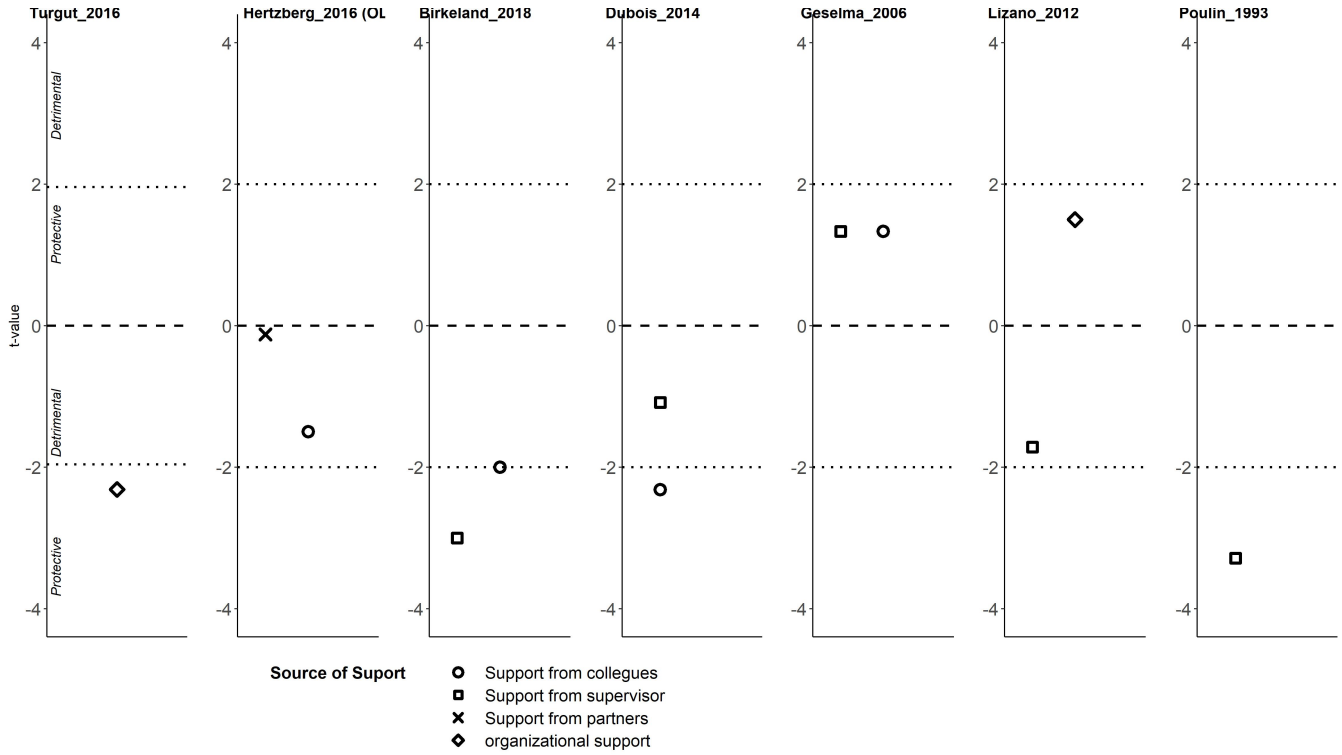


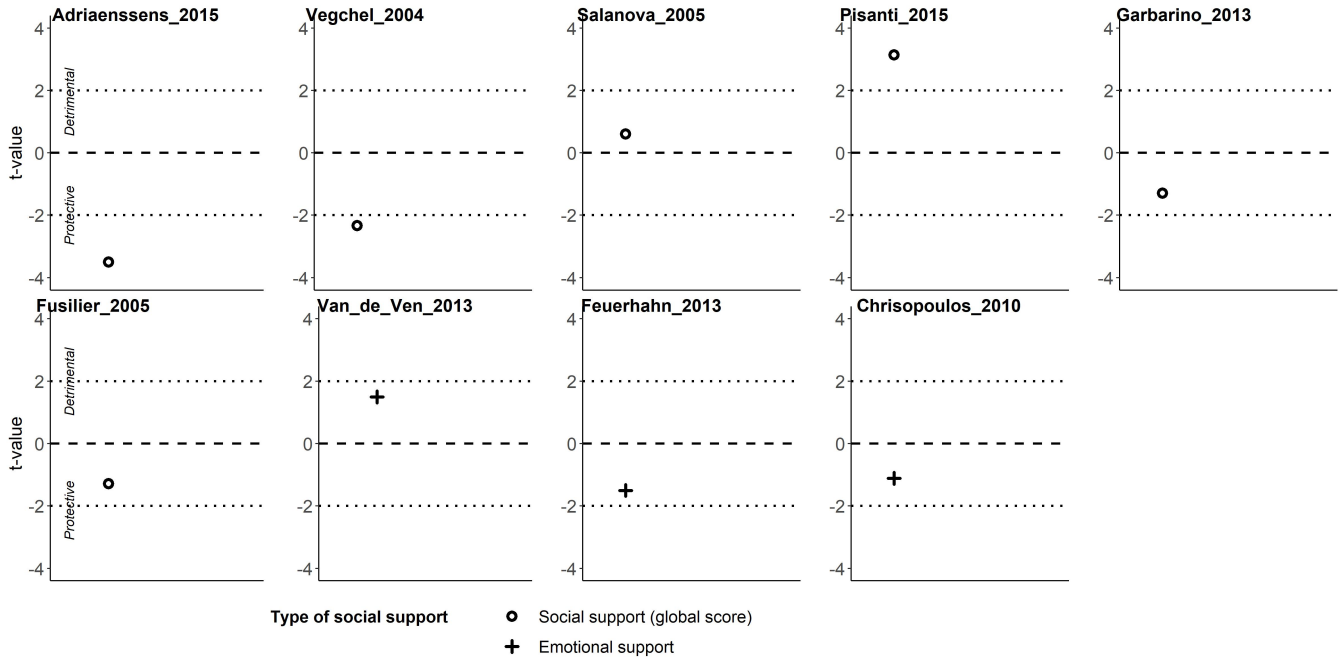
Job resources

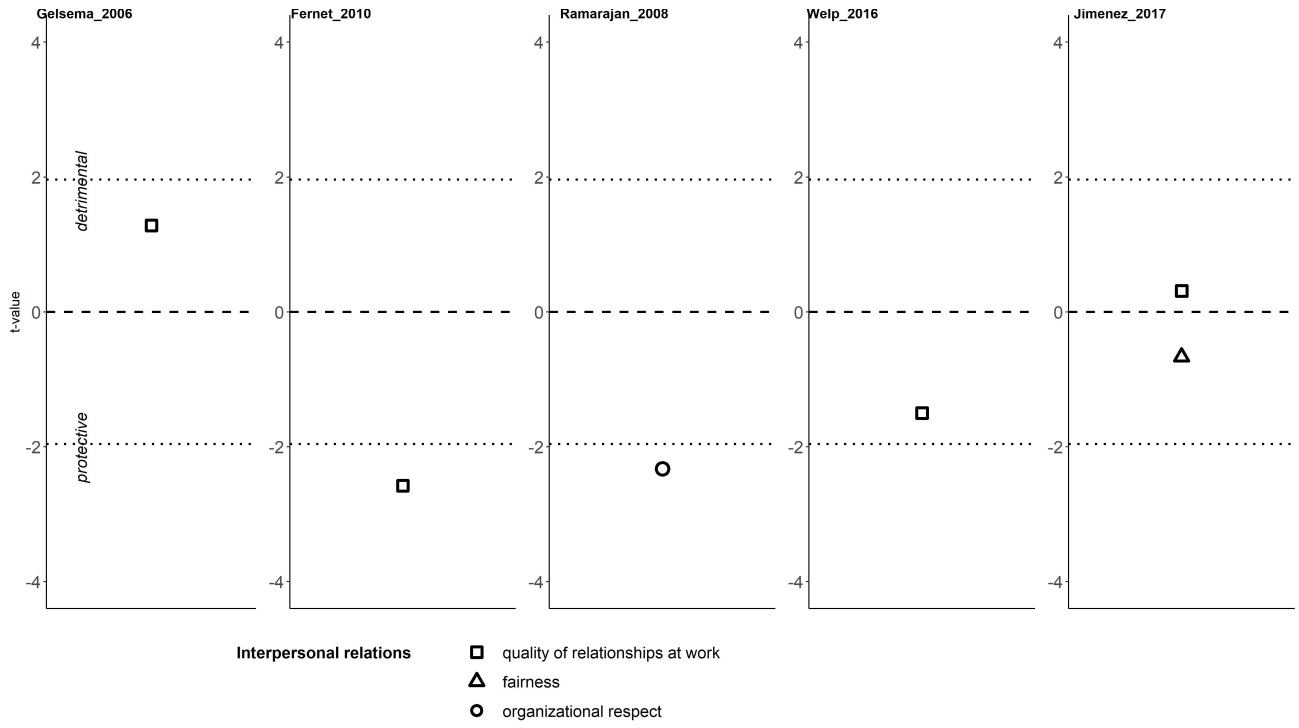


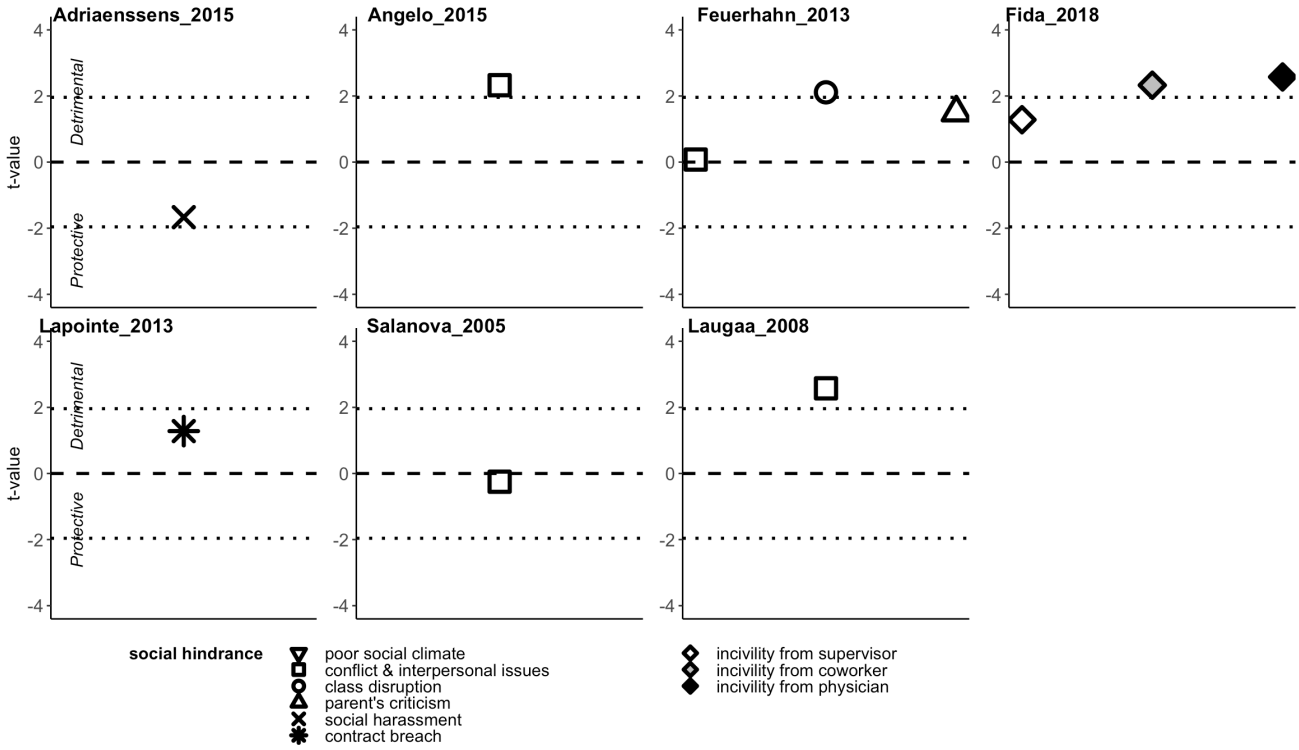


Interactions at work

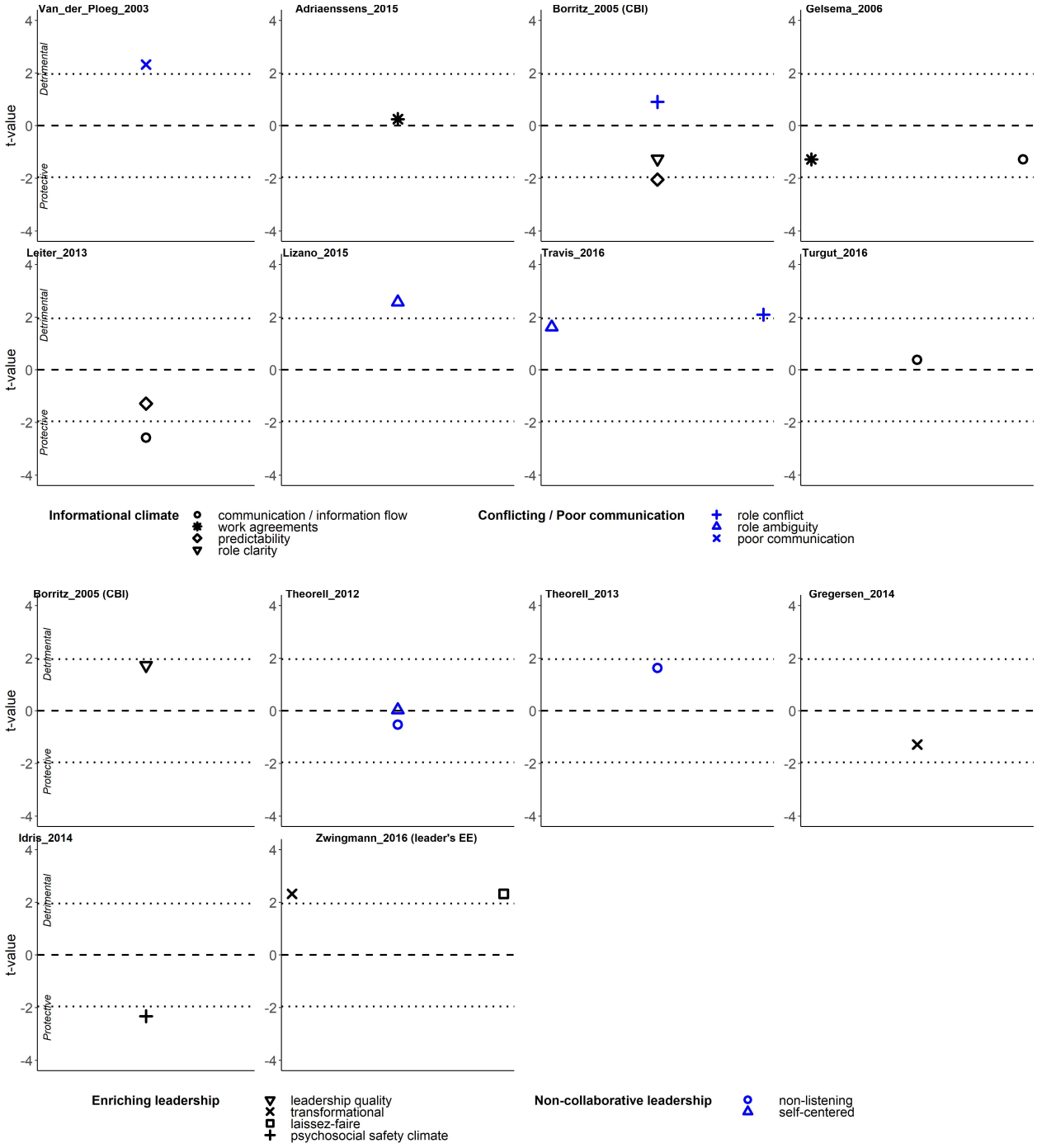




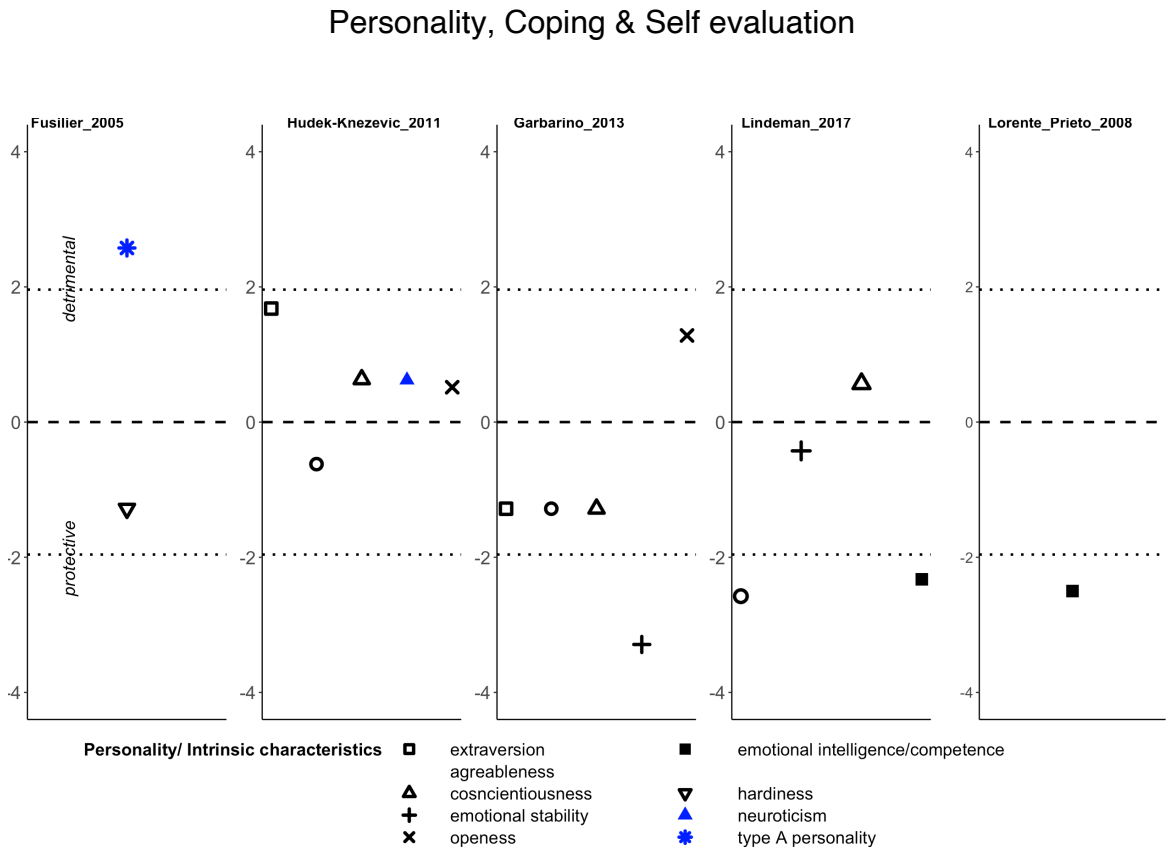

Gelsema_2006
Fernet_2010
Ramarajan_2008
Welp_2016
Jimenez_2017
t-value
detrimental
protective
4
2
0
-2
-4
Interpersonal relations
quality of relationships at work
fairness
organizational respect

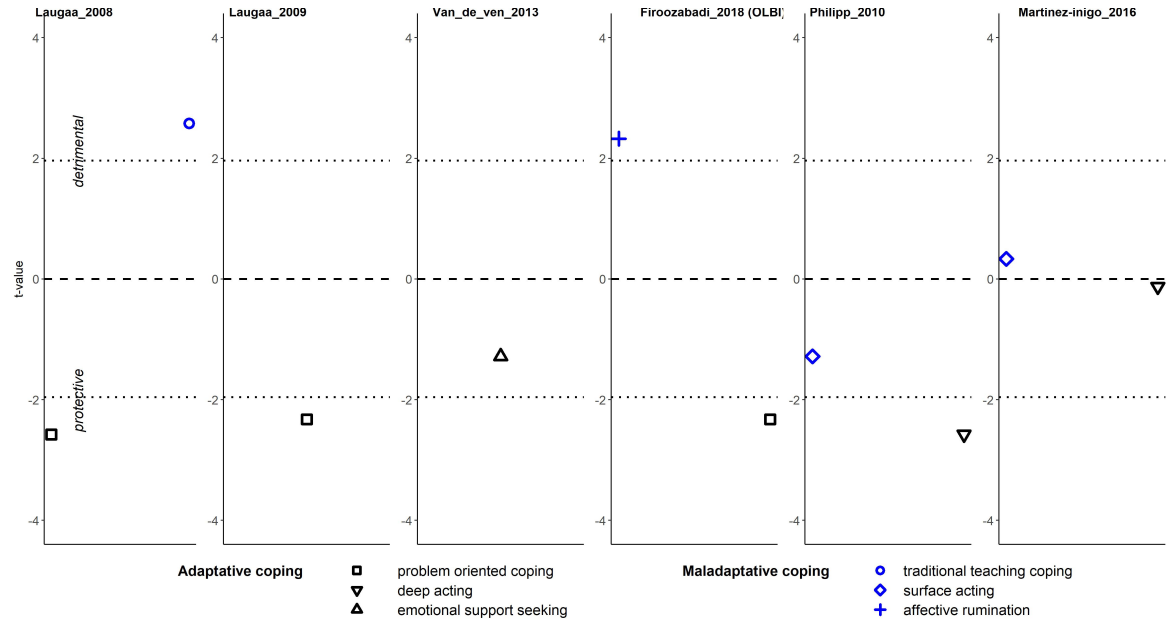

Adriaenssens_2015
Angelo_2015
Feuerhahn_2013
Fida_2018
Lapointe_2013
Salanova_2005
Laugaa_2008
t-value
Detrimental
Protective
4
2
0
-2
-4
social hindrance
poor social climate
conflict & interpersonal issues
class disruption
parent's criticism
social harassment
contract breach
incivility from supervisor
incivility from coworker
incivility from physician

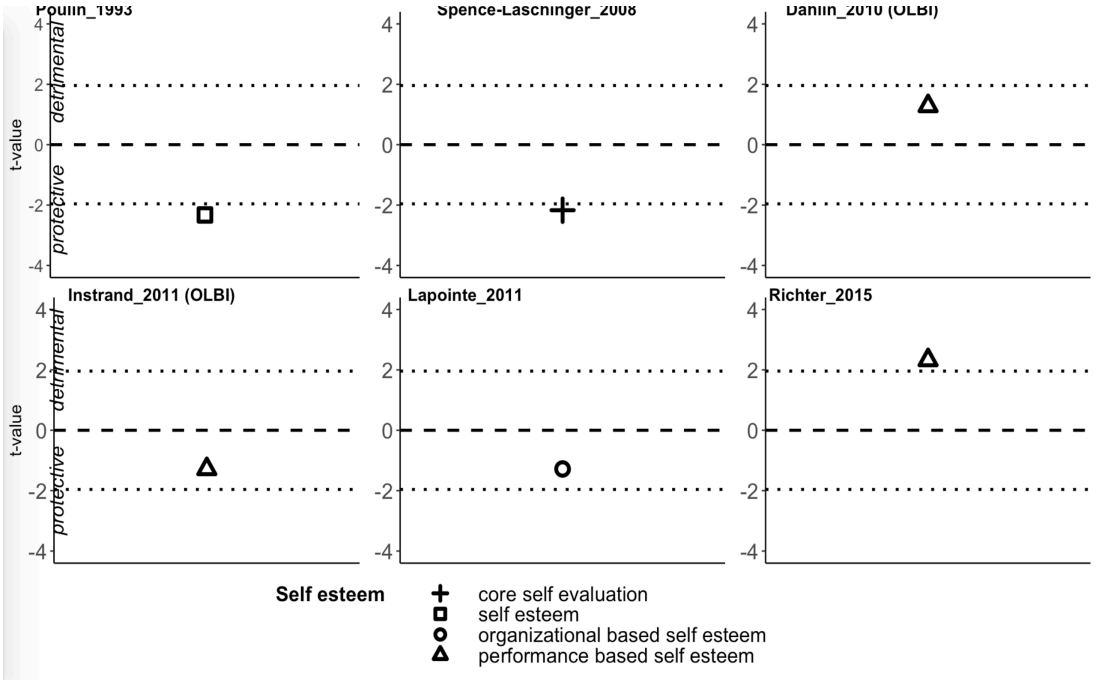
Communication and leadership

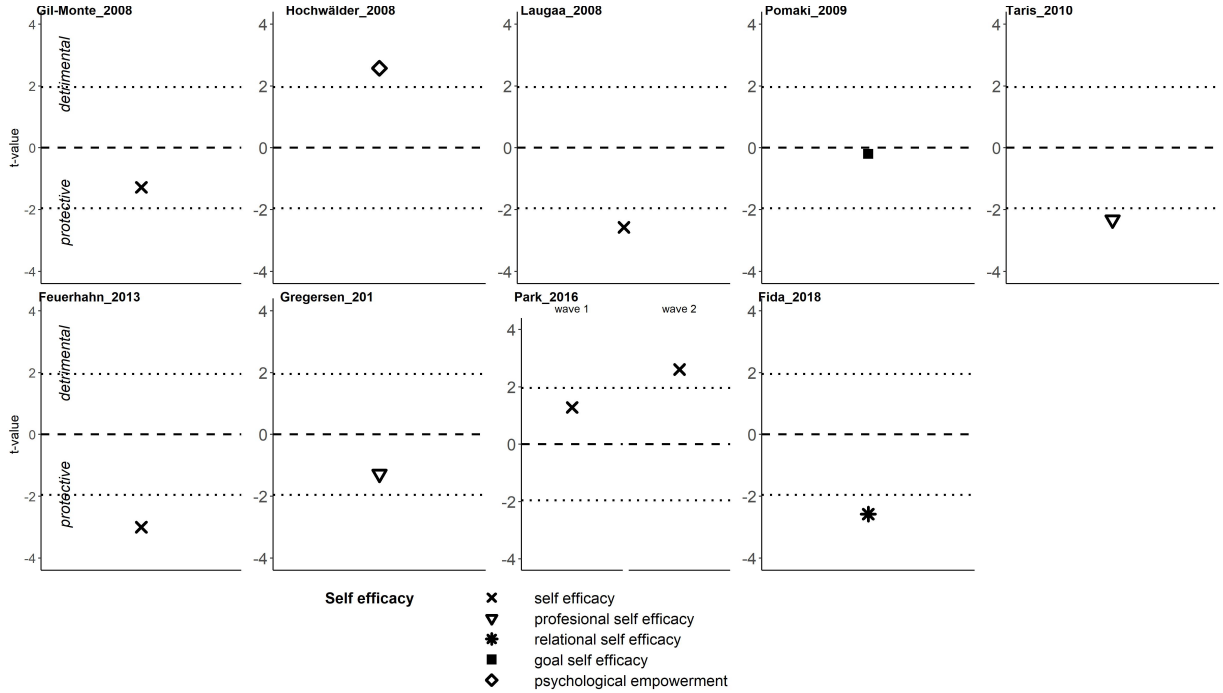


Individual characteristics

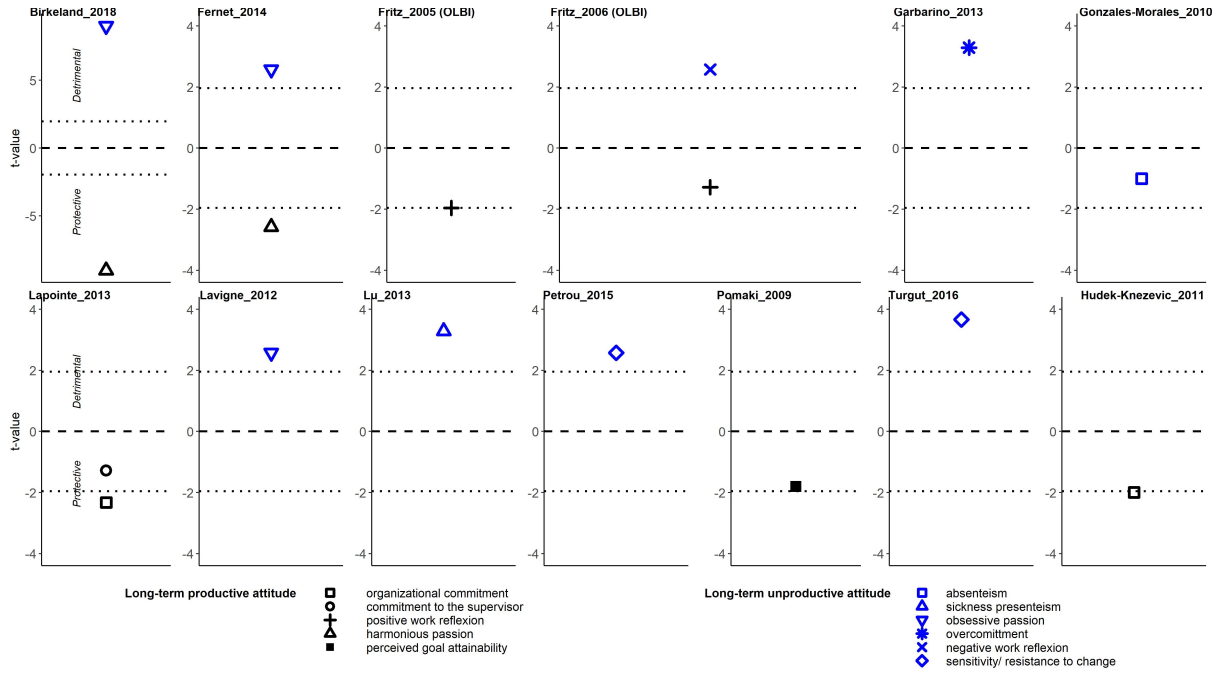


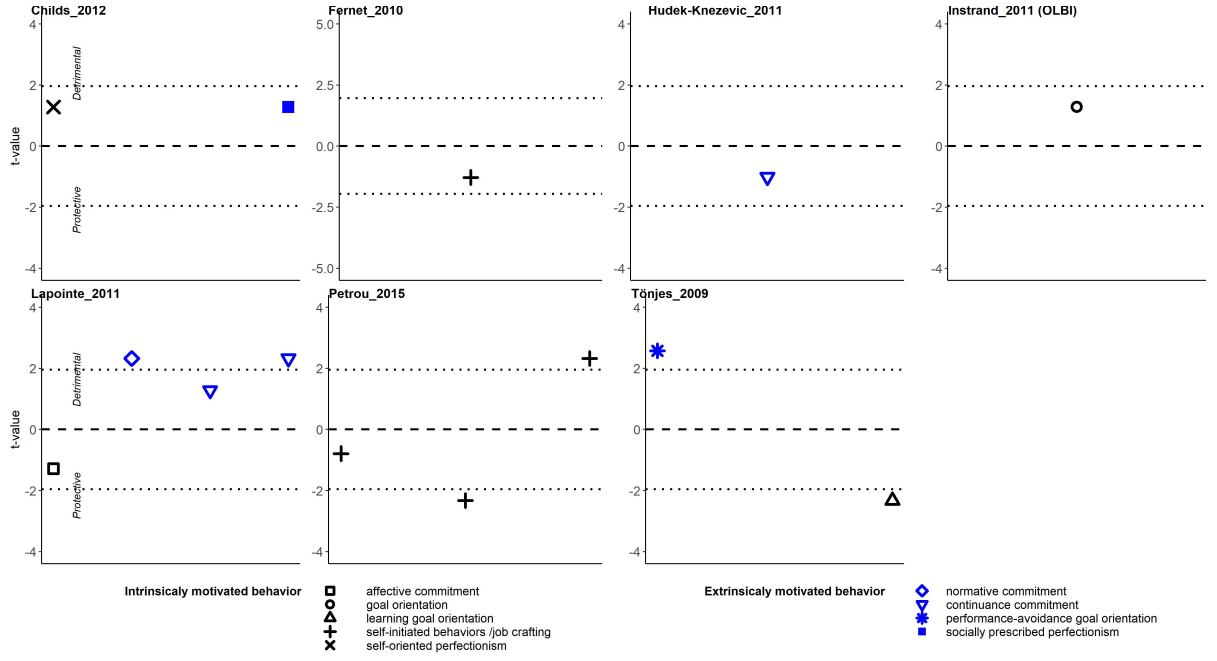



Poulin_1993
Spence-Laschinger_2008
Danlin_2010 (OLBI)
Instrand_2011 (OLBI)
Lapointe_2011
Richter_2015
t-value
detrimental
protective
Self esteem
core self evaluation
self esteem
organizational based self esteem
performance based self esteem

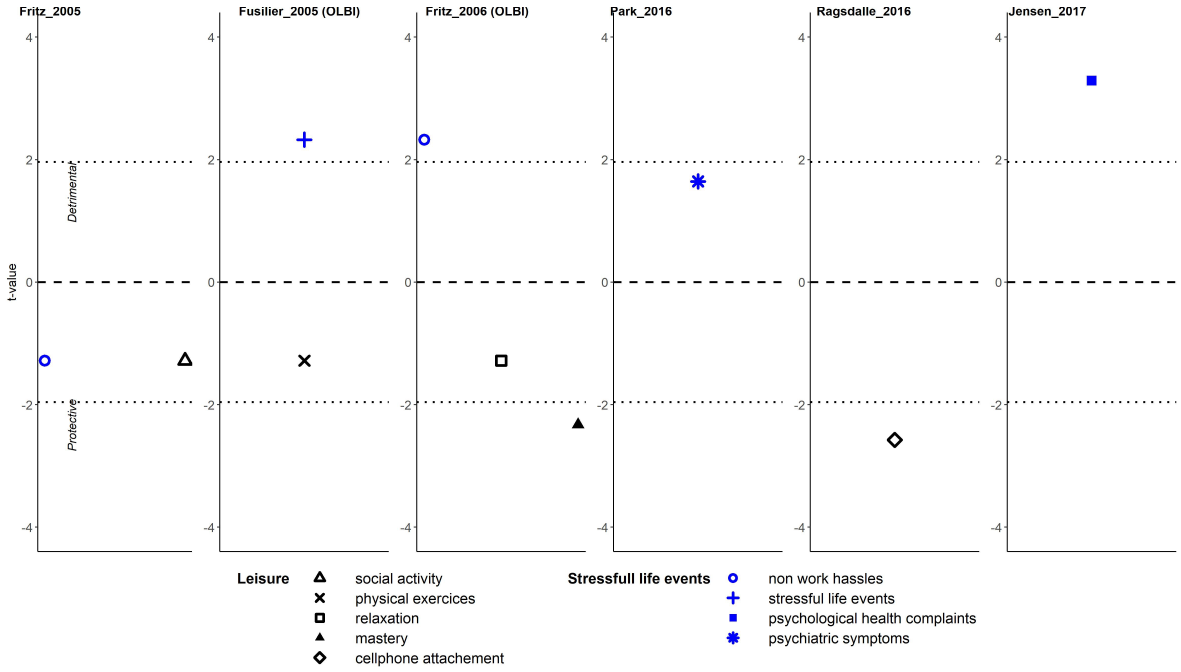

Gil-Monte_2008
Hochwälder_2008
Laugaa_2008
Pomaki_2009
Taris_2010
Feuerhahn_2013
Gregersen_201
Park_2016
wave 1
wave 2
Fida_2018
t-value
detrimental
protective
Self efficacy
self efficacy
profesional self efficacy
relational self efficacy
goal self efficacy
psychological empowerment

Job attitudes

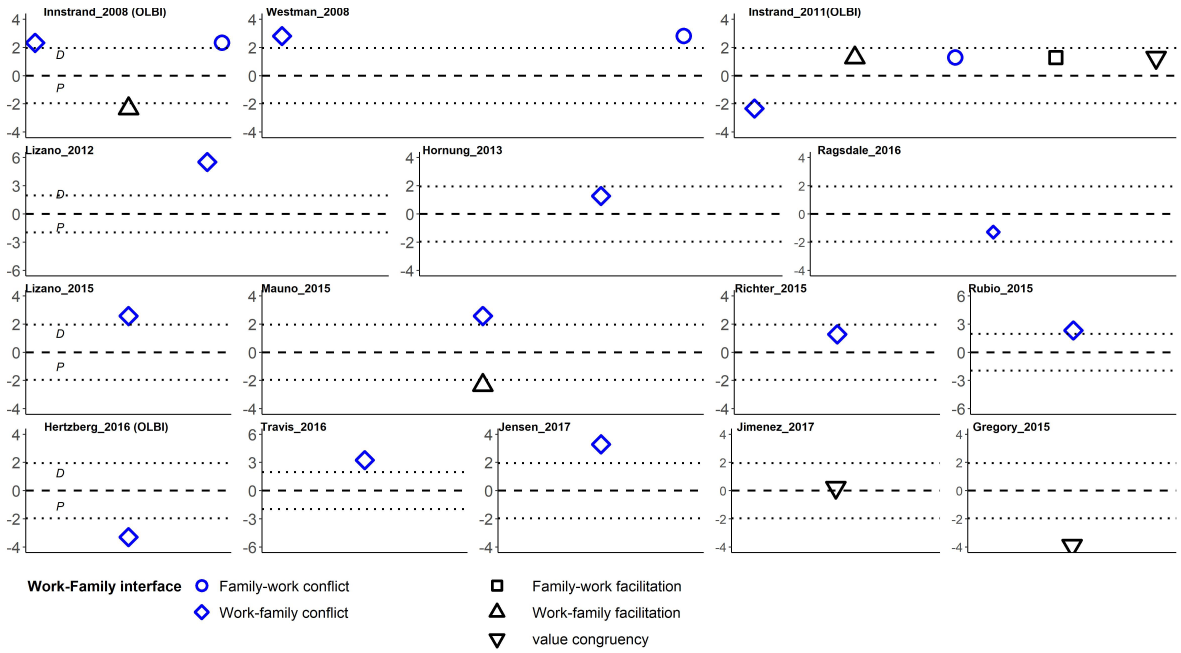




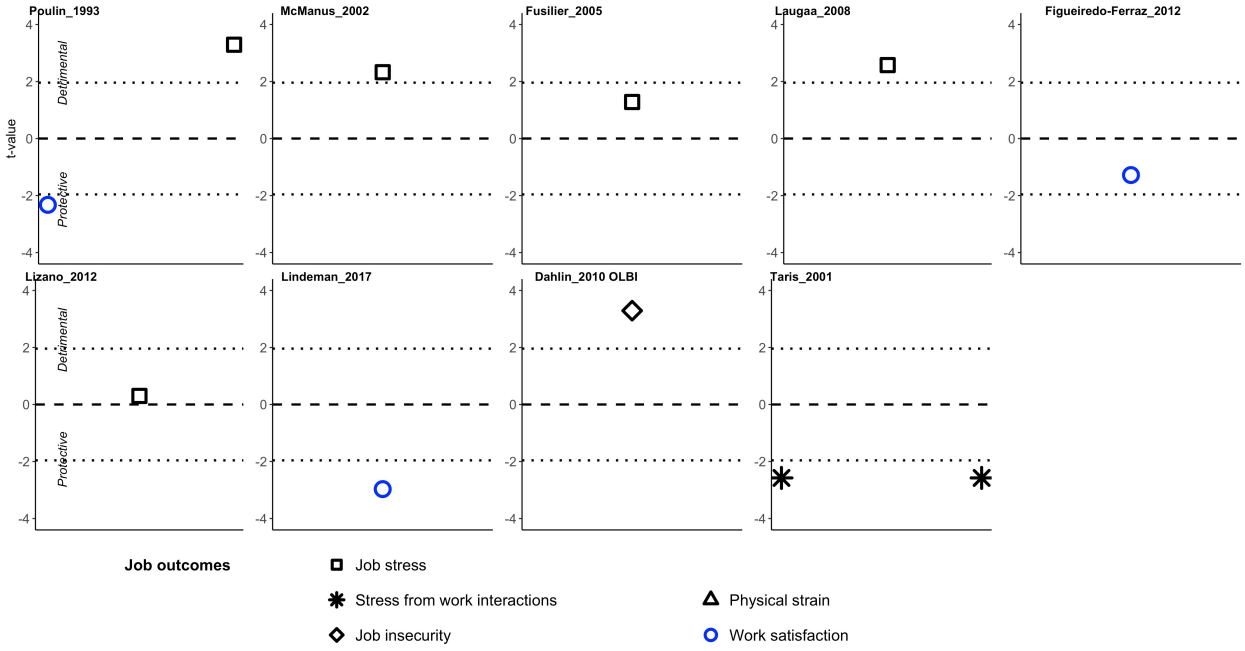
Personal events



Work-family interface



Job outcomes



Appendix C: Formulas and R-code examples

Formulas

Standard error from p-value (as recommended in Cochrane Handbook for systematic reviews of interventions):

SE = intervention of estimate / Z

Similarly, the following formula was entered in Excel (French):

=ABS(*effect estimate*/(LOI.NORMALE.STANDARD.INVERSE(p-value/2)))

R-codes examples

Detailed plot (Dahlin et al., 2010 – Job outcome)

```
ggdahlin<- ggplot(data = dahlin, mapping = aes(x = 1, y = standardized_beta),
                  group = dahlin$type) +
  geom_point(size = 5, shape = 23, fill = "white", stroke = 2) +
  ylim(-4, 4) +
  geom_abline(intercept = 0, slope = 0, lty = "dashed", size = 1) +
  geom_abline(intercept = 1.96, slope = 0, lty = "dotted", size = 1) +
  geom_abline(intercept = -1.96, slope = 0, lty = "dotted", size = 1) +
  ylab("Standardized coefficient's value") +
  theme_classic() +
  theme(axis.title.x=element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank(),
        axis.title.y = element_blank(),
        axis.text.y = element_text(size = 12))
```

General plot (Job outcome)

```
p<-ggplot(outcomes1, aes(x=t.values,y=N, color=factor)) +
  geom_point(size=3.5, position = position_jitter(w = 0.3, h = 0.3)) +
  geom_vline(xintercept = c(-1.96,0,1.96), colour="black", lty=c(2,1,2))+
  labs(title = "Job outcomes")+
  labs(x="t-value", y="Number of participants", col="", size=4)+
  scale_color_brewer(palette="Set1")+
  theme(plot.title = element_text(face = "bold", size=14),
        axis.title=element_text(size=14),
        legend.title = element_text(color = "blue", size = 12),
        legend.text = element_text(size = 12),
        legend.position = "right",
  )

p + annotate("text", x = 0.6, y = 4500, label = "Detrimental ==>  ", size=4)+
annotate("text", x = -0.53, y = 4500, label = "<== Protective",size=4)+
```

Sign test (or binomial test)

> binom.test(x,n, p=0.5, alternative = “less”, conf.level=0.95)