Prediction of Health-Related Outcomes and Turnover Intention with the Munich Employee Health Questionnaire (MEHQ)

Bettina Maria Zweck[a], Florian Pargent[b], Markus Bühner[b]

[a] Kantar Public Division, Kantar Deutschland GmbH, Munich, Germany

[b] Department of Psychology, Ludwig-Maximilians-Universität München, Munich, Germany

This study was financially supported by Kantar Deutschland GmbH.

Please address correspondences to:

Bettina Zweck

Kantar Public Division

Landsberger Straße 284

80687 Munich

Germany

bettina.zweck@kantarpublic.com

Abstract

Stress at work and its consequences are of growing interest. The Munich Employee Health Questionnaire (MEHQ; Zweck, 2017) was developed to measure both psychosocial hazards and outcomes considering health (strains, general health, sickness absence) and turnover intention. After reviewing psychosocial hazards that are supposed to be associated with stress, we show estimates of criterion validity for the MEHQ from cross-sectional and longitudinal (more than one year) analyses, based on online panel data from n = 1,327 German employees (sample sizes for longitudinal analyses between n = 383 and n = 444). Using methods from machine learning and predictive modeling in addition to linear regression, we could predict all outcomes except sickness absence both cross-sectionally and longitudinally. Highest predictive performance (estimated by 10 times repeated 10-fold cross-validation) was achieved for the outcome strains (cross-sectional: $R^2 \approx .5$, longitudinal: $R^2 \approx .4$). Longitudinal predictions could be greatly improved when outcomes measured at time 1 (among other covariates) were added as predictors. General health, strains, and turnover intention could be predicted using either the MEHQ domain sum scores or items as predictors directly. Classical linear regression models using sum scores showed predictive performance comparable to elastic net models with all items as predictors. No performance improvement was observed when using the nonlinear random forest. The MEHQ seems to be a competitive measure for predicting strains, general health, and turnover intention. Reasons for the low predictability of sickness absence are discussed.

Keywords: work-related stress, psychosocial hazards, turnover intention, longitudinal study, machine learning

Online supplemental materials: https://osf.io/cuaeh/

Analyses scripts: https://osf.io/t7a28/

Data: https://doi.org/10.5160/psychdata.zkba17mu29/

According to the Mental Health Action Plan 2013–2020 the “provision of healthy living and working conditions” is one of the current objectives of the World Health Organization (WHO) (WHO, 2013; p.17). Within its core Convention on Occupational Safety and Health, 1981 (No.155) and its accompanying Recommendation (No.164), also the International Labour Organization (ILO) formulated its aim of protecting workers’ physical and mental health (ILO, 2016). Regarding Europe, the European Agency for Safety and Health at Work (EU-OSHA) emphasizes in its “Priorities for occupational safety and health 2013–2020” (EU-OSHA, 2013) the importance of promoting good health and preventing work-related stress. Even if stress research has a long history (Bliese, Edwards, & Sonnentag, 2017) starting around the 1930s (Laird, 1933), the objectives of these important institutions underline that stress at work is still a serious problem. During the past decades, workplaces have strongly changed due to globalization and digitization (De Lange, 2005)–and therefore new hazards are rising: Information and communication technologies like the internet and smartphones allow people to work everywhere and always. The boundaries between work and private life disappear (ILO, 2016) and possibilities to recover are rare (Barber & Santuzzi, 2015). To persist against the global competition resulting from globalization, downsizing or temporary work are one of the actions taken by companies (Kawachi, 2008; Sverke, Hellgren, & Näswall, 2002). Thus, job insecurity is rising–alongside declines in both physical and mental health (Sverke et al., 2002).

Studies support the current seriousness of the problem “work-related stress”: In an opinion survey with experts from 54 countries across the world (n = 324), over 90% think that work-related stress is of concern in their country (ILO, 2016). Results from surveys on working conditions and health conducted in Europe, Chile, USA, Canada, Columbia, Argentina, Japan or Korea show a similar picture (ILO, 2016).

The consequences of work-related stress can be serious. It can affect health-related physical outcomes like cardiovascular diseases (e.g., Kivimäki et al., 2006; Rosário, Fonseca, Nienhaus, & Torres da Costa, 2016; Rosengren et al., 2004), musculoskeletal disorders (e.g., Deeney & O'Sullivan, 2009) or sleep disturbances (e.g., Rosário et al., 2016). Relationships between work-related stressors and mental outcomes like depression or burnout were reported (e.g., Blackmore et al., 2007; Bonde, 2008; Melamed, Armon, Shirom, & Shapira, 2011; Nübling et al., 2013; Rugulies, Aust, & Madsen, 2017). Additionally, an association between work-related stressors and behavioral outcomes like turnover and sickness absence was investigated in the past (e.g., Avey, Luthans, & Jensen, 2009; Head et al., 2007; Magnavita & Garbarino, 2013; Niedhammer, Chastang, Sultan-Taïeb, Vermeylen, & Parent-Thirion, 2013; Rosário et al., 2016; Rugulies, Aust, & Pejtersen, 2010; Slany et al., 2014). However, work related stress is not always related to sickness absence: When the pressure is growing, this can force people to go to work nonetheless (Lesuffleur, Chastang, Sandret, & Niedhammer, 2014; Smulders & Nijhuis, 1999)–even if they are ill, they do not take the day off. Being at work while being ill is called presenteeism (e.g., Cooper & Dewe, 2008).

Work-related stress is not only associated with health issues, it also has big economic consequences. According to Hassard, Teoh, Visockaite, Dewe, and Cox (2018), the global cost of work-related stress is up to $ 187 billion per year, although such estimates greatly differ depending on which indicators are included (Brun & Lamarche, 2006).

There are–in addition to the global interest of WHO and ILO–also national initiatives to combat work-related stress. Due to the Framework Directive (89/391/EEC), European countries are encouraged to ensure health and safety of workers. In Denmark, a law on the prevention of psychosocial stress was passed already in 1974. Since 2013, every employer in Germany is obligated to conduct a psychosocial risk assessment to evaluate conditions that can cause work-related stress pursuant to § 5, German Occupational Safety and Health Act

(Janetzke & Ertel, 2017). There are also national examples outside Europe where health protection is to some extent legally binding: In Chile and Peru, the protection of mental health is included in the Constitution as an individual's right. In Mexico, psychosocial risk factors for work-related-stress are defined in detail within the Occupational Safety and Health (OSH) legislation (ILO, 2016).

Defining psychosocial risk factors is important, as these are arguably the sources of experiencing work-related stress. Those psychosocial risk factors can be called *psychosocial hazards* (ILO, 2016) . Research on stress at work has identified many different psychosocial hazards so far. In their summary of psychosocial hazards, ILO (2016) refers to Cox, Griffiths, and Rial-González (2000) and their research on work-related stress. However, Cox et al. (2000) and many other reviews or meta-analytic articles (e.g., Bonde, 2008; Stansfeld & Candy, 2006) do not consider specific aspects of new forms of work. Facing major economic and technical changes, psychosocial hazards such as "telepressure" (Barber & Santuzzi, 2015) or information overload (A. R. Lee, Son, & Kim, 2016) should also be taken into account. Furthermore, research has identified aspects of personality that are linked to the experience of stress: People with high values in resilience seem to be able to withstand pressure and to cope successfully with adversity (Fletcher & Sarkar, 2013). In contrast, a "risky" personality pattern seems to be linked to a higher burnout risk (Bauer et al., 2006). According to Bauer et al. (2006) this behavioral pattern can be described with intense involvement and lack of dissociation from work-related problems, limited enjoyment of life and reduced mental resilience with regard to pressure and strains. Moreover, aspects of the personal environment such as life events are not listed by Cox et al. (2000), but should be considered when discussing psychosocial hazards. Life events are described as incidents that significantly interfere with ongoing life on a temporary or permanent basis (Cleland, Kearns, Tannahill, & Ellaway, 2016). Examples for life events are birth of a child, marriage, new home, or death of a close person (Cleland et al., 2016). Research showed that these life events can be linked to mental and physical health (Cleland et al., 2016; Rosengren et al., 2004). Already in the 1960s Holmes and Rahe (1967) developed a scale to measure ‘stressful’ life events.

In Table 1, we have listed psychosocial hazards mentioned by Cox et al. (2000) and have added aspects that should also be considered as psychosocial hazards based on other research. It should be noted that–as working conditions constantly change–new psychosocial hazards will probably emerge in the future. Similar to Cox et al. (2000), we sorted specific burdens by domains. We further added the categories “new forms of work” and “aspects outside working conditions”.

Table 1

Psychosocial hazards sorted by domains

JOB & TASK DESIGN	
Completeness of tasks	Short work cycles and fragmented work (Cox et al., 2000) (S)
Control / Decision Latitude	Low control, low decision latitude (Cox et al., 2000) (S)
Variability	Monotonous and repetitive work (Cox et al., 2000) (S)
Role in organization	Role ambiguity (Cox et al., 2000) (S)
Qualification	Under or over promotion (status incongruity), underuse of skills, role insufficiency (individual's ability not fully used) (Cox et al., 2000) (S)
Emotional demands	Dealing with other people and their problems (Leka, Griffiths, & Cox, 2003) (S); Hiding emotions ("emotion-rule dissonance") (Holman, Martinez-Inigo, & Totterdell, 2008) (S)
ORGANIZATION OF WORK	
Working time (quantitative)	Work overload, work underload, shift working (Cox et al., 2000) (S); overtime, long working hours (Sparks, Cooper, Fried, & Shirom, 1997) (M), (Spurgeon, Harrington, & Cooper, 1997) (R); no (scheduled) breaks (Waongenngarm, Areerak, & Janwantanakul, 2018) (R)
Working time (qualitative) / intensity of work / Workflow	High levels of time pressure, working to deadlines, unpredictable hours (Cox et al., 2000) (S); interruptions (Zijlstra, Roe, Leonora, & Krediet, 1999); mental overload, (psychological) job demands (Bonde, 2008) (R), (Stansfeld & Candy, 2006) (M)

INTERPERSONAL RELATIONSHIPS AT WORK	
Cooperation / Social support	Lack of social support (Cox et al., 2000) (S); social isolation (Johnson & Hall, 1988)
Conflicts	Interpersonal conflicts (Cox et al., 2000) (S); bullying, harassment (Leka et al., 2003) (S); managerial bullying (Colligan & Higgins, 2005) (S)
Leadership	Low participation in decision making, lack of feedback on performance, poor pay, sense of inequity (Cox et al., 2000) (S); no trust and reciprocity (aspects of *social capital*) (Kawachi, 1999); organizational injustice (Elovainio et al., 2009) (L), (Kivimäki et al., 2004) (L)
WORK ENVIRONMENT / WORK EQUIPMENT	
Physical working environment	Problems regarding the reliability, availability, suitability, and maintenance, or repair of both equipment and facilities; noise, unfavorable temperature (Cox et al., 2000) (S); unfavorable lighting (Krüger, 2016) (S); occupational noise (Kersten & Backé, 2015); lift heavy loads, rapid physical activity, awkward arm and body position (R. Karasek et al., 1998)
NEW FORMS OF WORK	
Home-work interface	Conflicting demands of work and home, dual career problems (Cox et al., 2000) (S); imbalance between work and family (J. L. Wang, Lesage, Schmitz, & Drapeau, 2008); constant availability (Pangert, Pauls, & Schüpbach, 2016) (S); insufficient recovery time at home (Wendsche & Lohmann-Haislah, 2016) (S); "telepressure" (Barber & Santuzzi, 2015)
Globalization / Technical progress in general / Information overload	Job insecurity, fear of redundancy (Cox et al., 2000) (S); technical progress, which often implies learning new skills in unpaid free time (Chandola, 2010) (S); information overload (for example due to new information and communication technologies) (A. R. Lee et al., 2016)
ASPECTS OUTSIDE WORKING CONDITIONS	
Commitment and values	Work of low social value (Cox et al., 2000) (S); low (organizational/occupational) commitment (K. Lee, Carswell, & Allen, 2000) (M), (Mathieu & Zajac, 1990) (M); low job involvement (Brown, 1996) (M)
Aspects of the personality	No detachment (Fritz, Yankelevich, Zarubin, & Barger, 2010), (Sonnentag, Binnewies, & Mojza, 2010) (L), (M. Wang et al., 2013), (Wendsche & Lohmann-Haislah, 2016) (S); risk type: risky pattern (Bauer et al., 2006); type A behavior pattern (Edwards, Baglioni, & Cooper, 1990), (Edwards et al., 1991); presenteeism (Bergström et al., 2009) (L); low resilience (Fletcher & Sarkar, 2013) (R)
Aspects of the personal environment	Low emotional support at home (Cox et al., 2000) (S); poor emotional support outside work (Nordin, Westerholm, Alfredsson, & Åkerstedt, 2012) (L); major life events (Cleland et al., 2016) (L), (Rosengren et al., 2004)

Note. The proposed classification of psychosocial hazards is on content and not on statistical procedures. Overlaps of categories are possible. Possible interactions are not considered. Meta-Analysis (M); review (R); longitudinal study (L); secondary literature like book, booklet or report (S).

Different models were developed to describe *why* there is a link between psychosocial hazards and stress-related outcomes. The effort-reward imbalance (ERI; Siegrist, 1996) and the demand/control model (DC model; R. A. Karasek, 1979), which was expended by the dimension of social support later on (DCS model; Johnson & Hall, 1988), are among the most popular ones (Nübling et al., 2013).

The ERI model assumes that high-effort/low-reward conditions at work can be related to stressful experiences (Siegrist, 1996). The DC model states that working conditions are supposed to have negative mental or physical consequences when high demands at work come along with low decision latitude. Low social support can increase the risk of negative outcomes (Johnson & Hall, 1988). Research focusing on the assumptions of these two models could demonstrate significant effects in line with the assumptions (e.g., Melamed et al., 2011; Rugulies et al., 2017).

Measuring psychosocial hazards and its outcomes

As ERI and DC(S) model belong to the most popular models in the area of stress research, many studies on stress research are based on the questionnaires related to these models: The ERI questionnaire (Siegrist, 1996) and the Job Content Questionnaire (JCQ; R. Karasek et al., 1998).

One recently developed questionnaire that is not based on one specific theory is the Copenhagen Psychosocial Questionnaire (COPSOQ; T. S. Kristensen, Hannerz, Høgh, & Borg, 2005). The COPSOQ uses key elements from the ERI and DCS model and supplements them with other psychosocial hazards like work-privacy-conflict. Presumably due to its broader range of measured psychosocial hazards, Nübling et al. (2013) showed in a cross-sectional study in Germany that burnout can be better predicted by the COPSOQ than by the ERI questionnaire. Comparing predictive performance measures of the COPSOQ for different outcomes with each other, predictive performance for burnout was R^2 = .35–and herewith worse than for job satisfaction (R^2 = .51). General health could be predicted worst (R^2 = 11).

Currently, the COPSOQ is one of the most commonly used questionnaires for the assessment of psychosocial hazards. This questionnaire has already been translated into various languages and has been used in numerous studies worldwide (e.g., Dupret, Bocéréan, Teherani, Feltrin, & Pejtersen, 2012; Li et al., 2010; Moncada et al., 2014; Nübling & Hasselhorn, 2010; Setti, d'Errico, Di Cuonzo, Fiabane, & Argentero, 2017; Stauder et al., 2017). However, the dimension “working environment”, the aspect “completeness of tasks” and some aspects of new working conditions “like information overload” were not part of the COPSOQ version that was used in Germany around 2016 (Nübling, Lincke, Vomstein, & Haug, 2016). The outcome “turnover” was also not included. However, the dimension “working environment” and the outcome “turnover intention” were added in a later COPSOQ version (Nübling et al., 2018). Unfortunately, we could not identify any peer-reviewed studies regarding this new version. Accordingly, all statements regarding the COPSOQ in the current article refer to the original COPSOQ study in which its development is described (T. S. Kristensen et al., 2005), its German validation study (Nübling & Hasselhorn, 2010), and the questionnaire used in Germany in 2016 (Nübling et al., 2016).

Apart from the new German COPSOQ version of 2018 (Nübling et al., 2018), turnover (intention) is not assessed by other German or English questionnaires that measure psychosocial hazards listed by ILO (2016) and WHO (Leka & Jain, 2010). As turnover can also be a result of work-related stress (e.g., Michie, 2002), we believe that this criterion is important to measure in stress questionnaires. This would not be a problem of content validity, as antecedents for work-related stress and turnover (intention) are supposed to overlap: In a longitudinal COPSOQ study based on nurses in China, in which turnover intention was added as an additional outcome, it was demonstrated that turnover intention could be predicted by psychosocial hazards integrated in the COPSOQ (Li et al., 2010). This finding is supported by turnover research indicating that antecedents for turnover (intention) are quite similar to those for health-related outcomes: Some identified antecedents for turnover (intention) are work overload, low scope, role ambiguity, poor pay, low social support, poor leadership styles, injustice, job insecurity, interpersonal conflicts, low physical comfort, and low organizational commitment (Avey et al., 2009; Griffeth, Hom, & Gaertner, 2000; Houkes, Janssen, de Jonge, & Bakker, 2003; Suadicani, Bonde, Olesen, & Gyntelberg, 2013; Sverke et al., 2002; Waldman, Carter, & Hom, 2015; Wasti, 2003). Based on the turnover research and the results of Li et al. (2010), we assume it is reasonable to measure both health-related outcomes and turnover intention within one questionnaire.

The lack of questionnaires in 2015 that exhaustively measure psychosocial hazards and stress related outcomes (including turnover) was the initial reason to develop the Munich Employee Health Questionnaire (MEHQ; Zweck, 2017), the primary measure dealt with in the present study.

For the MEHQ, relevant psychosocial hazards found in the literature (see Table 1) were taken into account. In addition to the background literature, two more sources were used in defining the item pool of the MEHQ: A representative survey based on n = 20,036 employees in Germany (Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, 2012) and a qualitative survey based on n = 15 experts[1] conducted by the first author of this article. The resulting item pool was tested in a cognitive pretest in three waves with n = 42 people in total

[1] The definition for “experts” referred to persons who had contact points with the topic "work-related stress" within their current position at work. It was also necessary that they had been in their current job position for at least one year.

to ensure that regardless of education, gender or age, the items are understood by each person.

After cognitive pretesting, the item pool was quantitatively examined and validated online in two cross-sectional independent waves with employees in Germany.

In addition to psychosocial hazards, the MEHQ measures the outcomes general health, strains (like burnout symptoms, sleep disturbances), sickness absence, and turnover intention. Turnover *intention* was included instead of real turnover for the following reasons: It can be assumed that people who already suffer from work-related stress could be too exhausted (as exhaustion is associated with burnout; Freudenberger, 1974; Maslach & Jackson, 1981) to actively pursue a job change. Additionally, people who have a lot of pressure at work might have less time to apply for another job (e.g., Chandola, 2010). Thinking about quitting (without going through with it) might be a more sensitive indicator for work-related stress. Nevertheless, it should be noted that turnover intention is an important predictor for real turnover (Griffeth et al., 2000; Mai, Ellis, Christian, & Porter, 2016). Thus, we are more in line with the aims of the WHO and ILO that are about the prevention of work-related stress (ILO, 2016; WHO, 2013). Having an early indicator to prevent real behavior can be crucial as people who have already quit their jobs cannot be brought back easily. Additionally, this early indicator can help employers to prevent deviant behaviors that are supposed to be linked with turnover intention (Mai et al., 2016).

To date, many studies within the field of stress research have been conducted cross-sectionally (e.g., Bliese et al., 2017; Rosário et al., 2016; Theorell & Hasselhorn, 2005)–the same for COPSOQ. To our knowledge, there are only a few COPSOQ studies that were conducted longitudinally (e.g., Albertsen, Rugulies, Garde, & Burr, 2010; Aust, Rugulies, Finken, & Jensen, 2010; Borritz et al., 2005; Borritz et al., 2010; Dicke et al., 2018; T. R. Kristensen, Jensen, Kreiner, & Mikkelsen, 2010; Li et al., 2010; Rugulies et al., 2010;

Rugulies et al., 2007)–most of them with focus on specific working populations, so it is unclear whether those results can be generalized.

Even if causal inferences cannot be tested by longitudinal research designs (Zapf, Dormann, & Frese, 1996), they allow for an investigation of the relationship between psychosocial hazards and stress related outcomes over time. Hence, there is a need for longitudinal research designs based on representative samples, especially when speaking about how "completely" a questionnaire can measure stress-related outcomes–in other words: How good is its criterion validity?

The present study

Aims of the present study

The aim of this study is to evaluate the criterion validity of the MEHQ based on both a cross-sectional and a longitudinal design. More specifically, our general research objective is to investigate to what extent strains, general health, turnover intention, and sickness absence can be predicted by psychosocial hazards measured with the MEHQ. Zweck (2017) analyzed criterion validity with classical methods like correlations and linear regression in the cross-sectional setting. In this study, we try to maximize the predictive performance by applying regularized linear and nonlinear models from predictive modeling and machine learning in addition to classical linear regression. In machine learning, it is considered of utmost importance to provide realistic estimates for the expected accuracy of predictive models on independent observations. In-sample performance estimates, which are heavily used in psychology, are too optimistic and should not be used in predictive modeling (Yarkoni & Westfall, 2017). Instead, we will report performance estimates from cross-validation.

Applying predictive modeling methods to the problem of work-related stress is of high practical relevance: In light of the protection mandate of ILO, WHO and EU-OSHA

(EU-OSHA, 2013; ILO; 2016; WHO, 2013) any prediction of stress-related outcomes by psychosocial hazards should be as accurate as possible. Although predictive modeling methods are predestined to be used in this context, we are not aware of any comparable studies on psychosocial hazards and stress-related outcomes using similar methods. It is unclear whether the use of nonlinear algorithms in contrast to linear models can lead to more accurate predictions. Moreover, many machine learning algorithms can effectively use a larger number of predictor variables than is typically seen in psychological research. When the goal is to maximize predictive performance, a promising strategy for improving the predictions of stress-related outcomes is to directly use the questionnaire items of the MEHQ as predictor variables, instead of the domain sum scores. A similar comparison was conducted by Seeboth and Mõttus (2018), suggesting that small improvements are possible by using items. In contrast, sum scores might be preferable in guiding interventions, as they provide a more easily interpretable summary of important psychological concepts.

These arguments result in the following detailed research questions:

Q1) How accurately can we predict the four outcomes of work-related stress: strains, general health, turnover intention, and sickness absence based on psychosocial hazards measured with the MEHQ?

Q2) Do non-linear machine learning algorithms like the random forest give better predictive performance in comparison to regularized linear models like the elastic net or classical unregularized linear regression?

Q3) Can predictive performance be increased by adding demographic covariates in the cross-sectional setting or by adding demographic variables and outcomes measured at t1 as covariates in the longitudinal setting?[2]

[2] We analyzed a condition without (Q1) and with (Q3) covariates because we were 1) interested in what is the predictive performance of only the MEHQ compared to the MEHQ plus covariates? 2) we are not

Q4) Can predictive performance be increased by using the questionnaire items instead of the sum scores as predictor variables?

In summary, the current study builds on previous research by (a) using a variety of psychosocial hazards and stress-related outcomes in (b) both a longitudinal as well as a cross-sectional design (c) based on a quota sample to approximate a representative sample of employees, and (d) applying both linear and nonlinear machine learning methods with resampling in addition to classical linear regression.

Method

Procedure and participants

During 2016 and 2017, a total of three waves were conducted with the MEHQ in Germany. All samples were taken from a German online access panel supported by the market and social research Institute Kantar. Panel members receive credit points for each completed questionnaire that they can redeem for coupons or non-cash gifts.

Wave 1 lasted from May 11th to May 23rd and wave 2 from August 25th to September 9th in 2016. Participants of wave 1 or wave 2 were interviewed again in wave 3, which was conducted more than one year later from November 30th to December 13th in 2017. As 476 participants of wave 1 or wave 2 were no longer part of the panel when wave 3 was starting, we could not invite them to participate in wave 3. Wave 1 and wave 2 were part of the development process of the questionnaire: Wave 1 was used to select the final item pool, whereas the final structure was identified in wave 2 (more information can be taken from Zweck, 2017).

---

sure which information (demographic covariates, health or turnover intention) are available in practice e.g. as part of an employee survey.

Wave 1 and 2 consisted of independent samples, as participants from wave 1 were excluded from wave 2. For the cross-sectional analyses of the present study, both datasets were analyzed jointly. We label wave 1 and wave 2 as timepoint t1 and wave 3 as timepoint t2. The time lag between t1 and t2 of slightly over one year is in line with De Lange, Taris, Kompier, Houtman, and Bongers (2004) who reported the best model fit estimating the relationship between working conditions and health-related outcomes based on a 1-year time lag. More importantly, the basic reference point for all items of the MEHQ is 12-months (participants are instructed to think about the last twelve months when answering MEHQ items). By choosing a time lag between t1 and t2 greater than 12 months (namely at least 14 months), we avoid overlapping time intervals.

For wave 1 and 2 we used targets regarding age, education and gender. Targets for employees were taken from the information provided by the Federal Statistical Office in Germany. By keeping the distribution of important demographic variables in the general population, this nearly representative sample approximates a true random sample of German employees.

Exclusion criteria: We only included participants with complete responses. In the longitudinal condition, we removed participants who did not have the same job at t2. To avoid systematic bias caused by effects of satisficing (Krosnick, 1999), we further excluded *speeders* and *straightliners*[3] from all analyses (syntax on how to identify speeders and straightliners can be found in our analyses scripts at https://osf.io/t7a28/). This resulted in

[3] Speeders in web-surveys are characterized by extremely fast responses, typically those who select a response without having read the question first. The response style of straightlining is described by giving identical ratings to a series of statements (Zhang & Conrad, 2014). Here again, it can be supposed that these persons have not read the question when giving their answers. Therefore, it can be assumed that answers of persons responding in this way are not interpretable.

sample sizes of *n* = 1,327 for cross-sectional conditions (original sample size: *N* = 1,523) and *n* = 444 for longitudinal conditions (original sample size: *N* = 609), except for the longitudinal prediction of sickness absence with *n* = 383. The dropout rates were 4.7% for wave 1, 4.2% for wave 2, and 3.2% for wave 3.

Measures

Survey respondents answered the MEHQ. The German MEHQ including all questions used in wave 3 can be found in the online supplemental materials. An English translation can be taken from Appendix A.

The MEHQ consists of 82 items for the core module (including psychosocial hazards and stress-related consequences) and 11 additional items for a manager module that can be completed by respondents with leadership responsibility. In this article, we focus on the core module because the sample size for the manager module was too small for our statistical analyses. As the original study on the development process and validation of the MEHQ (Zweck, 2017) is only published in German, we will report the most important psychometric indices here. These were computed based on the combined sample of wave 1 and 2.

Psychosocial hazards: Psychosocial hazards were measured on fully labeled five-point-Likert scales. Three different sets of labels were used:

- 1 (*Not at all)*, 2 *(Low)*, 3 *(Average)*, 4 *(Strong)*, 5 *(Very strong)*
- 1 *((Almost) Never)*, 2 *(Rarely),* 3 *(Sometimes),* 4 *(Often),* 5 *((Almost) Always)*
- 1 *(Strongly disagree),* 2 *(Disagree)*, 3 *(Neither agree nor disagree)*, 4 *(Agree)*, 5 *(Strongly agree)*.

According to exploratory factor analyses (Zweck, 2017), psychosocial hazards of the MEHQ can be “summarized” analogues to the domains shown in Figure 1. Additionally, measured outcomes are presented here. An overview of which items belong to which domain is presented in Appendix B.

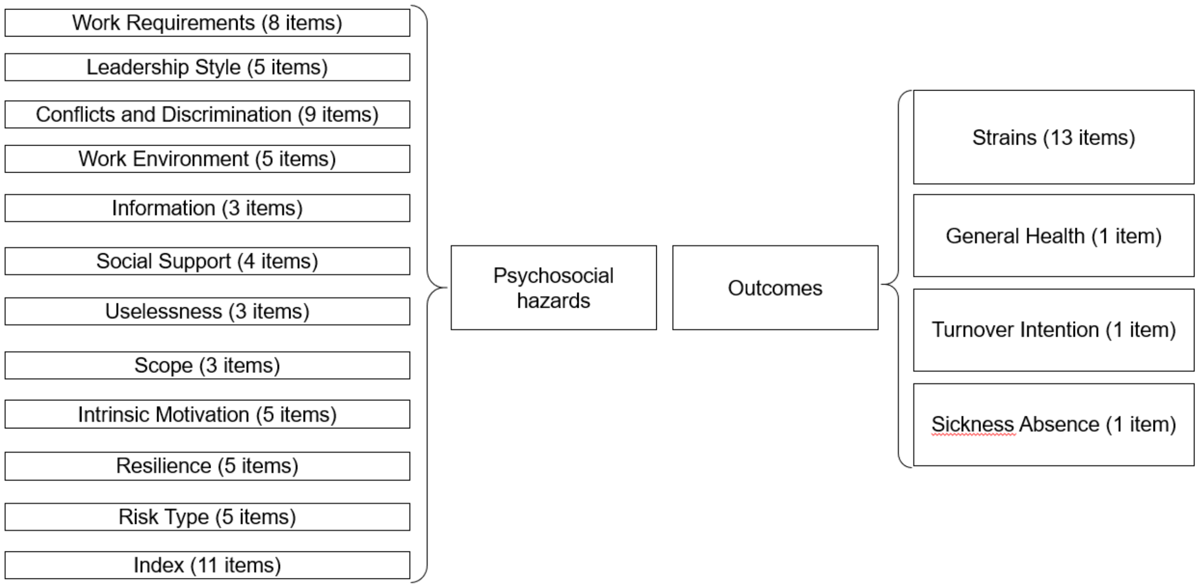

Figure 1. Domains and outcomes of the MEHQ.

95% confidence intervals for Cronbach's alpha (α) and McDonalds Omega (Ω) for the domains of the MEHQ can be stated as follows:

- Work Requirements (α: [.88, .90], **Ω**: [.88, .90])
- Leadership Style (α: [.89, .91], **Ω**: [.89, 0.91])
- Conflicts and Discrimination (α: [.86, .88], **Ω**: [.86, .88])
- Work Environment (α: [.79, 83] **Ω**: [.79, .83])
- Information (α: [.83, .86], **Ω**: [.83, .86])
- Social Support (α: [.87, .89], **Ω**: [.87, .89])
- Uselessness (α: [.77, .81], **Ω**: [.77, .81])
- Scope (α: [.45, .54], **Ω**: [.46, .55])
- Intrinsic Motivation (α: [.83, .85], **Ω**: [.83, .86])
- Resilience (α: [.64, .70], **Ω**: [.64, 70])
- Risk Type (α: [.67, 72], **Ω**: [.68, .73])

- Index for measuring psychosocial hazards based on 11 items. Although they do not form a homogenous factor, all items are part of the questionnaire as they are theoretically relevant for measuring psychosocial hazards.

A confirmatory factor analysis jointly modeling the 11 correlated dimensions (index items are not included) revealed the following model fit: $\text{Chi}^2 = 4700.785$ ($df = 1375$, $p < .001$), RMSEA = .043 (90% CI: [.041, .044]), CFI = .970, SRMR = .052. The model was fit with the WLSMV estimator in the R package lavaan (Rosseel, 2012).

Outcomes. The MEHQ also includes items to measure four different outcomes of work-related stress (see Figure 1):

Strains were measured as the sum score of 13 items representing stress-related physical and mental consequences.

General health was measured on a 10-point scale with labeled endpoints ranging from 1 *(Very poor health)* to 10 *(Very good health)*: "When you think about the last 12 months, how do you rate your physical and mental health? 1 point means 'very poor health', 10 points mean 'very good health'. The levels in between allow you to scale your assessment."

Turnover Intention was measured with a single item ("I think about changing my job.") developed by Zweck (2017) using a five-point fully labeled Likert scale with the following verbal labels: 1 *(Strongly disagree)*, 2 *(Disagree)*, 3 *(Neither agree nor disagree),* 4 *(Agree)*, 5 *(Strongly agree).*

In the first wave, *sickness absence* was measured by asking for the number of sick days because of health reasons with the following six categories: *0 days*, *1–5 days*, *6–10 days*, *11–15 days*, *16–20 days*, *more than 20 days*. In the second wave, an open question was used instead of the closed format. In the third wave, sick days were measured with an open-ended question like in wave 2 ("How many days have you been unable to work in the past 12 months due to health problems–physical or mental? If you are not sure of the exact number, just estimate."). While complete responses were generally requested for all questions in each wave, participants were allowed not to report their absence days in wave 3. It could be possible that participants of wave 3 were not working (e.g. retired or unemployed) and therefore were not able to answer the question regarding sickness absence. This resulted in the smaller sample size for longitudinal analyses of sickness absence.

To enable a direct comparison of sickness absence between waves, open answers from waves 2 and 3 were collapsed to the six categories from wave 1.

Covariates. For our study, we consider the demographic covariates gender, age, as well as the control variables tenure, time duration between measurements, and outcomes measured at t1. We report gender (binary) as previous studies indicate gender related differences regarding psychosocial hazards and stress experiences (Amercian Psychological Association, 2010; Bogg & Cooper, 1994; Lesuffleur et al., 2014; Purvanova & Muros, 2010; Rugulies, Bültmann, Aust, & Burr, 2006; Rugulies, Norborg, Sørensen, Knudsen, & Burr, 2009).

We added *age* (categories: *Less than 25 years*, *25–34 years*, *35–44 years*, *45–54 years*, *More than 55 years*) as a covariate because it is also assumed to be a relevant socio-demographic factor in the assessment of work-related stress (Marinaccio et al., 2013). As the measured effects of the relationship between psychosocial hazards and strains can be related to the length of exposure to psychosocial hazards (e.g., De Lange, 2005), we also controlled for *tenure* and *duration*. Tenure was measured with a question about how many years the participants have been working for their current employer (categories: *Less than 1 year*, *1 year to less than 3 years*, *3 years to less than 5 years*, *5 years to less than 10 years*, *10 years and more*). In longitudinal analyses, the duration variable was based on the number of days between the first and second measurement of the participant. Guided by other longitudinal studies on the prediction of stress-related outcomes with psychosocial hazards (e.g., Albertsen et al., 2010; Borritz et al., 2005), we also considered all outcomes measured at t1 (strains, health, turnover intention, sickness absence) as covariates in one condition of our longitudinal analyses.

Statistical methods

Linear regression and the elastic net. Classical linear regression is probably the most common predictive model in psychological science. However, estimates from unregularized linear models are unstable in small samples with many predictors. To increase prediction accuracy, regularized linear models like ridge regression and the LASSO (Tibshirani, 1996) reduce the variance of the resulting predictions with only a small increase in bias by shrinking the magnitude of regression coefficients towards zero. Ridge regression retains all predictors while the LASSO returns zero estimates for some coefficients, effectively selecting a set of important predictors. An algorithm that combines both regularization strategies is the elastic net (Zou & Hastie, 2005). The elastic net is a linear model, so coefficients that are unequal to zero can be interpreted as in classical linear regression.

Random forest. In contrast to the elastic net, which typically models linear relationships without including interaction terms, the random forest (Breiman, 2001) is a highly nonlinear algorithm that shows good performance in a wide range of machine learning problems (Fernández-Delgado, Cernadas, Senén, & Amorim, 2014). In a random forest, hundreds of decision trees are grown on bootstrap samples of the original dataset. For a new observation, the predictions from all trees are aggregated, which stabilizes the highly variable predictions from single trees. Due to its tree structure, the random forest also natively incorporates nonlinear interactions between many predictors.

Baseline model. To evaluate the performance of predictive models, comparing different algorithms is often a necessity. A simple method to judge whether an algorithm performs "better than chance" is the comparison with a baseline model that does not use any predictor variables and can only make uninformed guesses. In the current study, the baseline model constantly predicts the mean of the outcome variable in the training set.

Cross-validation. Predictive modeling relies on performance measures to quantify the similarity between predictions of an algorithm and the true labels. We treat all outcomes as regression problems, in which case the standard performance measure in machine learning is the mean squared error, $MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$ with y_i being the observed and $\hat{y}_i$ the predicted outcome of subject i. However, as psychologists are more familiar with the coefficient of determination, $R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{N}(y_i-\bar{y})^2}$ with $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$, we report R^2 as primary measure while results for the MSE can be found in the online supplemental materials.

In psychology, R^2 is commonly computed on the same data also used for model fitting. Unfortunately, in-sample R^2 can greatly overestimate the predictive out-of-sample accuracy of the model in small samples with many predictors (Larson, 1931; Yarkoni & Westfall, 2017; Yin & Fan, 2001). A better strategy to estimate predictive performance on new observations is to use resampling methods like ten-fold cross-validation (10-CV; Kohavi, 1995). In 10-CV, the data is randomly partitioned into 10 equally sized parts. Each part once serves as a test set whose observations are predicted based on a model estimated on a training set (the union of the remaining 9 parts). Compared to the in-sample R^2, the average performance across the ten CV folds is a more realistic estimate for the performance of the full model, which is trained on the whole data and would be applied in practical applications. To stabilize performance estimates in small samples, 10-CV can be repeated with different partitions of the data to average across several cross-validation runs.

Note that when evaluated on test sets, R^2 can take on negative values (which frequently happens when particularly flexible models are "overfitted" to small samples). Thus, cross-validated R^2 should not be interpreted as a ratio of explained variance. Similar to the idea of comparing a model's predictive performance to the baseline model, the interpretation of negative R^2 is that the predictive model provides worse predictions for new observations than a simple model, which constantly predicts the mean outcome value in the test set.

When estimating the performance of algorithms that require hyperparameter tuning, nested resampling has to be performed inside the cross-validation scheme (Bischl, Mersmann, Trautmann, & Weihs, 2012). In our study, 10-CV with 10 repetitions is used to estimate predictive performance in the outer loop while simple 10-CV is used in the inner loop for hyperparameter tuning. For the elastic net, the parameter *alpha* determines the mixing of the elastic net penalty with LASSO regularization for *alpha* = 0 and ridge regression for *alpha* = 1. A second parameter *lambda* controls the amount of regularization. For *lambda* = 0, the elastic net is equivalent to classical linear regression. *Alpha* was tuned on a regular grid with resolution 50. *Lambda* was tuned internally in another loop (10-CV on a grid of resolution 400 to find the *lambda* with the best mean MSE). In a random forest, the ratio parameter *mtry.perc* determines the number of randomly drawn predictor variables at each split of a tree. A second parameter *min.node.size* determines the minimum number of observations in a tree node to attempt further splitting. Both parameters were jointly tuned on a two-dimensional grid of size 20 (*mtry.perc* between 0 and 0.5) times 3 (*min.node.size* equal to 5, 10 or 15). One thousand trees were fitted for each random forest. Optimal hyperparameter configurations for *alpha*, *mtry.perc*, and *min.node.size* used within the training sets of the outer loop were chosen based on the best mean MSE in the respective test sets of the inner loop.

Predictive modeling conditions

Separate predictive models were fitted for each of the four outcomes. In the cross-sectional condition (cross-sectional), outcomes at t1 were predicted by psychosocial hazards measured by the MEHQ from the same time point. In the longitudinal condition (longitudinal), outcomes at t2 were predicted by psychosocial hazards measured by the MEHQ from t1.

We compared models with different sets of predictor variables. In each model, we either included the sum scores of the 11 MEHQ domains in addition to the sum score of the index (sum scores), all 66 MEHQ psychological hazard items (items) or both sum scores and items (combined). The combined condition was included as using both aggregated and unaggregated scores as predictors might increase predictive performance and is common in machine learning. For an assessment of how much predictive performance can be attributed to the MEHQ alone, all models were fitted either with (covariates included) or without covariates (covariates excluded). Covariates were *gender*, *age*, and *tenure* in the cross-sectional analyses and *gender*, *age*, *tenure*, *duration*, and all *outcomes* measured at t1 in the longitudinal analyses. We also computed models containing only the respective covariates (covariates only) as an additional control condition.

Software

The *mlr* framework (Bischl et al., 2016) was used to interface different machine learning algorithms in the open source statistical software R (R Core Team, 2018) and run all benchmark analyses. Elastic net models were trained with the glmnet package (Friedman, Hastie, & Tibshirani, 2010). The ranger package (Wright & Ziegler, 2017) was used to train random forests. Benchmark analyses were run in parallel on the Linux Cluster of the Leibniz Supercomputing Centre in Garching, Germany via the batchtools package (Lang, Bischl, & Surmann, 2017). Plots were created with ggplot2 (Wickham, 2016). Reproducible analyses scripts are available under https://osf.io/t7a28/. We also published all data necessary to reproduce the reported results as a scientific use file (Zweck & Pargent, 2019).

Results

Descriptive statistics of covariates, which demonstrate a high similarity especially between the quoted samples of wave 1 and wave 2, are presented in Table 2.

Table 2

Distribution of covariates for each wave

	Wave 1	Wave 2	Wave 3
n (after exclusion criteria)	446	881	444
Gender (% women)	48	48	49
Age (% in each category)			
≤ 24 years	7	11	4
25–34 years	23	20	13
35–44 years	26	22	22
45–54 years	26	28	35
≥ 55 years	18	19	26
Tenure (% in each category)			
< 1 year	9	11	6
1–2 years	17	18	13
3–4 years	16	14	14
5–9 years	20	20	25
≥ 10 years	37	37	42
Duration (Median days t2 – t1)	567	457	

Note. For waves 1 and 2, descriptive statistics were computed for all employees included in the cross-sectional analyses. For wave 3 and the variable *duration*, only those participants are reported who were included in the longitudinal analyses for the outcomes strains, general health, and turnover intention. For the longitudinal analyses of sickness absence, only n = 383 participants could be included.

Correlations between the measured outcomes (Table 3) show that strains were most stable over time whereas sickness absence was the least stable outcome. The highest correlation between outcomes was observed between strains and general health. The second highest correlation was found between strains and turnover intention.

Table 3

Correlations between outcomes

	Strains	General health	Turnover intention	Sickness absence
Strains	.79	.54	.44	.38
General health	.57	.65	.23	.46
Turnover intention	.36	.13	.58	.09
Sickness absence	.33	.37	.11	.50

Note. Pearson correlations between outcomes. Upper diagonal: t1 (n = 1,327). Lower diagonal: t2 (n = 444, for sickness absence: n = 383). Diagonal: correlation between t1 and t2 (n = 444, for sickness absence: n = 383).

Predictive performance estimates of R^2 for all analysis conditions are presented separately for each outcome in Figures 2–5 (Figure 2: Strains, Figure 3: General health, Figure 4: Turnover intention, Figure 5: Sickness absence). We display the mean R^2 estimates from 10 times repeated 10-CV along with the standard deviation (above and below) across all 100 test sets. Predictive algorithms are indicated by different colors. Tables with the corresponding numeric estimates for R^2 and MSE along with in-sample estimates of predictive performance in the training sets can be found in the online supplemental materials.

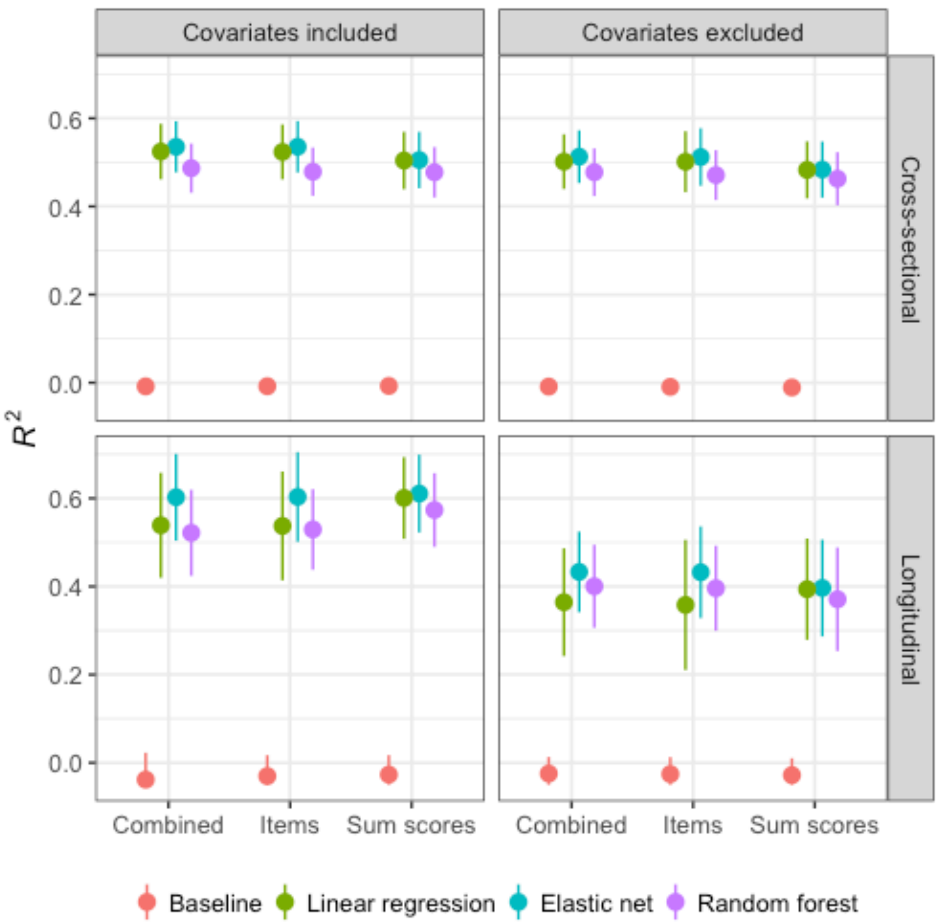

Figure 2. Benchmark results for strains from 10-fold cross-validation with 10 repetitions (*M* +/– 1 *SD*). Cross-sectional performance of linear regression only using covariates: *M* = 0.06, *SD* = 0.05. Longitudinal performance of linear regression only using covariates: *M* = 0.60, *SD* = 0.09. Cross-sectional: *n* = 1,327. Longitudinal: *n* = 444 (383 for sickness absence).

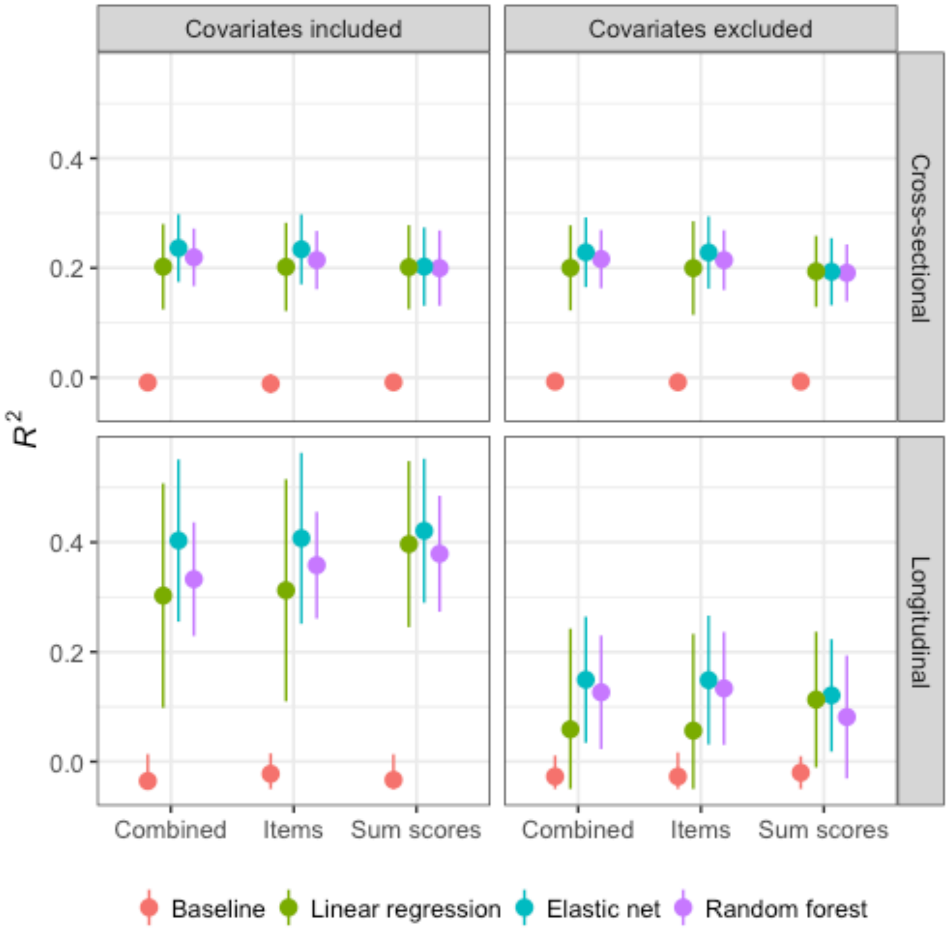

Figure 3. Benchmark results for general health from 10-fold cross-validation with 10 repetitions (*M* +/- 1 *SD*). Cross-sectional performance of linear regression only using covariates: *M* = -0.01, *SD* = 0.02. Longitudinal performance of linear regression only using covariates: *M* = 0.43, *SD* = 0.14. Cross-sectional: *n* = 1,327. Longitudinal: *n* = 444 (383 for sickness absence).

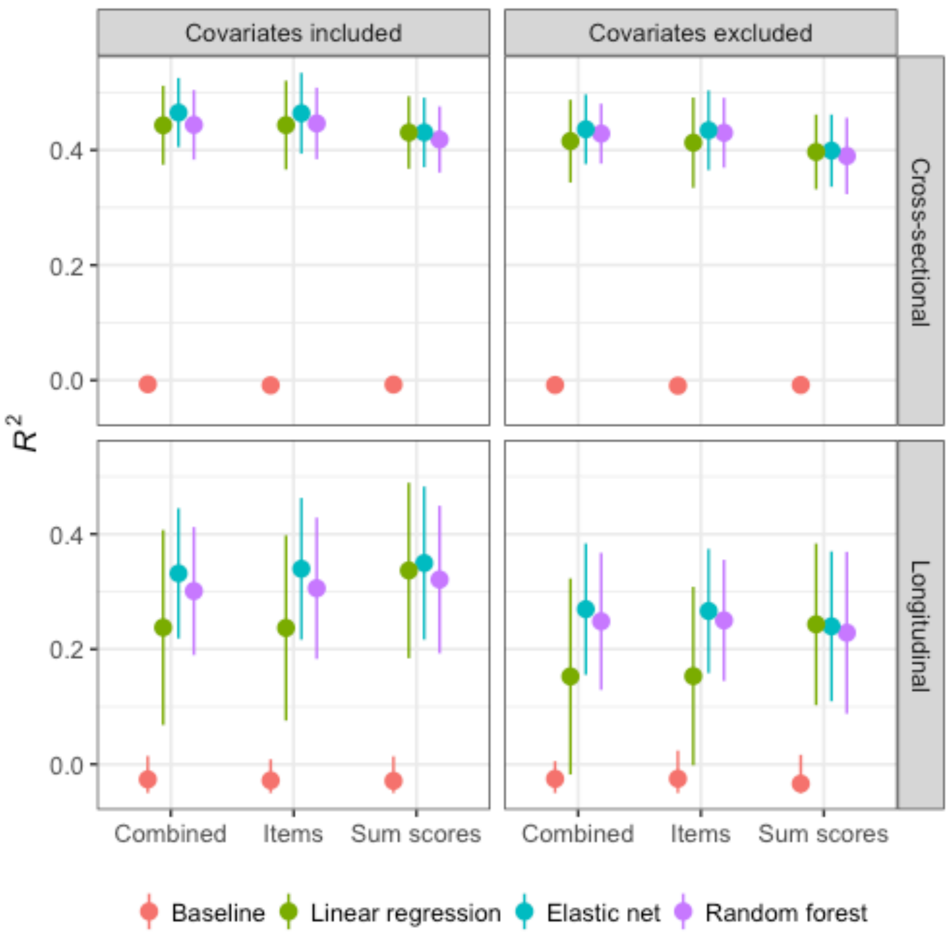

Figure 4. Benchmark results for turnover intention from 10-fold cross-validation with 10 repetitions (*M* +/- 1 *SD*). Cross-sectional performance of linear regression only using covariates: *M* = 0.08, *SD* = 0.05. Longitudinal performance of linear regression only using covariates: *M* = 0.31, *SD* = 0.16. Cross-sectional: *n* = 1,327. Longitudinal: *n* = 444 (383 for sickness absence).

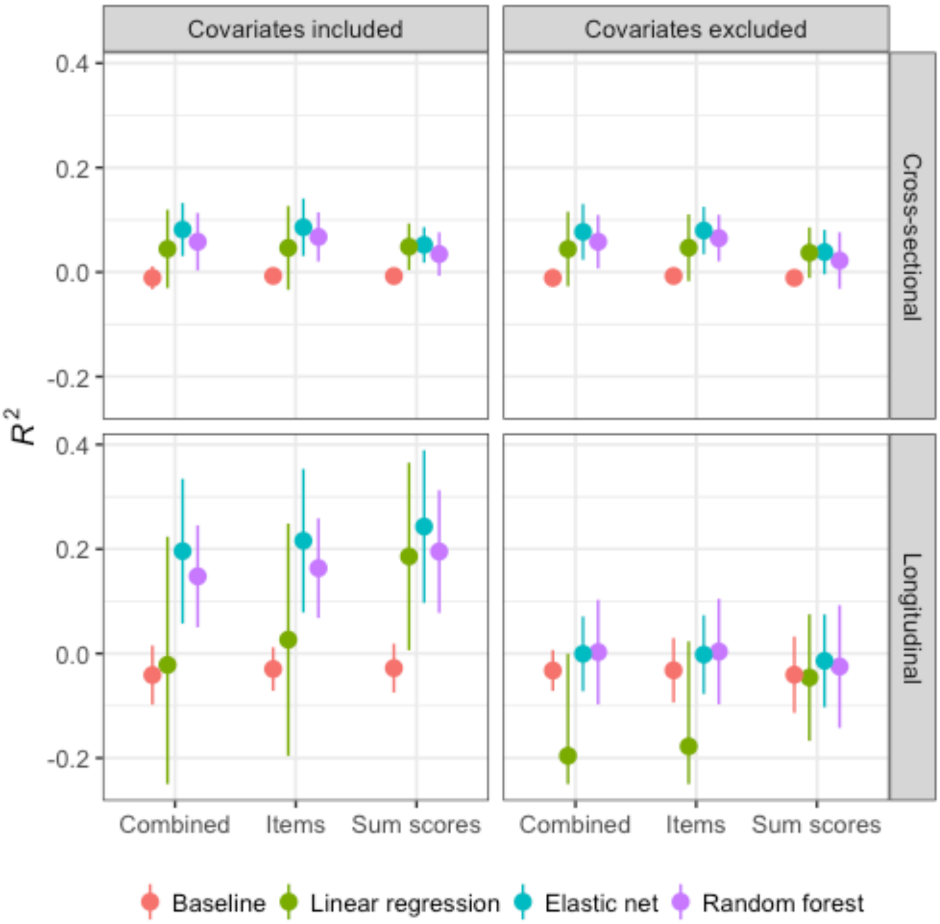

Figure 5. Benchmark results for sickness absence from 10-fold cross-validation with 10 repetitions (*M* +/- 1 *SD*). Cross-sectional performance of linear regression only using covariates: *M* = -0.0001, *SD* = 0.03. Longitudinal performance of linear regression only using covariates: *M* = 0.23, *SD* = 0.17. Cross-sectional: *n* = 1,327. Longitudinal: *n* = 444 (383 for sickness absence).

Q1) Performance of predicting outcomes of work-related stress with psychosocial hazards

Above chance predictions (in comparison with the baseline model) were achieved in all conditions, except for longitudinal predictions of sickness absence without adding covariates. Highest predictive performance, for models using only psychosocial hazards as predictors, was observed for the outcome strains (cross-sectional: $R^2 \approx .5$, longitudinal: $R^2 \approx$

.4), followed by turnover intention (cross-sectional: $R^2 \approx .4$, longitudinal: $R^2 \approx .25$), general health (cross-sectional: $R^2 \approx .2$, longitudinal: $R^2 \approx .1$), and sickness absence (cross-sectional: $R^2 \approx .05$, longitudinal: $R^2 \approx 0$).

For models without covariates, higher predictive performance was estimated for cross-sectional predictions in comparison with longitudinal predictions. Without adding covariates, it was not possible to predict sickness absence in the longitudinal condition better than chance. Here, performance estimates for all models were very close to an R^2 of zero and did not reliably beat the baseline model.

Q2) Comparison of predictive algorithms

In general, classical linear regression was competitive with both the elastic net and the random forest. Sometimes, classical linear regression yielded slightly inferior performance when single items or items in combination with sum scores were used as predictor variables. However, notable performance deficits for linear regressions were only observed in the longitudinal conditions when using items as predictors, probably due to the much smaller sample sizes. Predictive performance estimates for the elastic net and the random forest were quite similar. When only considering the point estimates of predictive performance (ignoring the variance of CV estimates across folds), a small advantage was observed for the elastic net in all above chance conditions.

Q3) Increasing predictive performance by adding covariates

In cross-sectional conditions, adding the covariates age, gender, tenure, and duration as predictors did not notably increase predictive performance for any outcome. In contrast, adding covariates in the longitudinal condition that also included all outcomes measured at t1 led to a big performance boost (e.g., linear regression for general health using sum scores, $\Delta R^2 \approx .3$). For general health and sickness absence, longitudinal performance with covariates was even higher compared to the cross-sectional conditions. Our control analyses show that similar longitudinal predictive performance can be achieved when providing only the set of covariates as predictors. This finding is also reflected by the coefficient estimates of the final elastic net models trained on the full datasets (see online supplemental materials). In longitudinal models including covariates as predictors, coefficients corresponding to outcomes measured at t1 have the highest estimates and most coefficients of both items or sum scores have zero estimates.

Q4) Comparison of sum scores and items as predictor variables

Predictive performance was quite similar for models with different sets of predictor variables (see Figures 2–5). In conditions without covariates, slightly better performance (when only considering point estimates) was observed when items were used as predictors instead of sum scores. The combined condition with both items and sum scores always showed nearly identical performance to the items condition.

Discussion

Although stress at work is a subject of global interest (ILO, 2016; WHO, 2013), current questionnaires do not provide an exhaustive measurement of psychosocial hazards. The most prominent COPSOQ (T. S. Kristensen et al., 2005) does not measure all relevant factors identified by the literature (e.g., information overload; A. R. Lee et al., 2016). Zweck (2017) developed the Munich Employee Health Questionnaire to exhaustively measure relevant aspects for workplace stress. In the original publication of the MEHQ, only cross-sectional evidence of criterion validity is presented. For this study, we analyzed information from a second point in time at least 14 months after the cross-sectional surveys in order to establish criterion validity based on a longitudinal design in comparison with the respective cross-sectional results.

Ten-fold cross-validation was applied to estimate the performance of both linear and nonlinear predictive models (baseline, classical linear regression, elastic net, random forest) to predict four outcomes of work-related stress (general health, strains, turnover intention, sickness absence) in two design conditions (cross-sectional, longitudinal). We used different types of predictors based on psychosocial hazards measured by the MEHQ (sum scores, items, combined) along with different control settings (covariates excluded, covariates included, only covariates). We showed that in conditions using only psychosocial hazards as predictors, all outcomes except sickness absence could be predicted with satisfying accuracy both in the cross-sectional and in the longitudinal settings (Q1). The performance of classical linear regression models with sum scores as predictor variables was competitive with regularized linear and nonlinear machine learning algorithms (Q2). Predictive performance could be greatly improved by including outcomes measured at t1 in the longitudinal setting (Q3). Using single items as predictor variables instead (or in addition to) sum scores did not lead to a practically relevant performance increase (Q4).

Most previous studies assessing psychological risks and outcomes at work were only conducted cross-sectionally (Bliese et al., 2017; Rosário et al., 2016; Theorell & Hasselhorn, 2005). Besides assessing a broad range of psychosocial hazards in a longitudinal design, our study has several further advantages over previous work: First, we use a nearly representative sample of German employees. In contrast, many longitudinal studies conducted with the COPSOQ are based on healthcare sector samples (e.g., Aust et al., 2010; T. R. Kristensen et al., 2010; Li et al., 2010). Naturally, predictive performance decreased in the longitudinal design with covariates excluded. Nevertheless, the resampled R^2 for strains was still quite high (cross-sectional: $R^2 \approx .5$, longitudinal: $R^2 \approx .4$). In a cross-sectional analysis based on the COPSOQ within the Gutenberg Health Study (Nübling et al., 2013), only job satisfaction could be predicted to a comparable amount ($R^2 = .51$). Predictions for health-related outcomes like burnout and general health were less accurate (burnout: $R^2 = .35$, general health: $R^2 = .11$). Based on these results, the MEHQ shows even superior predictive performance in the longitudinal setting compared to the cross-sectional studies of the COPSOQ regarding health-related outcomes.

One reason for this increased performance could be the broader range of aspects measured by the MEHQ. In addition to the COPSOQ (T. S. Kristensen et al., 2005; Nübling & Hasselhorn, 2010), the MEHQ measures aspects like information overload, completeness of task, working environment or major life events – shortly, nearly all aspects identified as relevant (see Table 1). Thus, the content validity of the MEHQ is superior compared to other questionnaires on the topic.

Within the current study, we focused not only on the prediction of health-related outcomes, but also on turnover intention as it has high economical relevance (Brun & Lamarche, 2006; Hassard et al., 2018) and partially similar antecedents (Griffeth et al., 2000). Turnover intention has not been the focus of questionnaires measuring stress at work to date, but the correlation between strains and turnover intention (r = .44 at t1 and r = .36 at t2) underlines the relevance of turnover intention in the field of stress at work and its health-related outcomes. We could show its relevance and its functioning within the MEHQ.

In summary, we found satisfying criterion validity of the MEHQ with regard to the outcomes general health, turnover intention and strains. However, we could only predict sickness absence in longitudinal models where covariates–and therefore sickness absence at t1–were included. In longitudinal models, where only the MEHQ was considered for prediction, the predictive performance for sickness absence was low (cross-sectional setting) or close to zero (longitudinal setting).

Predicting sickness absence

Despite some studies reporting significant associations of some psychosocial hazards with sickness absence (e.g., Head et al., 2006; Niedhammer et al., 2013; Rugulies et al., 2010; Slany et al., 2014)–even if associations were found to be small (Magnavita & Garbarino, 2013), we found low (cross-sectionally) or non-existent (longitudinally) predictive performance for sickness absence when using only psychosocial hazards as predictors. Several reasons might explain these results:

Pressure at work and long working hours, which are both related to bad health (Cox et al., 2000), might reduce absence in the short run (Chandola, 2010; Lesuffleur et al., 2014; Smulders & Nijhuis, 1999). People who work many hours are rarely sick, but when they are sick, they are absent for a longer time (Lesuffleur et al., 2014). Thus, the *duration* of sickness absence might be important. The study of Reeuwijk et al. (2015) revealed that the duration of more than 15 sick days seems to be the best operationalization for predicting sickness absence. In fact, many studies that found a significant effect on sickness absence only looked at longer durations. For example, one study conducted with the COPSOQ only looked at sickness absence for 3 weeks or more (Rugulies et al., 2010). The finding that an increasing burnout level is positively related to longer sickness absence duration (Borritz, Rugulies, Christensen, Villadsen, & Kristensen, 2006) could be an additional reason for focusing on long sickness absence. In the present study, by contrast, we analyzed sickness absence based on different categories (*0 days*, *1–5 days*, *6–10 days*, *16–20 days*, *More than 20 days*).

On a different note, the frequency of being absent from work might be more important than the duration. Lesuffleur et al. (2014), found slightly stronger effects when analyzing the *frequency* of being off to work than sickness absence *duration* based on a huge sample of nearly 47,000 employees in France.

Furthermore, there seem to be many non-stress-related factors that determine whether an employee is ill and absent from work. A person who exercises a lot might be little stressed, but might have sick days due to sports injuries. Thus, we argue that sickness absence might not be an adequate criterion for evaluating criterion validity of a questionnaire assessing psychosocial hazards.

Insights from predictive modeling

Previous research on work-related stress has mostly relied on regression analyses with generalized linear models (e.g., Burr, Albertsen, Rugulies, & Hannerz, 2010; Li et al., 2010; Nübling et al., 2013). To the best of our knowledge we present the first study analyzing the relationships between psychosocial hazards and stress related outcomes with predictive modeling and machine learning methods. As the linear relationship between psychosocial hazards and stress-related outcomes was questioned in the past (e.g., Zapf et al., 1996), we included both linear (elastic net) and non-linear (random forest) algorithms.

Surprisingly, the linear elastic net performed slightly better than the random forest that automatically models nonlinear functional relationships and interactions. This finding might suggest that our questionnaire data mostly contains linear main effects. Alternatively, the sample sizes of n = 1,327 in the cross-sectional and n = 444 or n = 383 in the longitudinal setting were too small for the random forest to incorporate any reliable nonlinear trends into the model. Further studies with bigger samples might tell whether this is a consistent pattern when working with psychological questionnaire data on work-related stress.

Empowered by the capability of regularized predictive models to incorporate a large number of predictor variables, we also investigated whether predictions could be improved by using all MEHQ items as predictors directly, in contrast to the traditional strategy of using the aggregated sum scores. Overall, we did not find any big performance increase by using items. However, when items without covariates were combined with the elastic net, we observed slightly better performance for all outcomes compared to using sum scores with classical linear regression (cross-sectional: mean ΔR^2 = .036, longitudinal: mean ΔR^2 = .035, but note the high standard deviation between CV folds). This observation is in line with a recent study discussing increased predictive performance when using items from Big Five questionnaires to predict 40 different life domain outcomes based on elastic net models

(Seeboth & Mõttus, 2018). Their reported absolute performance increase (mean $\Delta R^2 = .013$) is comparable to our results. However, we doubt that such small observed performance increases are currently worth to be pursued in practice, as they would be accompanied by several complications in modeling and in the interpretation of results:

First, to handle the larger amount of predictor variables, regularized predictive models–which are still new to many applied psychologists–are typically needed. Using those models also requires knowledge about resampling methods like 10-CV, as in-sample performance estimates (i.e. evaluating the performance on the same data used to fit the model) can be way too optimistic (Yarkoni & Westfall, 2017). This is especially important when using nonlinear models like the random forest, as is demonstrated by our in-sample fits reported in the online supplemental materials.

Second, applying algorithms like the random forest to make new predictions in practice (e.g., predict negative stress-related outcomes for an employee with problematic working conditions) requires access to the full trained model, as there is no simple prediction equation that can be published in a paper or a manual. Running the model to make a prediction might require more methodological know-how as can be expected from most current human resource departments.

Third, interpreting results is more complicated with a large number of items compared to a small number of intuitive domain sum scores. While the elastic net still provides an interpretable regression coefficient for each predictor variable (see online supplemental materials), such information is not directly available for the random forest and would require methods from interpretable machine learning (Molnar, 2019). Sum scores also have a clear advantage when practitioners want to use the results of the predictive model to speculate about possible interventions. According to GDA (2017)[4], interventions affect workplace categories rather than single psychosocial hazards. Thus, the aggregation level of domain sum scores facilitates theorizing about interventions that might reduce the respective psychosocial hazards, which in turn could reduce stress-related outcomes in work places.

Implications for measuring work-related stress

This study makes a few important contributions to stress research regarding how to measure “stress at work”. First, we show that it is possible to predict health-related outcomes *and* turnover intention with the MEHQ over more than one year, based on a nearly representative sample of German employees. Regarding stress-related health outcomes, the MEHQ seems to provide higher predictive performance than the widely used COPSOQ. One might expect that the broader range of aspects in the MEHQ causes a longer interview duration, which would be a disadvantage for the practical application of the questionnaire within employee surveys. However, this is not the case. Median interview duration for the MEHQ is below 15 minutes (Zweck, 2017), which is lower than the interview duration for the COPSOQ indicated with 20 minutes (Richter & Schütte, 2014). Note however, that our panel sample was very familiar with answering surveys online. Interview duration might be longer if employees with less online experience answer the MEHQ.

The MEHQ is one of the first validated work related “stress-questionnaires” for Germany that assesses turnover intention. This is surprising, considering the literature on similar antecedents for turnover intention and other stress-related outcomes (e.g., Li et al., 2010). Indeed, our results show that psychosocial hazards predict both health-related

[4] The GDA (Gemeinsame Deutsche Arbeitsschutzstrategie) is a union of the federal government, the federal states and the accident insurance agencies to ensure security and health for employees in Germany. Therefore, the GDA published practical recommendations for conducting a psychosocial risk assessment in line with the legal specifications in Germany.

outcomes and turnover intention. Both assessments are relevant because stress-related health consequences as well as turnover create costs for the company (Brun & Lamarche, 2006; Hassard et al., 2018). Second, we showed that turnover intention can be predicted well (better than general health) by using antecedents like information overload that have not been in the focus of turnover research to date. Turnover research (e.g., Griffeth et al., 2000; Hom, Lee, Shaw, & Hausknecht, 2017) considers job satisfaction, as one of the best antecedents for turnover. We did not measure this variable, nevertheless, the MEHQ has a good predictive performance for turnover intention. The current study can contribute to turnover research by showing that there are important antecedents of turnover intention such as information overload that have not been comprehensively investigated in the field of turnover research so far.

Our findings suggest that questionnaires in the human resource sector on work-related stress should assess both health and turnover intention outcomes. However, this might not be advisable or practical within the context of employee surveys. Health-related questions within an employee survey could be defined as sensitive questions due to their intrusiveness and threat of disclosure (Tourangeau & Yan, 2007), which as a consequence could produce socially desirable answers, item nonresponses, or break-offs. Thus, we recommend avoiding asking a wide range of sensitive health-related questions in the practical use to ensure high response accuracy as well as reasonable response rates. This is reflected in regulations in Germany, where assessing work-place related hazards is required but assessing employee health is not (GDA, 2017). In our analyses, we found that adding outcomes measured at t1 as predictors in longitudinal analyses can greatly increase predictive performance. For general health and sickness absence, longitudinal predictive power in conditions with t1 outcomes included as covariates was even higher than in the cross-sectional settings. The high stability of outcomes over time is supported by further control analyses that showed that comparable predictive performance can be achieved in the longitudinal setting with only using the set of covariates (including outcomes at t1) as predictors. This finding is in line with earlier studies suggesting that the effect of psychosocial hazards on work-related stress largely vanishes when controlling for earlier measurements of work-related stress (e.g., Albertsen et al., 2010; Borritz et al., 2005). Thus, when earlier estimates of stress-related outcomes are available, they should be considered when the primary goal is to achieve a maximum predictive performance for future well-being of employees. However, this does not mean that measuring psychosocial hazards is not necessary: First, hazards are theoretically important as they provide accessible starting points when designing intervention studies to reduce work-related stress (Aust et al., 2010). Second, when stress-related outcomes cannot be assessed directly, work-related hazards might be the only option to predict stress-related outcomes in the workplace. And third, as mentioned earlier, we do not recommend asking sensitive questions such as health-related outcomes within the context of employee surveys.

By showing that prediction of health-related outcomes and turnover intention is possible *only* with the psychosocial hazards measured by the MEHQ, we provide one possible strategy on how countries and companies can contribute to the protection mandate of EU-OSHA, ILO and WHO (EU-OSHA, 2013; ILO, 2016; WHO, 2013) through monitoring the well-being of employees without violating their privacy by collecting sensitive health data.

Limitations and suggestions for future research

When discussing the findings from the current study, several limitations should be taken into account. First, as the MEHQ was only tested within a quoted sample of German employees, the results are only generalizable for this group so far. In contrast, other questionnaires like the COPSOQ have been validated in several languages and are used internationally (Dupret et al., 2012; Li et al., 2010; Moncada et al., 2014; Nübling &

Hasselhorn, 2010; Setti et al., 2017; Stauder et al., 2017). Thinking of the international use of the MEHQ, we assume that the MEHQ does not lose predictive power as job characteristics seem to be comparable across boundaries at least in industrial nations (R. Karasek et al., 1998). Nevertheless, one should validate an English version of the MEHQ to statistically ensure that the predictive performance reported in the current study is comparable for other countries. We provide a currently unvalidated English version in Appendix A and encourage other researchers to perform validation studies of the MEHQ in English speaking countries.

Second, we used only self-reported data, which could make our results susceptible to common method variance (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; Zapf et al., 1996). As a consequence, the relationships between the psychosocial hazards and the outcomes measured by the MEHQ might be overestimated. Future research could use objective measurements of stress such as medical outcomes like blood pressure (Landsbergis, Dobson, Koutsouras, & Schnall, 2013) as criteria instead. Here, it would be particularly important to link psychosocial hazards to medical outcomes *over time*, as only a few studies adopted this approach before (Bliese et al., 2017).

Third, underestimation of the relationship between psychosocial hazards and the outcomes in the MEHQ is also possible. Persons who are heavily exhausted because of stress at work (as exhaustion is associated with burnout; Freudenberger, 1974; Maslach & Jackson, 1981) might not have participated in this voluntary survey because of those stress related outcomes.

Fourth, as we had to rely on non-experimental data in this study, we cannot make any causal claims about the relationship between psychosocial hazards and outcomes measured by the MEHQ. We can make predictions based on psychosocial hazards measured by the MEHQ, but without knowing whether there might be important unobserved variables, reciprocal or reversed relationships (De Lange et al., 2004; Meier & Spector, 2013; Zapf et al., 1996). The only design to make causal claims would be a randomized intervention study in a longitudinal setting to check whether an intervention focused on reducing psychosocial hazards has the intended effect on reducing observed outcomes of work-related stress (e.g., Bliese et al., 2017; Michie & Williams, 2003).

Fifth, we combined wave 1 and wave 2 for our analyses although both waves were conducted in different time periods (wave 1 in May 2016 and wave 2 in August and September 2016). However, we controlled for different time intervals between t1 and t2 by adding the variable “duration” as a covariate. Our combined analysis is further justified by descriptive statistics of covariates, which suggests that wave samples are comparable.

Sixth, the MEHQ does not measure productivity. Productivity loss was recognized to be related to psychosocial hazards (Van den Berg, Elders, de Zwart, & Burdorf, 2009) and could be important from an economic point of view as productivity loss is also related to costs associated with workplace stress (Hassard et al., 2018). Therefore, it would be interesting for future research to investigate how accurate productivity can be predicted with the MEHQ.

Conclusion

With the present study, we demonstrate satisfying criterion validity for the MEHQ based on a quoted sample of German employees. Using both linear and nonlinear methods from predictive modeling and machine learning, we could predict health-related outcomes and turnover intention both cross-sectionally and over a one-year time period. However, longitudinal predictions better than chance were not possible for sickness absence without adding covariates as predictors. Longitudinal performance could be greatly increased by adding outcomes measured at t1 as covariates. Predicting outcomes with classical linear regression based on sum scores was competitive with machine learning models based on either domain sum scores or MEHQ items directly. When not including covariates, using regularized linear models in combination with all items as predictors slightly improved predictive performance compared to using classical linear regression based on sum scores. The nonlinear random forest was not superior to classical as well as regularized linear models. By providing a validated instrument to measure psychosocial risks at workplaces, we contribute to calls of the WHO and ILO to improve workplace health.

References

Albertsen, K., Rugulies, R., Garde, A. H., & Burr, H. (2010). The effect of the work environment and performance-based self-esteem on cognitive stress symptoms among Danish knowledge workers. *Scandinavian Journal Of Public Health, 38*(Suppl 3), 81-89. doi:10.1177/1403494809352104

Amercian Psychological Association. (2010). *Stress in America. Stress and gender*. Retrieved from https://www.apa.org/news/press/releases/stress/2010/gender-stress.pdf

Aust, B., Rugulies, R., Finken, A., & Jensen, C. (2010). When workplace interventions lead to negative effects: Learning from failures. *Scandinavian Journal Of Public Health, 38*(Suppl 3), 106-119. doi:10.1177/1403494809354362

Avey, J. B., Luthans, F., & Jensen, S. M. (2009). Psychological capital: A positive resource for combating employee stress and turnover. *Human Resource Management, 48*(5), 677-693. doi:10.1002/hrm.20294

Barber, L. K., & Santuzzi, A. M. (2015). Please respond ASAP: Workplace telepressure and employee recovery. *Journal of Occupational Health Psychology, 20*(2), 172-189. doi:10.1037/a0038278

Bauer, J., Stamm, A., Virnich, K., Wissing, K., Müller, U., Wirsching, M., & Schaarschmidt, U. (2006). Correlation between burnout syndrome and psychological and psychosomatic symptoms among teachers. *International Archives Of Occupational And Environmental Health, 79*(3), 199-204. doi:10.1007/s00420-005-0050-y

Bergström, G., Bodin, L., Hagberg, J., Lindh, T., Aronsson, G., & Josephson, M. (2009). Does sickness presenteeism have an impact on future general health? *International Archives Of Occupational And Environmental Health, 82*(10), 1179-1190. doi:10.1007/s00420-009-0433-6

Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E., . . . Jones, Z. M. (2016). mlr: Machine learning in r. *Journal of Machine Learning Research, 17*(170), 1-5. Retrieved from http://www.jmlr.org/

Bischl, B., Mersmann, O., Trautmann, H., & Weihs, C. (2012). Resampling methods for meta-model validation with recommendations for evolutionary computation. *Evolutionary Computation, 20*(2), 249-275. doi:10.1162/EVCO_a_00069

Blackmore, E. R., Stansfeld, S. A., Weller, I., Munce, S., Zagorski, B. M., & Stewart, D. E. (2007). Major depressive episodes and work stress: Results from a national population survey. *American Journal of Public Health, 97*(11), 2088-2093. doi:10.2105/AJPH.2006.104406

Bliese, P. D., Edwards, J. R., & Sonnentag, S. (2017). Stress and well-being at work: A century of empirical trends reflecting theoretical and societal influences. *Journal of Applied Psychology, 102*(3), 389-402. doi:10.1037/apl0000109

10.1037/apl0000109.supp (Supplemental)

Bogg, J., & Cooper, C. L. (1994). An examination of gender differences for job satisfaction, mental health, and occupational stress among senior U.K. civil servants. *International Journal of Stress Management, 1*(2), 159-172. doi:10.1007/BF01857609

Bonde, J. P. E. (2008). Psychosocial factors at work and risk of depression: A systematic review of the epidemiological evidence. *Occupational And Environmental Medicine, 65*(7), 438-445. doi:10.1136/oem.2007.038430

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. doi:10.1097/01.jom.0000175155.50789.98

Borritz, M., Christensen, K. B., Bültmann, U., Rugulies, R., Lund, T., Andersen, I., . . . Kristensen, T. S. (2010). Impact on burnout and psychosocial work characteristics on future long-term sickness absence, prospective results of the Danish PUMA study among human service workers. *Journal of Occupational and Environmental Medicine, 52*(10), 964-970. doi:10.1097/JOM.0b013e3181f12f95

Borritz, M., Rugulies, R., Christensen, K. B., Villadsen, E., & Kristensen, T. S. (2006). Burnout as a predictor of self-reported sickness absence among human service workers: prospective findings from three year follow up of the PUMA study. *Occupational And Environmental Medicine, 63*(2), 98-106. doi:10.1136/oem.2004.019364

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32. doi:10.1023/A:1010933404324

Brown, S. P. (1996). A meta-analysis and review of organizational research on job involvement. *Psychological Bulletin, 120*(2), 235-255. doi:10.1037/0033-2909.120.2.235

Brun, J.-P., & Lamarche, C. (2006). *Assessing the costs of work stress*. Retrieved from http://www.mtpinnacle.com/pdfs/Assessing-Costs-of-WorkStress.pdf

Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. (2012). *BIBB/BAuA-Erwerbstätigenbefragung 2012*. Retrieved from https://www.baua.de/DE/Themen/Arbeitswelt-und-Arbeitsschutz-im-Wandel/Arbeitsweltberichterstattung/Arbeitsbedingungen/BIBB-BAuA-2012.html

Burr, H., Albertsen, K., Rugulies, R., & Hannerz, H. (2010). Do dimensions from the Copenhagen Psychosocial Questionnaire predict vitality and mental health over and above the job strain and effort-reward imbalance models? *Scandinavian Journal Of Public Health, 38*(Suppl 3), 59-68. doi:10.1177/1403494809353436

Chandola, T. (2010). *Stress at work* Retrieved from https://www.thebritishacademy.ac.uk/sites/default/files/Stress%20at%20Work.pdf

Cleland, C., Kearns, A., Tannahill, C., & Ellaway, A. (2016). The impact of life events on adult physical and mental health and well-being: Longitudinal analysis using the GoWell health and well-being survey. *BMC Research Notes, 9*, 1-9. doi:10.1186/s13104-016-2278-x

Colligan, T. W., & Higgins, E. M. (2005). Workplace stress: Etiology and consequences. *Journal of Workplace Behavioral Health, 21*(2), 89-97. doi:10.1300/J490v21n02_07

Cooper, C., & Dewe, P. (2008). Well-being: Absenteeism, presenteeism, costs and challenges. *Occupational Medicine, 58*(8), 522-524. doi:10.1093/occmed/kqn124

Cox, T., Griffiths, A., & Rial-González, E. (2000). *Research on work-related stress.* Luxembourg: Office for Official Publications of the European Communities.

De Lange, A. H. (2005). *What about causality? Examining longitudinal relations between work characteristics and mental health.* (Doctoral dissertation), Radboud University, Nijmegen. Retrieved from http://hdl.handle.net/2066/19615

De Lange, A. H., Taris, T. W., Kompier, M. A. J., Houtman, I. L. D., & Bongers, P. M. (2004). The relationships between work characteristics and mental health: Examining normal, reversed and reciprocal relationships in a 4-wave study. *Work & Stress, 18*(2), 149-166. doi:10.1080/02678370412331270860

Deeney, C., & O'Sullivan, L. (2009). Work related psychosocial risks and musculoskeletal disorders: Potential risk factors, causation and evaluation methods. *Work: Journal of Prevention, Assessment & Rehabilitation, 34*(2), 239-248. doi:10.3233/WOR-2009-0921

Dicke, T., Marsh, H. W., Riley, P., Parker, P. D., Guo, J., & Horwood, M. (2018). Validating the Copenhagen Psychosocial Questionnaire (COPSOQ-II) using Set-ESEM:

Identifying psychosocial risk factors in a sample of school principals. *Frontiers In Psychology, 9*, 584-584. doi:10.3389/fpsyg.2018.00584

Dupret, E., Bocéréan, C., Teherani, M., Feltrin, M., & Pejtersen, J. H. (2012). Psychosocial risk assessment: French validation of the Copenhagen Psychosocial Questionnaire (COPSOQ). *Scandinavian Journal Of Public Health, 40*(5), 482-490. doi:10.1177/1403494812453888

Edwards, J. R., Baglioni, A. J., Jr., & Cooper, C. L. (1990). Examining the relationships among self-report measures of the Type A behavior pattern: The effects of dimensionality, measurement error, and differences in underlying constructs. *Journal of Applied Psychology, 75*(4), 440-454. doi:10.1037/0021-9010.75.4.440

Edwards, J. R., Baglioni, A. J., Jr., & Cooper, C. L. (1991). 'Examining the relationships among self-report measures of the Type A behavior pattern: The effects of dimensionality, measurement error, and differences in underlying constructs': Correction. *Journal of Applied Psychology, 76*(5), 657-657. doi:10.1037/0021-9010.76.5.657

Elovainio, M., Ferrie, J. E., Gimeno, D., Vogli, R. D., Shipley, M., Brunner, E. J., . . . Kivimäki, M. (2009). Organizational justice and sleeping problems: The Whitehall II study. *Psychosomatic Medicine, 71*(3), 334-340. doi:10.1097/PSY.0b013e3181960665

European Agency for Safety and Health at Work (EU-OSHA). (2013). *Priorities for occupational safety and health research in Europe: 2013-2020*. Retrieved from https://osha.europa.eu/de/tools-and-publications/publications/reports/priorities-for-occupational-safety-and-health-research-in-europe-2013-2020

Fernández-Delgado, M., Cernadas, E., Senén, B., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *The Journal of Machine Learning Research, 15*(1), 3133-3181. Retrieved from http://www.jmlr.org/

Fletcher, D., & Sarkar, M. (2013). Psychological resilience: A review and critique of definitions, concepts, and theory. *European Psychologist, 18*(1), 12-23. doi:10.1027/1016-9040/a000124

Freudenberger, H. J. (1974). Staff Burn-Out. *Journal of Social Issues, 30*(1), 159-165. doi:10.1111/j.1540-4560.1974.tb00706.x

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1), 1-22. doi:10.18637/jss.v033.i01

Fritz, C., Yankelevich, M., Zarubin, A., & Barger, P. (2010). Happy, healthy, and productive: The role of detachment from work during nonwork time. *Journal of Applied Psychology, 95*(5), 977-983. doi:10.1037/a0019462

Gemeinsame Deutsche Arbeitsschutzstrategie. (2017). *Arbeitsschutz in der Praxis. Empfehlungen zur Umsetzung der Gefährdungsbeurteilung psychischer Belastung.* Retrieved from https://www.gda-psyche.de/SharedDocs/Downloads/DE/empfehlungen-zur-umsetzung-der-gefaehrdungsbeurteilung-psychischer-belastung.pdf?__blob=publicationFile&v=1

Griffeth, R. W., Hom, P. W., & Gaertner, S. (2000). A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of Management, 26*(3), 463-488. doi:10.1177/014920630002600305

Hassard, J., Teoh, K. R. H., Visockaite, G., Dewe, P., & Cox, T. (2018). The cost of work-related stress to society: A systematic review. *Journal of Occupational Health Psychology, 23*(1), 1-17. doi:10.1037/ocp0000069

Head, J., Kivimäki, M., Martikainen, P., Vahtera, J., Ferrie, J. E., & Marmot, M. G. (2006). Influence of change in psychosocial work characteristics on sickness absence: The Whitehall II study. *Journal of Epidemiology and Community Health, 60*(1), 55-61. doi:10.1136/jech.2005.038752

Head, J., Kivimäki, M., Siegrist, J., Ferrie, J. E., Vahtera, J., Shipley, M. J., & Marmot, M. G. (2007). Effort-reward imbalance and relational injustice at work predict sickness absence: The Whitehall II study. *Journal Of Psychosomatic Research, 63*(4), 433-440. doi:10.1016/j.jpsychores.2007.06.021

Holman, D., Martinez-Inigo, D., & Totterdell, P. (2008). Emotional labour and employee well-being: An integrative review. In N. Ashkanasy & C. Cooper (Eds.), *Research companion to emotion in organizations* (pp. 301-315). Northampton: Edward Elgar Publishing.

Holmes, T. H., & Rahe, R. H. (1967). The Social Readjustment Rating Scale. *Journal Of Psychosomatic Research, 11*(2), 213-218. doi:10.1016/0022-3999(67)90010-4

Hom, P. W., Lee, T. W., Shaw, J. D., & Hausknecht, J. P. (2017). One hundred years of employee turnover theory and research. *Journal of Applied Psychology, 102*(3), 530-545. doi:10.1037/apl0000103

Houkes, I., Janssen, P. P. M., de Jonge, J., & Bakker, A. B. (2003). Specific determinants of intrinsic work motivation, emotional exhaustion and turnover intention: A multisample longitudinal study. *Journal of Occupational & Organizational Psychology, 76*(4), 427-450. doi:10.1348/096317903322591578

International Labour Organization (ILO). (2016). *Workplace stress. A collective challenge.* Retrieved from http://www.ilo.org/wcmsp5/groups/public/@ed_protect/@protrav/@safework/documents/publication/wcms_473267.pdf

Janetzke, H., & Ertel, M. (2017). *Psychosocial risk management in a European comparison.* Retrieved from https://www.baua.de/EN/Service/Publications/Report/F2339.pdf?__blob=publicationFile&v=3

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. doi:10.2105/AJPH.78.10.1336

Karasek, R., Brisson, C., Kawakami, N., Houtman, I., Bongers, P., & Amick, B. (1998). The Job Content Questionnaire (JCQ): An instrument for internationally comparative assessments of psychosocial job characteristics. *Journal of Occupational Health Psychology, 3*(4), 322-355. doi:10.1037/1076-8998.3.4.322

Karasek, R. A. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Science Quarterly, 24*(2), 285-308. doi:10.2307/2392498

Kawachi, I. (1999). Social capital and community effects on population and individual health. *Annals Of The New York Academy Of Sciences, 896*, 120-130. doi:10.1111/j.1749-6632.1999.tb08110.x

Kawachi, I. (2008). Globalization and workers' health. *Industrial Health, 46*(5), 421-423. doi:10.2486/indhealth.46.421

Kersten, N., & Backé, E. (2015). Occupational noise and myocardial infarction: Considerations on the interrelation of noise with job demands. *Noise & Health, 17*(75), 116-122. doi:10.4103/1463-1741.153403

Kivimäki, M., Ferrie, J. E., Head, J., Shipley, M. J., Vahtera, J., & Marmot, M. G. (2004). Organisational justice and change in justice as predictors of employee health: The Whitehall II study. *Journal of Epidemiology and Community Health, 58*(11), 931-937. doi:10.1136/jech.2003.019026

Kivimäki, M., Virtanen, M., Elovainio, M., Kouvonen, A., Väänänen, A., & Vahtera, J. (2006). Work stress in the etiology of coronary heart disease - a meta-analysis. *Scandinavian Journal Of Work, Environment & Health, 32*(6), 431-442. Retrieved from http://www.sjweh.fi/

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Ijcai, 14*(2), 1137-1145. Retrieved from https://www.ijcai.org/

Kristensen, T. R., Jensen, S. M., Kreiner, S., & Mikkelsen, S. (2010). Socioeconomic status and duration and pattern of sickness absence. A 1-year follow-up study of 2331 hospital employees. *BMC Public Health, 10*(1), 643-653. doi:10.1186/1471-2458-10-643

Kristensen, T. S., Hannerz, H., Høgh, A., & Borg, V. (2005). The Copenhagen Psychosocial Questionnaire -A tool for the assessment and improvement of the psychosocial work environment. *Scandinavian Journal Of Work, Environment & Health, 31*(6), 438-449. doi:10.5271/sjweh.948

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50*(1), 537. doi:10.1146/annurev.psych.50.1.537

Krüger, J. (2016). *Psychische Gesundheit in der Arbeitswelt: Beleuchtung*. Retrieved from https://www.baua.de/DE/Angebote/Publikationen/Berichte/F2353-4b.html

Laird, D. A. (1933). The influence of noise on production and fatigue, as related to pitch, sensation level, and steadiness of the noise. *Journal of Applied Psychology, 17*(3), 320-330. doi:10.1037/h0072423

Landsbergis, P. A., Dobson, M., Koutsouras, G., & Schnall, P. (2013). Job strain and ambulatory blood pressure: A meta-analysis and systematic review. *American Journal of Public Health, 103*(3), e61-e71. doi:10.2105/AJPH.2012.301153

Lang, M., Bischl, B., & Surmann, D. (2017). Tools for R to work on batch systems. *The Journal of Open Source Software, 2*(10). doi:10.21105/joss.00135

Larson, S. C. (1931). The shrinkage of the coefficient of multiple correlation. *Journal of Educational Psychology, 22*(1), 45-55. doi:10.1037/h0072400

Lee, A. R., Son, S.-M., & Kim, K. K. (2016). Information and communication technology overload and social networking service fatigue: A stress perspective. *Computers in Human Behavior, 55*(Part A), 51-61. doi:10.1016/j.chb.2015.08.011

Lee, K., Carswell, J. J., & Allen, N. J. (2000). A meta-analytic review of occupational commitment: Relations with person- and work-related variables. *Journal of Applied Psychology, 85*(5), 799-811. doi:10.1037/0021-9010.85.5.799

Leka, S., Griffiths, A., & Cox, T. (2003). *Work organisation & stress*. Retrieved from https://www.who.int/occupational_health/publications/en/oehstress.pdf

Leka, S., & Jain, A. (2010). *Health impact of psychosocial hazards at work: An overview.* Retrieved from http://apps.who.int/iris/bitstream/handle/10665/44428/?sequence=1

Lesuffleur, T., Chastang, J.-F., Sandret, N., & Niedhammer, I. (2014). Psychosocial factors at work and sickness absence: results from the French national SUMER survey. *American Journal Of Industrial Medicine, 57*(6), 695-708. doi:10.1002/ajim.22317

Li, J., Fu, H., Hu, Y., Shang, L., Wu, Y., Kristensen, T. S., . . . Hasselhorn, H. M. (2010). Psychosocial work environment and intention to leave the nursing profession: Results from the longitudinal Chinese NEXT study. *Scandinavian Journal Of Public Health, 38*(Suppl 3), 69-80. doi:10.1177/1403494809354361

Magnavita, N., & Garbarino, S. (2013). Is absence related to work stress? A repeated cross-sectional study on a special police force. *American Journal Of Industrial Medicine, 56*(7), 765-775. doi:10.1002/ajim.22155

Mai, K. M., Ellis, A. P. J., Christian, J. S., & Porter, C. O. L. H. (2016). Examining the effects of turnover intentions on organizational citizenship behaviors and deviance behaviors: A psychological contract approach. *Journal of Applied Psychology, 101*(8), 1067-1081. doi:10.1037/apl0000115

Marinaccio, A., Ferrante, P., Corfiati, M., Di Tecco, C., Rondinone, B. M., Bonafede, M., . . . Iavicoli, S. (2013). The relevance of socio-demographic and occupational variables for the assessment of work-related stress risk. *BMC Public Health, 13*, 1157-1157. doi:10.1186/1471-2458-13-1157

Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of Occupational Behavior, 2*(2), 99-113. doi:10.1002/job.4030020205

Mathieu, J. E., & Zajac, D. M. (1990). A review and meta-analysis of the antecedents, correlates, and consequences of organizational commitment. *Psychological Bulletin, 108*(2), 171-194. doi:10.1037/0033-2909.108.2.171

Meier, L. L., & Spector, P. E. (2013). Reciprocal effects of work stressors and counterproductive work behavior: A five-wave longitudinal study. *Journal of Applied Psychology, 98*(3), 529-539. doi:10.1037/a0031732

Melamed, S., Armon, G., Shirom, A., & Shapira, I. (2011). Exploring the reciprocal causal relationship between job strain and burnout: A longitudinal study of apparently healthy employed persons. *Stress and Health: Journal of the International Society for the Investigation of Stress, 27*(4), 272-281. doi:10.1002/smi.1356

Michie, S. (2002). Causes and management of stress at work. *Occupational And Environmental Medicine, 59*(1), 67-72. doi:10.1136/oem.59.1.67

Michie, S., & Williams, S. (2003). Reducing work related psychological ill health and sickness absence: a systematic literature review. *Occupational And Environmental Medicine, 60*(1), 3-9. doi:10.1136/oem.60.1.3

Molnar, C. (2019). *Interpretable machine learning. A Guide for Making Black Box Models Explainable.* Retrieved from https://christophm.github.io/interpretable-ml-book/

Moncada, S., Utzet, M., Molinero, E., Llorens, C., Moreno, N., Galtés, A., & Navarro, A. (2014). The Copenhagen Psychosocial Questionnaire II (COPSOQ II) in Spain - a tool for psychosocial risk assessment at the workplace. *American Journal Of Industrial Medicine, 57*(1), 97-107. doi:10.1002/ajim.22238

Niedhammer, I., Chastang, J.-F., Sultan-Taïeb, H., Vermeylen, G., & Parent-Thirion, A. (2013). Psychosocial work factors and sickness absence in 31 countries in Europe. *European Journal of Public Health, 23*(4), 622-629. doi:10.1093/eurpub/cks124

Nordin, M., Westerholm, P., Alfredsson, L., & Åkerstedt, T. (2012). Social support and sleep: Longitudinal relationships from the WOLF study. *Psychology, 3*(12A), 1223-1230. Retrieved from http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A824887&dswid=-7626

Nübling, M., & Hasselhorn, H. M. (2010). The Copenhagen Psychosocial Questionnaire in Germany: From the validation of the instrument to the formation of a job-specific database of psychosocial factors at work. *Scandinavian Journal Of Public Health, 38*(Suppl 3), 120-124. doi:10.1177/1403494809353652

Nübling, M., Lincke, H. J., Vomstein, M., & Haug, A. (2016). Befragung zu psychosozialen Faktoren am Arbeitsplatz. Deutsche Standard-Version des COPSOQ (Copenhagen

Psychosoical Questionnaire). Retrieved from https://www.copsoq.de/assets/COPSOQ-Standard-Fragebogen-FFAW.pdf

Nübling, M., Seidler, A., Garthus-Niegel, S., Latza, U., Wagner, M., Hegewald, J., . . . Letzel, S. (2013). The Gutenberg Health Study: Measuring psychosocial factors at work and predicting health and work-related outcomes with the ERI and the COPSOQ questionnaire. *BMC Public Health, 13*, 538-538. doi:10.1186/1471-2458-13-538

Nübling, M., Vomstein, M., Nolle, I., Lindner, A., Haug, A., & Lincke, H. J. (2018). Mitarbeiterbefragung zu psychosozialen Faktoren am Arbeitsplatz. Deutsche Standardversion des COPSOQ (Copenhagen Psychosocial Questionnaire). Retrieved from https://www.copsoq.de/assets/COPSOQ-Fragebogen-mit-Skalenzuordnung-040618-download.pdf

Pangert, B., Pauls, N., & Schüpbach, H. (2016). *Die Auswirkungen arbeitsbezogener erweiterter Erreichbarkeit auf Life-Domain-Balance und Gesundheit*. Retrieved from https://www.baua.de/DE/Angebote/Publikationen/Berichte/Gd76.pdf?__blob=publicationFile

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879-903. doi:10.1037/0021-9010.88.5.879

Purvanova, R. K., & Muros, J. P. (2010). Gender differences in burnout: A meta-analysis. *Journal of Vocational Behavior, 77*(2), 168-185. doi:10.1016/j.jvb.2010.04.006

R Core Team. (2018). R: A language and environment for statistical computing. Retrieved from https://www.R-project.org/

Reeuwijk, K. G., Robroek, S. J. W., Niessen, M. A. J., Kraaijenhagen, R. A., Vergouwe, Y., & Burdorf, A. (2015). The prognostic value of the Work Ability Index for sickness absence among office workers. *Plos One, 10*(5), 1-13. doi:10.1371/journal.pone.0126969

Richter, G., & Schütte, M. (2014). Porträts exemplarisch ausgewählter Analyseinstrumente und -verfahren. In Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (Ed.), *Gefährdungsbeurteilung psychischer Belastung* (pp. 187-263). Berlin: Erich Schmidt Verlag.

Rosário, S., Fonseca, J. A., Nienhaus, A., & Torres da Costa, J. (2016). Standardized assessment of psychosocial factors and their influence on medically confirmed health outcomes in workers: A systematic review. *Journal of Occupational Medicine & Toxicology, 11*, 1-13. doi:10.1186/s12995-016-0106-9

Rosengren, A., Hawken, S., Ôunpuu, S., Sliwa, K., Zubaid, M., Almahmeed, W. A., . . . Yusuf, S. (2004). Association of psychosocial risk factors with risk of acute myocardial infarction in 11119 cases and 13648 controls from 52 countries (the INTERHEART study): Case-control study. *The Lancet, 364*(9438), 953-962. doi:10.1016/S0140-6736(04)17019-0

Rosseel, Y. (2012). Lavaan: An R package for structural equatation modeling and more. Version 0.5-12 (BETA). *Journal of Statistical Software, 48*(2), 1-36. doi:10.18637/jss.v048.i02

Rugulies, R., Aust, B., & Madsen, I. E. H. (2017). Effort–reward imbalance at work and risk of depressive disorders. A systematic review and meta-analysis of prospective cohort studies. *Scandinavian Journal Of Work, Environment & Health, 43*(4), 294-306. doi:10.5271/sjweh.3632

Rugulies, R., Aust, B., & Pejtersen, J. H. (2010). Do psychosocial work environment factors measured with scales from the Copenhagen Psychosocial Questionnaire predict register-based sickness absence of 3 weeks or more in Denmark? *Scandinavian Journal Of Public Health, 38*(3 Suppl), 42-50. doi:10.1177/1403494809346873

Rugulies, R., Bültmann, U., Aust, B., & Burr, H. (2006). Psychosocial work environment and incidence of severe depressive symptoms: Prospective findings from a 5-year follow-up of the Danish work environment cohort study. *American Journal Of Epidemiology, 163*(10), 877-887. doi:10.1093/aje/kwj119

Rugulies, R., Christensen, K. B., Borritz, M., Villadsen, E., Bültmann, U., & Kristensen, T. S. (2007). The contribution of the psychosocial work environment to sickness absence in human service workers: Results of a 3-year follow-up study. *Work & Stress, 21*(4), 293-311. doi:10.1080/02678370701747549

Rugulies, R., Norborg, M., Sørensen, T. S., Knudsen, L. E., & Burr, H. (2009). Effort-reward imbalance at work and risk of sleep disturbances. Cross-sectional and prospective results from the Danish Work Environment Cohort Study. *Journal Of Psychosomatic Research, 66*(1), 75-83. doi:10.1016/j.jpsychores.2008.05.005

Seeboth, A., & Mõttus, R. (2018). Successful explanations start with accurate descriptions: Questionnaire items as personality markers for more accurate predictions. *European Journal of Personality*. doi:10.1002/per.2147

Setti, I., d'Errico, A., Di Cuonzo, D., Fiabane, E., & Argentero, P. (2017). Validation and psychometric properties of the Italian Copenhagen Psychosocial Questionnaire II - short version. *BPA - Applied Psychology Bulletin (Bollettino di Psicologia Applicata), 65*(280), 48-57. Retrieved from https://www.giuntios.it/catalogo/rivista/bollettino-di-psicologia-applicata

Siegrist, J. (1996). Adverse health effects of high-effort/low-reward conditions. *Journal of Occupational Health Psychology, 1*(1), 27-41. doi:10.1037/1076-8998.1.1.27

Slany, C., Schütte, S., Chastang, J.-F., Parent-Thirion, A., Vermeylen, G., & Niedhammer, I. (2014). Psychosocial work factors and long sickness absence in Europe. *International Journal Of Occupational And Environmental Health, 20*(1), 16-25. doi:10.1179/2049396713Y.0000000048

Smulders, P. G. W., & Nijhuis, F. J. N. (1999). The Job Demands-Job Control model and absence behaviour: Results of a 3-year longitudinal study. *Work & Stress, 13*(2), 115-131. doi:10.1080/026783799296093

Sonnentag, S., Binnewies, C., & Mojza, E. J. (2010). Staying well and engaged when demands are high: The role of psychological detachment. *Journal of Applied Psychology, 95*(5), 965-976. doi:10.1037/a0020032

Sparks, K., Cooper, C., Fried, Y., & Shirom, A. (1997). The effects of hours of work on health: A meta-analytic review. *Journal of Occupational & Organizational Psychology, 70*(4), 391-408. doi:10.1111/j.2044-8325.1997.tb00656.x

Spurgeon, A., Harrington, J. M., & Cooper, C. L. (1997). Health and safety problems associated with long working hours: A review of the current position. *Occupational And Environmental Medicine, 54*(6), 367-375. doi:10.1136/oem.54.6.367

Stansfeld, S., & Candy, B. (2006). Psychosocial work environment and mental health - a meta-analytic review. *Scandinavian Journal Of Work, Environment & Health, 32*(6), 443-462. Retrieved from http://www.sjweh.fi/

Stauder, A., Nistor, K., Zakor, T., Szabó, A., Nistor, A., Ádám, S., & Thege, B. K. (2017). Quantifying multiple work-related psychosocial risk factors: Proposal for a composite indicator based on the COPSOQ II. *International Journal Of Behavioral Medicine, 24*(6), 915-926. doi:10.1007/s12529-017-9651-6

Suadicani, P., Bonde, J. P., Olesen, K., & Gyntelberg, F. (2013). Job satisfaction and intention to quit the job. *Occupational Medicine, 63*(2), 96-102. doi:10.1093/occmed/kqs233

Sverke, M., Hellgren, J., & Näswall, K. (2002). No security: A meta-analysis and review of job insecurity and its consequences. *Journal of Occupational Health Psychology, 7*(3), 242-264. doi:10.1037/1076-8998.7.3.242

Theorell, T., & Hasselhorn, H. M. (2005). On cross-sectional questionnaire studies of relationships between psychosocial conditions at work and health - are they reliable? *International Archives Of Occupational And Environmental Health, 78*(7), 517-522. Retrieved from https://link.springer.com/journal/420

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 267-288. doi:10.1111/j.2517-6161.1996.tb02080.x

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*(5), 859-883. doi:10.1037/0033-2909.133.5.859

Van den Berg, T. I. J., Elders, L. A. M., de Zwart, B. C. H., & Burdorf, A. (2009). The effects of work-related and individual factors on the Work Ability Index: A systematic review. *Occupational And Environmental Medicine, 66*(4), 211-220. doi:10.1136/oem.2008.039883

Waldman, D. A., Carter, M. Z., & Hom, P. W. (2015). A multilevel investigation of leadership and turnover behavior. *Journal of Management, 41*(6), 1724-1744. doi:10.1177/0149206312460679

Wang, J. L., Lesage, A., Schmitz, N., & Drapeau, A. (2008). The relationship between work stress and mental disorders in men and women: Findings from a population-based study. *Journal of Epidemiology and Community Health, 62*(1), 42-47. doi:10.1136/jech.2006.050591

Wang, M., Liu, S., Liao, H., Gong, Y., Kammeyer-Mueller, J., & Shi, J. (2013). Can't get it out of my mind: Employee rumination after customer mistreatment and negative mood in the next morning. *Journal of Applied Psychology, 98*(6), 989-1004. doi:10.1037/a0033656

Waongenngarm, P., Areerak, K., & Janwantanakul, P. (2018). The effects of breaks on low back pain, discomfort, and work productivity in office workers: A systematic review of randomized and non-randomized controlled trials. *Applied ergonomics, 68*, 230-239. doi:10.1016/j.apergo.2017.12.003

Wasti, S. A. (2003). Organizational commitment, turnover intentions and the influence of cultural values. *Journal of Occupational and Organizational Psychology, 76*(3), 303-321. doi:10.1348/096317903769647193

Wendsche, J., & Lohmann-Haislah, A. (2016). *Psychische Gesundheit in der Arbeitswelt: Detachment*. Retrieved from https://www.baua.de/DE/Angebote/Publikationen/Berichte/F2353-3c.html

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag.

World Health Organization (WHO). (2013). *Mental Health Action Plan 2013-2020*. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/89966/9789241506021_eng.pdf;jsessionid=ADC4DD0EFB5D32FF9AF675754AE0555B?sequence=1

Wright, M. N., & Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software, 77*(1), 1-17. doi:10.18637/jss.v077.i01

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100-1122. doi:10.1177/1745691617693393

Yin, P., & Fan, X. (2001). Estimating R-squared shrinkage in multiple regression: A comparison of different analytical methods. *Journal of Experimental Education, 69*(2), 203-224. doi:10.1080/00220970109600656

Zapf, D., Dormann, C., & Frese, M. (1996). Longitudinal studies in organizational stress research: A review of the literature with reference to methodological issues. *Journal of Occupational Health Psychology, 1*(2), 145-169. doi:10.1037/1076-8998.1.2.145

Zhang, C., & Conrad, F. (2014). Speeding in web surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Methods, 8*(2), 127-135. doi:10.18148/srm/2014.v8i2.5453

Zijlstra, F. R. H., Roe, R. A., Leonora, A. B., & Krediet, I. (1999). Temporal factors in mental work: Effects of interrupted activities. *Journal of Occupational & Organizational Psychology, 72*(2), 163-185. doi:10.1348/096317999166581

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67*(2), 301-320. doi:10.1111/j.1467-9868.2005.00503.x

Zweck, B. M. (2017). *Entwicklung und Validierung eines Fragebogens zur psychischen Gefährdungsbeurteilung.* (Doctoral dissertation), Ludwig-Maximilians-Universität München. Retrieved from https://edoc.ub.uni-muenchen.de/21057/

Zweck, B. M., & Pargent, F. (2019). *Munich Employee Health Questionnaire (MEHQ): Primärdaten der Erhebungswellen 1 und 2 (2016) und 3 (2017) (Version 1)* [Daten und Dokumentation]. Trier: Psychologisches Datenarchiv PsychData des Leibniz-

Zentrums für Psychologische Information und Dokumentation ZPID.
doi.org/10.5160/psychdata.zkba17mu29

Appendix A

English Version of the Munich Employee Health Questionnaire

English translation of the German Munich Employee Health Questionnaire (MEHQ) version that was used in wave 3 (collected from November 30th to December 13th in 2017). Variables are named with "_3" to indicate the third wave. Data of wave 3 together with previous waves 1 and 2 can be requested for scientific use at DOI: 10.5160/psychdata.zkba17mu29

Procedure of the translation process

1) English translation of the German MEHQ by an American citizen with fluent German skills. She was instructed to use everyday language to make sure that the items are understood by people irrespective of their educational background → version 1.
2) Check of the American English translation (version 1) by an employee of a market and research institute located in London to ensure that the questionnaire is understood as well by people using British English → version 2.
3) Check of version 2 by our research team to ensure that the translated items are still comparable to the original German ones → few corrections → version 3.
4) Check of version 3 by another employee of the market and research institute located in London who is professionally in the field of questionnaire design → version 4 – final version.

All variables with "S" or "V" in their name refer to demographic and additional occupational information. The core module of the MEHQ starts with section A. Items of the core module presented in Appendix A with _3 (section A, B, C, D, E, F, G, H, OM, Z) were similar during all three waves.

Variables with “rec” are recoded variables. Positive formulated variables were recoded so that the highest value expresses the highest negative burden (exception: resilience items H2, H7, H9).

In order to reach participants from wave 1 or wave 2 again in the longitudinal setting (wave 3) we adapted some wordings and filter questions. These changes are highlighted in the questionnaire. Note that variables of the questionnaire are sometimes not displayed in ascending order due the questionnaire construction process (for more information, please see Zweck, 2017).

To facilitate online programming, we also included scripter notes like for example “single coded” in the questionnaire.

Please consider that the English version of the MEHQ has not been validated yet. The validated German version of the MEHQ can be found in the online supplemental materials.

GENERAL INTRODUCTION MUNICH EMPLOYEE HEALTH QUESTIONNAIRE (MEHQ)

Welcome to the survey, "Health at the Work!"
You might remember your participation from last year. First and foremost, thank you for that. The issue of "health at work" is still very relevant. Therefore, it would be great if you could once again support us by answering the following questions. The survey will take about 15 minutes of your time. You can check the progress bar at the top of your screen throughout the survey to see how far along you are. You will be asked about different aspects of your work as well as health issues. The purpose of the survey is to identify areas of your work that may affect your health, which is why your personal opinion is very important.
Only a summary of the results will be published. Therefore, no conclusions can be drawn from individual participants. Of course, the data analysis will be kept strictly confidential and will remain anonymous.
Thank you for your participation!
To go to the next screen and start the survey, please click on the lower right arrow.

PERSONAL AND WORKPLACE INFORMATION

To begin, we ask that you please answer a few details about yourself and your workplace. Please continue by clicking on the right arrow.

SP1_a_3: What is your sex?		**Single coded**
1	Female	
2	Male	

SP2_a_3: What is your age?			**Single coded**
1	Under 15 years	→ GO TO SCREEN OUT	
2	15–17 years		
3	18–24 years		
4	25–34 years		
5	35–44 years		
6	45–54 years		
7	55–64 years		
8	65 years and above		

SP3a_3: Which school-leaving qualification did you achieve?		**Single coded**
1	Qualification from a special school	
2	Lower secondary school-leaving qualification	
3	Qualification from a polytechnic secondary school (8th grade POS)	
4	Specialized lower secondary school-leaving qualification	
5	Intermediate secondary school-leaving qualification	
6	Qualification from a polytechnic secondary school (10th grade POS)	
7	Specialized upper secondary school-leaving qualification	
8	Upper secondary school-leaving qualification	
9	Qualification from an extended secondary school (EOS)	
10	Foreign school-leaving qualification	
11	Other school-leaving qualification	
12	No school-leaving qualification	

FILTER: ASK ONLY IF **SP3a_3**=10		
SP3b_3: Please try to assign your foreign school-leaving qualification according to the (main) German qualifications.		**Single coded**
1	Lower secondary school-leaving qualification	
2	Intermediate secondary school-leaving qualification	
3	Upper secondary school-leaving qualification	

V1_3_Code1, V1_3_Code2, V1_3_Code3: Has any personal or occupational change occurred within the last 12 months? [1]		**Open**
Please type your answer in the text box.		
V1_3_Codes = 1	Don't know	

[1] Question only used in wave 3.

SA8_3: Are you currently employed, i.e. having a paid job of any kind?			**Single coded**
1	Yes		
2	No [1]	→ GO TO **SA6_3**	

[1] In wave 1 and wave 2, these persons were screened out.

SA1_3: To which industry does the company you currently work for belong?		**Single coded**
1	Agriculture, forestry, horticulture, fisheries	
2	Industrial sector	
3	Crafts	
4	Trade and retail	
5	Business-related services	
6	Personal services	
7	Other industry: **Open (SA1_96_3)*	
8	Don't know	
9	No answer	

SA9_3: Approximately how many employees are engaged in your company? Please do not refer to your company as a whole but to your local place of work.		**Single coded**
1	1–9 employees	
2	10–49 employees	
3	50–249 employees	
4	250 or more employees	
5	Don't know	
6	No answer	

SA7_3: What is your current occupation? Please state the exact title in the text box. For example, do not write electrician, but electrical fitter; not saleswoman, but saleswoman for shoes.	**Open**

SA10_3: What type of education or training is usually required for this type of work?		**Single coded**
1	No completed vocational training is required	
2	Completed vocational training	
3	Advanced training (e.g. master craftsperson, skilled worker or any other type of an advanced training degree)	
4	University degree or degree from another institution of higher education	
5	Don't know	
6	No answer	

SA3a_3: What is your current occupational status?		**Single coded**
1	Blue-collar worker *(e.g. a person whose work tasks are mainly characterized by physical work)* [1]	
2	White-collar worker *(e.g. a person whose work tasks are mainly characterized by mental work)* [1]	
3	Civil servant	
4	Self-employed [2]	
5	Freelancer [2]	
6	Working for self-employed relative [2]	
7	In training	

[1] We added this explanation only in the English translation due to a comment of one of our English translators.
[2] In wave 1 and 2, these persons were screened out.

SA4_3: How many years have you been working for your current employer?		**Single coded**
1	Less than 1 year	
2	From 1 year up to, but less than 3 years	
3	From 3 years up to, but less than 5 years	
4	From 5 years up to, but less than 10 years	
5	10 years or more	

SA5_3: On average, how many hours do you work per week, including overtime?	**Numeric**
Min = 1 | Max = 150	
Please type your answer in the text box.	

SA11_3: Are you working in shifts or in a similar type of working arrangement where your start and finish times change?		**Single coded**
1	Yes – working in shifts.	
2	Not shift work but my start and finish times do vary	
3	Neither – *I have regular (contracted) hours* [1]	

[1] We added this explanation only in the English translation due to a comment of one of our English translators.

SA6_3: Do you have leadership responsibility, i.e. are you the supervisor of one or more employees? If you are not working currently, please think about your last job. [1]		**Single coded**
1	Yes	
2	No	

[1] We added "If you are not currently working (…)" only in wave 3.

MAIN BLOCK

Now we will begin with the main part of the survey. Here we will ask for your assessment regarding different areas of your work. If you are not currently working, please base your answer on your last job. To start the main section, please click on the right arrow. [1]

[1] We added "If you are not currently working (…)" only in wave 3.

A: Work Environment					**Matrix**
Please indicate how strongly you feel burdened at work through the following aspects. A '1' means "not burdened at all," whereas a '5' indicates "very strongly burdened". The numbers in between can be used to scale your assessment. Please take the past 12 months up to the present day into account. If you have been at your current position for a shorter duration, please take only this period of time into consideration. How strongly do you feel affected by …					
	Not at all (1)	Low (2)	Average (3)	Strongly (4)	Very strongly (5)
A1_3 ... unfavorable lighting conditions at your workplace, such as too bright or too weak lighting?	☐	☐	☐	☐	☐
A2_3 ... uncomfortable working posture or strenuous movements?	☐	☐	☐	☐	☐
A3_3 ... inadequate work equipment such as bad tools, machines, computers?	☐	☐	☐	☐	☐
A5_3 ... unpleasant temperatures such as cold or heat?	☐	☐	☐	☐	☐
A6_3 ... noise such as noises from coworkers, a construction site, loud machines?	☐	☐	☐	☐	☐

B: Work requirements					**Matrix**
When thinking about the past 12 months at work, how often ...					
	(Almost) Never (1)	Rarely (2)	Sometimes (3)	Often (4)	(Almost) Always (5)
B1_3 ... were you interrupted at work, e.g. by coworkers, calls, e-mails?	☐	☐	☐	☐	☐
B5_3 ... did you have to work under uncomfortably strong pressure, such as deadlines or performance pressure?	☐	☐	☐	☐	☐
B6_3 ... did you have to hide your feelings?	☐	☐	☐	☐	☐
B7_3 ... did you have to neglect personal stuff because of your work?	☐	☐	☐	☐	☐
B11_3 ... did so much happen all at once that you could hardly handle it?	☐	☐	☐	☐	☐
B12_3 ... did you reach the limits of your performance abilities?	☐	☐	☐	☐	☐
B13_3 ... did you not have enough time to complete all your tasks?	☐	☐	☐	☐	☐
B14_3 ... did your employer contact you about work issues outside of your normal working hours?	☐	☐	☐	☐	☐
B15_3 ... did you work more than 6 hours without taking a break (meant here are breaks longer than 15 minutes)?	☐	☐	☐	☐	☐
B20_3 ... did unforeseen things happen to which you had to react quickly?	☐	☐	☐	☐	☐

C: Responsibility at work					**Matrix**
Now we will ask you to indicate to what extent you agree with the following statements regarding your work. A "1" stands for "strongly disagree", whereas a "5" means "strongly agree". The numbers in between can be used to scale your assessment. To what extent do you agree with the following statements? Please take the past 12 months to present day into account.					
	Strongly disagree (1)	Disagree (2)	Neither agree nor disagree (3)	Agree (4)	Strongly agree (5)
C1_3 I am told in detail how to do my job.	☐	☐	☐	☐	☐
C2_3 In my work, I am easily replaceable.	☐	☐	☐	☐	☐
C4_3_rec My job is designed so that I can perform a complete work task/producing an entire product from start to finish.	☐	☐	☐	☐	☐
C5_3_rec In my work, I can decide for myself when to do which tasks.	☐	☐	☐	☐	☐
C6_3_rec I can see from the results whether my work was done well or not.	☐	☐	☐	☐	☐

D: Qualification					**Matrix**
To what extent do you agree with the following statements regarding your work? Please take the past 12 months to present day into account.					
	Strongly disagree (1)	Disagree (2)	Neither agree nor disagree (3)	Agree (4)	Strongly agree (5)
D1_3_rec In my work, I can fully apply what I am good at.	☐	☐	☐	☐	☐
D2_3_rec My work is varied.	☐	☐	☐	☐	☐
D3_3_rec I can always learn new things in my work.	☐	☐	☐	☐	☐
D5_3 I'm afraid I won't be able to keep up with all of the new technology at work.	☐	☐	☐	☐	☐

E: Role clarity and information					**Matrix**
To what extent do you agree with the following statements regarding your work? Please take the past 12 months to present day into account.					
	Strongly disagree (1)	Disagree (2)	Neither agree nor disagree (3)	Agree (4)	Strongly agree (5)
E1_3 At work, I feel flooded with unimportant information.	☐	☐	☐	☐	☐
E2_3 In my work, unclear demands are made.	☐	☐	☐	☐	☐
E4_3 I have to do things in my work that I think are unnecessary.	☐	☐	☐	☐	☐
E5_3_rec I have all the information I need to do my job well.	☐	☐	☐	☐	☐
E6_3_rec I know exactly which things fall into my area of responsibility.	☐	☐	☐	☐	☐
E8_3_rec I am clear what is expected of me at work.	☐	☐	☐	☐	☐

F: Workplace in general					**Matrix**
To what extent do you agree with the following statements? Please take the past 12 months to present day into account.					
	Strongly disagree (1)	Disagree (2)	Neither agree nor disagree (3)	Agree (4)	Strongly agree (5)
F1_3 In my work, I do things that are not accepted by some parts of society.	☐	☐	☐	☐	☐
F5_3 I think about changing my job.	☐	☐	☐	☐	☐
F6_3 I am worried that I will lose my job.	☐	☐	☐	☐	☐
F7_3_rec I am currently working in my dream job.	☐	☐	☐	☐	☐
F9_3_rec I find my work to be meaningful.	☐	☐	☐	☐	☐
F12_3_rec I am proud to work for my employer.	☐	☐	☐	☐	☐
F13_3 I'm worried that I might be transferred to another position or another work location against my will.	☐	☐	☐	☐	☐
F14_3_rec I enjoy my work.	☐	☐	☐	☐	☐
F15_3 At work there are situations that burden me emotionally.	☐	☐	☐	☐	☐
F16_3_rec What I do at work is valued.	☐	☐	☐	☐	☐

G1_3: Colleagues and supervisors	**Single coded**
Now we would like to ask you a few questions about your colleagues and supervisors. To narrow down the questions, please indicate which of the following applies to you:	
1	I have colleagues and supervisors.
2	I have colleagues, but no supervisors.
3	I have supervisors, but no colleagues.
4	I have neither colleagues nor supervisors.

FILTER: ASK ONLY section G2 IF **G1_3**=1,2; if **G1_3**= 3: ask only G2_5_3_rec; if **G1_3**=4: skip G-section and proceed with Z1_3 (for those with leadership responsibility; SA6_3=1) and with P1_3 for those without leadership responsibility (SA6_3=2)

G2a: Colleagues and sense of community | **Matrix**

To what extent do you agree with the following statements regarding your work?
Please take the past 12 months to present day into account.

	Strongly disagree (1)	Disagree (2)	Neither agree nor disagree (3)	Agree (4)	Strongly agree (5)
G2_1_3_rec The work-related cooperation between my direct colleagues and me is very good.	☐	☐	☐	☐	☐
G2_2_3_rec I can talk openly with my direct colleagues about work problems.	☐	☐	☐	☐	☐
G2_3_3 Sometimes intense arguments between my colleagues and me can occur.	☐	☐	☐	☐	☐
G2_4_3_rec I get support from my direct colleagues when I need it.	☐	☐	☐	☐	☐
G2_5_3_rec In my job, I feel part of a community.	☐	☐	☐	☐	☐
G2_6_3 There are situations in which my direct colleagues treat me in a dismissive way.	☐	☐	☐	☐	☐

FILTER: ASK ONLY IF **G1_3**=1,3

G2b: Supervisors | **Matrix**

To what extent do you agree with the following statements?
Please take the past 12 months to present day into account.

	Strongly disagree (1)	Disagree (2)	Neither agree nor disagree (3)	Agree (4)	Strongly agree (5)
G2_8_3_rec My supervisor gives me clear feedback regarding my work.	☐	☐	☐	☐	☐
G2_9_3 My supervisor acts as if my job satisfaction is unimportant to him or her.	☐	☐	☐	☐	☐
G2_10_3_rec In case of disagreements or conflicts, my supervisor strives try to find a good solution.	☐	☐	☐	☐	☐
G2_11_3_rec I get support from my supervisor when I need it.	☐	☐	☐	☐	☐
G2_7_3 There are situations in which my supervisor treats me in a dismissive way.	☐	☐	☐	☐	☐
G2_13_3 I feel as if I am being kept under surveillance by my supervisor.	☐	☐	☐	☐	☐
G2_15_3 My supervisor treats me unfairly.	☐	☐	☐	☐	☐
G2_16_3_rec My supervisor is always there for me if I need to talk.	☐	☐	☐	☐	☐
G2_17_3_rec When making decisions, my supervisor factors in my opinion.	☐	☐	☐	☐	☐
G2_19_3 Sometimes intense arguments between my supervisors and me can occur.	☐	☐	☐	☐	☐

FILTER: ASK ONLY IF **SA6_3**=1

Z1: Management tasks 1 | **Matrix**

You've indicated that you have leadership responsibility.
If you think back on the past 12 months at work, how often ...

	(Almost) Never (1)	Rarely (2)	Sometimes (3)	Often (4)	(Almost) Always (5)
Z1_3_3 ... was it hard to find the time to do both your management tasks and your 'normal' daily tasks?	☐	☐	☐	☐	☐
Z1_5_3 ...did you have to act against your own convictions?	☐	☐	☐	☐	☐
Z1_6_3 ...have you had concerns that your employees might be better qualified than you are?	☐	☐	☐	☐	☐
Z1_7_3 ...did you feel put in a "sandwich position," meaning you had to meet the expectations of both your supervisor and your employees at the same time?	☐	☐	☐	☐	☐
Z1_8_3 ...did you have to give bad news to your employees?	☐	☐	☐	☐	☐

FILTER: ASK ONLY IF **SA6_3**=1

Z2: Management tasks 2 | **Matrix**

Relative to the last 12 months until today: How much do you agree with the following statements?

	Strongly disagree (1)	Disagree (2)	Neither agree nor disagree (3)	Agree (4)	Strongly agree (5)
Z2_1_3_rec I feel comfortable in a management role.	☐	☐	☐	☐	☐
Z2_2_3_rec I can rely on my employees.	☐	☐	☐	☐	☐
Z2_3_3_rec I feel up to the challenge of being in a management position.	☐	☐	☐	☐	☐
Z2_4_3_rec I feel accepted by my employees.	☐	☐	☐	☐	☐
Z2_5_3 I worry about the high level of responsibility as a manager.	☐	☐	☐	☐	☐
Z2_6_3 I worry about the demands of my employees, such as their salary demands.	☐	☐	☐	☐	☐

P1_3: Personal opinion | **Open**

In general terms, when you think about your work, which aspects make you most unhappy? These may be aspects that we have already asked about in this questionnaire, but they can also be entirely new aspects.

Please type your answer in the text box.

-98	Don't know

H: Personal Background					**Matrix**
Now we will ask about your personal background. Unlike the other questions, please do not just think about your work, but as well about your private situation. To what extent do you agree with the following statements? Please take the past 12 months to present day into account again.					
	Strongly disagree (1)	Disagree (2)	Neither agree nor disagree (3)	Agree (4)	Strongly agree (5)
H1_3 I only stop when I have done everything perfectly.	☐	☐	☐	☐	☐
H2_3 My friends and family are a great support to me when I am not doing well.	☐	☐	☐	☐	☐
H3_3 There have been private things that have bothered me a lot, such as marriage, building a house, illness, separation in a relationship.	☐	☐	☐	☐	☐
H4_3 It's hard for me to say no.	☐	☐	☐	☐	☐
H5_3_rec In my free time I can easily switch off from work.	☐	☐	☐	☐	☐
H6_3 At my work, I'm very afraid of not being good enough.	☐	☐	☐	☐	☐
H7_3 I can work well under pressure.	☐	☐	☐	☐	☐
H8_3 I'm reluctant to give up.	☐	☐	☐	☐	☐
H9_3 I have hobbies that I enjoy.	☐	☐	☐	☐	☐
H10_3 I go to work, although due to my state of health I should have called in sick.	☐	☐	☐	☐	☐
H11_3 I always feel that I have not done enough, even though I've been trying very hard.	☐	☐	☐	☐	☐

HEALTH

In the final part of the survey, we ask you to answer some questions about your health. As already mentioned, your statements are kept strictly confidential and evaluated anonymously.
Please click on the right arrow to start.

O1_3_rec: General health											**Left-right matrix**
When you think about the last 12 months, how do you rate your physical and mental health? 1 point means "very poor health," 10 points mean "very good health." The levels in between allow you to scale your assessment.											
Very poor health	1	2	3	4	5	6	7	8	9	10	Very good health

O2new_3: How many days have you been unable to work in the past 12 months due to health problems–physical or mental? If you are not sure of the exact number, just estimate.	**Numeric**
Max = 365	
-99	No answer [1]

[1] "No answer" was included only in wave 3 (otherwise, participants without a job in could not have answered this question).

OM: General condition					**Matrix**
Now we will ask you a few questions about your general well-being in and out of work. If you think back to the last 12 months: How often ...					
	(Almost) Never (1)	Rarely (2)	Sometimes (3)	Often (4)	(Almost) Always (5)
OM2_3 ... were you sad?	☐	☐	☐	☐	☐
OM3_3 ... did you have stomach or abdominal pain?	☐	☐	☐	☐	☐
OM5_3 ... could you not fall asleep properly or sleep properly?	☐	☐	☐	☐	☐
OM7_3 ... did you feel really annoyed?	☐	☐	☐	☐	☐
OM8_3 ... did you feel tired or physically exhausted?	☐	☐	☐	☐	☐
OM9_3 ... did you feel like not going to work at all?	☐	☐	☐	☐	☐
OM10_3 ... were you dizzy?	☐	☐	☐	☐	☐
OM11_3 ... was it hard for you to concentrate on something?	☐	☐	☐	☐	☐
OM12_3 ... did you feel the need to rest and withdraw?	☐	☐	☐	☐	☐
OM14_3 ... did you have back pain or a muscle tenseness?	☐	☐	☐	☐	☐
OM15_3 ... did you have the feeling of not being able to calm down?	☐	☐	☐	☐	☐
OM16_3 ... did you have a headache?	☐	☐	☐	☐	☐
OM18_3 ... did you think: "I can't do this anymore?"	☐	☐	☐	☐	☐

O3_3: Was work stress responsible for the fact that you were temporarily unable to work within the last 12 months?	**Single coded**
1	Yes, exclusively
2	Yes, partly
3	No not at all
4	Don't know

P2_3: When you think about the general mental and physical health at work, is there anything your employer could improve? If so, what would that be?	**Open**
Please type your answer in the text box.	
-98	Don't know

COMMENTS

P4_3: You are almost at the end of the survey. What is your opinion? Is this questionnaire well suited for use in your industry?	**Single coded**
1	Yes
2	No, because: **Open (P4_2_3)*
3	Don't know

P5_3: And is this questionnaire well suited for the assessment of your current occupation?		**Single coded**
1	Yes	
2	No, because: **Open (P5_2_3)*	
3	Don't know	

P6_3: Do you think that this questionnaire covers all relevant aspects of your work that might impact on your health? [1]		**Single coded**
1	Yes	
2	No, because: **Open (P6_2_3)*	
3	Don't know	

[1]We changed the wording in comparison to the German version a little bit due to a comment of one of our English translators.

P3_3: Do you have any other general comments about this questionnaire? If not, please click on the right arrow to finish the survey.	**Open**

CLOSING

Not back

Thank you very much for your participation!

You can now close the browser window.

Appendix B

Assignment of items to domains

Table B1

Assignment of items to domains

Domain/Aspect	**Items**	**Number of items**
Work requirements	B5, B6, B7, B11, B12, B13, B15, B20	8
Leadership style	G2_8_rec, G2_10_rec, G2_11_rec, G2_16_rec, G2_17_rec	5
Conflicts and discrimination	F1, F13, G2_3, G2_6, G2_7, G2_9, G2_13, G2_15, G2_19	9
Work environment	A1, A2, A3, A5, A6	5
Information	E5_rec, E6_rec, E8_rec	3
Social support	G2_1_rec, G2_2_rec, G2_4_rec, G2_5_rec	4
Uselessness	E1, E2, E4	3
Scope	C1, C2, C5_rec	3
Intrinsic motivation	D2_rec, D3_rec, F7_rec, F9_rec, F14_rec	5
Resilience	H1, H2, H7, H8, H9	5
Risk type	H4, H5_rec, H6, H10, H11	5
Index	B1, B14, C4_rec, C6_rec, D1_rec, D5, F6, F12_rec, F15, F16_rec, H3	11
Strains	OM2, OM3, OM5, OM7, OM8, OM9, OM10, OM11, OM12, OM14, OM15, OM16, OM18	13
General health	O1_rec	1
Sickness absence	O2	1
Turnover intention	F5	1

Note. Items marked with "_rec" are items that have been recoded according to negative psychosocial burden. Exception: Resilience (not recoded because of statistical reasons see Zweck, 2017).