Aus dem Institut und der Poliklinik für Arbeits-, Sozial- und Umweltmedizin der
Ludwig-Maximilians-Universität zu München
Direktor: Prof. Dr. med. Dennis Nowak

Internalized goal conflict at work influences job satisfaction, work engagement, and depressive symptoms: Evidence from a panel study of early-career physicians

Dissertation
zum Erwerb des Doktorgrades der Medizin
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von Matthias Guth
aus München

2022

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Berichterstatter:	Prof. Dr. phil. Dipl.-Psych. Matthias Weigl
Mitberichterstatter:	Prof. Dr. med. Peter Hermanek PD Dr. med. Cornelius Schüle
Dekan:	Prof. Dr. med. Thomas Gudermann
Tag der mündlichen Prüfung:	27.01.2022

Abstract

Internalized goal conflict at work describes the psychological response to perceived conflicts between work-related objectives. Compared to more senior colleagues, early-career physicians might lack effective mental routines for priority-setting and might be in a structurally weaker position when rejecting demands from others. This study postulates that goal conflict causes work-related stress in early-career physicians and, in turn, contributes to negative work attitudes and impaired health.

Goal conflict and its possible effects were analyzed in a panel study of early-career physicians in Germany (N = 590; baseline survey conducted in 2004; three follow-up surveys; follow-up period: nine years, eight months). A goal conflict scale was constructed using items from an established occupational psychology research tool. The scale demonstrated acceptable internal consistency reliability, and results from factor analyses suggested unidimensionality. Fixed-effects and random-intercept regression models were developed to limit the influence of possible confounders.

On average, goal conflict was lowest in the last follow-up survey when participants were most experienced. Regression models evaluating both inter- and intra-participant variation in work experience also indicated that perceived goal conflict decreased with mounting work experience. Goal conflict was associated with reduced job satisfaction, decreased work engagement, and higher depressive symptoms scores. These results depended neither on the statistical estimation approach used (i.e., fixed-effects versus random-intercept regressions) nor on the set of individual and professional characteristics specified as covariates.

Overall, the study demonstrates that internalized goal conflict constitutes an important facet of work-related stress in early-career physicians. The concept has implications for health care management and medical education. Furthermore, addressing goal conflict may help senior physicians to better support and supervise their younger colleagues. Additional quantitative and qualitative research is warranted to deepen the understanding of both the theoretical foundations and the implications of internalized goal conflict in health care.

Titel und Zusammenfassung in deutscher Sprache

Internalisierte Zielkonflikte am Arbeitsplatz beeinflussen Arbeitszufriedenheit, Arbeitsengagement und depressive Symptome: Evidenz aus einer Panelbefragung ärztlicher Berufsanfängerinnen und Berufsanfänger

Internalisierte Zielkonflikte am Arbeitsplatz beschreiben die psychische Reaktion auf die Wahrnehmung von Widersprüchen zwischen arbeitsbezogenen Zielsetzungen. Verglichen mit erfahreneren Kolleginnen und Kollegen kann es ärztlichen Berufseinsteigenden an effektiven mentalen Routinen zur Prioritätensetzung mangeln. Ferner befinden sie sich in einer strukturell schwächeren Position, um Forderungen anderer zurückzuweisen. Diese Studie postuliert, dass durch Zielkonflikte arbeitsbezogener Stress bei ärztlichen Berufseinsteigenden entsteht, der wiederum einen Nährboden für eine negative Einstellung zur Arbeit und für negative gesundheitliche Folgen darstellt.

Zielkonflikte und ihre möglichen Effekte wurden in einer Panelbefragung ärztlicher Berufsanfängerinnen und Berufsanfänger in Deutschland untersucht (N = 590; erste Befragung im Jahr 2004; drei Folgeerhebungen; Beobachtungsdauer: neun Jahre und acht Monate). Aufbauend auf Items eines etablierten arbeitspsychologischen Forschungsinstruments wurde eine Zielkonfliktskala entwickelt. Die Skala wies eine akzeptable Reliabilität im Sinne der internen Konsistenz auf. Ergebnisse von Faktorenanalysen legten die Eindimensionalität der Skala nahe. Um den Einfluss möglicher Störfaktoren zu begrenzen, wurden Fixed-Effects- und Random-Intercept-Regressionsmodelle entwickelt.

Im Durchschnitt zeigten sich in der letzten Folgeerhebung die niedrigsten Messwerte für Zielkonflikte. Auch Regressionsmodelle, die Unterschiede im Umfang der gesammelten Arbeitserfahrung zwischen den Befragten sowie zwischen den Beobachtungszeitpunkten zur Schätzung heranzogen, deuteten darauf hin, dass Zielkonflikte mit steigender Arbeitserfahrung an Relevanz verloren. Zielkonflikte waren ferner mit reduzierter Arbeitszufriedenheit, verringertem Arbeitsengagement sowie mit erhöhten Messwerten für depressive Symptome assoziiert. Diese Ergebnisse zeigten sich unabhängig von der verwendeten statistischen Methodik (d. h. Fixed-Effects- versus Random-Intercept-Regressionen) sowie unter Verwendung unterschiedlicher privater und berufsbezogener Merkmale als Kovariaten.

Insgesamt zeigt die Studie, dass internalisierte Zielkonflikte eine wichtige Facette von arbeitsbezogenem Stress bei ärztlichen Berufseinsteigenden darstellen. Das Konzept besitzt Implikationen für das Gesundheitsmanagement und die medizinische Ausbildung. Ferner kann die

Beschäftigung mit Zielkonflikten erfahrenen Ärztinnen und Ärzten dabei helfen, jüngere Mitglieder ihrer ärztlichen Teams besser zu unterstützen und zu supervidieren. Weitere quantitative und qualitative Forschungsarbeiten sind notwendig, um sowohl die theoretischen Grundlagen als auch die Implikationen von internalisierten Zielkonflikten im Gesundheitswesen noch tiefgreifender zu analysieren.

Table of contents

Introduction

Stress at work can be an important factor that negatively influences personal well-being and mental health. Compared with the general population, early-career physicians have reported a lower quality of life as well as higher rates of fatigue and depressive symptoms (Dyrbye et al., 2014; Mata et al., 2015). In health care and other sectors, the concept of burnout has been widely used to study the consequences of chronic exposure to work-related psychological stressors (Shirom, 2003). In their well-established definition, Maslach, Schaufeli, and Leiter (2001) described burnout as a three-dimensional phenomenon characterized by emotional exhaustion, depersonalization, and a reduced sense of personal accomplishment.

Empirical evidence suggests that burnout is common among early-career physicians across different countries and medical specialties (Blanchard et al., 2010; Dyrbye et al., 2014; Starmer, Frintner, & Freed, 2016; Weidner, Phillips, Fang, & Peterson, 2018). For example, Dyrbye et al. (2014) reported a burnout rate of 50% in U.S. residents – i.e., physicians during postgraduate specialty training – compared to 31% in a similarly aged sample of college graduates from the general population.

Work engagement: The positive antipode to burnout

As an addition to the concept of burnout, the construct of work engagement was developed to describe the multifaceted interplay between the working environment and the emotions, perceptions, and attitudes of individual employees (Maslach, Schaufeli, & Leiter, 2001; Schaufeli, Bakker, & Salanova, 2006; Schaufeli, Salanova, González-Romá, & Bakker, 2002). Maslach, Schaufeli, and Leiter (2001) defined work engagement as the positive antipode to burnout. Their concept hinges on outlining the counterparts of the three dimensions of burnout: energy (as the antipode to exhaustion), involvement (as the antipode to depersonalization), and efficacy (as the antipode to a reduced sense of personal accomplishment). The Utrecht Work Engagement Scale (Schaufeli, Bakker, & Salanova, 2006; Schaufeli, Salanova, González-Romá, & Bakker, 2002) operationalizes the concept of work engagement. It is based on the three pillars of vigor, dedication, and absorption. Vigor is related to the enlarged burnout model's energy–exhaustion dimension and describes high levels of energy and mental resilience on the job. Dedication captures aspects of the depersonalization–involvement dimension of burnout and refers to experiencing a sense of significance and the feeling of being strongly involved in one's job. Absorption is more distinct from the primary conception of burnout and describes high levels of concentration and immersion in one's work.

The concept of work engagement aims at depicting a significant facet of employee work attitudes; empirical studies found links between low work engagement and high turnover intentions, a high risk of long-term sickness absence due to mental illness, and a high risk for the onset of clinical depression (Imamura et al., 2016; Roelen et al., 2015; Schaufeli & Bakker, 2004).

Depression as a response to chronic stress in physicians

Burnout and, by implication, its positive antipode work engagement are considered to be related to, but not identical with, mental health concepts such as depression (Maslach, Schaufeli, & Leiter, 2001; Tyssen, 2018; West, Dyrbye, & Shanafelt, 2018). In particular, burnout is explicitly conceptualized as being work-related, while depression is not restricted to a specific context. However, the clinical presentation of burnout and depression can be similar (Bianchi, Schonfeld, & Laurent, 2015).

Lowered mood, decreased energy and activity, and reduced interest and capacity for enjoyment are core features of depression as defined in the 10th edition of the World Health Organization's International Statistical Classification of Diseases and Related Health Problems (World Health Organization, 2019). According to the Global Burden of Disease Study (James et al., 2018), depressive disorders were the third most important overall source of years lived with disability worldwide and the most important source of years lived with disability among all mental disorders studied. Likewise, depressive disorders seem to be of high relevance for physicians and, in particular, for early-career physicians. Mata et al. (2015) conducted a meta-analysis of 54 cross-sectional and longitudinal studies that investigated the point or period prevalence of depression or depressive symptoms among residents and reported an average prevalence estimate of 29%. The median increase in depressive symptoms after starting to work as a physician was estimated at 16% based on a sub-analysis of seven longitudinal surveys that studied this question.

Chronic work-related stress in physicians and professional performance

Adverse working conditions can not only result in impaired mental health among physicians. By negatively affecting professional performance, unfavorable working conditions in health care may also pose a risk to patients. There is mounting empirical evidence for such negative effects of chronic work-related stress. Teoh, Hassard, and Cox (2019) conducted a systematic review of 21 studies that analyzed the relationship between measures of physicians' working conditions and indicators of the quality of patient care. Their review suggests that adverse working conditions have a negative impact on measures of clinical excellence (e.g., decreased disease management performance) and patient safety (e.g., increased rates of treatment errors).

In another systematic review that was based on the concept of burnout and included results from a meta-analysis that pooled 47 cross-sectional and longitudinal studies, Panagioti et al. (2018) found that physician burnout was associated with reduced patient satisfaction ratings and was related to indicators of low professionalism (e.g., low adherence to treatment guidelines or provision of suboptimal information to patients). In addition, burnout seemed to be associated with a higher risk of patient safety incidents (e.g., therapeutic or diagnostic incidents or adverse drug events). Panagioti et al. (2018) also conducted a subgroup analysis based on 21 studies that focused on early-career physicians, and their findings suggest that the association of burnout with low professionalism is stronger in this subgroup compared to the overall sample featuring studies of physicians irrespective of career-level. Analogously, studies demonstrated positive associations between work engagement and professional performance for physicians and employees in other professional fields (Halbesleben & Wheeler, 2008; Loerbroks, Glaser, Vu-Eickmann, & Angerer, 2017; Xanthopoulou, Bakker, Demerouti, & Schaufeli, 2009).

However, studying the effects of chronic stress on professional performance can pose methodological challenges. For example, most of the presented results appear to be driven by studies that relied on the surveyed physicians' self-reports of patient safety incidents. In the Panagioti et al. (2018) review, no significant association with burnout was found in a subgroup analysis of studies that used more objective patient safety measures such as medical records or surveillance data. Tyssen (2018) uncovered a similar pattern when reviewing evidence on the relationship between burnout and patient safety. Varying explanations for this discrepancy have been put forward. Panagioti et al. (2018) suspected that reporting systems fail to adequately capture the full range of relevant patient safety incidents. Conversely, Tyssen (2018) argued that the subjective perception of impaired functioning in burnout might not necessarily be linked to an actual reduction in the quality of medical care. Operationalizing indicators for professional performance is not the only methodological challenge faced by researchers when studying chronic stress in physicians. In particular, the Teoh, Hassard, and Cox (2019) review also revealed considerable heterogeneity with respect to how working conditions were operationalized and measured in the different studies. While some indicators focused on single aspects of workplace-related stress (e.g., the burden of challenging encounters), others attempted to capture more abstract and broader constructs (e.g., total quantitative workload or cognitive demands). Overall, these empirical analyses illustrate that the operationalization of both the response (e.g., performance measures) and the explanatory variable (e.g., indicators of adverse working conditions) warrants particular attention when studying the psychosocial working conditions of physicians and their ramifications.

Conceptualizing work-related stress: The job demands-resources model

Grounding constructs in theory strengthens the validity of empirical research. Widely used theoretical models that relate sources of workplace-related stress (i.e., stressors) to adverse psychological outcomes (i.e., types of mental strain) include the job demand-control model introduced by Karasek (1979); its derivative, the job demand-control-support model proposed by Johnson and Hall (1988); and the job demands-resources (JD-R) model developed by Demerouti, Bakker, Nachreiner, and Schaufeli (2001). According to the job demand-control model (Karasek, 1979), job demands (i.e., potential stressors) need to be analyzed in tandem with job decision latitude (i.e., control) in order to evaluate the causes of work-related mental strain (e.g., depression or exhaustion). Consequently, a working environment causing high mental strain is characterized by a combination of high job demands and strong constraints that restrict employees when making decisions (i.e., low control). Johnson and Hall (1988) argued that social support at the workplace constitutes a third crucial factor in addition to job demands and job control. Their job demand-control-support model assumes that low social support exacerbates the adverse effects of high job demands.

Demerouti, Bakker, Nachreiner, and Schaufeli (2001) integrated factors such as job control and social support into a more general framework known as the JD-R model. Based on the concept of burnout, the model suggests that excessive job demands and insufficient job resources result in chronic work-related stress. They defined job resources and job demands as broad constructs. In their basic application, they proposed six types of job resources: feedback, rewards, job control, participation, job security, and supervisor support. Job demands were operationalized using the following categories: physical workload, time pressure, recipient contact, physical environment, and shift work. In an approach aimed at subdividing job demands, Cavanaugh, Boswell, Roehling, and Boudreau (2000) proposed two categories: challenge-related stressors (e.g., high responsibility) and hindrance-related stressors (e.g., job insecurity). According to their concept, stress resulting from challenge-related stressors can be perceived as positively stimulating and might be associated with positive outcomes. Conversely, hindrance-related stress can be felt as frustrating and might be associated with negative outcomes. This division of job demands into a positively and a negatively perceived category resembles the more general classification of stress into positive eustress and negative distress (Lazarus, 1993). It is furthermore related to the transactional model of stress (Lazarus & Folkman, 1984) that emphasizes the importance of studying subjective judgments about stressors and resources. As a universal framework, the JD-R model is not restricted to specific jobs or sectors. However, Demerouti, Bakker, Nachreiner, and Schaufeli (2001) suggested that the relevance of specific types of job resources and job demands might depend on the kind of job studied.

Stress at the workplace of physicians: Contradictory objectives and goal conflict

When applying the JD-R model to the workplace of physicians, a wide range of potentially relevant job demands must be considered. Job demands faced by physicians include structural factors such as long working hours or reduced sleep during night shifts. In addition, stress amongst physicians might result from general characteristics of working in medicine such as exposure to emotionally demanding situations, uncertainty, or the need to act in time-critical circumstances. Furthermore, the interactive physician-patient relationship is particularly important for the practice of medicine; patients co-produce medical services (e.g., by providing information or adhering to recommendations), and medical care is typically produced as well as consumed at the same time (Büssing & Glaser, 2002).

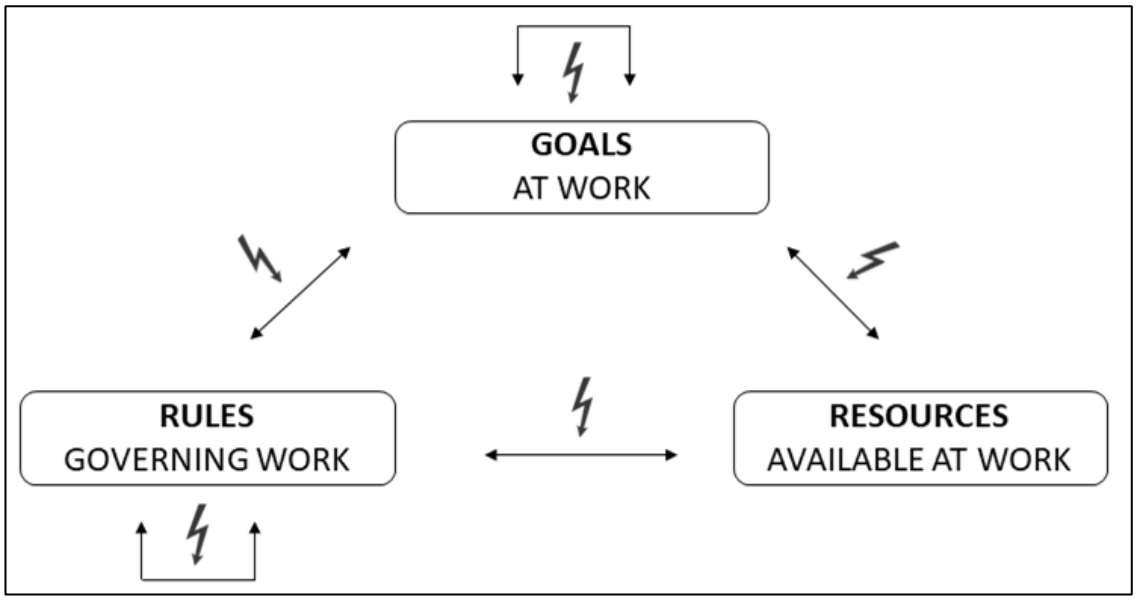
Overall, practicing medicine implies balancing multiple missions such as scientific excellence, compassionate care, and economic rationality. Additionally, physicians need to meet demands articulated by various stakeholders such as patients, superiors, and hospital administrators. In their everyday working lives, these different missions and demands can come into conflict with each other. For instance, economic rationality might suggest reducing a patient's length of stay in a hospital after surgery. Conversely, observing post-surgery patients for longer and managing delayed complications in the hospital can increase patient safety.

The concept of *goal conflict at work* (Moldaschl, 1991a, 1991b, 2007) focuses on such situations and identifies contradictory work-related objectives as a potential source of workplace-related stress (i.e., a type of job demand). Moldaschl (1991a, 1991b, 2007) embedded goal conflict at work into the broader concept of *demand conflict at work*. This idea is based on analyses of work goals in connection with resources available at the workplace and with rules governing work processes. Figure 1 summarizes the concept, and Table 1 gives examples of the different types of conflict that can occur. The three-pillar structure based on goals, rules, and resources can be a helpful framework and can guide a systematic analysis of sources of workplace-related stress. However, the classification of conflicts into their different types is, to a certain extent, arbitrary. Rules and resources can, in many situations, be framed as goals as well. For instance, the rule and legal obligation to inform a patient about all possible complications arising from a procedure (Table 1, example 2) can also be interpreted as the goal of respecting a patient's autonomy by obtaining informed consent. Analogously, limited resources are, to a certain extent, a consequence of the principle of economic rationality. The respective goal would be to use the available resources as efficiently as possible. By implication, conflicts between different work goals can be considered as an umbrella concept that encompasses many of the conflicts involving rules and resources.

The theory of contradictory objectives and resulting goal conflict has been used in previous empirical studies that have analyzed job demands faced by physicians and other professionals. For instance, Cavanaugh, Boswell, Roehling, and Boudreau (2000) operationalized the presence of contradictory objectives in terms of the inability of employees to understand what is expected of them and classified it as a hindrance stressor. Likewise, Schneider, Hornung, Weigl, Glaser, and Angerer (2017) used contradictory objectives as one of several hindrance-type work demands faced by physicians. Furthermore, Büssing and Glaser (2002) integrated the concept of goal conflict into an occupational health psychology research tool they developed to evaluate working conditions in hospitals.

Analogous to the logic of the transactional model of stress (Lazarus & Folkman, 1984), individual characteristics and cognitions might influence the degree of stress arising from contradictory objectives. Therefore, analyses of workplace-related stress should differentiate between external workplace characteristics and associated internal reactions. The following convention was adopted for this study: On the external dimension, the term *contradictory objectives* refers to the potential stressor (i.e., the job demand). On the internal dimension, *goal conflict* describes the resulting stress as an individual response.

Figure 1: Demand conflict at the workplace of early-career physicians



Note. Summary of sources of demand conflict; adopted and translated from Moldaschl (2007)

Table 1: Demand conflict at the workplace of early-career physicians

#	Conflict between A and	B	Example
1	**Goal**	**Goal**	The head of the department instructs the resident to quickly discharge the patient after surgery in order to be able to admit another patient (goal 1 = economic rationality). However, the resident wants to monitor the patient longer in order to be able to react quickly in case of complications (goal 2 = increase patient safety).
2	**Goal**	**Rule**	The resident wants to calm down a nervous patient prior to a routine procedure (goal = cater to the patient's emotional needs). However, it is her/his legal obligation (rule) to inform her/him about all possible complications, including some very rare and serious ones.
3	**Goal**	**Resource**	The resident wants to inform the patient about a serious diagnosis in an undisturbed environment (goal = appropriate communication with patients), but no suitable room (resource) is available.
4	**Rule**	**Rule**	The head of the radiology department instructs the radiology assistants to perform MRI scans of surgical patients only if a surgical resident has informed the patient about possible complications and has obtained the patient's consent in writing (= rule 1). However, the chief of surgery instructs the surgical residents to send patients directly to the radiology department (= rule 2). In her/his opinion, the radiology residents should inform patients about the risks of the radiological procedure as they are experts in this field.
5	**Rule**	**Resource**	The resident is obliged to inform the patient about all relevant complications of a procedure and obtain her/his consent in writing (rule). However, the necessary form is not available (resource).

Note. Possible manifestations of demand conflict at work. Examples developed by the author based on the theoretical concept outlined by Moldaschl (1991a, 1991b, 2007)

Concepts related to internalized goal conflict: Similarities and differences

Goal conflict at work is related to several other concepts that focus on inconsistent or contradictory work-related norms and principles. In particular, relevant theoretical overlap exists with the framework of misfit of organizational and personal work standards (Edwards, 2008; Tanner, Bamberg, Kersten, Kozak, & Nienhaus, 2017), with the model of conflict based on divergent professional identities (DiBenigno, 2017; Fiol, Pratt, & O'Connor, 2009), and with the notions of moral dilemma and moral distress (Fourie, 2015; Jameton, 1984; Kälvemark, Höglund, Hansson, Westerholm, & Arnetz, 2004).

Based on Edwards's (2008) general framework of person-environment fit, Tanner, Bamberg, Kersten, Kozak, and Nienhaus (2017) identified a misfit between organizational and personal work standards as a potential source of stress and stress-related adverse health outcomes in physicians. In a longitudinal survey of German physicians, they showed that a misfit between organizational and personal work standards was associated with depressive symptoms, both cross-sectionally as well as longitudinally. Conflicts involving rules in Moldaschl's (1991a, 1991b, 2007) framework of demand conflict at work resemble the concept of misfit of organizational and personal work standards. Personal work standards can be used to derive corresponding work-related goals. Analogously, adherence to organizational standards and their guiding principles can represent a relevant goal for employees. Therefore, this idea of a misfit between organizational and personal work standards represents a subdomain of internalized goal conflict in a broader sense. However, the concept of internalized goal conflict focuses on external work demands and not on internal, personal preferences as sources of negative stress.

Personal work standards are also linked to professional identities (DiBenigno, 2017; Fiol, Pratt, & O'Connor, 2009), another concept related to goal conflict at work. However, while the framework of goal conflict focuses on internalized conflicts between different goals (i.e., within-person conflict), professional identities seem to be particularly important when analyzing conflicts between different professional groups within one organization (i.e., between-persons conflict). DiBenigno (2017) analyzed conflicts between mental health care providers and commanders in the U.S. army regarding issues like health data protection or treatment recommendations that involved taking individual patients out of their regular job. Mental health care providers tended to identify themselves as health care professionals who prioritized their patients' individual well-being. At the same time, commanders leaned towards seeing themselves as unit leaders that prioritized the larger group's functioning over the individual needs of single soldiers. Taking a similar approach, Fiol, Pratt, and O'Connor (2009) evaluated the working environment in a U.S. hospital and observed comparable conflicts between medical staff and hospital administrators. While

health care providers viewed themselves as responsible for the quality of patient care, hospital administrators described themselves as being in control of the organizational future of the hospital as a whole. The idea of internalized goal conflict also differs from conflicts rooted in differing personal identities with respect to the scope of conflicts covered. Goals can be in contradiction to each other even if they are based on the same professional identity concept. Taking the timing of a patient's discharge after surgery again as an example illustrates this. Earlier discharge can not only achieve the goal of using hospital beds more efficiently but can also lower the risk of hospital-acquired infections. In this situation, physicians who identify themselves first and foremost as professionals responsible for patient safety could face a dilemma (i.e., decreased risk of hospital-acquired infections versus better treatment in cases of complications after surgery).

Internalized goal conflict is also linked to the concepts of moral dilemma and moral distress (Fourie, 2015; Jameton, 1984; Kälvemark, Höglund, Hansson, Westerholm, & Arnetz, 2004). Coined by Jameton (1984), the two terms have been used in various contexts as frameworks for describing the implications of decision-making in health care with respect to questions that have an ethical dimension (Fourie, 2015; Kälvemark, Höglund, Hansson, Westerholm, & Arnetz, 2004). Fourie (2015) defined moral distress as a psychological response to morally challenging situations. In a qualitative study, Kälvemark, Höglund, Hansson, Westerholm, and Arnetz (2004) analyzed dilemma situations experienced by Swedish physicians, nurses, and pharmacists. For their study, Kälvemark, Höglund, Hansson, Westerholm, and Arnetz (2004) applied the concept of moral distress to all stressful situations that involved an ethical dimension. However, some of the dilemmas outlined in the study seem to use a narrower definition of an ethical issue than others. For instance, whether to provide urgently needed health care to an uninsured person might unambiguously be considered an ethical question. In other cases, whether a decision involves an ethical dimension or not will largely depend on the context. Interrupting a conversation with a patient when a colleague calls might constitute an ethical issue if a severe diagnosis is communicated for the first time. However, if the call occurs during a routine consultation, questions about how to act seem more to revolve around priority-setting between contradictory goals of similar moral value. Compared to the idea of dilemma situations causing moral distress, the concept of internalized goal conflict allows describing stressful clinical situations without the added prerequisite that moral reasoning – i.e., arguing about which alternatives can be considered as morally right or wrong – needs to be of relevance in the specific situation.

Overall, the concept of goal conflict constitutes a comprehensive framework for analyzing job demands in medicine and can be applied to various challenging situations encountered by early-career physicians in hospitals.

Outline and contributions of this study

This study adds to the existing literature on job demands and work-related stress in early-career physicians (cf. Dyrbye et al., 2014; Mata et al., 2015; Panagioti et al., 2018). It is based on the theoretical conception of internalized goal conflict at work proposed by Moldaschl (1991a, 1991b, 2007) and used a corresponding empirical research tool developed by Büssing and Glaser (1999, 2000, 2002). In contrast to previous studies that have analyzed goal conflict together with other types of job demands (cf. Cavanaugh, Boswell, Roehling, & Boudreau, 2000; Hornung, Rousseau, Glaser, Angerer, & Weigl, 2010; Schneider, Hornung, Weigl, Glaser, & Angerer, 2017), the study focused explicitly on goal conflict.

The study used longitudinal panel data from a survey of 590 medical job entrants in Germany to evaluate whether goal conflict represented a relevant facet of job-related stress in early-career physicians. Possible associations between goal conflict, work attitudes, and mental health were tested. In addition, the relationship between work experience and goal conflict was evaluated. Fixed-effects and random-intercept regression models were developed to limit the influence of possible confounders. Established research instruments were used to operationalize the examined constructs. The psychometric properties of all computed scales were evaluated using factor analytic approaches and reliability tests.

Hypotheses

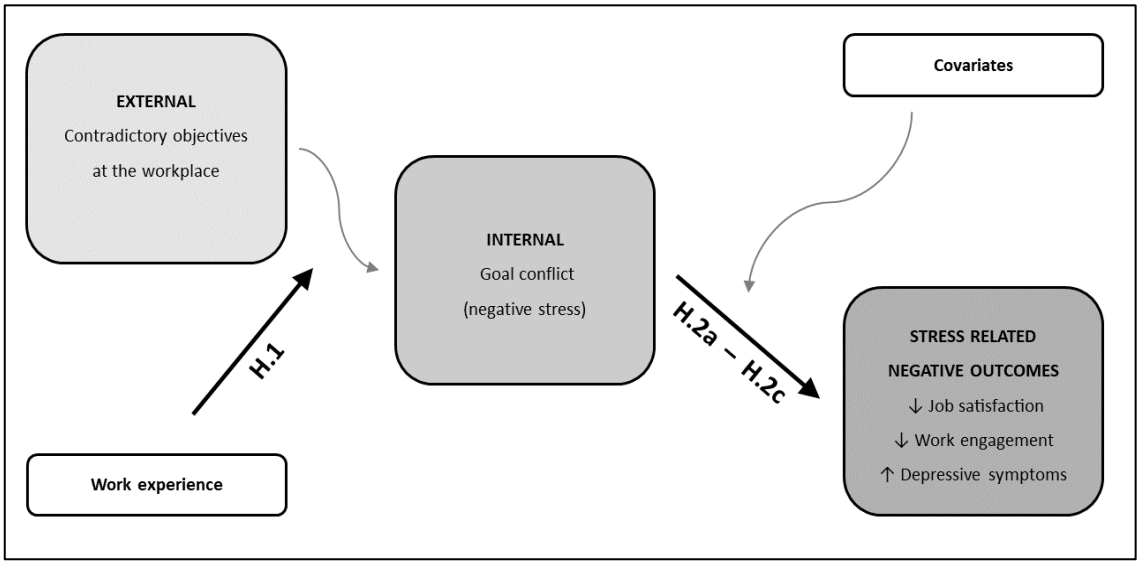
Figure 2 summarizes the hypotheses of this study. Hypothesis H.1 postulates that increased work experience is related to decreased goal conflict. Hypotheses H.2a and H.2b postulate that goal conflict negatively influences work attitudes defined both broadly as job satisfaction (H.2a) and more specifically as work engagement (H.2b). Hypothesis H.2c postulates that goal conflict is associated with depressive symptoms as a measure of impaired mental health.

H.1: Work experience and goal conflict

How psychosocial work demands are perceived can depend on the individual employee; the same work demand can be a negative stressor for some individuals, while others may regard it as a neutral characteristic or even as a positively challenging aspect of their work (Büssing & Glaser, 2002). In workplace situations that involve contradictory objectives, some physicians might experience goal conflict, while others might not. Over the course of their careers, physicians can potentially develop capabilities and mental routines that help them balance opposing work demands in more and more situations. Compared to more experienced colleagues, early-career physicians might lack such skills. Therefore, work experience should be a predictor of the degree of goal conflict perceived by early-career physicians:

H.1: Increased work experience is associated with decreased goal conflict.

Figure 2: Conceptual model, key constructs, and hypotheses of this study



H.2a – 2c: Goal conflict, job satisfaction, work engagement, and depressive symptoms

Practicing medicine inevitably involves balancing contradictory objectives in various circumstances. If physicians struggle in these situations, the resulting goal conflict should represent a relevant manifestation of work-related stress and be related to significant adverse effects. Three such stressor-strain relationships were tested in this study:

1) *Job satisfaction* was selected as an outcome as it represents a broad and intuitive measure of work-related attitudes. Empirical evidence suggests that it is positively associated with job performance (Judge, Thoresen, Bono, & Patton, 2001) and negatively related to physicians' intentions to leave their job (Hann, Reeves, & Sibbald, 2010). Therefore, negative stress induced by goal conflict should be associated with low job satisfaction:

*H.*2a*: Increased goal conflict is associated with decreased job satisfaction.*

2) Work engagement presents a construct intended to measure a chronic work-related outcome rather than a short-term or momentary state (Schaufeli, Bakker, & Salanova, 2006). It was selected as an outcome to identify the long-term effects of goal conflict on job attitudes. In addition, work engagement can be considered the antipode of the widely used concept of burnout (Schaufeli, Salanova, González-Romá, & Bakker, 2002). No instrument to directly measure burnout was included in the baseline survey. However, low work engagement should be an indicator of burnout. Furthermore, work engagement, as a concept associated with positive psychology (Schaufeli, Salanova, González-Romá, & Bakker, 2002; Seligman & Csikszentmihalyi, 2000), was included in order to study whether reducing goal conflict strengthens positive attitudes towards work. As chronic negative stress should negatively influence work attitudes, increased goal conflict scores should be associated with decreased work engagement:

H.2b Increased goal conflict is associated with decreased work engagement.

3) Depression was chosen as an outcome as it presents a common manifestation of impaired mental health among early-career physicians (Mata et al., 2015). Chronic negative stress, such as goal conflict, should increase the prevalence and the intensity of *depressive symptoms*:

H.2c Increased goal conflict is associated with increased depressive symptoms.

Methods

This study used panel data from a longitudinal survey of early-career physicians in Germany. At baseline, the participating physicians were in the early phases of postgraduate training and worked in hospitals. Three-follow up surveys were conducted. The total study duration was nine years and eight months. Participants received paper-based survey questionnaires, and self-assessment measures were used for goal conflict, the studied outcomes, and relevant covariates. Data were analyzed based on fixed-effects and random-intercept regression models.

The following sections describe the study methodology in detail. In the next section, inclusion criteria are presented. After that, the study design, the data collection process, and the procedure used to obtain the study sample are outlined. Subsequently, the research instruments are presented. The remaining sections summarize the statistical modeling approach and outline the procedures used for statistical analyses and regression model estimation.

Inclusion criteria

Data from the physician directory of the Bavarian Chamber of Medical Doctors were used to construct the pool of physicians that were invited to participate in the study. The directory can be considered a comprehensive registry of all physicians in the region, as every physician practicing in the German federal state of Bavaria is legally obliged to register with the chamber (BayHKaG, 2002). The initial pool consisted of 1000 early career physicians from the region of Munich, Germany. All physicians included in the pool worked in direct patient care. In addition, all of the included physicians were employed by hospitals at the time of address data extraction.

In Germany, physicians start specialist training (residency) in their chosen field directly after graduation from medical school, and specialist training programs take place at accredited institutions such as hospitals (DAAD, 2020). As the study focused on early-career physicians, only physicians in their second or third year of postgraduate training were included in the initial pool. Selecting participants with at least a year of work experience made sure that participants were sufficiently experienced to provide meaningful and valid evaluations of their working conditions. No additional criteria were applied. In particular, participants were included regardless of the medical specialty in which they worked.

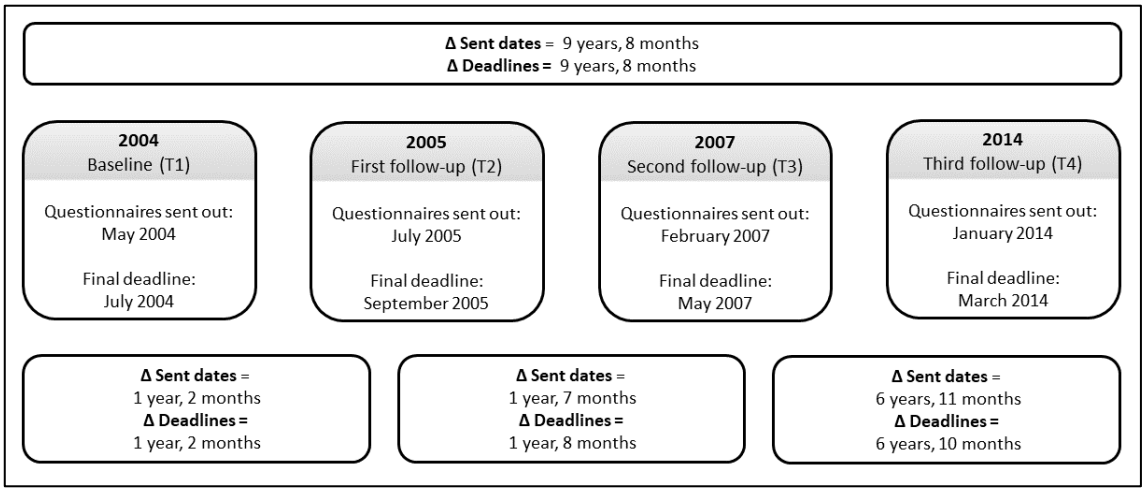
Study design

Figure 3 outlines the study design. The baseline survey (T1) was conducted in 2004. The participants were followed-up around one, around three, and around 10 years after (i.e., in 2005, 2007, and 2014). The total study duration was nine years and eight months.

Short time lags between the first three surveys aimed at evaluating the first career phase of postgraduate specialty training in detail. The first time lag of around one year was chosen as previous studies have successfully used similar time spans to study psychosocial working conditions and their effects (Jackson, Schwab, & Schuler, 1986; Tokuyama, Nakao, Seto, Watanabe, & Takeda, 2003). An interval of around three years between T1 and T3 was chosen to survey participants at the end of postgraduate specialty training, as participants were in their second or third program year at baseline and specialty training programs in Germany usually last for a total of five to six years (DAAD, 2020).

The lag of around 10 years between the baseline survey and the last follow-up was chosen to survey participants at a later career stage at which they have developed substantial professional expertise. 10 years were suggested as the minimal time necessary to reach expert status in various fields such as medicine, sports, the arts, and science (Ericsson, 2004; Ericsson, Krampe, & Tesch-Roemer, 1993).

Figure 3: Study overview



Data collection procedure

All 1000 early-career physicians in the initial pool received a paper-based invitation together with the study materials via regular mail. The invitation and all study materials were written in German. Response deadlines were extended two times for each of the four surveys to increase participation rates. All participants who did not return the questionnaire prior to the deadline extension received a reminder and were informed about the new deadline. In addition, respondents received a small payment via bank transfer for each completed questionnaire (€25 for the T1 survey and €30 for each of the T2 – T4 surveys).

All 621 participants who had returned the questionnaire of the baseline survey in 2004 received the study materials for all three follow-up surveys. Before sending out the follow-up surveys, address data were updated from the official registry of the Bavarian Chamber of Medical Doctors. In addition to the questionnaire, the study kit included the following items: an invitation letter, an information document about the study, an informed consent form, and a bank details form for the transfer of the small payment for returned questionnaires. All questionnaires were pre-coded with a pseudonymized participant identifier that permitted to match responses across the baseline and the three follow-up surveys. In addition, two envelopes were provided: a pre-stamped return envelope and a separate, smaller envelope for the filled-out questionnaire. Participants were instructed to insert only the questionnaire in the smaller envelope. Afterward, the sealed smaller envelope had to be put in the second envelope together with the signed informed consent form and the bank details form. Upon arrival in the study center, questionnaire envelopes were separated from the returned forms that contained personal data. In a second, separate step, data entry was performed by a study team member who was not aware of the participants' identities. All members of the study team who analyzed the data had no possibility to link survey responses to specific participants.

The study was organized by the Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine of Ludwig Maximilian University Munich, Germany (T1 – T3, i.e., 2004, 2005, and 2007) and the Institute of Occupational and Social Medicine of Heinrich Heine University Düsseldorf, Germany (T4, i.e., 2014). Approval of the Ethics Committee of Ludwig Maximilian University was granted prior to the baseline survey; the Ethics Committee of Heinrich Heine University approved the study prior to the third follow-up survey in 2014.

Study sample

Figure 4 states response rates for all four surveys. At baseline, N = 621 physicians returned questionnaires (response rate 62%). Ninety percent of all baseline respondents returned the first follow-up survey (N = 561). Respective response rates were 85% for the second follow-up (N = 525) and 72% for the final follow-up (N = 450).

Respondents were excluded from subsequent analyses if responses were missing for the goal conflict scale. Participants were also excluded if they failed to provide responses for all of the other key study variables (i.e., job satisfaction, work engagement, depressive symptoms, and – for the second and third follow-up – work experience). The study sample obtained by using this procedure was composed of all participants who provided data to test at least one of the four hypotheses. Table 2 lists the number of excluded participants for each of the four surveys. At baseline (T1), N = 31 sets of observations were excluded. Respective figures were N = 24 for the first follow-up (T2), N = 38 for the second follow-up (T3), and N = 17 for the final follow-up (T4). In total, N = 2.047 sets of observations were included in the study sample. Fifty-one percent of these sets were provided by female participants and 49% by male participants.

Figure 4: Study sample

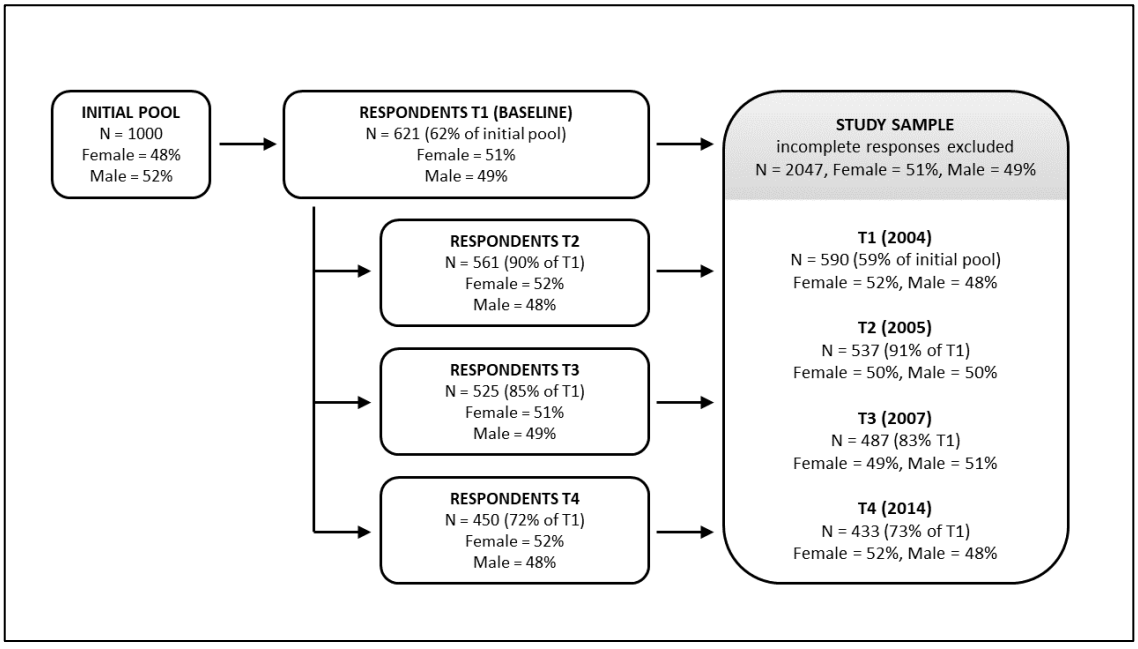


Table 2: Excluded participants

	Baseline (T1)	First follow-up (T2)
Total	**31**	**24**
Missing	31 = JS & WEN & DS	22 = GC & JS 01 = GC 01 = GC & JS & WEN
	Second follow-up (T3)	**Third follow-up (T4)**
Total	**38**	**17**
Missing	37 = GC & JS & WEN 01 = GC	08 = GC 03 = GC & WEX 02 = GC & JS & WEN 02 = GC & JS & WEN & WEX 01 = GC & JS 01 = GC & JS & WEX

Note. JS = job satisfaction; WEN = work engagement; DS = depressive symptoms; GC = Goal conflict; WEX = work experience

Goal conflict scale

The study's goal conflict scale was constructed using items from a more extensive research instrument, the Work Analysis Instrument for Hospitals (Büssing & Glaser, 1999, 2000, 2002). The "Tätigkeits- und Arbeitsanalyseverfahren für das Krankenhaus" (TAA-KH), as it is referred to in German, is an occupational health psychology research tool developed explicitly for hospitals. It was designed to account for specifics of jobs involving direct patient care and, in particular, to analyze the working environment of nurses.

The TAA-KH intends to provide tools for a comprehensive analysis of the working environment in hospitals. It assesses negative and positive aspects of all relevant tasks encountered by employees working in direct patient care and additionally analyzes organizational aspects on the level of organizational units (typically wards) and the entire hospital such as leadership characteristics. The instrument's proposed use cases are occupational psychology research and human resources

management. Its results are intended to provide guidance for improving working conditions in hospitals so that workplaces better fulfill the requirements of employees (i.e., to humanize work) and, in turn, efficacy and efficiency are increased.

The TAA-KH contains a self-rating questionnaire that is referred to as TAA-KH-S. In addition to the TAA-KH-S, the complete TAA-KH features an expert-rating tool for shift observations and guidelines for interviews and the assessment of organizational characteristics.

The TAA-KH-S, which contains the items used for this study's goal conflict scale, is divided into five parts. The first three parts operationalize positive aspects of tasks encountered by hospital employees: Rewarding challenges (e.g., chances to obtain new qualifications), transparency, decision latitude, participation (e.g., the possibility to influence shift planning), and job resources (e.g., social support). The remaining two parts address negative characteristics such as social and organizational stressors (e.g., lack of necessary materials) and – based on Moldaschl's (1991a, 1991b, 2007) concept – manifestations of demand conflict at work.

The instrument's developers have made two versions of the self-rating questionnaire available: a long version (442 items, time necessary to complete the survey: 45 to 60 minutes) and a shorter screening version (182 items, time necessary to complete the survey: 20 to 30 minutes). Detailed test instructions are provided in the handbook to increase the instrument's objectivity.

The reliability and the construct validity of the long version were analyzed empirically in two test user groups (N = 536 and N = 476) by the instrument's developers. As a measure of reliability, Cronbach's α is reported for all TAA-KH-S scales in the instrument's handbook. Sufficiently high values (i.e., $\alpha \geq 0.6$) are stated for 59 of the total 62 scales. For the scale made up of the six goal conflict items included in the TAA-KH-S, a value of $\alpha = 0.78$ (first test group) and $\alpha = 0.81$ (second test group) is reported in the instrument's handbook. Factor analyses and convergent validity analyses were performed by the instrument's developers to evaluate the construct validity of the TAA-KH-S scales. Results from factor analyses reported in the instrument's handbook suggest unidimensionality for most scales. Convergent validity analyses performed by the instrument's developers tested correlations of the TAA-KH-S scales with other positive and negative measures (e.g., psychosomatic complaints). For the demand conflict scales, all empirically estimated correlations reported in the instrument's handbook are consistent with theoretical predictions (i.e., positive correlations with negative outcomes and vice versa).

Maximizing content validity was a central objective of the item development process. The process was grounded in occupational psychology theory (e.g., the concept of demand conflict) and involved iterative feedback from nurses, physicians, and hospital administrators. In total, the

instrument's developers tested the long version of the questionnaire with 2193 participants in 15 different hospitals. The development of the screening version, which is based on the long version, involved tests with an additional 280 participants from three hospitals.

Parts of the screening version of the TAA-KH-S were used in the survey questionnaires of the presented study. Most of the included items did not focus on goal conflict but on other workplace characteristics and work demands (e.g., employee participation in organizational decision making or exposure to social stressors). These supplementary items were not analyzed in the presented study.

At baseline (T1), 65 TAA-KH-S items were included in the survey questionnaire. Six of these items focused on goal conflict. For the first follow-up (T2), 95 items from the TAA-KH-S were used. Additional scales from the TAA-KH-S were added (e.g., a scale that measured variety in work tasks) while other scales were shortened or eliminated (e.g., the items that intended to evaluate the staffing of the participant's department). In this process, two of the six items intended to measure goal conflict were removed from the survey questionnaire. In total, four items measuring goal conflict were stated in the survey questionnaire of the first follow-up.

The TAA-KH-S questionnaire used for the second follow-up was identical to the questionnaire used for the first follow-up. The identical four items focused on goal conflict and 91 other TAA-KH-S items were included. For the third follow-up (T4), the overall questionnaire was shortened. In total, 63 TAA-KH-S items, including four items measuring goal conflict, were used in the third-follow up questionnaire.

This study's goal conflict scale used TAA-KH-S items that measured this construct. Following a similar approach, subsets of the TAA-KH-S items have previously been used to study specific research questions. Based on data from the last three surveys of this study, Schneider, Hornung, Weigl, Glaser, and Angerer (2017) used 32 TAA-KH-S items to measure a range of job demands and resources. Using data from the first three surveys of this study, Weigl, Hornung, Petru, Glaser, and Angerer (2012) constructed a work overload scale (based on four TAA-KH-S items), a job autonomy scale (based on seven TAA-KH-S items) and a professional support scale (based on three TAA-KH items). In a separate study on work redesign in U.S. and German hospitals, Hornung, Rousseau, Glaser, Angerer, and Weigl (2010) used 23 TAA-KH-S items to measure work characteristics. Self-translated versions of the items were used for the U.S. survey. In another study, Hornung, Weigl, Glaser, and Angerer (2013) constructed two scales measuring work overload and patient demands using TAA-KH-S items.

Table 3 presents the item wording of the six TAA-KH-S items measuring goal conflict and the versions used in the four surveys of this study. The items are focused on operationalizing two major manifestations of goal conflict: quality-versus-quantity dilemmas and dilemmas involving contradictory instructions.

For the baseline survey conducted in 2004 (T1), the original TAA-KH-S items that were intended to survey nurses were adapted for surveying physicians. For the first follow-up survey conducted in 2005 (T2), the wording was shortened and generalized. In addition, the goal conflict items were positioned in one continuous block, while in the baseline survey, items from all TAA-KH-S scales were presented in a mixed order. The goal conflict items featured in the second follow-up survey conducted in 2007 (T3) were identical to the items used for the first follow-up survey. For the third follow-up survey conducted in 2014 (T4), the indefinite pronoun “one” was changed to the personal pronoun “I” in all statements.

The response format was identical in all four surveys and consisted of a five-level Likert-type scale ranging from 1 = “No, not at all” to 5 = “yes, indeed.” The goal conflict score was computed separately for each survey based on items 1 – 4, the four items that were included in all four questionnaires. The two additional goal conflict items included in the baseline survey were not used for score computation. The arithmetic mean of the four items was used as the goal conflict score for each of the surveys. Observations were only used if participants had answered at least 75% of the scale’s items (i.e., a maximum of one item was missing). If a participant did not meet this criterion, its goal conflict score was considered missing for the respective survey.

Table 3: Goal conflict items

Item #	**TAA-KH Screening** Wording (Büssing & Glaser, 2002)	**Baseline (T1)** Wording (position within section)	**First follow-up (T2)** Wording (position within section)	**Second follow-up (T3)** Wording (position within section)	**Third follow-up (T4)** Wording (position within section)
Item 1	**When working on this ward one gets assigned tasks that are not compatible with each other**	**When working in this department one gets assigned tasks that are not compatible with each other (12/65)**	**Again and again one gets assigned tasks that are not compatible with each other (74/95)**	**Again and again one gets assigned tasks that are not compatible with each other (74/95)**	**Again and again I get assigned tasks that are not compatible with each other (49/63)**
Item 2	**When working on this ward one has to take care of that many patients so that the quality of nursing care suffers**	**When working in this department one has to take care of that many patients so that the quality of the work one does suffers (26/65)**	**Again and again one has so much to do so that the quality of the work one does suffers (75/95)**	**Again and again one has so much to do so that the quality of the work one does suffers (75/95)**	**Again and again I have so much to do so that the quality of the work I do suffers (50/63)**
Item 3	**When working on this ward one receives instructions from superiors about which one knows that they do not lead to the expected result**	**When working in this department one receives instructions from superiors about which one knows that they do not lead to the expected result (38/65)**	**Again and again one receives instructions from superiors about which one knows that they do not lead to the expected result (76/95)**	**Again and again one receives instructions from superiors about which one knows that they do not lead to the expected result (76/95)**	**Again and again I receive instructions from superiors about which I know that they do not lead to the expected result (51/63)**
Item 4	**On this ward one has to fulfill tasks for which one is actually not responsible**	**In this department one has to fulfill tasks for which one is actually not responsible (47/65)**	**Again and again one has to fulfill tasks for which one is actually not responsible (77/95)**	**Again and again one has to fulfill tasks for which one is actually not responsible (77/95)**	**Again and again I have to fulfill tasks for which I am actually not responsible (52/63)**
Item 5	When working on this ward different superiors give conflicting orders	When working in this department different superiors give conflicting orders (57/65)	not included in the survey	not included in the survey	not included in the survey
Item 6	If one wants to do his work in a thorough and accurate manner one often does not complete it	If one wants to do his work in this department in a thorough and accurate manner one often does not complete it (63/65)	not included in the survey	not included in the survey	not included in the survey

Note. Author's own translations from German; wording modifications underlined by the author; items printed in bold were used for goal conflict score computation

Outcomes and covariate measures

The outcomes work engagement and depressive symptoms were measured using established scales. Job satisfaction, work experience, and covariates were measured using single items.

Job satisfaction

Job satisfaction was measured using the German equivalent of the item "Overall, how happy are you with you your current job?" Response options were: 1 = "unsatisfied," 2 = "rather not satisfied," 3 = "partly satisfied," 4 = "satisfied," and 5 = "very satisfied." While measuring overall job satisfaction with a single-item measure has a long tradition, several scales have been developed in an attempt to measure job satisfaction more reliably and account for several facets of the construct (Dunham & Herman, 1975; Wanous, Reichers, & Hudy, 1997). However, based on a meta-analysis comparing single-item measures of job satisfaction with some of these job satisfaction scales, Wanous, Reichers, and Hudy (1997) concluded that job satisfaction seems to be a sufficiently narrow construct to be measured with acceptable reliability using a single item.

Work engagement

Work engagement was measured using the German version of the Utrecht Work Engagement Scale (UWES) developed by Schaufeli, Salanova, González-Romá, and Bakker (2002). The original version of the scale consists of 17 items. The scale attempts to operationalize the three dimensions of work engagement: Vigor (six items), dedication (five items), and absorption (six items).

Based on a retrospective analysis of surveys that used the 17-item version in 10 different countries (N = 14521), Schaufeli, Bakker, and Salanova (2006) proposed a shorter version of the instrument with nine items (UWES-9). The shorter scale features three items for each of the three work engagement dimensions. Schaufeli, Bakker, and Salanova (2006) reported high internal consistency reliability for the shorter scale based on analyses of their international sample (median $\alpha = 0.92$) and stated results from factor analyses that showed that a unidimensional factor structure (i.e., overall work engagement as a single latent construct) fitted their data equally well as a three-factor structure based on the different dimensions.

To date, the UWES is available in 31 languages (Schaufeli, 2020). Sautier et al. (2015) conducted a psychometric analysis of the German version of the nine-item scale. Analogously to the results reported for the international sample by Schaufeli, Bakker, and Salanova (2006), their analyses (N =179) showed high internal consistency ($\alpha = 0.94$) and a unidimensional factor structure.

In this study, the 17-item version was used in the baseline survey (T1), in the first follow-up survey (T2), and in the second follow-up survey (T3). For the third follow up survey, the nine-item version was used to reduce the time necessary for completing the survey questionnaire. Table 4 states the wording of the nine items used in all four surveys. For reasons unknown to the author, item number 12 of the long scale ("I can continue working for very long periods at a time") was used as a ninth item in the third follow-up questionnaire. This choice is inconsistent with the short version as proposed by the scale developers (Sautier et al., 2015; Schaufeli, 2020; Schaufeli, Bakker, & Salanova, 2006). The official version includes item number 14 of the long scale ("I get carried away when I'm working") as the ninth item of the short scale. The other eight UWES items used in the third follow-up survey were identical with the official UWES-9 scale. In summary, the nine-item scale used in the third follow-up survey (T4) consisted of four items measuring the vigor dimension, three items measuring the dedication dimension, and two items measuring the absorption dimension of work engagement.

The work engagement scale used for the presented analyses was computed based on the nine items included in all four surveys. Similarly to the approach used for the goal conflict scale, the arithmetic mean of the nine responses was used as the total score; the score was only computed if a participant had answered at least 75% of the scale's items in the respective survey (i.e., a maximum of two items were missing). As proposed in the official UWES versions (Sautier et al., 2015; Schaufeli, 2020; Schaufeli, Bakker, & Salanova, 2006; Schaufeli, Salanova, González-Romá, & Bakker, 2002), the response format was a Likert-type scale ranging from 0 = "never" to 6 = "always."

Table 4: Work engagement items

Original table is protected by copyright and was removed for the final publication of this text. Schaufeli (2020) features the scale.

Depressive symptoms

The State-Trait-Depression scales (STDS) were used as a measure of depressive symptoms (Krohne, Schmukle, Spaderna, & Spielberger, 2002; Spaderna, Schmukle, & Krohne, 2002; Spielberger, Ritterband, Reheiser, & Brunner, 2003). While other common measures of depression, such as the Beck Depression Inventory (BDI) proposed by Beck, Steer, Ball, and Ranieri (1996), aim at assessing clinical depression, the STDS attempt to measure depressive symptoms in non-clinical populations. Due to this focus, using the STDS should have facilitated the detection of mild depressive symptoms among the studied early-career physicians. The STDS items operationalize aspects of the cognitive-affective dimension of depression, i.e., negative emotions and associated thought patterns. Somatic symptoms – e.g., change in appetite or sleeping – that are typically integrated into depression scales used for clinical assessment (cf. Beck, Steer, Ball, & Ranieri, 1996) are not evaluated by the STDS.

The subdivision of the STDS into two independents scales, the state and the trait scale, is another distinctive feature. In the vein of concepts such as cognitive vulnerability to depression (Teasdale, 1988), the trait scale attempts to operationalize persistent depressive tendencies. Conversely, the state scale assesses the intensity of current depressive symptoms at the time of survey. The state and the trait scale consist of 10 items each. For both scales, five of these items are negatively worded and describe negative feelings and thinking patterns (dysthymia). In contrast, the remaining five items are positively worded and attempt to operationalize a positive antipode to depression (euthymia).

Krohne, Schmukle, Spaderna, and Spielberger (2002) evaluated the German and the English STDS versions using a German and an American sample of university students (total N = 461 and total N = 714). They reported that internal consistency was high and comparable for both versions (for the state scale $\alpha = 0.88$ and $\alpha = 0.85$, for the trait scale $\alpha = 0.91$ and $\alpha = 0.90$) and that two relevant factors mapping the euthymia and dysthymia subscales were found when analyzing both versions. However, for both the trait and the state scale, they estimated high correlations between the two subscales and concluded that in both cases, the 10 items could be combined in one overall scale. In addition, Krohne, Schmukle, Spaderna, and Spielberger (2002) reported substantial positive correlations between the STDS and other depression scales such as the BDI. These findings suggest that high STDS scores correspond to clinical depression.

In this study, the German version of the state scale was used to compute a total score. Table 5 lists the wording of the 10 items of this scale. Wording or item positioning was not changed during the study (T1 – T4). For all 10 items, participants had to endorse statements about how they feel at the time of survey on a four-level Likert-type scale (1 = “not at all” to 4 = “very”). As suggested

by Krohne, Schmukle, Spaderna, and Spielberger (2002), the five items measuring euthymia were reverse-scored, and in turn, the arithmetic mean was calculated as an overall measure of depressive symptoms. Again, observations were only used if participants had answered at least 75% of the scale's items in the respective survey (i.e., had answered a minimum of eight items).

Table 5: Depressive symptoms scale

Original table is protected by copyright and was removed for the final publication of this text. Spaderna, Schmukle, and Krohne (2002) feature the German STDS items used in questionnaires (T1 - T4); Spielberger, Ritterband, Reheiser, and Brunner (2003) feature their English equivalents.

Work experience

Table 6 summarizes the work experience measures used in this study. Pre-study job tenure (T1) represented total work experience as a physician prior to the baseline survey (single item). For the first follow-up (T2), work experience could not be evaluated as no such measure was included in the questionnaire. Total work experience at T1 was subtracted from total work experience at T3 (single item) to compute observational tenure at T3. At T4, total work experience was not assessed. However, participants were asked to specify whether they had pauses from working as a physician in the period between this survey and 2007 (i.e., the year of the previous survey). If participants stated that they were working without interruptions, observational tenure at T4 was calculated by adding 83 months — the time between the sent-out dates of the second and the third follow-up survey — to observational tenure as computed at T3. If interruptions were specified, they were subtracted from this value.

Table 6: Work experience measures

	Baseline (T1)	**First follow-up (T2)**
Measure	**Pre-study tenure**	no work experience measure included in survey
Computation	work experience [T1]	
	Second follow-up (T3)	**Third follow-up (T4)**
Measure	**Observational tenure**	**Observational tenure**
Computation	work experience [T3] - work experience [T1]	no interruption: observational tenure [T3] + 83 months interruption: observational tenure [T3] + 83 months - total duration of interuption(s)

Note. Work experience = total work experience as a physician

Private and professional characteristics

Age and gender were assessed using single items. In the baseline survey (T1), participants specified the medical specialty in which they trained based on a list of 27 predefined medical specialties with an additional option to enter free text.

In the three follow-up surveys (T2 – T4), participants stated the type of organization in which they worked. At T2 and T3, three response options were given: "Hospital," "private practice/outpatient service," and "other." At T4, a decision tree was used. Type of organization was set to "hospital" if participants stated that they worked in health care (item 1/3), their job involved the provision of medical care to patients (item 2/3), and they were working in the hospital sector (item 3/3). Type of organization was set to "private practice/outpatient service" if participants specified that they worked in health care (item 1/3), in the provision of medical care to patients (item 2/3), and in an outpatient service of any type (item 3/3). "Other" was set for all different response combinations. A binary indicator for a leadership position was computed based on responses to a list of positions (T1 – T4). This indicator was set to positive if respondents selected the German equivalents to one of the following positions: "attending physician" or "chief of service." It was set to negative if "resident physician" was specified. The translations provided here are based on guidelines published by the University of Zurich (2019). If participants selected the option "other," the indicator was set to missing. As positions in general practice and outpatient services (e.g., private practices and health centers) often do not correspond with the hospital hierarchy levels specified as response options, "other" was possibly selected by a substantial share of physicians that worked outside hospitals.

A binary indicator for participants who had completed specialist training was computed based on the corresponding item included in the second and third follow-up survey. The inclusion criteria specified that participants were in their second or third year of specialty training at baseline. Therefore, the indicator for completed specialist training was irrelevant for the baseline survey and the first follow-up survey and was not included in these two survey questionnaires. A binary indicator for part-time contracts was computed for all four surveys. A variable representing the total number of job changes since starting to work as a physician was computed based on responses to the corresponding item included in the third follow-up survey.

Psychometric analyses and statistical modeling approach

The following section describes the measures and criteria that were used to evaluate the psychometric qualities of the scales featured in this study. Subsequently, the statistical modeling approach based on fixed-effects and random-intercept regressions is presented.

Psychometric analysis

Cronbach's α (Cronbach, 1951) was computed for the goal conflict scale, the work engagement scale, and the depressive symptoms scale to assess reliability. As coefficient α increases if the average correlation between scale items increases, it is considered a measure of internal consistency. Coefficient α can take on values up to 1. For research settings, coefficients between 0.7 and 0.8 were suggested as acceptable lower cut-off values (Bland & Altman, 1997; Lance, Butts, & Michels, 2006; Tavakol & Dennick, 2011). In addition to internal consistency reliability tests, factor analyses were conducted to evaluate construct validity. The factor structures of the three scales were assessed based on the estimated eigenvalues. Eigenvalues describe the share of the total scale variance that can be attributed to the computed factors (cf. DeVellis, 2017). A unidimensional factor structure characterizes scales that measure a single latent construct (i.e., such scales have only one relevant factor). Two criteria were used to determine the number of relevant factors: The cut-off for factor retention suggested by Kaiser (1960) and a comparison with simulated data, as suggested by Horn (1965). Kaiser (1960) proposed to retain all factors that explain a higher share of the total variance than the average item, i.e., that have an eigenvalue > 1. Horn (1965) suggested performing a so-called parallel analysis, i.e., a comparison with eigenvalues computed based on a simulated dataset of uncorrelated variables. The original and the simulated dataset should contain an identical number of observations and variables (i.e., items). Following this approach, factors are only retained if their estimated eigenvalues exceed those computed using the simulated data.

Statistical modeling approach

The statistical models used to analyze the presented panel data were based on fixed-effects and random-intercept regressions. By reproducing the grouped structure of panel data, these approaches aim at limiting the influence of possible confounders. In the dataset of the presented study, each participant's responses constituted a group of observations within the data; each group contained a maximum of four different sets of observations – one set for each completed survey. When estimating regression coefficients, classical ordinary least squares regression models ignore the grouped structure of panel data and pool the complete dataset (cf. Gelman & Hill, 2007).

Fixed-effects regressions are an alternative approach. Their key feature is the inclusion of a binary indicator variable for each participant (cf. Wooldridge (2016) and Stock and Watson (2019)). The indicator variables are participant-specific and absorb variance stemming from constant individual confounders (e.g., personality traits). A disadvantage of fixed-effects regression models is that stable covariates (e.g., age and gender) cannot be specified as all variance caused by such constant factors is equally absorbed by the participant indicators.

Multilevel random-intercept regression models – also referred to as mixed-effects models – constitute an alternative and are not subject to this restriction (cf. Gelman & Hill, 2007). In multilevel models, components of the principal (first-level) regression are designated as so-called random-effects and are estimated using an additional regression level. In the case of random-intercept models, the random effect is a participant-specific intercept that is estimated based on partial pooling of the data. Whether the resulting coefficient estimates are more similar to analogous no-pooling estimation results – i.e., results from ordinary least square based estimation – or to analogous complete-pooling estimation results – i.e., results from fixed-effects based estimation – depends on the number of observations in each group and the variance within and between groups. If all other influences are held constant, the results are more similar to an analogous no-pooling estimate for participants with relatively few observations and more similar to an analogous complete-pooling estimate for participants with a relatively high number of observations.

All random-intercept regression models used in the presented study included two goal conflict measures computed based on each participant's individual scores: each participant's mean goal conflict score and the corresponding survey-specific deviations from this overall average. Coefficients for both components were estimated in all models. Using the deviation from the individual average is conceptually related to using a no-pooling estimate. Likewise, this approach controls for confounding caused by unobservable variables that are both stable and participant-specific. In this study, variance caused by such confounders was absorbed by the coefficients estimated for each participant's average goal conflict score.

In addition, some regression models featured binary indicators for follow-up surveys. No indicator for the baseline survey was specified to avoid multicollinearity. Analogous to participant indicators in fixed-effects regression models – that absorbed variance caused by participant-specific confounders – the survey indicators absorbed variance caused by unobservable time-specific factors that affected all participants to an equal extent (e.g., changes in employment law).

Statistical computation and output reporting

The following section outlines how the statistical analyses were conducted and presents the guidelines used for output reporting.

Statistical computation

Stata version 13.0 (StataCorp, 2013) was used for variable and scale computations, for descriptive statistics, and for psychometric analyses. Aggregated information on missing values was extracted with the Stata package mvpatterns (Weesie, 2020). Cronbach's α was computed with the Stata command "alpha," and factor analyses were conducted with the Stata command "factor." The Stata package fapara (Ender, 2020) was used for parallel analysis, and the replication parameter for dataset simulation was set to 1000.

Fixed-effects and random-intercept regression models were estimated with R version 3.5.3 (R Core Team, 2019). The R command "lm" was used to specify fixed-effects models. Random-intercept models were implemented with the R package lme4 (Bates, Mächler, Bolker, & Walker, 2015). All random-intercept models were fitted using the restricted maximum likelihood approach. Test statistics were computed with the R packages sjstats (Lüdecke, 2019) and pbkrtest (Halekoh & Højsgaard, 2014). P-values for all random-intercept models were calculated based on the Kenward-Roger approximation (Kenward & Roger, 1997). The package haven (Wickham & Miller, 2019) was used to import the Stata dataset into R, and the package stargazer (Hlavac, 2018) was used to export output tables from R.

Significance level and decimal places

The significance level was set to 5%. In the text, estimates that fulfilled this criterion (i.e., $p < 0.05$) were referred to as *significant*. Estimates were referred to as *very significant* if they fulfilled the criterion $p < 0.001$. In output tables, significant estimates were marked with asterisks according to the following rule: * if $p < 0.05$, ** if $p < 0.01$ and *** if $p < 0.001$. Generally, two decimal places were stated when referring to results in tables and the text. As an exception to this rule, three decimal places were specified for p-values if they were lower than 0.01.

Results

An overview of the descriptive statistics of the study sample is presented in the first section. This outline is followed by summaries of the results of two dropout analyses and the psychometric analyses. Subsequently, estimates based on models that analyzed the association between work experience and goal conflict are reported (hypothesis H.1). The remaining sections focus on associations between goal conflict and job satisfaction, work engagement, and depressive symptoms (hypotheses H.2a–2c).

Descriptive statistics

Table 7 provides a summary of descriptive statistics. Overall the study sample included N = 2047 observations (T1 – T4 combined).

Sample characteristics

At baseline, participants were, on average, 30.53 years old. Both at T1 and when aggregating over the complete study duration, slightly more women participated in the study than men. At baseline, 52% of the participants were women. In the overall sample, female participants contributed 51% of all observations.

Goal conflict

The mean goal conflict score (scale range: 1 – 5) increased from a baseline value of 2.89 to a value of 3.00 at T2. At T3, the mean score was 2.98. The mean score then dropped to a value of 2.51 at T4. This tendency was mirrored by the deviation from the individual average score that was positive at T1 to T3 and negative at T4. Overall, the mean of the goal conflict score was 2.86.

Work attitude and mental health measures

The overall average of the job satisfaction score (range: 1 – 5) was 3.46. Mean job satisfaction was similar at T1, T2, and T3 and then increased at T4. Work engagement, measured on a scale that ranged from 0 to 6, was, on average, 3.62 in the overall sample. Compared to baseline, it decreased at T2 and further at T3. It then increased at T4 when the highest average work engagement score was computed. The average score of the depressive symptoms scale (range: 1 – 4) remained relatively stable during the study. Overall, its mean value was 1.82.

Work experience

At T1, pre-study job tenure (i.e., work experience prior to the baseline survey) was, on average, 31.85 months (= 2.65 years). Observational tenure (i.e., work experience collected during the study) was computed only at T3 (mean value = 31.58 months, equal to 2.63 years) and T4 (mean value = 106.52 months, equal to 8.88 years). Considerable variation was observed with respect to this variable; its standard deviation was 12.13 months (= 1.01 years) at T3 and 19.31 months (= 1.61 years) at T4. These findings mirror the relative rotation flexibility of medical specialist training programs in Germany that, in most cases, allow for program interruptions and job changes.

Professional characteristics

The indicator for a leadership position was irrelevant during the participants' residency (< 1% of all observations at T1 and T2). At T3, 4% of the participants had reached a medical leadership position in a hospital. The share of participants who had reached a leadership position increased to 58% at T4. The indicator's validity is limited for physicians working outside hospitals (e.g., in private practices or health centers) as professional roles in these types of organizations do not necessarily correspond to the hospital hierarchy; the indicator was missing for 1% of the participants at T3 and 41% of the participants at T4 when many participants worked outside hospitals.

At the end of the study, 93% of the participants had completed specialist training. The share of participants working part-time increased from 4% (at T1 and T2) to 10% at T3 and further to 36% at T4. Overall, 29% of all participants stated that they worked part-time in at least one of the surveys (50% of all female participants and 7% of all male participants). All surveys combined, 12% of all observations were contributed by participants working part-time at the respective point of data collection. The average number of job changes during the study was 1.84 in the total sample. This variable was measured in the 2014 survey. Analogous to the other variables, its overall mean was calculated as a weighted average. Therefore, the measure accounted for the fact that participants who completed all surveys contributed more observations than those who missed surveys. Negligible variation was observed when the mean number of job changes during the study was computed separately for T1 to T4 based only on the group of participants that completed the respective survey.

Table 7: Descriptive statistics

	2004 (T1) *N = 590*	**2005 (T2)** *N = 537*	**2007 (T3)** *N = 487*	**2014 (T4)** *N = 433*	**All observations (T1 - T4)** *N = 2047*
Age Mean (SD); in years	30.53 *(2.73)*	31.61 *(2.65)*	33.20 *(2.68)*	40.13 *(2.84)*	33.48 *(4.49)*
Gender	Female = 52% Male = 48%	Female = 50% Male = 50%	Female = 49% Male = 51%	Female = 52% Male = 48%	Female = 51% Male = 49%
Goal conflict (score) Mean (SD); scale range: 1-5	2.89 *(0.76)*	3.00 *(0.87)*	2.98 *(0.88)*	2.51 *(0.85)*	2.86 *(0.86)*
Goal conflict (deviation) Mean (SD)	0.03 *(0.50)*	0.12 *(0.52)*	0.10 *(0.54)*	-0.31 *(0.60)*	0.00 *(0.56)*
Goal conflict (ind. average) Mean (SD); scale range: 1-5	2.86 *(0.66)*	2.87 *(0.67)*	2.88 *(0.65)*	2.82 *(0.63)*	2.86 *(0.65)*
Job satisfaction Mean (SD); scale range: 1-5	3.39 *(0.96)*	3.38 *(0.95)*	3.39 *(1.03)*	3.71 *(0.88)*	3.46 *(0.97)*
Work engagement Mean (SD); scale range: 0-6	3.63 *(0.99)*	3.59 *(0.99)*	3.55 *(1.01)*	3.74 *(1.03)*	3.62 *(1.00)*
Depressive symptoms Mean (SD); scale range: 1-4	1.83 *(0.48)*	1.79 *(0.48)*	1.84 *(0.50)*	1.80 *(0.49)*	1.82 *(0.49)*
Observational tenure Mean (SD); in months	/	not included in survey	31.58 *(12.13)*	106.52 *(19.31)*	64.88 *(40.44)*
Pre-study tenure Mean (SD); in months	31.85 *(14.12)*	31.73 *(13.89)*	31.64 *(12.77)*	31.98 *(13.85)*	31.80 *(13.68)*
Leadership position	< 1%	< 1%	4%	58%	9%
Specialist training completed	not included in survey	not included in survey	12%	93%	50%
Working part-time	4%	4%	10%	36%	12%
Total job changes at T4 Mean (SD); in months	1.84 *(1.45)*	1.85 *(1.47)*	1.84 *(1.44)*	1.83 *(1.42)*	1.84 *(1.44)*

Note. SD = standard deviation

Medical specialties

Table 8 illustrates that the participants worked in various branches of medicine. Most participants trained in internal medicine (27%) followed by surgery (13%) and anesthesiology (11%).

Table 8: Medical specialties at baseline

	Frequency	Percentage
Internal medicine	160	27%
Surgery	79	13%
Anesthesiology	63	11%
Pediatrics	40	7%
Obstetrics and gynaecology	39	7%
Family medicine	36	6%
Neurology	33	6%
Psychiatry	23	4%
Dermatology	19	3%
Radiology	15	3%
Other specialties	83	14%

Note. Top 10 medical specialties based on baseline survey data from 2004; N = 590; not summing up to 100% due to rounding

Workplace type

Table 9 reports the workplace type of participants. All participants worked in hospitals at baseline (inclusion criterion). In the follow-up surveys, a substantial number of participants reported that they work in outpatient services (e.g., private practices, health centers) and in other types of organizations (e.g., pharmaceutical industry, public health authorities). In the last follow-up survey, 54% of all participants worked in hospitals, 35% worked in outpatient services, and 12% in other fields. These results reflect that experienced physicians in Germany often work outside hospitals, for example, as self-employed physicians in general practices (Bundesärztekammer, 2018). Over the course of the study, participants working in hospitals at the time of survey contributed 86% of all sets of observations. 11% of all sets were contributed by physicians working in outpatient services and 3% by physicians working in other types of organizations.

Table 9: Workplace type (baseline and follow-up surveys)

	Baseline (T1)	First follow-up (T2)	Second follow-up (T3)	Third follow-up (T4)
Hospital	100% (inclusion criterion)	94%	89%	54%
OS	0%	5%	10%	35%
Other	0%	< 1%	1%	12%

Note. OS = outpatient service (e.g., private practice, health center); not summing up to 100% for some surveys due to rounding

Dropout analyses

Two dropout analyses were conducted to assess the risk of systematic bias due to non-random sampling or non-random attrition; the study sample was compared with the initial address pool, and dropouts were compared with participants that completed all four surveys.

Comparison of the study sample with the initial pool

Data on gender were available for the initial address pool. In the initial pool, N = 482 physicians were female (48%) and N = 518 were male (52%). Table 10 lists the results of chi-squared tests that compared the gender distribution in the baseline study sample with the gender distribution in the group of physicians who did not return questionnaires and with the group of physicians whose responses were removed due to incomplete data. The gender distribution differed significantly between the study sample and a combined group of non-responders and respondents who provided incomplete data. Significant differences were also found when comparing the study sample with non-responders only. These two results suggest that women were slightly overrepresented in the study sample. No significant gender distribution differences were found when comparing the respondents who provided incomplete data with either the study sample or the non-responder group. However, the low number of observations in the incomplete data group (N = 18 men and N = 13 women) reduced the statistical power of these two tests.

Table 10: Gender distribution comparison at baseline

Compared Groups	χ2	p	Significance level
Did not respond or incomplete response & Study sample	7.04	0.008	**
Did not respond & Study sample	6.56	0.01	*
Incomplete response & Study sample	1.12	0.29	n.s.
Did not respond & Incomplete response	0.02	0.89	n.s.

Note. p = p-value; ** = p < 0.01, * = p < 0.05, n.s. = not significant

Assessment of the risk of bias due to non-random attrition

A second dropout analysis was conducted to assess the risk of bias due to non-random attrition during the study. Table 11 states the results of this analysis. In the study sample, N = 334 participants participated in all four surveys. Responses from N = 256 participants were missing for at least one survey.

Statistical comparisons were made between these two groups. Between-group differences were evaluated for all nine key study variables assessed in the baseline survey: age, gender, goal conflict, job satisfaction, work engagement, depressive symptoms, pre-study tenure, leadership position, and working part-time. Differences in mean values were evaluated using t-tests with two-tailed p-values. For categorical variables, chi-squared tests were used to compare frequencies.

As reported in Table 11, no significant differences were found between the two groups at baseline.

Table 11: Attrition analysis

	Participants who completed all surveys *N = 334*	**Participants with missing data for at least one survey** *N = 256*	**Group comparison** *H_0 = No difference*
Age at T1 Mean (SD), in years	30.40 *(2.75)*	30.71 *(2.69)*	t = -1.38 *(p = 0.17, df = 588)*
Gender N (%)	Female: 170 (51%) Male: 164 (49%)	Female: 135 (53%) Male: 121 (47%)	$\chi 2$ = 0.20 *(p = 0.66, df= 1)*
Goal conflict at T1 Mean (SD)	2.87 *(0.77)*	2.92 *(0.76)*	t = -0.70 *(p = 0.48, df = 588)*
Job satisfaction at T1 Mean (SD)	3.41 *(0.96)*	3.36 *(0.96)*	t = 0.58 *(p = 0.56, df = 587)*
Work engagement at T1 Mean (SD)	3.62 *(0.99)*	3.63 *(0.98)*	t = -0.04 *(p = 0.97, df = 588)*
Depressive symptoms at T1 Mean (SD)	1.85 *(0.48)*	1.82 *(0.49)*	t = 0.76 *(p = 0.45, df = 586)*
Pre-study tenure Mean (SD), in months	31.79 *(14.57)*	31.92 *(13.54)*	t = -0.10 *(p = 0.92, df = 588)*
Leadership position at T1 N (%)	0 (0%)	1 (< 1%)	$\chi 2$ = 1.31 *(p = 0.25, df= 1)*
Working part-time at T1 N (%)	13 (4%)	9 (4%)	$\chi 2$ = 0.05 *(p = 0.82, df= 1)*

Note. The group comparison column states results from t-tests (t, with two-tailed p-values) or chi-squared tests ($\chi 2$, for nominal scales); p = p-value, SD = standard deviation, df = degrees of freedom

Psychometric analyses

The following analyses were conducted to evaluate the psychometric properties of the used survey instruments. Cronbach's alpha was computed as a measure of internal consistency for the goal conflict scale and for the scales used to measure the outcomes work engagement and depressive symptoms. In addition, factor analyses were performed for all three scales.

Internal consistency

Table 12 summarizes the computed coefficients α for the three scales. For the goal conflict scale, a value of $\alpha = 0.71$ was computed at baseline. Higher values were computed for the three follow-up surveys (range $\alpha = 0.81 – 0.83$) for which the item wording was simplified and the four items were presented in one continuous section. The coefficient obtained for 2004 might have been deflated due to the more specific item wording used in the first survey. Table 13 illustrates how the wording was altered using item 2, the item that was most markedly changed and had the lowest item-rest correlation in 2004 (respective correlation coefficient = 0.47). The scope of the first version of this item was limited to quantitative demands stemming from the number of patients for which physicians had to care for; the exclusion of physicians that faced quality-versus-quantity dilemmas that were similar but involved other types of tasks (e.g., administrative, research, or education-related) could have increased error variance. Overall, the goal conflict scale demonstrated acceptable internal consistency at baseline and in all three follow-up surveys ($\alpha > 0.70$). The broader wording used in later surveys performed better than the version used for the baseline survey.

For all four surveys, high values for α were computed for the work engagement scale (range $\alpha = 0.90 – 0.92$). The fact that one of the nine items used to construct the work engagement scale differed from the official version of the UWES-9 does not seem to have affected reliability; internal consistency estimates computed for the presented data were similar to those reported for the official version (cf. Sautier et al. (2015) and Schaufeli, Bakker, and Salanova (2006)). Likewise, high Cronbach's alpha values ranging from $\alpha = 0.88$ to 0.90 were computed for the depressive symptoms scale.

Table 12: Cronbach's alpha of the scales used in the survey questionnaires

SCALE	T1 *Baseline*	T2 *Follow-up 1*	T3 *Follow-up 2*	T4 *Follow-up 3*
Goal Conflict	α = 0.71	α = 0.83	α = 0.82	α = 0.81
Work Engagement	α = 0.90	α = 0.91	α = 0.92	α = 0.92
Depressive Symptoms	α = 0.88	α = 0.89	α = 0.90	α = 0.89

Note. α = Cronbach's alpha

Table 13: Wording of item 2 of the goal conflict scale

Survey	Wording
Baseline (T1)	When working in this department one has to take care of that many patients so that the quality of the work one does suffers
Follow-up surveys (T2 – T4)	Again and again one has/I have [T2 & T3 / T4] so much to do so that the quality of the work one does/I do [T2 & T3 / T4] suffers

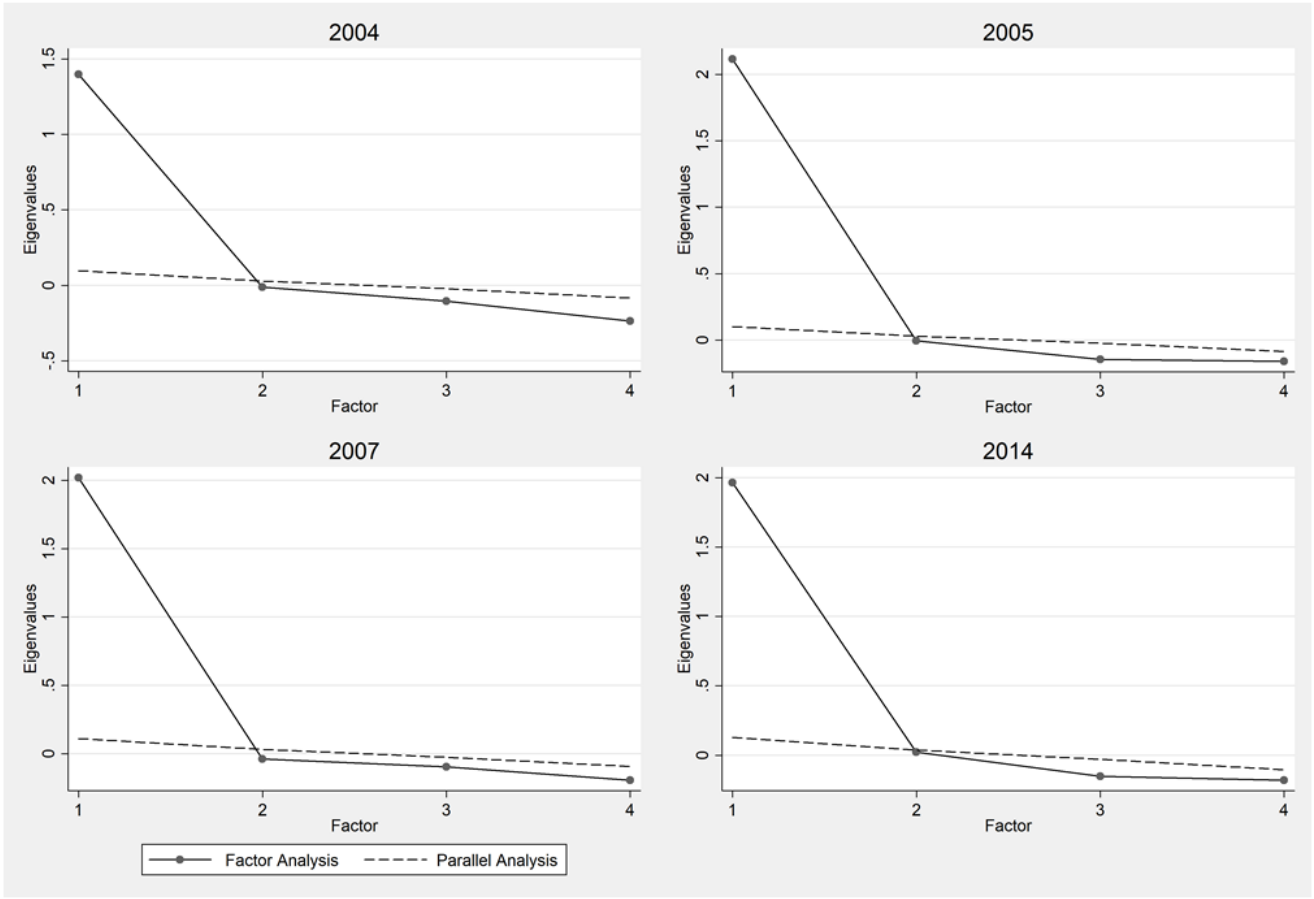
Note. Author's own translations from German; wording modifications underlined by the author

Factor analyses

Results of the factor analyses performed for the goal conflict scale are presented in Figure 5. In the baseline survey and all three follow-up surveys, only one factor had an eigenvalue > 1. Comparing the actual eigenvalues with the parallel analysis eigenvalues showed that for only one factor the actual eigenvalue was higher than the simulated eigenvalue. This pattern was found consistently in the analyses of all four surveys. Overall, these findings suggest that the goal conflict scale measured one single latent construct.

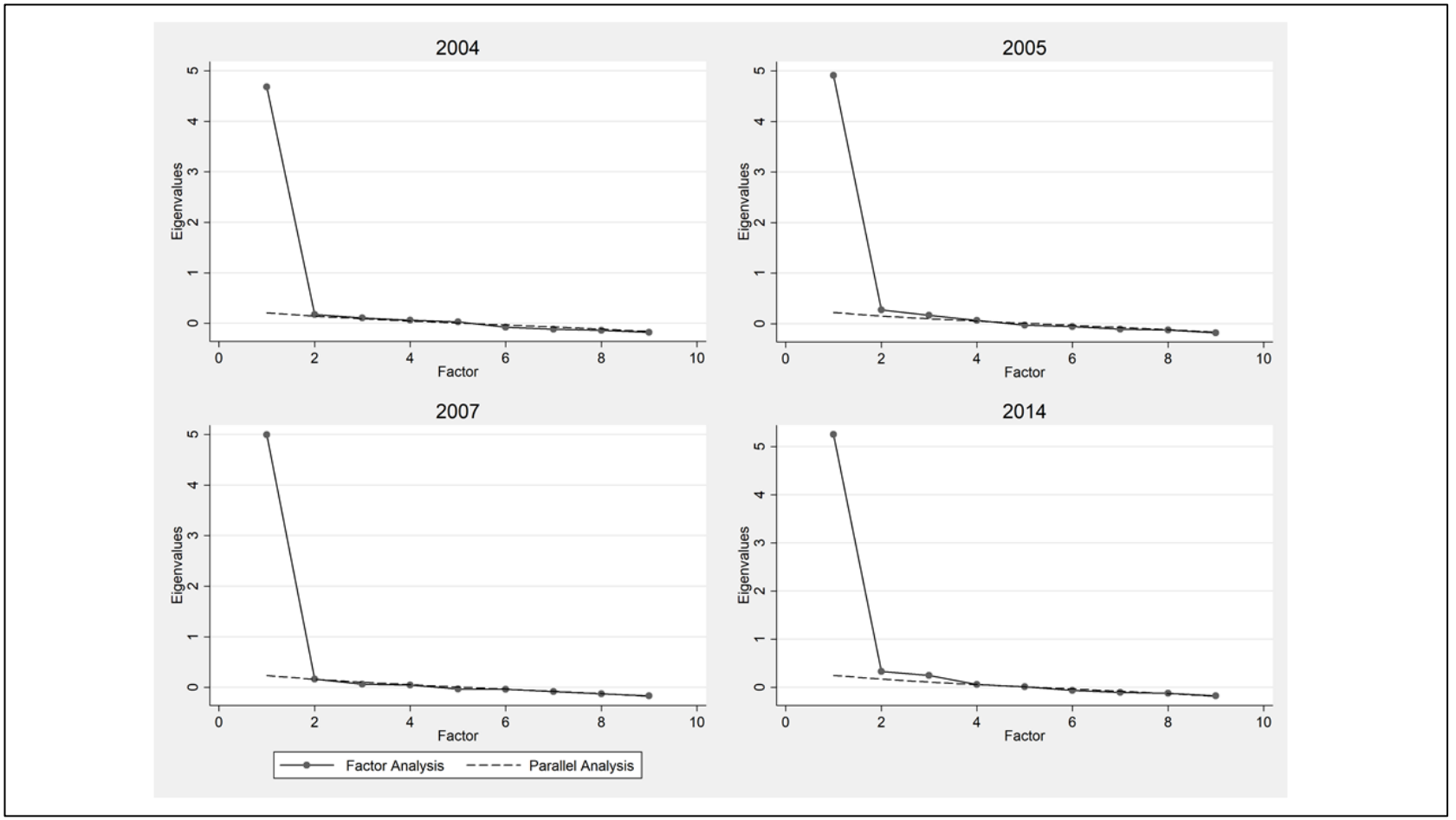
Figure 6 summarizes factor analyses performed for the work engagement scale. Only one factor had an actual eigenvalue that was distinctly higher than the respective simulated value and fulfilled the > 1 criterion. This result was found in the analysis of the baseline survey and the analyses of all three follow-up surveys. Analogous analyses for the depressive symptoms scale are reported in Figure 7. Once more, similar results were found for the baseline survey and the three follow-up surveys, and only one factor had an actual eigenvalue > 1. However, comparing the actual eigenvalues with the simulated eigenvalues pointed to a potential second factor; the factor loading patterns seemed to map the two subscales for euthymia and dysthymia that together constitute the depressive symptoms scale.

Figure 5: Factor analyses of the goal conflict scale (baseline and follow-up surveys)



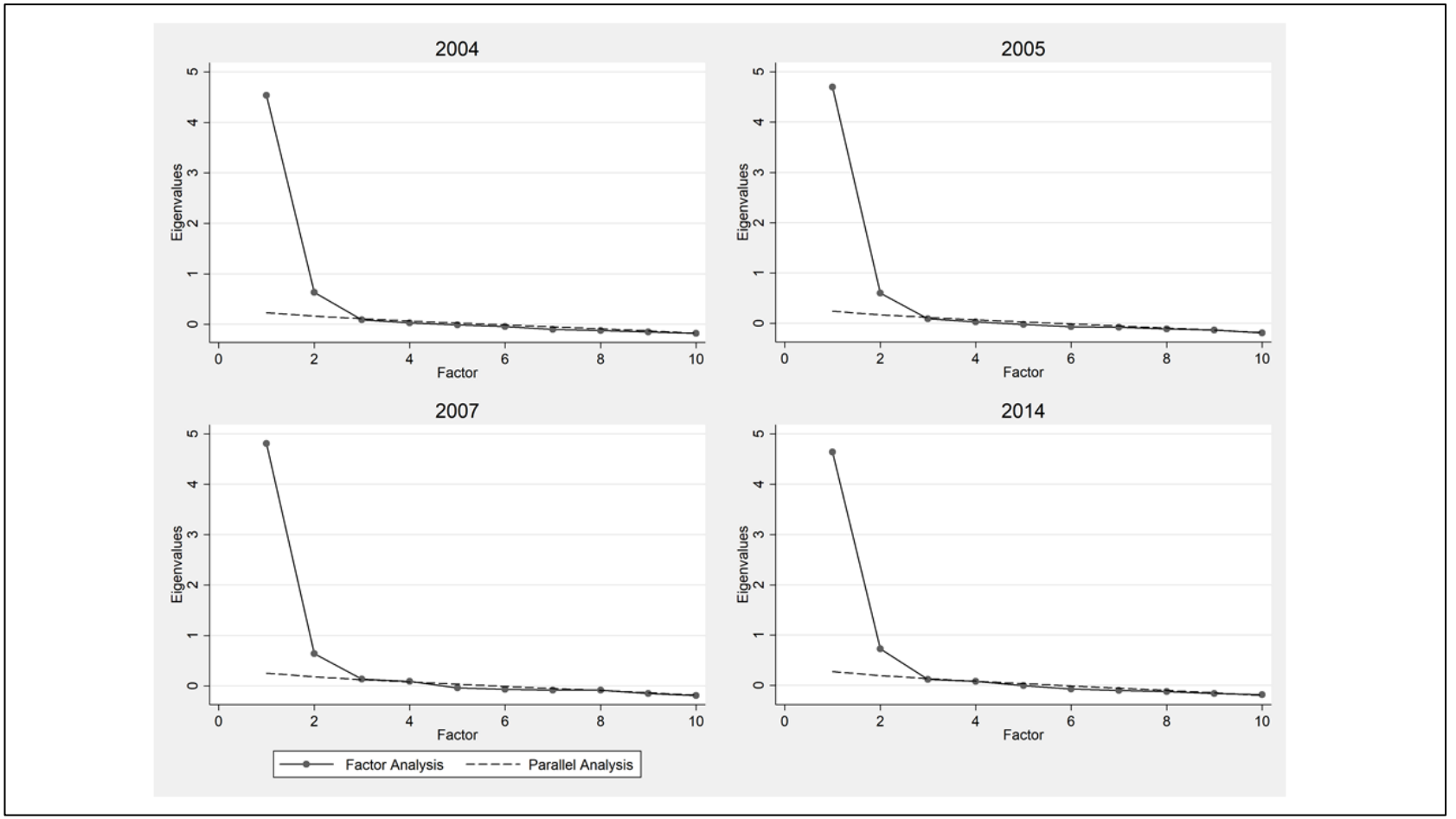
Note. The actual eigenvalues obtained for the studied sample are represented by solid dots and are connected by a solid line; a dashed line represents simulated eigenvalues generated from random data (parallel analysis)

Figure 6: Factor analyses of the work engagement scale (baseline and follow-up surveys)



Note. The actual eigenvalues obtained for the studied sample are represented by solid dots and are connected by a solid line; a dashed line represents simulated eigenvalues generated from random data (parallel analysis)

Figure 7: Factor analyses of the depressive symptoms scale (baseline and follow-up surveys)



Note. The actual eigenvalues obtained for the studied sample are represented by solid dots and are connected by a solid line; a dashed line represents simulated eigenvalues generated from random data (parallel analysis)

H.1: Work experience and goal conflict

Ten models were estimated to test whether perceived goal conflict decreased with increasing work experience. Figure 8 illustrates the basic idea behind these models. Observational tenure between T1 and T3 was computed for each participant based on its responses in the T3 survey and then associated with goal conflict at T3. Likewise, observational tenure between T1 and T4 was measured at T4 and then related to goal conflict at this point. Therefore, all models described in this section were restricted to participants that completed the 2007 survey or the 2014 survey (or both). The number of observations per model ranged between N = 590 and N = 871.

Overview of the tested models

A synopsis of the variables used in the 10 models is given in Table 14. All models that were labeled with even numbers featured an indicator for the year 2014 that captured overall time trends. Otherwise, each of these models was identical to the model that was labeled with the preceding odd number (e.g., the indicator was the only difference between Model 2 and Model 1). Model 1 and Model 2 were fixed-effects models that used observational tenure as a single explanatory variable. Model 3 and Model 4 were random-intercept models that replicated Model 1 and Model 2. In addition, pre-study job tenure was included to control for work experience prior to the start of the study. Model 5 and Model 6 also featured personal characteristics as covariates. The list of control variables was further extended for Model 7 and Model 8; these models also included all professional characteristics except for the indicator for a leadership position. The latter was added in Model 9 and Model 10. The indicator for a leadership position was integrated separately due to its high number of missing values. Consequently, Model 9 and Model 10 estimates were based on N = 590 observations only compared to N = 736 observations for Model 7 and Model 8 estimates (ΔN = 146 observations, 20% of the total for Model 7 and Model 8).

Figure 8: Model outline for H.1

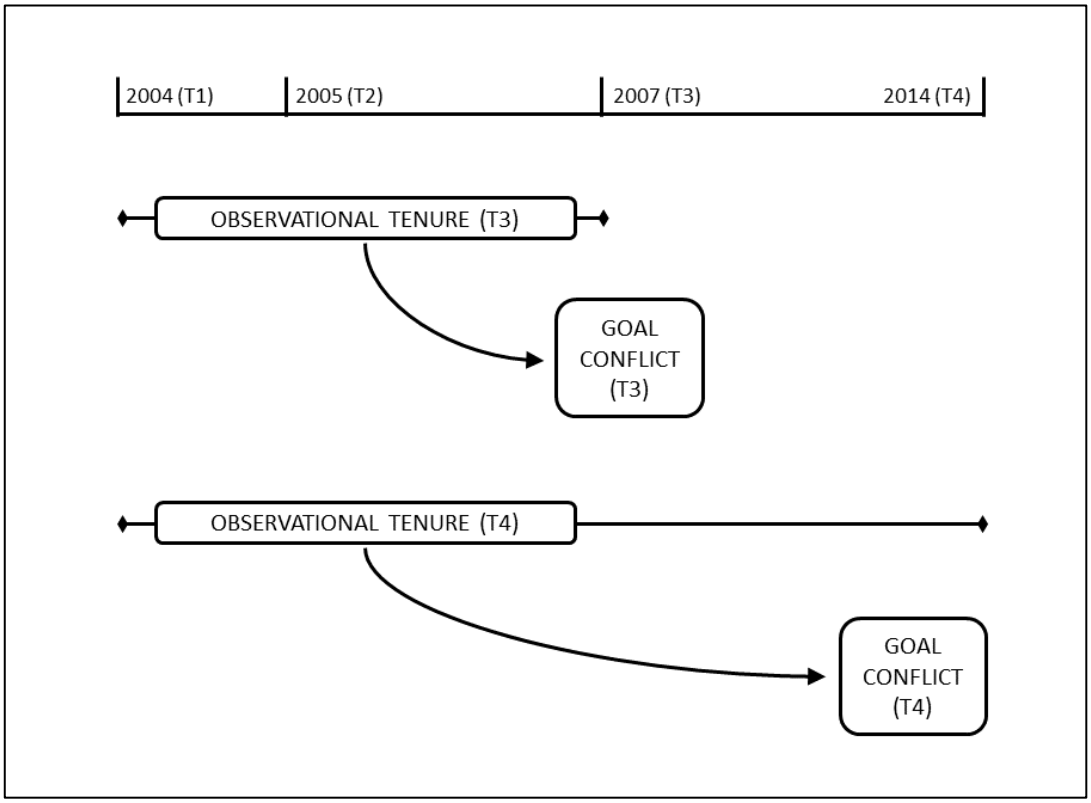


Table 14: Model overview for H.1

Model Number	**#1**	**#2**	**#3**	**#4**	**#5**	**#6**	**#7**	**#8**	**#9**	**#10**
Surveys used	*T3 - T4*	*T3 - T4*	*T3 - T4*	*T3 - T4*	*T3 - T4*	*T3 - T4*	*T3 - T4*	*T3 - T4*	*T3 - T4*	*T3 - T4*
Type	*FE*	*FE*	*RI*	*RI*	*RI*	*RI*	*RI*	*RI*	*RI*	*RI*
Observational tenure										
Pre-study tenure										
Personal characteristics										
Professional characteristics *except leadership position*										
Leadership position										
Year-2014-indicator										

Note. FE = fixed-effects model; RI = random-intercept model; dark grey box = variable(s) stated in the first column was/were used in the model stated in the first row; white box = variable(s) stated in the first column was/were not used in the model stated in the first row

Coefficient estimates for work experience

Coefficient estimates for the variables representing work experience are stated in the output summaries presented in Table 15 and Table 16. The point estimates for the coefficients of observational tenure were negative in all five models that did not feature the indicator for the year 2014 (i.e., Model 1, Model 3, Model 5, Model 7, and Model 9). These estimates were significant in four of these five models and very significant in three of these five models; only the estimate computed in Model 9 was not significant. These findings suggest that increased work experience was associated with decreased goal conflict. Verbally, the point estimate of -0.01 computed in Model 1 – Model 3 implies that a one-month increase in observational tenure was associated with a 0.01-units decrease in goal conflict (scale range 1 – 5), all other factors within the framework of the models left unchanged.

Once the indicator for the year 2014 was added to the models, it captured the average increase of work experience between 2007 and 2014 (Model 2, Model 4, Model 6, Model 8, and Model 10). Coefficient estimates for the indicator were negative, mapping the overall decrease in perceived goal conflict during the study. Except for the estimate in Model 10, all coefficient estimates for the indicator were either very significant (Model 2, Model 4, and Model 6) or significant (Model 8). In the models with the indicator for the year 2014, the coefficient of the observational tenure variable described the effect of individual deviations from the average increase in work experience. Compared with estimates in the models without this indicator, the point estimates for the coefficients of observational tenure changed their direction in the models with the indicator. The then positive estimates might suggest that participants with an above-average increase in work experience perceived a higher degree of goal conflict. However, only one of these coefficient estimates was significant (Model 2), while the remaining four estimates were not significant (Model 4, Model 6, Model 8, Model 10). Models 3 – Model 10 also featured pre-study job tenure; this variable represented work experience prior to the start of the study. All coefficient estimates for this variable were not significant.

Covariates

Coefficient estimates for most of the specified covariates were not significant. However, the coefficient estimates for the indicator representing female participants were significant and negative in Model 5 and Model 6, and the coefficient estimates for the indicator representing participants who worked part-time were significant and negative in Model 7 and Model 9. However, the coefficient estimate for gender was not significant in the higher-order models (Model 7 – Model 10), and the coefficient estimate for participants who worked part-time was not significant in the models with the indicator for 2014 (Model 8 and Model 10).

Table 15: Work experience and goal conflict (1/2)

	Model 1 *Fixed-Effects*	**Model 2** *Fixed-Effects*	**Model 3** *Random-Intercept*	**Model 4** *Random-Intercept*
Observational tenure	**-0.01***** p < 0.001	**0.01**** p = 0.002	**-0.01***** p < 0.001	< 0.01 p = 0.10
Pre-study tenure			< 0.01 p = 0.18	< 0.01 p = 0.13
Year 2014		**-1.38***** p < 0.001		**-0.69***** p < 0.001
Surveys used	T3,T4	T3,T4	T3,T4	T3,T4
Observations	871	871	871	871

Note. Depending variable: Goal conflict score; independent variable names are printed in bold in the left column; significant estimates are printed in bold, *** = p < 0.001, ** = p < 0.01, * = p < 0.05

Table 16: Work experience and goal conflict (2/2)

	Model 5 *Random-Intercept*	**Model 6** *Random-Intercept*	**Model 7** *Random-Intercept*	**Model 8** *Random-Intercept*	**Model 9** *Random-Intercept*	**Model 10** *Random-Intercept*
Observational tenure	**-0.01***** p < 0.001	< 0.01 p = 0.13	**- < 0.01*** p = 0.03	< 0.01 p = 0.46	- < 0.01 p = 0.13	< 0.01 p = 0.91
Pre-study tenure	< 0.01 p = 0.17	0.01 p = 0.06	0.01 p = 0.08	0.01 p = 0.06	< 0.01 p = 0.20	< 0.01 p = 0.17
Gender Female = 1	**-0.22**** p = 0.001	**-0.16*** p = 0.02	-0.12 p = 0.13	-0.12 p = 0.15	-0.07 p = 0.42	-0.07 p = 0.41
Part-time			**-0.22**** p = 0.007	-0.14 p = 0.11	**-0.22*** p = 0.02	-0.17 p = 0.11
Year 2014		**-0.68***** p < 0.001		**-0.45*** p = 0.03		-0.26 p = 0.27
Covariates without significant estimates	Age at T1	Age at T1	Age at T1 Specialist Job changes	Age at T1 Specialist Job changes	Age at T1 Specialist Job changes Leadership position	Age at T1 Specialist Job changes Leadership position
Surveys used Observations	T3, T4 871	T3, T4 871	T3, T4 736	T3, T4 736	T3, T4 590	T3, T4 590

Note. Depending variable: Goal conflict score; independent variable names are printed in bold in the left column, “covariates without significant estimates” lists additional covariates that were included in the model, p was > 0.05 for these estimates, “job changes" is the total number of job changes between 2004 and 2014 (reported in 2014); significant estimates are printed in bold, *** = p < 0.001, ** = p < 0.01, * = p < 0.05

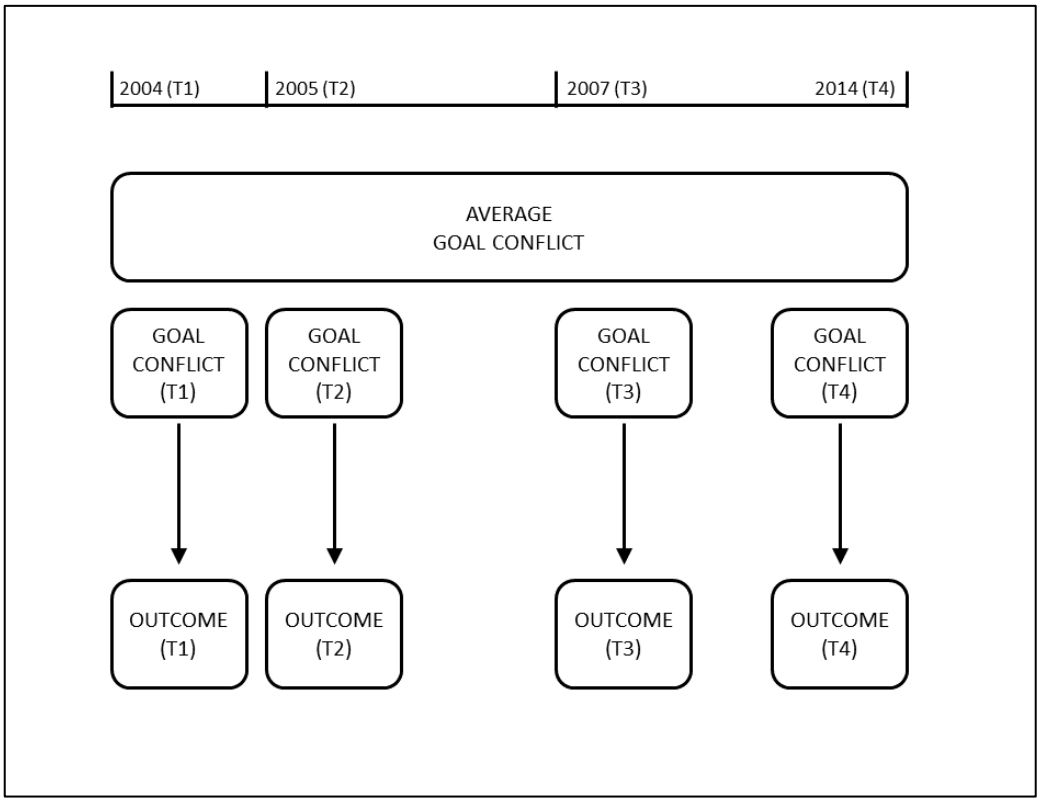
H.2a – 2c: Goal conflict, job satisfaction, work engagement, and depressive symptoms

Models used to estimate the association between goal conflict and job satisfaction, work engagement, and depressive symptoms (hypotheses H.2a – 2c) shared a common structure. Figure 9 outlines the concept. Goal conflict was measured at T1 – T4 and related to the outcomes as measured at the same time (e.g., goal conflict at T1 was related to job satisfaction at T1). In addition, the average goal conflict score during the study was computed for each participant and was also related to the outcome measures. In most of the specifications, the models used the entire study sample and included observations from 2004 to 2014.

Overview of the tested models

The variable pattern of the 11 models tested for each of the three hypotheses H.2a – 2c is described in Table 17. Model 1 was a fixed-effects model that used the total goal conflict score as a single explanatory variable. Model 2 was the most basic random-intercept model. Analogous to the higher-order random-intercept models, Model 2 featured two variables representing goal conflict: each participant's individual average goal conflict score based on the entire study duration (to evaluate between-participants variation) and each participant's time-specific deviation from this average (to evaluate within-participant variation).

Figure 9: Model outline for H.2a – 2c



The higher-order models were all random-intercept models. They augmented Model 2 by integrating additional covariates. For Model 3, the variables age and gender were added. Variables for professional characteristics were integrated into the framework in two steps: Model 4 featured the variable representing the total number of job changes and the indicators for a leadership position and for participants who worked part-time. Model 5 also featured the indicator for participants who had completed specialist training and the two variables representing work experience (observational tenure and pre-study job tenure). Model 5 was estimated based only on data provided by participants that completed the 2007 and the 2014 survey, as both the variable representing observational tenure and the indicator for participants who had completed specialist training were only measured in those two surveys.

Model 6 combined Model 3 with Model 4. Model 7 was a combination of Model 3 and Model 5. Analogous to Model 5, the estimation of Model 7 was based only on data provided by participants of the 2007 and 2014 surveys.

Model 8 – Model 11 were variations of the basic model that specified interaction terms in an attempt to characterize the relationship between goal conflict and the tested outcomes in a more nuanced way. The interaction terms were chosen based on the coefficients estimated for the models described above. Interaction terms with goal conflict were specified for all covariates with significant coefficient estimates in at least one of those seven models (i.e., Model 1 – 7): Goal conflict * Age at baseline (Model 8), Goal conflict * Gender (Model 9), Goal conflict * Job changes (Model 10), and Goal conflict * Leadership position (Model 11). For the interaction terms, goal conflict was specified as the time-specific deviation from each participant's individual average goal conflict score.

All seven models without interaction terms and all four models with interaction terms featured indicator variables for survey years that captured general time trends affecting all participants to an equal extent. Models 5 and 7 featured only an indicator for 2014, as they were based only on observations provided by participants of the 2007 and 2014 surveys. All other models featured indicators for 2005, 2007, and 2014. In these models, observations from the 2004 survey were identified by setting those three indicators equal to zero.

Table 17: Model overview for H.2a – 2c

Model Number	**#1**	**#2**	**#3**	**#4**	**#5**	**#6**	**#7**	**#8**	**#9**	**#10**	**#11**
Surveys used	*T1 - T4*	*T1 - T4*	*T1 - T4*	*T1 - T4*	*T3 - T4*	*T1 - T4*	*T3 - T4*	*T1 - T4*	*T1 - T4*	*T1 - T4*	*T1 - T4*
Type	*FE*	*RI*	*RI*	*RI*	*RI*	*RI*	*RI*	*RI*	*RI*	*RI*	*RI*
Goal conflict (GC) *Score*											
Goal conflict (GC) *Deviation/Average*											
Personal characteristics								Age	Gender		
Job changes (JC), Part-time, Leadership position (LP)										JC	LP
Work experience, Specialist training completed											
Interaction terms								GC * Age	GC * Gender	GC * JC	GC * LP
Year Indicators					2014		2014				

Note. FE = fixed-effects model; RI = random-intercept model; dark grey box = (all) variable(s) stated in the first column was/were used in the model stated in the first row; light grey box = some variables stated in the first column were used in the model stated in the first row, variables used are stated within the box; white box = variable(s) stated in the first column was/were not used in the model stated in the first row

H.2a in detail: Goal conflict and job satisfaction

Table 18 and Table 19 are summaries of the seven models without interaction terms that tested whether changes in goal conflict were associated with changes in job satisfaction (Model 1 – Model 7). Model 1 – Model 3 used observations from the baseline survey and all three follow-up surveys (N = 2042). Two main factors contributed to the reduction of the number of observations used for estimating Model 4 and Model 6 (N = 1418). Firstly, as the variable representing the total number of job changes was measured at T4, the models featuring this covariate were estimated based only on data provided by participants that completed the 2014 survey. Secondly, including the indicator for leadership position reduced the number of observations in the respective models (see above). The number of observations was further reduced in Model 5 and Model 7 (N = 588), as both observational tenure and the indicator for participants who had completed specialist training were measured only in 2007 and 2014.

Coefficient estimates for goal conflict

All 13 coefficient estimates for the variables representing goal conflict in the seven models without interaction terms were negative and very significant. These findings suggest that increased goal conflict was associated with decreased job satisfaction. The fixed-effects point estimate for the coefficient of the variable representing the total goal conflict score was -0.57 (Model 1). In the most basic random-intercept model (Model 2), the analogous estimate for the coefficient of the variable representing goal conflict deviation was identical. Overall, the coefficient estimates for goal conflict deviation ranged between -0.56 (Model 4 and Model 6) and -0.62 (Model 5) in the models without interaction terms. Coefficient estimates for average goal conflict ranged between -0.45 (Model 5) and -0.53 (Model 3). The coefficient estimates for goal conflict deviation were slightly higher than the estimates for average goal conflict (Model 2 – Model 7). Verbally, a coefficient estimate of -0.57 (Model 1 – Model 3) implies that a one-point increase in goal conflict (scale range: 1 – 5) was associated with a 0.57-points decrease in job satisfaction (scale range 1 – 5), ceteris paribus.

Covariates and year indicators

In the seven models without interaction terms, the coefficient estimates for the covariates gender, total job changes, age at baseline, and leadership position were significant in some or all specifications. The estimates for the coefficients of the indicator representing female participants and the variable representing the total number of job changes were negative and significant in all models that included either one or both of these covariates (Model 3 – Model 7). A negative and significant coefficient estimate for age at baseline was computed in two of the three models

featuring this variable (Model 3, Model 6), with Model 7 being the exception. The estimate for the coefficient of the indicator for a leadership position was significant and positive in the models restricted to participants of the 2007 and 2014 surveys (Model 5 and 7). In contrast, coefficient estimates for this covariate were not significant in models that included more observations (Model 4 and 6). The estimate for the coefficient of the indicator for the year 2014 was significant in Model 2 and Model 3, while the respective estimate was not significant in all other models without interaction terms. The estimates for the coefficients of all other indicators for survey years were not significant.

Models with interaction terms

Estimation results for the four models that featured interaction terms (Model 8 – Model 11) are stated in Table 20. Significant and negative coefficient estimates were computed for the interaction terms Goal conflict * Age at baseline (Model 8) and Goal conflict * Job changes (Model 10). No significant estimates were computed for the interaction terms Goal conflict * Gender (Model 9) and Goal conflict * Leadership position (Model 11). Coefficients for interaction terms must be interpreted together with the additional coefficients of the two variables that were interacted. Taking Model 10 as an example, a one-point increase in goal conflict deviation (scale range 1 – 5) was associated with a 0.43-points decrease in job satisfaction (scale range 1 – 5) for a participant who had never changed its job, ceteris paribus ([1*-0.43] + [0*-0.08] + [1*0*-0.06]). For a participant who had changed its job once, the same 1-point increase in goal conflict deviation was associated with a 0.57-points decrease in job satisfaction ([1*-0.43] + [1*-0.08] + [1*1*-0.06]). The estimated negative effect was further aggravated for participants with more job changes (-0.71 for two job changes, -0.85 for three job changes, and so forth). Overall, the significant coefficient estimates for the interaction terms in Model 8 and Model 10 suggest that the negative association between goal conflict and job satisfaction was intensified if participants were either older or had more job changes (or both). The coefficient estimates for average goal conflict were negative and very significant in all four models with interaction terms and comparable to analogous estimates from the previously described models without interaction terms.

Table 18. Goal conflict and job satisfaction (1/3)

	Model 1 *Fixed-Effects*	**Model 2** *Random-Intercept*	**Model 3** *Random-Intercept*	**Model 4** *Random-Intercept*
Goal Conflict Score	**-0.57***** p < 0.001			
Goal Conflict Deviation		**-0.57***** p < 0.001	**-0.57***** p < 0.001	**-0.56***** p < 0.001
Goal Conflict Average		**-0.51***** p < 0.001	**-0.53***** p < 0.001	**-0.50***** p < 0.001
Age at T1			**-0.04***** p < 0.001	
Gender Female = 1			**-0.14**** p = 0.002	
Job changes				**-0.09***** p < 0.001
Year 2014	0.09 p = 0.07	**0.10*** p = 0.04	**0.10*** p = 0.04	-0.08 p = 0.40
Covariates without significant estimates	Year 2005 Year 2007	Year 2005 Year 2007	Year 2005 Year 2007	Year 2005 Year 2007 Part-time Leadership position
Surveys used	T1 - T4	T1 - T4	T1 - T4	T1 - T4
Observations	2.042	2.042	2.042	1.418

Note . Depending variable: Job satisfaction; independent variable names are printed in bold in the left column, “goal conflict deviation” is the time-specific deviation from the individual average of the goal conflict score, “goal conflict average” is the individual average of the goal conflict score, “job changes" is the total number of job changes between 2004 and 2014 (reported in 2014), “covariates without significant estimates” lists additional covariates that were included in the model, p was > 0.05 for these estimates; significant estimates are printed in bold, *** = p < 0.001, ** = p < 0.01, * = p < 0.05

Table 19. Goal conflict and job satisfaction (2/3)

	Model 5 *Random-Intercept*	**Model 6** *Random-Intercept*	**Model 7** *Random-Intercept*
Goal Conflict Deviation	**-0.62***** p < 0.001	**-0.56***** p < 0.001	**-0.61***** p < 0.001
Goal Conflict Average	**-0.45***** p < 0.001	**-0.52***** p < 0.001	**-0.47***** p < 0.001
Age at T1		**-0.02*** p = 0.04	-0.03 p = 0.06
Gender Female = 1		**-0.15*** p = 0.01	**-0.19*** p = 0.02
Job changes	**-0.09**** p = 0.001	**-0.08***** p < 0.001	**-0.08**** p = 0.002
Leadership position	**0.28*** p = 0.01	0.11 p = 0.26	**0.24*** p = 0.04
Covariates without significant estimates	Part-time Specialist Observational tenure Pre-study tenure Year 2014	Part-time Year 2005 Year 2007 Year 2014	Part-time Specialist Observational tenure Pre-study tenure Year 2014
Surveys used	T3, T4	T1 - T4	T3, T4
Observations	588	1.418	588

Note. Depending variable: Job satisfaction; independent variable names are printed in bold in the left column, “goal conflict deviation” is the time-specific deviation from the individual average of the goal conflict score, “goal conflict average” is the individual average of the goal conflict score, “job changes" is the total number of job changes between 2004 and 2014 (reported in 2014), “covariates without significant estimates” lists additional covariates that were included in the model, p was > 0.05 for these estimates; significant estimates are printed in bold,
*** = p < 0.001, ** = p < 0.01, * = p < 0.05

Table 20: Goal conflict and job satisfaction (3/3)

	Model 8		Model 9		Model 10		Model 11
Goal conflict Deviation	0.50 p = 0.16	**Goal conflict** Deviation	**-0.52*** ** p < 0.001	**Goal conflict** Deviation	**-0.43*** ** p < 0.001	**Goal conflict** Deviation	**-0.59*** ** p < 0.001
Goal conflict Average	**-0.52*** ** p < 0.001	**Goal conflict** Average	**-0.52*** ** p < 0.001	**Goal conflict** Average	**-0.51*** ** p < 0.001	**Goal conflict** Average	**-0.51*** ** p < 0.001
Year 2014	**0.10*** p = 0.046	**Year 2014**	**0.10*** p = 0.045				
Age at T1	**-0.03*** ** p < 0.001	**Gender** Female = 1	**-0.12*** p = 0.01	**Job changes**	**-0.08*** ** p < 0.001	**Leadership position**	0.11 p = 0.23
Age at T1* **Goal conflict deviation**	**-0.04**** p = 0.003	**Gender*** **Goal conflict deviation**	-0.09 p = 0.13	**Job changes*** **Goal conflict deviation**	**-0.06**** p = 0.005	**Leadership position*** **Goal conflict deviation**	0.01 p = 0.93
Covariates without significant estimates	Year 2005 Year 2007	Covariates without significant estimates	Year 2005 Year 2007	Covariates without significant estimates	Year 2005 Year 2007 Year 2014	Covariates without significant estimates	Year 2005 Year 2007 Year 2014
Surveys used Observations	T1 - T4 2.042	Surveys used Observations	T1 - T4 2.042	Surveys used Observations	T1 - T4 1.611	Surveys used Observations	T1 - T4 1.843

Note. Depending variable: Job satisfaction; all models are random-intercept models; independent variable names are printed in bold in the left columns, “goal conflict deviation” is the time-specific deviation from the individual average of the goal conflict score, “goal conflict average” is the individual average of the goal conflict score, "job changes" is the total number of job changes between 2004 and 2014 (reported in 2014), interactions between two variables are stated as "variable_1_name*variable_2_name", “covariates without significant estimates” lists additional covariates that were included in the model, p was > 0.05 for these estimates; significant estimates are printed in bold, *** = p < 0.001, ** = p < 0.01, * = p < 0.05

H.2b in detail: Goal conflict and work engagement

A summary of the seven models without interaction terms that related goal conflict to work engagement is given in Table 21 and Table 22. Model 1 – Model 3 used a total of N = 2043 observations. The number of observations for Model 4 and Model 6 was N = 1421. Models 5 and 7 were estimated based on 589 observations provided by participants that completed the 2007 and 2014 surveys. The verbal interpretation of the coefficients is analogous to the H.2a models described in the previous section.

Coefficient estimates for goal conflict

All 13 coefficient estimates for variables representing goal conflict in the seven models without interaction terms were negative and very significant. These findings suggest that higher goal conflict scores were associated with lower work engagement scores. The fixed-effects coefficient estimate for total goal conflict was -0.27 (Model 1); the analogous coefficient estimate for goal conflict deviation in the most basic random-intercept model was identical (Model 2). Point estimates for the coefficient of goal conflict deviation ranged between -0.22 (Model 5 and Model 7) and -0.27 (Model 2 and Model 3). Estimates for the coefficient of average goal conflict ranged between -0.25 (Model 4 and Model 6) and -0.33 (Model 5). In four of the six random-intercept models without interaction terms, the coefficient estimate for average goal conflict was higher than the estimate for goal conflict deviation (Model 2, Model 3, Model 5, Model 7). Model 4 and Model 6 were the only exceptions; in these two models, coefficient estimates for average goal conflict and goal conflict deviation were identical.

Covariates and year indicators

Significant coefficient estimates were only computed for the covariate representing the total number of job changes. The negative coefficient estimates for this variable in all models that featured it suggest that more job changes were associated with less work engagement if all other factors were left unchanged (Model 4 – Model 7). All estimates for the coefficients of the indicators for survey years were not significant.

Models with interaction terms

Result summaries for the four models (Model 8 – Model 11) with interaction terms are presented in Table 23. Coefficient estimates for the four interaction terms were not significant. All four coefficient estimates for average goal conflict were negative and very significant.

Table 21: Goal conflict and work engagement (1/3)

	Model 1 *Fixed-Effects*	**Model 2** *Random-Intercept*	**Model 3** *Random-Intercept*	**Model 4** *Random-Intercept*
Goal Conflict Score	**-0.27***** p < 0.001			
Goal Conflict Deviation		**-0.27***** p < 0.001	**-0.27***** p < 0.001	**-0.25***** p < 0.001
Goal Conflict Average		**-0.30***** p < 0.001	**-0.30***** p < 0.001	**-0.25***** p < 0.001
Job changes				**-0.06*** p = 0.02
Covariates without significant estimates	Year 2005 Year 2007 Year 2014	Year 2005 Year 2007 Year 2014	Age at T1 Gender Year 2005 Year 2007 Year 2014	Part-time Leadership position Year 2005 Year 2007 Year 2014
Surveys used	T1 - T4	T1 - T4	T1 - T4	T1 - T4
Observations	2.043	2.043	2.043	1.421

Note. Depending variable: Work engagement; independent variable names are printed in bold in the left column, “goal conflict deviation” is the time-specific deviation from the individual average of the goal conflict score, “goal conflict average” is the individual average of the goal conflict score, “job changes" is the total number of job changes between 2004 and 2014 (reported in 2014), “covariates without significant estimates” lists additional covariates that were included in the model, p was > 0.05 for these estimates; significant estimates are printed in bold, *** = p < 0.001, ** = p < 0.01, * = p < 0.05

Table 22: Goal conflict and work engagement (2/3)

	Model 5 *Random-Intercept*	**Model 6** *Random-Intercept*	**Model 7** *Random-Intercept*
Goal Conflict: Deviation	**-0.22***** p < 0.001	**-0.25***** p < 0.001	**-0.22***** p < 0.001
Goal Conflict Average	**-0.33***** p < 0.001	**-0.25***** p < 0.001	**-0.32***** p < 0.001
Job changes	**-0.09**** p = 0.008	**-0.06*** p = 0.02	**-0.09**** p = 0.007
Covariates without significant estimates	Part-time Specialist Leadership position Observational tenure Pre-study tenure Year 2014	Age at T1 Gender Part-time Leadership position Year 2005 Year 2007 Year 2014	Age at T1 Gender Part-time Specialist Leadership position Observational tenure Pre-study tenure Year 2014
Surveys used	T3, T4	T1 - T4	T3, T4
Observations	589	1.421	589

Note. Depending variable: Work engagement; independent variable names are printed in bold in the left column, “goal conflict deviation” is the time-specific deviation from the individual average of the goal conflict score, “goal conflict average” is the individual average of the goal conflict score, “job changes" is the total number of job changes between 2004 and 2014 (reported in 2014), “covariates without significant estimates” lists additional covariates that were included in the model, p was > 0.05 for these estimates; significant estimates are printed in bold,
*** = p< 0.001, ** = p < 0.01, * = p < 0.05

Table 23: Goal conflict and work engagement (3/3)

	Model 8		Model 9		Model 10		Model 11
Goal conflict Deviation	0.24 p = 0.42	**Goal conflict** Deviation	**-0.28***** p < 0.001	**Goal conflict** Deviation	**-0.29***** p < 0.001	**Goal conflict** Deviation	**-0.26***** p < 0.001
Goal conflict Average	**-0.30***** p < 0.001	**Goal conflict** Average	**-0.30***** p < 0.001	**Goal conflict** Average	**-0.29***** p < 0.001	**Goal conflict** Average	**-0.27***** p < 0.001
						Year 2007	**-0.09*** p = 0.03
Age at T1	- < 0.01 p = 0.72	**Gender** Female = 1	0.02 p = 0.73	**Job changes**	-0.05 p = 0.05	**Leadership position**	0.14 p = 0.10
Age at T1* **Goal conflict deviation**	-0.02 p = 0.09	**Gender*** **Goal conflict deviation**	0.01 p = 0.86	**Job changes*** **Goal conflict deviation**	0.01 p = 0.74	**Leadership position*** **Goal conflict deviation**	0.08 p = 0.46
Covariates without significant estimates	Year 2005 Year 2007 Year 2014	Covariates without significant estimates	Year 2005 Year 2007 Year 2014	Covariates without significant estimates	Year 2005 Year 2007 Year 2014	Covariates without significant estimates	Year 2005 Year 2014
Surveys used Observations	T1 - T4 2.043	Surveys used Observations	T1 - T4 2.043	Surveys used Observations	T1 - T4 1.613	Surveys used Observations	T1 - T4 1.845

Note. Depending variable: Work engagement; all models are random-intercept models; independent variable names are printed in bold in the left columns, “goal conflict deviation” is the time-specific deviation from the individual average of the goal conflict score, “goal conflict average” is the individual average of the goal conflict score, "job changes" is the total number of job changes between 2004 and 2014 (reported in 2014), interactions between two variables are stated as "variable_1_name*variable_2_name", “covariates without significant estimates” lists additional covariates that were included in the model, p was > 0.05 for these estimates; significant estimates are printed in bold, *** = p < 0.001, ** = p < 0.01, * = p < 0.05

H.2c in detail: Goal conflict and depressive symptoms

Table 24 and Table 25 summarize the seven models without interaction terms that related depressive symptoms to goal conflict. N = 2032 observations were used to estimate the coefficients in Model 1 – Model 3 (surveys T1 – T4). For Model 4 and 6, the number of observations was N = 1415 (surveys T1 – T4). Model 5 and Model 7 were estimated based on data provided by participants that completed the 2007 and 2014 surveys (N = 587). Again, verbal interpretation of the coefficients follows the logic described above for the H.2a models.

Coefficient estimates for goal conflict

All 13 coefficient estimates for variables representing goal conflict in the seven models without interaction terms were positive and very significant. These findings suggest that increased goal conflict was associated with higher depressive symptoms scores. The fixed-effects coefficient estimate for the goal conflict score was 0.13 (Model 1). The same estimate was computed for the coefficient of the variable representing goal conflict deviation in Model 2 and Model 3. Point estimates computed for the coefficients of goal conflict deviation ranged from 0.12 (Model 4 and Model 6) to 0.16 (Model 5 and Model 7). Coefficient estimates for average goal conflict ranged from 0.17 (Model 5) to 0.23 (Model 3). In all six random-intercept models without interaction terms, the estimate for the coefficient of the variable representing average goal conflict was slightly higher than the estimate for the variable representing goal conflict deviation.

Covariates and year indicators

Significant and positive coefficient estimates were computed for the variable representing the total number of job changes in all models featuring it (Model 4 – Model 7). The coefficient of the indicator representing female participants was positive and significant in Model 3 and Model 6; in Model 7, no significant coefficient estimate was computed for this indicator. In the two models that were based exclusively on data provided by participants of the 2007 and the 2014 surveys, the estimate for the coefficient of the indicator for a leadership position was significant and negative (Model 5 and Model 7). No significant estimate was computed in the other models that featured this variable (Model 4 and Model 6). In Model 3, the coefficient estimate for age at baseline was significant and positive. Based on the higher-order models, no significant coefficient estimate was computed for this covariate (Model 6 and Model 7). In Model 1 – Model 4, the estimate for the coefficient of the indicator for 2005 was significant and negative. No significant estimate was computed for this indicator in Model 6. The estimates for the coefficients of all other indicators for survey years were not significant.

Models with interaction terms

Table 26 summarizes the four models that featured interaction terms (Models 8 – Model 11). Four interaction terms were specified: Goal conflict * Age at baseline, Goal conflict * Gender, Goal conflict * Job changes, and Goal conflict * Leadership position. No significant coefficient estimates were computed for these four interaction terms. In all four models with interaction terms, the estimates for the coefficients of the variable representing average goal conflict were positive and very significant.

Table 24: Goal conflict and depressive symptoms (1/3)

	Model 1 *Fixed-Effects*	**Model 2** *Random-Intercept*	**Model 3** *Random-Intercept*	**Model 4** *Random-Intercept*
Goal Conflict Score	**0.13***** p < 0.001			
Goal Conflict Deviation		**0.13***** p < 0.001	**0.13***** p < 0.001	**0.12***** p < 0.001
Goal Conflict Average		**0.21***** p < 0.001	**0.23***** p < 0.001	**0.19***** p < 0.001
Age at T1			**0.02**** p = 0.005	
Gender Female = 1			**0.11***** p < 0.001	
Job changes				**0.05***** p < 0.001
Year 2005	**-0.06**** p = 0.004	**-0.06**** p = 0.005	**-0.06**** p = 0.005	**-0.04*** p = 0.05
Covariates without significant estimates	Year 2007 Year 2014	Year 2007 Year 2014	Year 2007 Year 2014	Part-time Leadership position Year 2007 Year 2014
Surveys used Observations	T1 - T4 2.032	T1 - T4 2.032	T1 - T4 2.032	T1 - T4 1.415

Note. Depending variable: Depressive symptoms; independent variable names are printed in bold in the left column, “goal conflict deviation” is the time-specific deviation from the individual average of the goal conflict score, “goal conflict average” is the individual average of the goal conflict score, “job changes" is the total number of job changes between 2004 and 2014 (reported in 2014), “covariates without significant estimates” lists additional covariates that were included in the model, p was > 0.05 for these estimates; significant estimates are printed in bold, *** = p < 0.001, ** = p < 0.01, * = p < 0.05

Table 25: Goal conflict and depressive symptoms (2/3)

	Model 5 *Random-Intercept*	**Model 6** *Random-Intercept*	**Model 7** *Random-Intercept*
Goal Conflict: Deviation	**0.16***** p < 0.001	**0.12***** p < 0.001	**0.16***** p < 0.001
Goal Conflict Average	**0.17***** p < 0.001	**0.20***** p < 0.001	**0.18***** p < 0.001
Gender Female = 1		**0.09*** p = 0.01	0.08 p = 0.09
Job changes	**0.05**** p = 0.001	**0.05***** p < 0.001	**0.05**** p = 0.001
Leadership position	**-0.14*** p = 0.02	-0.02 p = 0.71	**-0.12*** p = 0.03
Covariates without significant estimates	Part-time Specialist Observational tenure Pre-study tenure Year 2014	Age at T1 Part-time Year 2005 Year 2007 Year 2014	Age at T1 Part-time Specialist Observational tenure Pre-study tenure Year 2014
Surveys used Observations	T3, T4 587	T1 - T4 1.415	T3, T4 587

Note. Depending variable: Depressive symptoms; independent variable names are printed in bold in the left column, “goal conflict deviation” is the time-specific deviation from the individual average of the goal conflict score, “goal conflict average” is the individual average of the goal conflict score, “job changes" is the total number of job changes between 2004 and 2014 (reported in 2014), “covariates without significant estimates” lists additional covariates that were included in the model, p was > 0.05 for these estimates; significant estimates are printed in bold, *** = p < 0.001, ** = p < 0.01, * = p < 0.05

Table 26: Goal conflict and depressive symptoms (3/3)

	Model 8		Model 9		Model 10		Model 11
Goal conflict Deviation	-0.07 p = 0.67	**Goal conflict** Deviation	**0.11*** ** p < 0.001	**Goal conflict** Deviation	**0.11*** ** p < 0.001	**Goal conflict** Deviation	**0.13*** ** p < 0.001
Goal conflict Average	**0.21*** ** p < 0.001	**Goal conflict** Average	**0.22*** ** p < 0.001	**Goal conflict** Average	**0.19*** ** p < 0.001	**Goal conflict** Average	**0.21*** ** p < 0.001
Year 2005	**-0.05** ** p = 0.005	**Year 2005**	**-0.05** ** p = 0.005	**Year 2005**	**-0.05*** p = 0.04	**Year 2005**	**-0.05** ** p = 0.006
Age at T1	**0.01*** p = 0.02	**Gender** Female = 1	**0.10** ** p = 0.001	**Job changes**	**0.04** ** p = 0.001	**Leadership position**	-0.05 p = 0.21
Age at T1* **Goal conflict deviation**	0.01 p = 0.19	**Gender*** **Goal conflict deviation**	0.04 p = 0.15	**Job changes*** **Goal conflict deviation**	0.01 p = 0.55	**Leadership position*** **Goal conflict deviation**	-0.03 p = 0.56
Covariates without significant estimates	Year 2007 Year 2014	Covariates without significant estimates	Year 2007 Year 2014	Covariates without significant estimates	Year 2007 Year 2014	Covariates without significant estimates	Year 2007 Year 2014
Surveys used	T1 - T4	Surveys used	T1 - T4	Surveys used	T1 - T4	Surveys used	T1 - T4
Observations	2.032	Observations	2.032	Observations	1.603	Observations	1.838

Note. Depending variable: Depressive symptoms; all models are random-intercept models; independent variable names are printed in bold in the left columns, “goal conflict deviation” is the time-specific deviation from the individual average of the goal conflict score, “goal conflict average” is the individual average of the goal conflict score, "job changes" is the total number of job changes between 2004 and 2014 (reported in 2014), interactions between two variables are stated as "variable_1_name*variable_2_name", “covariates without significant estimates” lists additional covariates that were included in the model, p was > 0.05 for these estimates; significant estimates are printed in bold, *** = p < 0.001, ** = p < 0.01, * = p < 0.05

Discussion

Based on longitudinal panel data, this study analyzed perceived goal conflict in early-career physicians and its ramifications over a period of around 10 years. Fixed-effects and random-intercept regression models were estimated to limit the influence of possible confounders. On average, goal conflict was lowest in the final follow-up survey when participants were most experienced. Regression models evaluating both inter- and intra-participant variation in work experience also indicated that perceived goal conflict decreased with mounting work experience. Goal conflict was associated with lower job satisfaction, decreased work engagement, and higher depressive symptoms scores.

The next sections discuss these results in detail. The first section is dedicated to the association between work experience and goal conflict (hypothesis H.1). Subsequently, associations between goal conflict and job satisfaction, work engagement, and depressive symptoms are reviewed (hypotheses H.2a – H.2c). The last sections summarize the strengths and limitations of the study, present suggestions for further research, and outline the practical implications of the findings.

Work experience and internalized goal conflict

A negative relationship between observational tenure – i.e., work experience accumulated during the study period – and perceived goal conflict was found in basic regression models. This finding suggests that increased work experience reduces perceived goal conflict in early-career physicians. Several mechanisms may be in play here. With increasing experience, physicians might build up cognitive routines that help them to address goal conflict in a more automatic and hence more subconscious and mentally less demanding manner. In addition, Pratt, Rockmann, and Kaufmann (2006) showed that early-career physicians customize their professional identities during their first years of working in medicine. The resulting identities may vary. Some physicians may put particular emphasis on the patient-physician relationship. Others might base their identity on their technical and scientific excellence or on being experienced case managers focused on process efficiency in a hospital. Regardless of the particular identity any individual physician might adopt, having a clearer self-concept could help in clarifying professional priorities; this could lead to a reduction in perceived goal conflict. Assuming that professional experience influences the degree of perceived goal conflict aligns with the definition of goal conflict as an individual response to an external demand (i.e., contradictory objectives). Analogous to factors influencing appraisal processes in the transactional model of stress (Lazarus & Folkman, 1984), work experience may alter how contradictory objectives at work are perceived. Previous studies specified contradictory objectives and the resulting experience of goal conflict as one of several job

demands within more general frameworks of work-related stress (cf. Cavanaugh, Boswell, Roehling, & Boudreau, 2000; Schneider, Hornung, Weigl, Glaser, & Angerer, 2017). The presented findings suggest that increased work experience reduces the importance of contradictory objectives when compared to other types of job demands.

However, analyzing the results in detail indicates that these interpretations should be qualified. The significance and direction of the estimated effect of work experience were altered once certain sets of covariates were added to the models. Firstly, no significant estimates were obtained for the effect of work experience once the leadership position indicator was included in the statistical models. However, the reduced statistical power due to the higher number of missing values is a relevant factor that needs to be considered when discussing this finding. Secondly, once the year indicator for 2014 was added as a covariate, the direction of the isolated effect of increasing work experience measured as higher observational tenure changed its direction. The positive point estimates obtained suggest that individuals with more work experience – i.e., fewer career interruptions – than the cohort's average perceived a greater degree of goal conflict. However, the positive point estimates obtained for the effect of observational tenure on goal conflict were no longer significant once additional person-specific controls were added. Significant negative associations were found for the year indicator for 2014. It is unclear how to interpret this finding. On average, the amount of work experience was higher in 2014 than in 2007. The result might, therefore, reflect an overall negative effect of increasing work experience on goal conflict. However, the result could also have been driven by other time-specific influences, such as overall improvements in average psychosocial working conditions. Consequently, the possibility that goal conflict was generally less prevalent in 2014 than in 2007 should be considered as an alternative explanation for this finding.

The estimated models do not clarify how mounting work experience may reduce goal conflict. As outlined above, the development of mental routines and coping strategies, as well as the customization of professional identities, are interesting aspects to consider in this context. However, other unmeasured factors, such as the accumulation of work-related resources, should equally be taken into account.

Goal conflict, job satisfaction, work engagement, and depressive symptoms

Increased goal conflict seems to be associated with lower job satisfaction, decreased work engagement, and increased depressive symptoms scores. For all three outcomes, very significant estimates were obtained that supported the initial conjectures. None of the three findings depended on the statistical estimation approach used (i.e., fixed-effects versus random-intercept regressions) or on the set of individual and professional characteristics specified as covariates.

Overall, the results suggest that internalized goal conflict does represent an important manifestation of negative work-related stress in early-career physicians and that perceived goal conflict is associated with relevant measures of mental strain in this group.

The random-intercept regression models estimated the effect of within-participant variation and the effect of between-participants variation in goal conflict separately. The estimates from the job satisfaction models indicate that the deviation from the person-specific average (i.e., within-participant variation) was associated with a more substantial effect when compared to the effect of the level of this person-specific average (i.e., between-participants variation). For work engagement and depressive symptoms, however, the opposite was the case, and between-participants variation was associated with a stronger effect than within-participant variation in most of the estimated model specifications. These divergent findings might indicate that short-term fluctuations in goal conflict might be especially relevant with respect to job satisfaction. Conversely, long-term effects of prolonged exposure to an environment rich in goal conflict might be particularly important when studying work engagement and depressive symptoms. However, the observed effect differences were small.

No significant effects were estimated for most of the covariates. The variable representing the total number of job changes was a noteworthy exception; significant coefficient estimates were obtained for this variable in all models without interaction terms that included this measure. However, this finding should be interpreted carefully as the evaluated outcomes (i.e., work attitudes and mental health status) could have influenced the propensity to change jobs. Taking the example of job satisfaction, physicians who were dissatisfied with their jobs might have had a higher tendency to change jobs. This line of reasoning implies that endogeneity due to a possible dual effect pathway from job satisfaction to job changes and vice versa may have influenced the coefficients estimated for the variable representing the total number of job changes. Similar arguments might play a role in the association of job changes with work engagement and depressive symptoms.

The other personal and professional characteristics specified as covariates had no consistent significant effect on all three outcomes. The coefficients for age and female gender were significant in several specifications of the job satisfaction model and the depressive symptoms model. These findings suggest that, at a constant level of goal conflict, higher age and female gender might be associated with lower job satisfaction and possibly also a higher prevalence of reported depressive symptoms in early-career physicians. In contrast, no significant coefficient estimates were obtained for these variables in the work engagement models. Some of the regression models also controlled for the potential effects of work experience, of occupying a leadership position, of completing specialist training, and of working part-time. No conclusive effects were estimated

for these covariates. However, data availability was limited for some of these professional characteristics. As a consequence, the statistical power of the respective analyses was reduced due to the lower number of observations.

Strengths and limitations

This study has several strengths and limitations. Its longitudinal design based on four assessments is a major strength. Perceived goal conflict in early-career physicians and its ramifications could be evaluated over a period of around 10 years. This extended observation period allowed the cohort to be tracked throughout postgraduate medical specialist training. In addition, participants were studied at subsequent career stages. Due to this design, factors influencing goal conflict as well as subsequent effects of goal conflict could be studied at different levels of professional experience. The statistical methodology based on fixed-effects and random-intercept regression models made full use of the design's advantages; it allowed to control for time-invariant and participant-specific confounders such as personality traits. The study was based on a large cohort of 590 physicians at baseline. Because of its positive effect on statistical power, this relatively large sample size facilitated the detection of significant effects.

The cohort included physicians from various specialties and domains of clinical medicine. The inclusion of participants with different clinical backgrounds increases the external validity of the study and suggests that internalized goal conflict seems to be a widespread phenomenon across the medical profession as a whole. However, the sample's restriction to early-career physicians in Germany might limit the generalizability of the study's findings. Medical education, health care systems, and working environments in hospitals vary across countries. These differences should be considered carefully in any attempt to draw inferences about the relevance of goal conflict for early-career physicians outside of Germany.

A dropout analysis compared the study sample with the initial address pool that was used to send out invitations. As no measure other than gender was available for the initial pool, the test was restricted to this single criterion. The results indicated that women were unintentionally overrepresented in the study sample, which could potentially have biased the estimation results. However, the difference between the share of female physicians in the initial pool and the overall sample was only three percentage points, suggesting that systematic bias caused by the overrepresentation of females in the study sample is unlikely. A second dropout analysis compared participants who dropped out with participants who completed all the surveys. As no significant differences between the two groups were observed, it is unlikely that non-random attrition during the study and related systematic bias were relevant factors influencing the estimation process.

Goal conflict was measured based on a scale derived from an established research instrument used in occupational health psychology to evaluate working environments in hospitals and their psychological implications (Büssing & Glaser, 2002). The scale demonstrated satisfactory psychometric qualities, notably acceptable measures of internal consistency and factor loading patterns that suggested unidimensionality. Nevertheless, to the best of the author's knowledge, this scale has not previously been used as an independent measure in longitudinal studies, and its wording was slightly altered during the study. These factors may have negatively influenced the psychometric quality of the scale, for example, with respect to its content validity.

All variables used in this study were based on self-reports. Methodologically common method bias must be considered if both the dependent and the independent variables are based on the same data source (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). There is also the risk that self-reports may generally be less accurate measures of outcomes related to mental strain when compared to other more objective indicators (Panagioti et al., 2018; Teoh, Hassard, & Cox, 2019; Tyssen, 2018). In particular, measuring depressive symptoms based on third-party assessments might be more indicative of actual mental health problems. However, third-party assessments of depressive symptoms (i.e., expert-based ratings) were not available for this study. In addition, it is more challenging to protect the anonymity of participants in studies based on third-party assessments, and privacy seems to be a significant concern of participants in studies focusing on depression among health care providers (Levine, Breitkopf, Sierles, & Camp, 2003).

Further research and practical implications

Further research could help deepen our understanding of goal conflict and its implications in several ways. Additional analyses in different countries could look at whether goal conflict is equally relevant in health care systems outside Germany. Similarly, studies that survey more experienced physicians could help to evaluate whether goal conflict remains an important phenomenon at later stages of medical careers. Occupational stress and burnout rates vary across medical specialties (Bernburg, Vitzthum, Groneberg, & Mache, 2016; West, Dyrbye, & Shanafelt, 2018). It would, therefore, be important to examine in additional studies whether goal conflict affects all branches of medicine to an equal extent.

Other studies could also include additional measures of mental strain – such as direct measures of burnout or additional indicators of impaired health among physicians – to further enrich our understanding of the implications of goal conflict. Furthermore, it would be interesting to assess whether goal conflict exerts a direct effect on professional performance by analyzing its relationship with quality of care outcomes such as measures of patient safety or of clinical excellence.

The presented study analyzed associations of goal conflict and the studied outcomes for each survey separately. In total, up to four cross-sectional associations were tested for each participant. In addition, the participant-specific average goal conflict score during the entire study duration was related to the three primary outcomes (job satisfaction, work engagement, and depressive symptoms). Lagged effects could be tested in subsequent studies to gain deeper insight into the longitudinal relationship between goal conflict and the outcomes (e.g., the relationship between goal conflict at baseline and depressive symptoms measured at the time of subsequent surveys). Additional psychometric studies could help to further evaluate content validity and reliability of the goal conflict scale constructed for this study. Such studies could also test whether the scale's item set could be improved. For instance, as the four-item scale is still relatively short, adding supplementary items could increase internal consistency without markedly threatening the scale's practicality and applicability (cf. DeVellis, 2017). Likewise, adding positively worded statements might further increase the scale's quality, as the version used for this study contained only negatively worded items (cf. Demerouti, Bakker, Nachreiner, & Schaufeli, 2001).

While the study demonstrates that internalized goal conflict seems to be an important facet of negative work-related stress – i.e., a potential source of mental strain – additional studies could focus on analyzing goal conflict in practical terms. In particular, qualitative or mixed-method studies could help to explore further what types of goal conflict are prevalent in clinical work. Moldaschl's (1991a, 1991b, 2007) classification that subdivides conflicts based on whether they involve goals, resources, or rules offers a useful starting point and could guide this type of research.

Additional research should also focus on factors that moderate the effects of perceived goal conflict on adverse outcomes such as impaired health. The theoretical framework of the JD-R model (Demerouti, Bakker, Nachreiner, & Schaufeli, 2001) suggests that job demands – such as contradictory objectives and resulting goal conflict – should be analyzed in conjunction with relevant resources. However, it is not yet clear what types of resources are most important when dealing with goal conflict. For example, one important research path would be to analyze whether job control – a cornerstone of the job demand-control model (Karasek, 1979) – increases or decreases the perception of goal conflict. On the one hand, job decision latitude could enable physicians to make compromises between different goals. On the other hand, it could intensify the feeling of being responsible for attaining both contradictory goals, leading to a consequent increase in goal conflict. As other domains of life can interact with work (Rothbard, 2001), factors not related to the workplace could also be relevant with respect to their impact on goal conflict. For instance, private resources might help to mitigate the adverse effects of goal conflict, and private responsibilities could give rise to additional conflicts between private and professional goals.

The results suggest that work experience might enable physicians to better cope with contradictory objectives at work. However, this longitudinal study could not clearly distinguish the effect of increasing average work experience from general time effects (e.g., overall improvements in working conditions). Further studies could address this limitation by analyzing goal conflict simultaneously for physicians with different levels of experience, for example, through cross-sectional assessments. Further research could also focus on understanding what types of experiences enable physicians to better deal with goal conflict and relate them to general adoption processes that take place in the early phases of medical careers such as identity customization (cf. Pratt, Rockmann, & Kaufmann, 2006). In addition, further research could attempt to clarify what types of leadership can positively influence the perception of goal conflict among junior physicians. In addition, it would be interesting to assess whether specific features of existing medical school curricula (e.g., medical ethics classes) help in preparing medical students for dealing with goal conflict in their subsequent professional careers. Likewise, retrospective or interventional studies could test whether programs that promote professional self-reflection, such as Balint groups, can reduce perceived goal conflict or mitigate its adverse effects.

The study adds to the extensive body of existing literature illustrating that excessive job demands can negatively influence the health status of physicians and their job attitudes (cf. Dyrbye et al., 2014; West, Dyrbye, & Shanafelt, 2018). Impaired health and job-related mental strain might increase absenteeism, and physicians who are dissatisfied with their working environments might consider leaving the practice of clinical medicine (Landon, Reschovsky, Pham, & Blumenthal, 2006; Maslach, Schaufeli, & Leiter, 2001). To improve the psychosocial working conditions of physicians, hospital administrators and politicians should attempt to develop strategies that reduce goal conflict and alleviate its effects (e.g., work redesign initiatives).

Goal conflict presents a significant challenge for medical education and clinical leadership. Medical schools should try to prepare their students for professional situations that potentially involve goal conflict. For instance, novel educational formats could facilitate discussions about contradictory objectives that physicians may encounter in their medical practice. The concept could also inform group formats that encourage professional reflection, such as structured peer group meetings among medical career entrants. Likewise, instead of transferring to subordinates those contradictory objectives that inevitably characterize the practice of medicine, senior physicians can attempt to prioritize goals for their less experienced colleagues. In addition, medical superiors should explicitly address goal conflict when supervising early-career physicians and can attempt to help early-career physicians develop adequate mental decision-making routines and coping strategies.

Conclusion

The presented study empirically analyzed internalized goal conflict among early-career physicians working in Germany. The concept of internalized goal conflict is based on the framework of demand conflict at work (Moldaschl, 1991a, 1991b, 2007) and is related to the idea of a misfit between organizational and personal work standards (Edwards, 2008; Tanner, Bamberg, Kersten, Kozak, & Nienhaus, 2017), to the model of conflicts based on divergent professional identities (DiBenigno, 2017; Fiol, Pratt, & O'Connor, 2009), and to the notions of moral dilemmas and moral distress (Fourie, 2015; Jameton, 1984; Kälvemark, Höglund, Hansson, Westerholm, & Arnetz, 2004).

While demand conflict and goal conflict are established frameworks in occupational health psychology, this study provides an important addition to previous research as it focuses on empirically analyzing goal conflict and its implications in a larger-scale study of medical professionals. Perceived goal conflict was operationalized using a four-item scale based on an existing broader survey instrument (Büssing & Glaser, 2002). The scale was used in a cohort study that surveyed 590 early-career physicians over a 10-year period covering postgraduate medical specialty training and the subsequent phase during which participants further established their professional careers.

The results suggest that increased goal conflict is associated with lower job satisfaction, reduced work engagement, and increased depressive symptoms. These findings neither depended on the statistical estimation approach used (i.e., fixed-effects versus random-intercept regression analysis) nor on the different covariates specified in the regression models (i.e., several private and professional characteristics that were used as control variables). In addition, the presented analysis indicates that increasing work experience might lower the degree of work-related goal conflict that early-career physicians perceive.

Further quantitative and qualitative research is warranted to deepen the understanding of both the theoretical foundations and the implications of internalized goal conflict in health care. The concept of goal conflict should inform future medical practice; addressing goal conflict may help senior physicians to better support and supervise their younger colleagues. Overall, this study points out that internalized goal conflict constitutes an important aspect of work-related stress faced by early-career physicians and is associated with negative work attitudes and impaired mental health.

Author contribution statement

The research presented in this thesis was conducted based on data from a longitudinal study of physicians that collected an extensive set of professional and private measures (Münchner Ärztestudie). The study was organized by the Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine of Ludwig Maximilian University in Munich and the Institute of Occupational and Social Medicine of Heinrich Heine University in Düsseldorf. The author was not involved in the planning of the surveys and in data collection. The study data were used for several other research projects, none of which dealt in detail with internalized goal conflict, the concept presented in this text. The author was provided with a study description sent to the Ethics committee of Heinrich Heine University that approved the last follow-up survey in 2014. Additionally, the author received a date table that included the sent-out dates for survey invitations, survey reminders, and response deadlines. Based on this information, the author wrote the "inclusion criteria," "study design," and "data collection procedure" sections of this text.

For data analyses, the author was provided with a pseudonymized data set. He also received the survey questionnaires and a key to match the questionnaires' items with variables in the data set. The author developed the hypotheses tested in this research project and selected the research instruments. The author chose the statistical methodology, developed the statistical models, and wrote the syntax used for statistical computation. All statistical analyses, including descriptive statistics, psychometrics, and regression model estimations, were performed by the author. The author created all figures and tables in this text. All sections of this text were written by the author.

Bibliography

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01

BayHKaG. (2002, accessed: 12/02/2020). *Gesetz über die Berufsausübung, die Berufsvertretungen und die Berufsgerichtsbarkeit der Ärzte, Zahnärzte, Tierärzte, Apotheker sowie der Psychologischen Psychotherapeuten und der Kinder- und Jugendlichenpsychotherapeuten*. [Law governing professional practice, professional associations, and specialized courts for physicians, dentists, veterinaries, and pharmacist, as well as psychotherapists and child and adolescent psychotherapists]. Retrieved from https://www.gesetze-bayern.de/Content/Document/BayHKaG/True

Beck, A. T., Steer, R. A., Ball, R. & Ranieri, W. F. (1996). Comparison of Beck Depression Inventories-IA and-II in Psychiatric Outpatients. *Journal of Personality Assessment, 67*(3), 588-597. doi:10.1207/s15327752jpa6703_13

Bernburg, M., Vitzthum, K., Groneberg, D. A. & Mache, S. (2016). Physicians' occupational stress, depressive symptoms and work ability in relation to their working environment: a cross-sectional study of differences among medical residents with various specialties working in German hospitals. *BMJ Open, 6*(6). doi:10.1136/bmjopen-2016-011369

Bianchi, R., Schonfeld, I. S. & Laurent, E. (2015). Burnout–depression overlap: A review. *Clinical Psychology Review, 36*, 28-41. doi:10.1016/j.cpr.2015.01.004

Blanchard, P., Truchot, D., Albiges-Sauvin, L., Dewas, S., Pointreau, Y., Rodrigues, M., Xhaard, A., Loriot, Y., Giraud, P., et al. (2010). Prevalence and causes of burnout amongst oncology residents: A comprehensive nationwide cross-sectional study. *European Journal of Cancer, 46*(15), 2708-2715. doi:10.1016/j.ejca.2010.05.014

Bland, J. M. & Altman, D. G. (1997). Statistics Notes: Cronbach's Alpha. *BMJ, 314*(7080), 572-572. doi:10.1136/bmj.314.7080.572

Bundesärztekammer. (2018, accessed: 12/02/2020). *Ärztestatistik zum 31.12.2018*. [German Medical Association - Physicians statistic - Reference date: 31/12/2018]. Retrieved from https://www.bundesaerztekammer.de/fileadmin/user_upload/downloads/pdf-Ordner/Statistik2018/Stat18AbbTab.pdf

Büssing, A. & Glaser, J. (1999). Tätigkeits- und Arbeitsanalyseverfahren für das Krankenhaus (TAA-KH). [The Work Analysis Instrument for Hospitals (TAA-KH)]. In H. Dunckel (Ed.), *Handbuch psychologischer Arbeitsanalyseverfahren* (1. ed., pp. 465–494). Zürich: vdf - Hoschulverlag an der ETH

Büssing, A. & Glaser, J. (2000). Four-stage process model of the core factors of burnout: The role of work stressors and work-related resources. *Work & Stress, 14*(4), 329-346. doi:10.1080/02678370110041884

Büssing, A. & Glaser, J. (2002). *Das Tätigkeits- und Arbeitsanalyseverfahren für das Krankenhaus - Selbstbeobachtungsversion (TAA-KH-S)*. [The Work Analysis Instrument for Hospitals - Self-rating version (TAA-KH-S)]. (1. ed.). Göttingen: Hogrefe

Cavanaugh, M. A., Boswell, W. R., Roehling, M. V. & Boudreau, J. W. (2000). An empirical examination of self-reported work stress among U.S. managers. *Journal of Applied Psychology, 85*(1), 65-74. doi:10.1037/0021-9010.85.1.65

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334. doi:10.1007/BF02310555

DAAD. (2020, accessed: 11/02/2020). *Medical specialist training: Becoming a medical specialist in Germany*. Retrieved from https://www.study-in.de/en/plan-your-studies/fields-of-study/human-medicine/medical-specialist-training_35929.php

Demerouti, E., Bakker, A. B., Nachreiner, F. & Schaufeli, W. B. (2001). The job demands-resources model of burnout. *Journal of Applied Psychology, 86*(3), 499-512. doi:10.1037/0021-9010.86.3.499

DeVellis, R. F. (2017). *Scale Development - Theory and Applications*. (4. ed.). Los Angeles: Sage

DiBenigno, J. (2017). Anchored Personalization in Managing Goal Conflict between Professional Groups: The Case of U.S. Army Mental Health Care. *Administrative Science Quarterly, 63*(3), 526-569. doi:10.1177/0001839217714024

Dunham, R. B. & Herman, J. B. (1975). Development of a female Faces Scale for measuring job satisfaction. *Journal of Applied Psychology, 60*(5), 629-631. doi:10.1037/0021-9010.60.5.629

Dyrbye, L. N., West, C. P., Satele, D., Boone, S., Tan, L., Sloan, J. & Shanafelt, T. D. (2014). Burnout Among U.S. Medical Students, Residents, and Early Career Physicians Relative to the General U.S. Population. *Academic Medicine, 89*(3), 443-451. doi:10.1097/ACM.0000000000000134

Edwards, J. R. (2008). Person–Environment Fit in Organizations: An Assessment of Theoretical Progress. *Academy of Management Annals, 2*(1), 167-230. doi:10.5465/19416520802211503

Ender, P. (2020). *fapara (Stata package) - Parallel analysis for pca and factor analysis*. Retrieved from https://stats.idre.ucla.edu/stata/ado/analysis/

Ericsson, K. A. (2004). Deliberate Practice and the Acquisition and Maintenance of Expert Performance in Medicine and Related Domains. *Academic Medicine, 79*(10), S70-S81. doi:10.1097/00001888-200410001-00022

Ericsson, K. A., Krampe, R. & Tesch-Roemer, C. (1993). The Role of Deliberate Practice in the Acquisition of Expert Performance. *Psychological Review, 100*(3), 363-406. doi:10.1037//0033-295X.100.3.363

Fiol, C. M., Pratt, M. G. & O'Connor, E. J. (2009). Managing Intractable Identity Conflicts. *The Academy of Management Review, 34*(1), 32-55. doi:10.5465/amr.2009.35713276

Fourie, C. (2015). Moral Distress and Moral Conflict in Clinical Ethics. *Bioethics, 29*(2), 91-97. doi:10.1111/bioe.12064

Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. (1 ed.). Cambridge: Cambridge University Press

Halbesleben, J. R. B. & Wheeler, A. R. (2008). The relative roles of engagement and embeddedness in predicting job performance and intention to leave. *Work & Stress, 22*(3), 242-256. doi:10.1080/02678370802383962

Halekoh, U. & Højsgaard, S. (2014). A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models - The R Package pbkrtest. *Journal of Statistical Software, 1*(9), 1-32. doi:10.18637/jss.v059.i09

Hann, M., Reeves, D. & Sibbald, B. (2010). Relationships between job satisfaction, intentions to leave family practice and actually leaving among family physicians in England. *European Journal of Public Health, 21*(4), 499-503. doi:10.1093/eurpub/ckq005

Hlavac, M. (2018). *stargazer: Well-Formatted Regression and Summary Statistics Tables - R package version 5.2.2*. Retrieved from https://cran.r-project.org/package=stargazer

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179-185. doi:10.1007/BF02289447

Hornung, S., Rousseau, D. M., Glaser, J., Angerer, P. & Weigl, M. (2010). Beyond top-down and bottom-up work redesign: Customizing job content through idiosyncratic deals. *Journal of Organizational Behavior, 31*(2-3), 187-215. doi:10.1002/job.625

Hornung, S., Weigl, M., Glaser, J. & Angerer, P. (2013). Is It So Bad or Am I So Tired? *Journal of Personnel Psychology, 12*(3), 124-131. doi:10.1027/1866-5888/a000088

Imamura, K., Kawakami, N., Inoue, A., Shimazu, A., Tsutsumi, A., Takahashi, M. & Totsuzaki, T. (2016). Work Engagement as a Predictor of Onset of Major Depressive Episode (MDE) among Workers, Independent of Psychological Distress: A 3-Year Prospective Cohort Study. *PLOS ONE, 11*(2). doi:10.1371/journal.pone.0148157

Jackson, S., Schwab, R. & Schuler, R. (1986). Toward an Understanding of the Burnout Phenomen. *Journal of Applied Psychology, 71*(4), 630-640. doi:10.1037/0021-9010.71.4.630

James, S. L., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., et al. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990 - 2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet, 392*(10159), 1789-1858. doi:10.1016/S0140-6736(18)32279-7

Jameton, A. (1984). *Nursing practice: the ethical issues*. (1. ed.). Englewood Cliffs: Prentice-Hall

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

Judge, T. A., Thoresen, C. J., Bono, J. E. & Patton, G. K. (2001). The job satisfaction–job performance relationship: A qualitative and quantitative review. *Psychological Bulletin, 127*(3), 376-407. doi:10.1037/0033-2909.127.3.376

Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement, 20*(1), 141-151. doi:10.1177/001316446002000116

Kälvemark, S., Höglund, A. T., Hansson, M. G., Westerholm, P. & Arnetz, B. (2004). Living with conflicts-ethical dilemmas and moral distress in the health care system. *Social Science & Medicine, 58*(6), 1075-1084. doi:10.1016/S0277-9536(03)00279-X

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

Kenward, M. G. & Roger, J. H. (1997). Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. *Biometrics, 53*(3), 983 - 997. doi:10.2307/2533558

Krohne, H. W., Schmukle, S. C., Spaderna, H. & Spielberger, C. D. (2002). The State-Trait Depression Scales: An International Comparison. *Anxiety, Stress, & Coping, 15*(2), 105-122. doi:10.1080/10615800290028422

Lance, C. E., Butts, M. M. & Michels, L. C. (2006). The Sources of Four Commonly Reported Cutoff Criteria: What Did They Really Say? *Organizational Research Methods, 9*(2), 202-220. doi:10.1177/1094428105284919

Landon, B. E., Reschovsky, J. D., Pham, H. H. & Blumenthal, D. (2006). Leaving Medicine: The Consequences of Physician Dissatisfaction. *Medical Care, 44*(3), 234-242. doi:10.1097/01.mlr.0000199848.17133.9b

Lazarus, R. S. (1993). From Psychological Stress to the Emotions: A History of Changing Outlooks. *Annual Review of Psychology, 44*(1), 1-22. doi:10.1146/annurev.ps.44.020193.000245

Lazarus, R. S. & Folkman, S. (1984). *Stress, appraisal, and coping*. (1. ed.). New York City: Springer Publishing Company

Levine, R. E., Breitkopf, C. R., Sierles, F. S. & Camp, G. (2003). Complications Associated With Surveying Medical Student Depression. *Academic Psychiatry, 27*(1), 12-18. doi:10.1176/appi.ap.27.1.12

Loerbroks, A., Glaser, J., Vu-Eickmann, P. & Angerer, P. (2017). Physician burnout, work engagement and the quality of patient care. *Occupational Medicine, 67*(5), 356-362. doi:10.1093/occmed/kqx051

Lüdecke, D. (2019). *sjstats: Statistical Functions for Regression Models (Version 0.17.6)*. Retrieved from https://cran.r-project.org/package=sjstats

Maslach, C., Schaufeli, W. B. & Leiter, M. P. (2001). Job Burnout. *Annual Review of Psychology, 52*(1), 397-422. doi:10.1146/annurev.psych.52.1.397

Mata, D. A., Ramos, M. A., Bansal, N., Khan, R., Guille, C., Di Angelantonio, E. & Sen, S. (2015). Prevalence of Depression and Depressive Symptoms Among Resident Physicians: A Systematic Review and Meta-analysis. *JAMA, 314*(22), 2373-2383. doi:10.1001/jama.2015.15845

Moldaschl, M. (1991a). *Frauenarbeit oder Facharbeit? Montagerationalisierung in der Elektroindustrie - Band 2*. [Women's work or skilled work? Rationalization of the assembly process in the electrical industry – Volume 2]. (1. ed.). Frankfurt/Main: Campus-Verlag

Moldaschl, M. (1991b). Widersprüchliche Arbeitsanforderungen - Psychische Belastung und doppelte Realität in der Produktion. [Demand conflict at work - Psychological stressors and double reality in manufacturing]. In *Mitteilungen 3* (1. ed., pp. 15-50). München: Sonderforschungsbereich 333 der Universität München - Entwickungsperspektiven von Arbeit

Moldaschl, M. (2007). Ressourcenorientierte Analyse von Belastung und Bewältigung in der Arbeit. [Resource-oriented analysis of stressors and coping at work]. In M. Moldaschl (Ed.), *Immaterielle Ressourcen. Nachhaltigkeit von Unternehmensführung und Arbeit I* (2 ed., pp. 285-321). München: Rainer Hampp Verlag

Panagioti, M., Geraghty, K., Johnson, J., Zhou, A., Panagopoulou, E., Chew-Graham, C., Peters, D., Hodkinson, A., Riley, R., et al. (2018). Association Between Physician Burnout and Patient Safety, Professionalism, and Patient Satisfaction: A Systematic Review and Meta-analysis. *JAMA Internal Medicine, 178*(10), 1317-1331. doi:10.1001/jamainternmed.2018.3713

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

Pratt, M. G., Rockmann, K. W. & Kaufmann, J. B. (2006). Constructing Professional Identity: The Role of Work and Identity Learning Cycles in the Customization of Identity among Medical Residents. *The Academy of Management Journal, 49*(2), 235-262. doi:10.2307/20159762

R Core Team. (2019). *R: A Language and Environment for Statistical Computing - Version 3.5.3* Retrieved from https://www.r-project.org

Roelen, C. A. M., van Hoffen, M. F. A., Groothoff, J. W., de Bruin, J., Schaufeli, W. B. & van Rhenen, W. (2015). Can the Maslach Burnout Inventory and Utrecht Work Engagement Scale be used to screen for risk of long-term sickness absence? *International Archives of Occupational and Environmental Health, 88*(4), 467-475. doi:10.1007/s00420-014-0981-2

Rothbard, N. P. (2001). Enriching or Depleting? The Dynamics of Engagement in Work and Family Roles. *Administrative Science Quarterly, 46*(4), 655-684. doi:10.2307/3094827

Sautier, L. P., Scherwath, A., Weis, J., Sarkar, S., Bosbach, M., Schendel, M., Ladehoff, N., Koch, U. & Mehnert, A. (2015). Erfassung von Arbeitsengagement bei Patienten mit hämatologischen Malignomen: Die psychometrischen Eigenschaften der deutschen Version der Utrecht Work Engagement Scale 9 (UWES-9). [Assessment of Work Engagement in Patients with Hematological Malignancies: Psychometric Properties of the German Version of the Utrecht Work Engagement Scale 9 (UWES-9)]. *Rehabilitation, 54*(05), 297-303. doi:10.1055/s-0035-1555912

Schaufeli, W. B. (2020, accessed: 28/01/2020). *Tests and manuals*. Retrieved from https://www.wilmarschaufeli.nl/tests/

Schaufeli, W. B. & Bakker, A. B. (2004). Job demands, job resources, and their relationship with burnout and engagement: a multi-sample study. *Journal of Organizational Behavior, 25*(3), 293-315. doi:10.1002/job.248

Schaufeli, W. B., Bakker, A. B. & Salanova, M. (2006). The Measurement of Work Engagement With a Short Questionnaire: A Cross-National Study. *Educational and Psychological Measurement, 66*(4), 701-716. doi:10.1177/0013164405282471

Schaufeli, W. B., Salanova, M., González-Romá, V. & Bakker, A. B. (2002). The Measurement of Engagement and Burnout: A Two Sample Confirmatory Factor Analytic Approach. *Journal of Happiness Studies, 3*(1), 71-92. doi:10.1023/A:1015630930326

Schneider, A., Hornung, S., Weigl, M., Glaser, J. & Angerer, P. (2017). Does it matter in the long run? Longitudinal effects and interactions in the differentiated job demands–resources model. *European Journal of Work and Organizational Psychology, 26*(5), 741-754. doi:10.1080/1359432X.2017.1347561

Seligman, M. E. P. & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist, 55*(1), 5-14. doi:10.1037/0003-066X.55.1.5

Shirom, A. (2003). Job-related burnout: A review. In J. C. Quick & L. E. Tetrick (Eds.), *Handbook of occupational health psychology* (1. ed., pp. 245-264). Washington, D.C.: American Psychological Association

Spaderna, H., Schmukle, S. C. & Krohne, H. W. (2002). Bericht über die deutsche Adaptation der State-Trait Depression Scales (STDS). [Report about the German adaptation of the State-Trait Depression Scales (STDS)]. *Diagnostica, 48*, 80-89. doi:10.1026//0012-1924.48.2.80

Spielberger, C. D., Ritterband, L. M., Reheiser, E. C. & Brunner, T. M. (2003). The nature and measurement of depression. *International Journal of Clinical and Health Psychology, 3*(2), 209-234. Retrieved from http://www.aepc.es/ijchp/articulos_pdf/ijchp-68.pdf

Starmer, A. J., Frintner, M. P. & Freed, G. L. (2016). Work–Life Balance, Burnout, and Satisfaction of Early Career Pediatricians. *Pediatrics, 137*(4). doi:10.1542/peds.2015-3183

StataCorp. (2013). *Stata Statistical Software: Release 13*. Retrieved from https://www.stata.com/

Stock, J. H. & Watson, M. W. (2019). *Introduction to econometrics*. (4 ed.). New York City: Pearson

Tanner, G., Bamberg, E., Kersten, M., Kozak, A. & Nienhaus, A. (2017). Misfit of organizational and personal work standards and its longitudinal effect on physicians' depressiveness. *Stress and Health, 33*(5), 665-675. doi:10.1002/smi.2753

Tavakol, M. & Dennick, R. (2011). Making sense of Cronbach's alpha. *Int J Med Educ, 2*, 53-55. doi:10.5116/ijme.4dfb.8dfd

Teasdale, J. D. (1988). Cognitive Vulnerability to Persistent Depression. *Cognition and Emotion, 2*(3), 247-274. doi:10.1080/02699938808410927

Teoh, K., Hassard, J. & Cox, T. (2019). Doctors' perceived working conditions and the quality of patient care: a systematic review. *Work & Stress*, 1-29. doi:10.1080/02678373.2019.1598514

Tokuyama, M., Nakao, K., Seto, M., Watanabe, A. & Takeda, M. (2003). Predictors of first-onset major depressive episodes among white-collar workers. *Psychiatry and Clinical Neurosciences, 57*(5), 523-531. doi:10.1046/j.1440-1819.2003.01158.x

Tyssen, R. (2018). What is the level of burnout that impairs functioning? *Journal of Internal Medicine, 283*(6), 594-596. doi:10.1111/joim.12769

University of Zurich. (2019, accessed: 14/02/2020). Fact Sheet: Academic and Medical Titles in German and English. Retrieved from https://www.kommunikation.uzh.ch/dam/jcr:00000000-517a-81b7-ffff-ffff83e864df/Akademische_Titel_in_Deutsch_und_Englisch.pdf

Wanous, J. P., Reichers, A. E. & Hudy, M. J. (1997). Overall job satisfaction: How good are single-item measures? *Journal of Applied Psychology, 82*(2), 247-252. doi:10.1037/0021-9010.82.2.247

Weesie, J. (2020). *mvpatterns (Stata package)*. Retrieved from http://www.stata.com/stb/stb61/ (within Stata)

Weidner, A. K. H., Phillips, R. L., Fang, B. & Peterson, L. E. (2018). Burnout and Scope of Practice in New Family Physicians. *The Annals of Family Medicine, 16*(3), 200-205. doi:10.1370/afm.2221

Weigl, M., Hornung, S., Petru, R., Glaser, J. & Angerer, P. (2012). Depressive symptoms in junior doctors: a follow-up study on work-related determinants. *International Archives of Occupational and Environmental Health, 85*(5), 559-570. doi:10.1007/s00420-011-0706-8

West, C. P., Dyrbye, L. N. & Shanafelt, T. D. (2018). Physician burnout: contributors, consequences and solutions. *Journal of Internal Medicine, 283*(6), 516-529. doi:10.1111/joim.12752

Wickham, H. & Miller, E. (2019). *haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files - R package version 2.1.1*. Retrieved from https://cran.r-project.org/package=haven

Wooldridge, J. M. (2016). *Introductory econometrics: a modern approach*. (6. ed.). Boston: Cengage

World Health Organization. (2019, accessed: 21/02/2020). *International Statistical Classification of Diseases and Related Health Problems - 10th Revision - 2019 Version*. Retrieved from https://icd.who.int/browse10/2019/en

Xanthopoulou, D., Bakker, A. B., Demerouti, E. & Schaufeli, W. B. (2009). Work engagement and financial returns: A diary study on the role of job and personal resources. *Journal of Occupational and Organizational Psychology, 82*(1), 183-200. doi:10.1348/096317908X285633

Appendix

Stata Syntax

Please refer to the script file on the CD attached to the print version of this thesis.

R Syntax

Please refer to the script file on the CD attached to the print version of this thesis.

German versions of the scales

The following tables present the German versions of the three scales used in the study.

Table 27: Goal conflict scale – German version

Original table is protected by copyright and was removed for the final publication of this text. Büssing and Glaser (2002) feature the items used for the scale.

Table 28: Work engagement scale – German version

Original table is protected by copyright and was removed for the final publication of this text. Schaufeli (2020) features the scale.

Table 29: Depressive symptoms scale – German version

Original table is protected by copyright and was removed for the final publication of this text. Spaderna, Schmukle, and Krohne (2002) feature the scale.

List of tables

List of figures

Danksagung

Mein großer Dank gilt Herrn Prof. Dr. Matthias Weigl, der die Erstellung dieser Arbeit mit viel Engagement und großer Expertise betreut hat und mir sowohl fachlich als auch menschlich eine große Unterstützung war. Herzlich bedanken möchte ich mich auch bei Hans Bauer, der mir wertvolle Anregungen und Impulse bezüglich der in dieser Arbeit verwendeten statistischen Modellierung gegeben hat. Bei Benjamin Lieberman, der mir bei einigen das Englische betreffenden sprachlichen Fragen sehr weitergeholfen hat, möchte ich mich ebenfalls ganz herzlich bedanken. Mein Dank gilt auch dem ganzen an der Münchner Ärztestudie beteiligten Team am Institut und der Poliklinik für Arbeits-, Sozial- und Umweltmedizin der Ludwig-Maximilians-Universität München und am Institut für Arbeits-, Sozial- und Umweltmedizin der Heinrich-Heine-Universität Düsseldorf. Außerdem danke ich den ärztlichen Kolleginnen und Kollegen, die an der Befragung teilgenommen haben und den Aufwand nicht scheuten, der für sie mit der Studie verbunden war.

Große Dankbarkeit empfinde ich gegenüber meinen Eltern Joan und Norbert, die mich stets sehr fürsorglich unterstützt und begleitet haben. An letzter und damit besonders hervorzuhebender Stelle möchte ich meiner Partnerin Pauline von ganzem Herzen danken; ihre liebevolle und fortwährende Unterstützung war auch bei der Anfertigung dieser Arbeit ein sehr wertvoller Rückhalt für mich.

Köln, im Februar 2022 Matthias Guth

Affidavit

Eidesstattliche Versicherung

Guth, Matthias

Name, Vorname

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel

Internalized goal conflict at work influences job satisfaction, work engagement, and depressive symptoms: Evidence from a panel study of early-career physicians

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

Köln, 13.02.2022 — Ort, Datum

Matthias Guth — Unterschrift Doktorandin bzw. Doktorand