Work-Family Conflict and Strain:

Revisiting Theory, Direction of Causality, and Longitudinal Dynamism

Abstract

Does work-family conflict (WFC) cause psychological strain or vice versa? How long do these effects take to unfold? What is the role of persistent WFC (or strain) levels in these processes? Prior research has left some of these questions open: Our systematic review reveals that WFC-strain studies have primarily used short (e.g., hours) or long (e.g., years) measurement lags, leaving mid-long lags underexplored. Moreover, while many work-family theories imply long-term effects, prior longitudinal research has often relied on cross-lagged panel models that assume effects to be solely within-person, not considering persistent between-person differences. We tested this assumption in five three-wave survey studies (*N* = 26,133) with varying lags (1 day, 1 week, 1 month, 6 months, 1 year) and found it to fail in all cases. Employing the random intercept crossed-lagged panel, a new approach in WFC research, our results indicate that the effects between WFC and strain (exhaustion, perceived stress, and affective rumination) depend primarily on longer-term WFC (or strain) levels. In contrast, short-term deviations from these levels (within-person effects) play a minor role. These findings suggest that the effects between WFC and strain may be more persistent than previously assumed, opening avenues for further theoretical and empirical development.

Keywords: work-family conflict, strain, within, between, contextual, cross-lagged panel model (CLPM), random intercept cross-lagged panel model (RI-CLPM)

Work-Family Conflict and Strain: Revisiting Theory, Direction of Causality, and Longitudinal Dynamism

With increasingly digital work and time-spatial flexibility, most workers will sooner or later experience some form of work-family conflict (WFC), defined as "a form of interrole conflict in which the role pressures from the work and family domains are mutually incompatible in some respect" (Greenhaus & Beutell, 1985, p. 77). Indeed, one in four women and one in five men in Europe report high levels of WFC (Borgmann, Kroll, et al., 2019). Affected individuals are at higher risk of psychological strain manifesting in feelings of exhaustion, sleep problems, depression, or other physical or psychosomatic symptoms such as head- or backache (hereafter referred to as strain) (Oshio et al., 2017).

To help intervene, research on WFC and strain has attempted to identify which of these two is at the start of the causal chain. Knowing this is essential as it is challenging to design effective interventions unless we better understand the causal order and type of effect between both constructs. For example, interventions aimed at managing work-life boundaries may be ineffective if it is mainly strain that leads to WFC and not vice versa. Similarly, if the effect is more long-term and contextual than within and short-lived, systemic approaches and holistic programs targeting physical, mental, and social domains might be needed to sustainably improve employees' well-being (Del Consuelo Medina et al., 2018). Accordingly, after decades of cross-sectional designs (Casper et al., 2007), WFC research has shifted to mainly longitudinal studies with more sophisticated analysis procedures to establish the causal direction (Allen et al., 2019; Lapierre & McMullan, 2016). In 2015, Nohe et al. (2015) could already rely on 33 longitudinal studies of the WFC-strain relationship to test the direction of causality between both constructs with a meta-analysis. They found that WFC and strain affect one another reciprocally. So, is the "chicken-and-egg problem" solved, and is the verdict about the direction of causality final?

Unfortunately, this might not be the case because of two main ambiguities in the literature. First, there is a broad agreement that *WFC levels* tend to be chronic (Maertz & Boyar, 2011; Smith et al., 2022), but the temporal nature of the *effects of WFC* is much less clear. Work-family theories are mostly silent on the specific time spans of the effects (Allen et al., 2019; Allen & French, 2023), and most studies estimate within-person effects between two consecutive waves of data collection (e.g., WFC leads to higher strain in the next time point and next time point only), making it hard to detect potential longer-term effects. Crucially, such analytical models leave the effects of a history of persistent WFC or strain levels from earlier waves (e.g., 3 years in a study with three annual data points) under-explored. This seems particularly relevant when studying severe instances of strain, which require—by definition—a "prolonged exposure to certain demands" (Demerouti et al., 2001, p. 500; Edú-Valsania et al., 2022).

Second, the analysis methods (primarily cross-lagged panel models, CLPM) in this literature assume that persistent, trait-like differences between persons play no role in the WFC-strain relationship (random effects assumption; Antonakis et al., 2021). However, newer research shows that WFC levels persist over time (Smith et al., 2022) and that these chronic levels play a major role in the relationship between WFC and personal resources (Ford et al., 2023). Given these recent findings, it may be prudent to reassess conclusions about the causal direction between WFC and strain drawn from earlier studies employing CLPM (Hamaker & Muthén, 2020).

We tackle these ambiguities in three ways. First, we address the potential mismatch between the theories employed in WFC-strain research and the analytical methods used to test them. Many methods assume the effect to be short-lived, depending only on the WFC level at the current or previous time. This assumption is not fully aligned with many common work-family theories, such as conservation of resources (COR) theory (Hobfoll, 1989), boundary theory (Ashforth et al., 2000), or the job demands-resources (JD-R) model (Demerouti et al., 2001). While theories rarely propose specific timeframes, longer-term effects are often an underlying assumption. Indeed, some theories explicitly state that the history of WFC matters (e.g., through accumulation or adaptation processes, loss/gain cycles). To address this potential theory-method mismatch, we systematically compared prior WFC-strain studies' research designs and methods against the effects proposed by commonly used work-family theories.

Second, we estimate reciprocal effects between WFC and strain using a random-intercept cross-lagged panel model (RI-CLPM) that extends the CLPM by relaxing assumptions about the effects of longer-term differences. This approach allows us to (a) separate trait-like between-person differences from the within effects (Hamaker et al., 2015); (b) specify which of these two effects prevails, i.e. the long-term trait-like baselines or within-person differences from the baselines (Gabriel et al., 2019, p. 972); and (c) assess lagged and not only cross-sectional effects as is the case in recent WFC research (Badawy & Schieman, 2020; Smith et al., 2022). By doing so, we can obtain more robust estimates of the effects between WFC and strain, which we operationalized as exhaustion, perceived stress, and daily affective rumination.

Third, we study a broad range of time lags. As Allen et al. (2019, p. 2) noted, the time between waves seems "often arbitrarily selected" and thus potentially misaligned with theory. While short lags (e.g., hours) or long lags (6 months or more) between waves are typical, mid-long lags (e.g., 1 week, 1 month) remain underresearched (Smith et al., 2022). To address this issue, we use five independent panel studies of the German workforce (N = 26,133), each with different lags: 1 day, 1 week, 1 month, 6 months, and 1 year. This is important because some central processes proposed by work-family theories (e.g., adaptation, accumulation, loss or gain spirals) may be left undetected in data with too short or too long lags (Allen et al., 2019; Matthews et al., 2014).

Taken together, we revisit the theoretical foundations, direction of causality, and temporal nature of the WFC-strain effects over five different lags using analytical methods that relax the random effects assumption. While we expect the WFC and strain levels to be mostly stable over time in all five panels (Allen et al., 2019; Smith et al., 2022), it is unclear if and how these stable differences in WFC and strain levels influence the WFC-strain effects.

Systematic Review of Longitudinal Studies Linking WFC and Strain

Our research starts with the observation that many WFC-strain studies have used theories that propose both longer-term and short-term effects but mainly tested those with analytical methods that assume that stable trait-like differences play no role in the analysis. To understand whether this applies more broadly and how the longitudinal features of studies depend on theory, we conducted a systematic literature review. Following Nohe et al.'s (2015) procedure, we searched PsycINFO, Web of Science, and PubMed using a combination of WFC-related terms (*work-family interference, work-family conflict, work-to-family conflict, work-life conflict, work-home interference, work interfering with family*) and longitudinal research terms (*longitudinal, lagged, and panel*). To ensure the coverage of the most recent studies, we also asked two email lists (Organizational Behavior and Occupational Health Psychology) for any new or forthcoming articles and checked the references of two recent reviews (Borgmann, Rattay, et al., 2019; Gunthier et al., 2020). The search yielded a total of 503 studies.

To be included, a study had to (a) be a quantitative peer-reviewed journal article, (b) use a non-experimental design, (c) use WFC and strain (e.g., exhaustion, depressive symptoms) or general health (e.g., physical health symptoms, perceived health) as focal predictor or outcome variables, (d) estimate longitudinal effects over at least two waves, and (e) use the same respondent for all measures and waves. Studies that combined WFC and family-work conflict (FWC) to one score were excluded, as were studies that only measured WFC or strain at just one wave. This screening led us to 95 studies, which two authors coded for their theoretical background, research design, and results. When codes differed, they were discussed and coded again and, where needed, checked by a third author.

Table 1 cross-tabulates methods and results by the seven work-family theories used most often in the reviewed studies (see Additional Online Material 9 [https://osf.io/wknqu] for additional descriptive information and Additional Online Material 2 [https://osf.io/un7hk] for the coding sheet). Most studies used two-wave data, except for COR studies that used two- (n = 14) and three-wave data (n = 12) equally. In general, nearly half of the studies collected data on an annual or two-year basis (n = 43), and 15 studies used lags of three years or longer.

We found no clear patterns between the theory and the method. The cross-lagged panel model (CLPM) was the most common modeling strategy (n = 29) with a clear margin. None of the CLPM studies reported testing the random effects assumption (Antonakis et al., 2021), leaving it unclear whether between-person differences matter and, if so, whether the CLPM is the appropriate modeling strategy. Regarding the direction of causality: Nine CLPM studies found a unidirectional effect (WFC leads to strain or vice versa), 17 studies claimed reciprocal causality, and three yielded non-significant results. Other modeling strategies supported unidirectional (n = 52) rather than reciprocal causality (n = 4).

[Insert Table 1 about here]

Our systematic review provides two key take-aways. First, we found no clear patterns between the theory, the research design, and the modeling approach. Second, most studies used analysis methods that rely on the random effects assumption without testing whether this assumption holds. Therefore, we will address the longitudinal effects between WFC and strain, focusing on why some theories might suggest longer-term effects and how these effects might be more appropriately modeled as trait-like differences instead of cross-lagged ones.

Longitudinal Effect Between WFC and Strain

Next, we present theoretical rationales for the potential directions of causality and the longitudinal dynamics (i.e., speed, duration, and stability of the effect over time). Then, we take an analytical look at the WFC-strain research, focusing on the theory-method match.

The Direction of the Effect Between WFC and Strain

A central argument for WFC causing strain (and not vice versa) is that stress induced by job demands spills over to the private domain and causes WFC, which then negatively affects subjective well-being (Spillover-Crossover Model, Bakker & Demerouti, 2013; Work-Home Resources Model, Ten Brummelhuis & Bakker, 2012). WFC increases the risk of burnout, especially if few resources are available (JD-R Model; Demerouti et al., 2001). Furthermore, boundary theory (Ashforth et al., 2000) proposes that WFC incidents are most likely to occur among persons whose jobs allow them to a) work at different locations and times and b) engage psychologically and physically with one role while still in another. WFC evolves because of boundary crossings and employees' low control of whether they combine or keep their work and personal lives separate (Kossek et al., 2012).

WFC may also be an outcome rather than a predictor of strain (Westman et al., 2004). People's strain levels may affect how they experience the interface between their work and family roles. WFC episodes may appear more intense or occur more frequently when people have been under stress. They are also more likely to recall high rather than low WFC incidents or to expect high WFC in the future (Kelloway et al., 1999; as cited in Nohe et al., 2015).

WFC and strain have also been proposed to have reciprocal relationships. WFC causes a loss of personal resources (e.g., energy, self-efficacy beliefs, emotional stability, time) and can trigger a loss cycle: WFC leads to losses in personal resources, which translates into higher levels of WFC and triggers further resource loss (COR theory; Hobfoll, 1989). Every further loss cycle leads to the depletion of resources that could be used to counteract future resource losses (Hobfoll et al., 2018). All in all, work-family theories provide rationales for three different directions of the effect (WFC causing strain or vice versa, and reciprocal effects). Thus, the direction of the effect is a matter that must be settled empirically.

Stability of WFC Levels and Persistence of WFC Effects Over Time

Recent evidence suggests that WFC levels are mostly stable for at least medium to long periods (Huyghebaert-Zouaghi et al., 2022; Smith et al., 2022). Accordingly, Smith et al. (2022) introduced the Stability and Change Model, which focuses on drivers of change and stability in WFC levels and recognizes that the latter aspect is central. They argue that WFC "exists around a stable, personal equilibrium, at least over mid- to long-term timeframes, because of corresponding stability in personal and situational drivers" (Smith et al., 2022, p. 3). They acknowledge that short-term changes from the usual WFC level are possible but should be "contextualized within the overarching stability" of WFC (Smith et al., 2022, p. 3). Indeed, WFC levels stabilize around usual levels after major life events because personality and other factors (e.g., demands and available resources) remain mostly unchanged (Headey, 2010, as cited in Smith et al., 2022).

Whereas it is clear that WFC *levels* tend to persist over time, there is ambiguity on the persistence of WFC *effects* over time (Ford et al., 2023) and whether this differs between strain types. An effect can vary between transient (sometimes called synchronous), where "there is an immediate strain reaction to [stressors] that diminishes when the [stressors] are removed" (Ford et al., 2014), and persistent, where strain is insensitive to short-term variations of WFC level but depends on the overall WFC level (baseline) instead. To illustrate this difference, Panel 1 in Figure 1 shows examples of both kinds of effects when the WFC level varies over time. When WFC is chronic, as in Panel 2, transient and persistent effects produce a similar strain pattern, hiding the fact that mechanisms through which a high strain level is maintained are different (transient vs. persistent effect).

[Insert Figure 1 about here]

To understand persistent and transient effects, consider an example from a different context: A person with a chronic alcohol problem is chronically drunk (transient effect, short term) and has chronic liver problems (persistent effect, long term). These variables correlate over time (cf. Panel 2 in Figure 1), but the effects of excessive alcohol drinking differ significantly between them. Intoxication is immediate and short-term, occurring within hours of drinking and subsiding just as quickly. Liver damage, however, accumulates over an extended period, with a single day of heavy drinking or abstinence making little immediate impact on liver health. Unlike the short-lived nature of intoxication, liver problems can last for an extended period, persisting even if alcohol consumption ceases.

Whereas the evidence for the persistence of WFC levels is strong, the persistence of effects involving WFC and various strain types remains understudied. It might be the case that WFC levels are stable because the causes of WFC are stable, as Smith et al. (2022) argue (like transient effects in Panel 2 of Figure 2). However, stable levels can also be a product of distinct WFC episodes, which have longer-term effects through accumulation (Maertz et al., 2019) (like persistent effects in Panel 1 of Figure 2). Thus, it remains unclear whether the WFC-strain effect is more like the alcohol consumption-being drunk effect or the alcohol consumption-liver damage effect.

Temporal Nature of the Effect Between WFC and Strain in Work-Family Theories

The commonly used work-family theories remain ambiguous regarding the temporal nature of the effects. We surveyed 30 scholars to find out how they interpret the time orientation of the commonly used work-family theories in Table 2. The findings reveal that while some theories hint at short- or long-term effects, they generally do not detail the progression of these effects over time. Specifically, there is a lack of clarity on how long WFC levels need to be elevated for significant strain to arise or how long strain remains heightened before normalizing after a phase of high WFC (Allen & French, 2023).

The studies in our systematic review varied greatly in terms of the kind of effects they proposed. At one extreme, episodic studies have focused on the acute reactions that WFC episodes might cause within a day (e.g., French & Allen, 2020; Shockley et al., 2022). These short-term effects might be changes in physiological and psychological functioning (e.g., heart rate, blood pressure, state fatigue, state negative affect; French & Allen, 2020). Such an effect might also unfold within hours (like the Transient effect in Figure 1): WFC first leads to changes in volatile personal resources (e.g., physical energy) and, in turn, to strain reactions a few hours later (Work-Home Resources Model; Ten Brummelhuis & Bakker, 2012).

The effects of WFC on strain might also take longer to unfold. If perceptions of high WFC persist even after a WFC episode is over, they might translate into exhaustion over time, particularly if the WFC episode remains unresolved (Maertz et al., 2019). If this is the case, it is not just a single episode of WFC but rather a cumulative history of unresolved episodes that results in strain reactions (like the persistent effect in Figure 1). Nevertheless, strain reactions are not inevitable. The initially negative and immediate effect might just decay over time. For example, "the longer one switches between a pair of roles, the more automatic or 'mindless' becomes the role transition" (boundary theory; Ashforth et al., 2000, p. 485). Consequently, an unexpected transition into a family (or work) role may not cause immediate strain reactions as it used to in the past (Ashforth et al., 2000, p. 486). Similar conclusions can also be derived from adaptation models (Zapf et al., 1996), which propose that a person experiences strain due to higher than usual WFC only until they adapt to the higher WFC level.

Instead of adaptation, the effects of chronic WFC may also be exacerbated over time. Such effects could be expected, for example, based on the loss cycles proposed by conservation of resources theory (COR, Hobfoll, 1989). Ford et al. (2023) tested this aspect of COR theory, finding that within a three-month period, WFC and resource levels appeared more consistent with stable traits than with the loss-cycle hypothesis. Still, it is unclear if loss cycles exist over other time frames. And if they do, does the overall timeframe (five WFC episodes within 2 weeks versus five episodes within 3 months) play a role, with shorter timeframes resulting in higher speed? In all, work-family theories do not provide clear guidance on the temporal dimensions of their proposed effects.

[Insert Table 2 about here]

Modeling Longitudinal Effects Between WFC and Strain

A cross-lagged panel model (CLPM) fitted to data from a three-wave design with a 1-year lag has been one of the most common empirical approaches in studies of WFC and strain dynamics over time (see Additional Online Material 2 [https://osf.io/un7hk]). This approach has recently faced two criticisms. First, the CLPM assumes stable between-person differences play no role in the analysis and can produce misleading results if this assumption does not hold (Hamaker et al., 2015). Despite the importance of the assumption, just one of the reviewed studies, Badawy and Schieman (2020), tested it but in a cross-sectional instead of a cross-lagged context, and one other study after our review (Ford et al., 2023) used a more advanced model that relaxed the assumption in a cross-lagged context. Second, the 1-year lag is mostly not justified and might lead to overlooking effects that occur within shorter lags (Allen et al., 2019). To better understand these challenges, we explain various longitudinal effects (within, between, contextual) and how they relate to the stability of variables (Smith et al., 2022). After that, we elaborate on the commonly used modeling methods and how the length of lags influences the observed effects.

Within, Contextual, and Between Effects

The various effects include within effects, contextual effects, between effects, or some combinations thereof (Antonakis et al., 2021). The *within effect* (also referred to as a within-person effect) refers to changes in a person's WFC level that lead to an immediate (or lagged) change in that person's strain level or vice versa. A key feature of a within effect is that WFC at one point in time influences strain at just one point in time, thereby implying that the history (context) of persistent WFC or strain levels does not matter: the effect is fully transient. The *contextual effect* is the opposite. Persons with high WFC tend to show high strain levels, and deviations from their typical WFC level have little effect on strain during the present or next time point. The effect is thus primarily driven by longer-term average WFC (or strain) levels: the effect is fully persistent. The first and third panels of Figure 2 demonstrate these extreme cases (entirely within and fully contextual effects) using simulated panel data (cf., Enders & Tofighi, 2007, Figure 1). The third type of effect is the *between effect* (also referred to as a between-person effect). The between effect is the sum of the two other effects (within and contextual) and is identical between the three panels in Figure 2.

[Insert Figure 2 about here]

Recognizing where the effect originates is critical to understanding when and why the CLPM can provide misleading results. A between effect may result from aggregation of the within effect on the between level (Panel 1 in Figure 2). For example, if there are stable between-person differences in WFC levels and WFC had a positive within effect on strain, we would see an effect of similar magnitude on the between-person level. Another possibility is that the between effect results from a contextual effect, capturing the effect that the average WFC level has on strain. As such, the finding that WFC levels tend to be stable (Smith et al., 2022) reveals little about the nature of WFC's effects on other variables unless we can differentiate the two mechanisms above. As effects are rarely entirely within (Panel 1 in Figure 2) or entirely contextual (Panel 3 in Figure 2), it is crucial to understand their proportions, that is, whether the effect is mostly within rather than contextual or vice versa (Panel 2 in Figure 2). This is also likely to vary between types of strain.

The Within Effect and Why the CLPM May Not Estimate It Properly

Choosing a proper analysis is essential. In episodic studies, it is important to separate the impact of a specific WFC episode from the day's overall WFC levels (Beal & Gabriel, 2019; Gabriel et al., 2019). Likewise, when studying WFC levels, it is crucial to distinguish between the effects of state and trait-like differences because “the theoretical interpretation would differ greatly depending on whether [this is done]” (Ford et al., 2023, p. 1223).

The critical challenge in the CLPM shown in Figure 3 is that it cannot correctly handle the contextual effect (Hamaker et al., 2015; Hamaker & Muthén, 2020). The method allows for within-person changes in WFC or strain from one point in time to the next but assumes that stable between-person differences do not matter. In the context of panel data analysis, this is often referred to as the random effects assumption (Antonakis et al., 2021), and it means that effects are homologous (i.e., identical) at the within and between levels (Beal & Gabriel, 2019). This is a strong assumption; CLPM will produce misleading results if it fails.

[Insert Figure 3 about here]

There are empirical reasons to believe that unmodeled trait-like differences may have driven the results of some prior WFC-strain studies. Hamaker et al. (2015, Model 1, p. 109) demonstrate a scenario like our Panel 3 in Figure 2 with stable between-person differences but no within effects. They show that applying CLPM to these data would indicate reciprocal effects of the same sign and similar magnitude. This is just the kind of results that the meta-analysis by Nohe et al. (2015) and many of the studies in our systematic review reports, raising the concern that prior conclusions of reciprocal effects might have been driven by stable between-person differences instead of reciprocal one time point to the next within-person effects. However, given that prior research has mostly not tested the random effects assumption, we do not know which one is the case.

The abovementioned problem can be solved by adding person-level random intercepts to the CLPM model, producing an RI-CLPM model (Hamaker et al., 2015; Mulder & Hamaker, 2021). The random intercepts capture the stable between-person variation and are allowed to be correlated with one another.[1] The starting point of RI-CLPM is the same

[1] Freeing the correlation among between variables is crucial; without it, the model becomes a random-effects model, as termed by econometricians (Hamaker & Muthén, 2020), failing to address the confounding of various effects (Antonakis et al., 2021).

confirmatory factor analysis model that the traditional latent variable CLPM (Figure 3) uses. However, instead of using the factors directly, the RI-CLPM models each factor as within and between parts (Figure 4). The within variables are akin to cluster-mean-centered variables in the multilevel model literature, and the between variables can be thought of as cluster means. The cross-lagged model uses the within variables producing clean estimates of the within effect, whose magnitude can be compared against the between-level covariance to understand whether the effects are primarily dynamic, primarily stable, or something in between.

[Insert Figure 4 about here]

The within and between/contextual effects can also be estimated with multilevel modeling or even regression with cluster-robust standard errors using cluster means of WFC (or strain) as a control variable (Antonakis et al., 2021). However, these approaches cannot estimate dynamic effects (i.e., persistence over time) because controlling a lagged dependent variable leads to dynamic panel bias and inconsistent estimates. To address this problem, we must use a dynamic panel model (e.g., Arellano-Bond model or RI-CLPM) (Xu et al., 2020).

Length of the Lag, Study Timeframe, and the Expected Dynamism Between the Variables

The lag length is another essential but often overlooked decision in WFC research (Allen et al., 2019; Taris & Kompier, 2003). On one extreme, experience sampling studies use lags measured in hours, implying short-term effects that become evident within hours (French & Allen, 2020). On the other extreme, panel studies may use lags as long as ten years, though a one-year lag is the most common (Allen et al., 2019). However, most WFC studies do not justify the choice of lags (Allen et al., 2019; Taris & Kompier, 2003). Indeed, as noted already by Kenny (1975), “Normally the lag between measurements is chosen because of convenience not theory, since theory rarely specifies the exact length of the causal lag” (p. 894).

Using a lag that is either too short or too long is problematic (Taris & Kompier, 2014). Too short lag leaves insufficient time for the effect to occur. If a lag is too long, the effect may already have disappeared (Allen et al., 2019; Matthews et al., 2014). Gollob and Reichardt

(1987) explain this with an aspirin example: Using a 2-minute lag would produce no effect on reducing headache, a 30-minute lag might indicate a substantial reduction, a 2- to 3-hour lag the most potent effect, and a 24-hour lag again no effect. In this sense, no one correct lag exists, but multiple lags are needed to understand the causal effect (Gollob & Reichardt, 1987, p. 82). For example, Ford et al.'s (2014) meta-analytical review of the relevance of time lags in longitudinal studies shows that effect sizes increase in lags up to 3 years and decline in lags longer than 3 years. What this means for the effect between WFC and strain is that we ultimately need to study the effect with varying lags to understand its dynamism.

Summary of Theory-Method Mismatch in WFC Research and Research Aims

Our review of theory and empirical studies suggests that researchers examining longitudinal effects of WFC levels typically assume long-term impacts. This results in a theory-method mismatch, akin to using "between-person methods to test within-person theories" (Gabriel et al., 2019, p. 972) in episodic studies. However, the issue is reversed here, with analytical methods for within-person effects applied to data potentially reflecting between-person (contextual) effects. Indeed, our review of the WFC-strain literature shows that there are two sets of WFC-strain studies: (1) cross-lagged models that overlook the possibility of trait-like differences and (2) cross-sectional models that estimate the cross-sectional effect either by decomposing the effects or eliminating trait-like differences but potentially missing the dynamic effects (e.g., Smith et al. 2022; Badawy & Schieman, 2020). Except for one recent example (Ford et al., 2023), no study has analyzed whether dynamic effects hold when the contextual effect is considered. As such, we present four research aims:

Research Aim 1: Does the random effects assumption hold when modeling the dynamics of WFC and strain levels? If the assumption does not hold, what would be the consequences of its violation?

Research Aim 2: How do the magnitudes of the within- and contextual effects compare?

Research Aim 3: What is the direction of causality between WFC and strain?

Research Aim 4: On which of the five potential lags and timeframes do the effects between WFC and strain levels occur?

Empirical Studies

To address our research aims, we conducted five panel studies of the German workforce with five different lags (1 day, 1 week, 1 month, 6 months, and 1 year). We used these data to estimate the direction of causality and longitudinal dynamism between WFC and strain using both CLPM and RI-CLPM. Using data with different lags lets us understand whether and how the effect varies across these lags and overall timeframes. Comparing results from two analysis approaches allows us to assess the impact of potential violations of the random effects assumption on the WFC-strain relationship and whether prior conclusions on reciprocal effects hold when trait-like differences in WFC and strain levels are considered.

Transparency and Openness

Our research was not preregistered and did not require ethics approval.[2] Data were analyzed using Mplus 8.5, STATA 18, and R 4.3.1. A testing manual of the entire analytical procedure, all associated Mplus files, and other research materials (i.e., items, prompts, response scales) are provided in Additional Online Material 1 (https://osf.io/qc3pf) for transparency and replication purposes. We followed the methodological checklist. The sampling plan, data exclusions, and measures are described in the following sections. Data are not available due to their proprietary nature.

Population and Sample

Our population is the German workforce. We used five independent three-wave surveys using lags of 1 day, 1 week, 1 month, 6 months, 1 year. The data were collected by

[2] Our study asked for respondents' perceptions of everyday phenomena (e.g., WFC, strain) and used no deception; thus, there was no need for ethics approval. The market research institutes collect only high-quality data from consenting participants. Panelists receive financial compensation for their efforts. The institutes ensure that all data protection regulations are met, guaranteeing anonymity and enabling longitudinal data linkage without compromising individual privacy.

three German market research institutes using their population-representative survey panels.[3] The 1-year and 6-month lag panels were collected within two larger research projects that used larger sample sizes and were part of other research (Brzykcy & Boehm, 2022; Goetz & Boehm, 2020; Schertler et al., in press). The 1-day, 1-week, and 1-month panels were collected specifically for this study. The 1-day lag panel was collected daily over a week at the end of the workday, between 5 p.m. and midnight, following the daily experience sampling methodology (Beal & Gabriel, 2019). To ensure comparability between the panels and to simplify statistical analysis, we used a 3-day window from Tuesday to Thursday when analyzing the daily data[4].

The complete data sets contain 1,598; 2,172; 2,159; 6,067; and 14,137 observations for 1-day, 1-week, 1-month, 6-month, and 1-year lags, respectively, as shown in Table 3, resulting in a total of 26,133 individuals. We excluded observations with careless responding patterns[5] (e.g., decreasing education level over time) (see Table 3). Among the qualifying individuals, 47% (weighted to take the differences in panel sizes into account) were women, and the average age was 45.5 (SD = 12.0). 34% had a university degree, 22% had a high school degree, and the remaining had a vocational or primary education degree.

[Insert Table 3 about here]

To assess potential attrition effects, we compared continuous survey participants (stayers) with those who dropped out using *t* tests (not reported). Attrition was consistent across panels with lags from 1 week to 1 year. Stayers tend to be slightly older, male, and

---

[3] Except for the panel with a one-year lag, data were collected in equally spaced intervals (July 2016; April 2017; April 2018; July 2020; February 2021; September 2021; mid-November 2021; mid-December 2022; mid-January 2022; November 18–21, 2022; November 25–28, 2022; December 2–5, 2022; May 8–12, 2023; panels with 1-year, 6-month, 1-month, 1-week, and 1-day lags, respectively).

[4] To ensure that the results are not specific to the choice of the window, we replicated all analyses also using the two other possible 2-day windows (i.e., Mon–Wed; Wed–Fri) and all five waves (i.e., Mon-Fri) with very similar results, which we provide in the Additional Online Material 5 (https://osf.io/jef9w)

[5] As a robustness check, we further analyzed the effects using the "moderate screening" strategy that Ward and Meade (2023) suggested. Because some careless respondent statistics have been shown to drop also many diligent respondents (Yentes, 2022; as cited in Ward & Meade, 2023), we report results with potentially careless respondents included. An alternative set of results with respondents excluded and a more detailed explanation of careless respondent analysis is available in the Additional Online Material 4 (https://osf.io/b3syv).

better educated at both attrition points. In the daily study, stayers were more educated, and leaders were more likely to drop out. However, the differences were minor (<5%) and thus inconsequential to our study. Item-level missingness was under .01%, making attrition the only missing data concern. We used FIML estimation to use all observations, not just those who completed responses to all waves.

Measures

Unless otherwise noted, all variables were measured on a 5-point agreement scale ranging from 1 (strongly disagree) to 5 (strongly agree) and back-translated from English to German[6]. The respondents were instructed to consider the current workday, past workweek, past month, past 6 months, or the present moment[7] (1-day panel, 1-week panel, 1-month panel, 6-month panel, and 1-year panel, respectively). The 1-day, 6-month, and 1-year panels used short-form scales of three items, while the two other panels included full versions of the scales. Refer to Additional Online Material 1 (https://osf.io/qc3pf) for the exact items and prompts.

Work-family conflict was assessed using three or five items from the WFC scale developed by Netemeyer et al. (1996). The sample WFC items included “The demands of my work interfere with my home and family life,” “The amount of time my job takes up makes it difficult to fulfill family responsibilities,” and “Things I want to do at home do not get done because of the demands my job puts on me” (reliability = .88–.94).

Exhaustion, described as “the essence of burnout” (Koeske & Koeske, 1989, p. 131), was our primary strain indicator (the only measure used in all panels) because it was the most common strain type in our review. We measured exhaustion with three or eight items from the

[6] The items were initially translated from English to German by the bilingual first author and then back-translated to English by another bilingual author. Discrepancies in translations were reviewed by three bilingual colleagues (two doctoral students and a postdoctoral fellow), leading to discussions among the authors until consensus was achieved. This process did not necessitate significant changes to the items.

[7] The 1-year panel did not use time referents because it was modeled based on prior WFC-strain research, which mostly does not use time referents. The 6-month panel originated from a larger study where time-referents were used and we decided to use this approach also in three shorter studies that were collected during writing the article. See the discussion section for more discussion on time referents in panel studies.

Oldenburg Burnout Inventory (Demerouti et al., 2010). The sample items include “During my work, I often feel emotionally drained,” “There are days when I feel tired before I arrive at work,” and “After my work, I usually feel worn out and weary” (reliability = .84–.90).

Perceived stress was added because it happened to be available in the larger 6-month panel, and we measured it again in the monthly, weekly, and daily surveys for comparison. Perceived stress was measured using three or ten items developed by Cohen et al. (1983). The sample items include “How often were you upset because of something that happened unexpectedly?”; “How often did you feel nervous or stressed?”; “How often did you feel that difficulties were piling up so high that you could not overcome them?”. The response anchors were 1 (*never*) to 5 (*very often*) (reliability = .87–.90).

Daily affective rumination was included in the 1-day panel because we wanted to have a strain measure that might be more transient and vary more dynamically with WFC levels (e.g., Junker et al., 2021 reports ICC=.56 and a significant within-individual effect on WFC after controlling for between-person differences). Daily affective rumination was assessed using the scale by Cropley et al. (2012). The original scale consisted of five items, but following Junker et al. (2021), we used just the three items with the highest factor loadings (i.e., “How often did you become tense today when you thought about work-related issues?”; “How often were you annoyed today by thinking about work-related issues when not at work?”; “How often were you irritated today by work issues when not at work?”). We used the German translations by Pauli et al. (2023). The scale response anchors were *very seldom or never* (1) to *very often or always* (5) (reliability = .91–.92).

Longitudinal Measurement Model and Measurement Invariance

Before the main analysis, we assessed the scales and their longitudinal measurement invariance with confirmatory factor analyses. Confirmatory factor analyses rarely fit the data perfectly, which was also the case in our study. We followed the current recommendations on model diagnostics (Kline, 2016, pp. 268–269) to check if a lack of perfect fit could be due to a severe local misspecification that would go undetected by the fit indices. For simplicity, we started by examining the longitudinal models of WFC and strain variables separately and, if necessary, making relevant adaptations to the individual measurement models before merging them into the final measurement model. We report the analysis workflow for the 1-year lag data and exhaustion in detail. The results for the four other datasets and strain indicators were similar. Table 6 shows the fit statistics for all models of WFC and exhaustion.

The longitudinal measurement model of WFC shows close but not exact model fit ($\chi^2(15) = 27.65$, $p < .05$; RMSEA = .01, SRMR = .01, CFI = .99, TLI = .99, AIC = 149039.95, BIC = 149328.99). After checking the correlation residuals, modification indices, and item wordings, we found that item 1 and item 3 (i.e., "The demands of my work interfere with my home and family life" and "Things I want to do at home do not get done because of the demands my job puts on me," respectively) correlate more than what the model implies. Unlike item 2, these two items focus on perceived job demands, and we allowed this dependency by specifying the error terms of these two items to be correlated. This model was no longer rejected by the test of exact fit ($\chi^2(12) = 13.29$, $p = .35$).

The longitudinal measurement model of exhaustion also yields a significant chi-square value ($\chi^2(15) = 28.15$, $p < .05$; RMSEA = .01, SRMR = .01, CFI = .99, TLI = .99). Using the same diagnostics as before, we found that based on the model, item 1 and item 2 are more strongly correlated than they should be. Both items refer to emotional exhaustion from work in retrospective (i.e., "During my work, I often feel emotionally drained" and "After my work, I usually feel worn out and weary," respectively). In contrast, item 3 asks for a prospective feeling about work (i.e., "There are days when I feel tired before I arrive at work"). Hence, we added an error term covariance between item 1 and item 2, after which the model passes the chi-square test ($\chi^2(12) = 12.68$, $p = .47$).

We estimated the full longitudinal measurement model with all factors (i.e., WF1-3 and EE1-3) using these modifications. The model showed a close but not exact model fit

(Configural model in Table 6, $\chi^2(96) = 353.42$). Because the exact fit hypothesis was rejected, we checked the modification indices to understand the source and extent of potential misspecification in the final model. None of the suggested modifications had clear theoretical interpretations, so we used this model as the starting point for further analysis[8].

Next, it was essential to establish longitudinal measurement invariance to ensure that changes in the variables over time can be attributed to changes in the factors instead of changes in how the scale works over time. The nested model testing sequence (Kline, 2016, pp. 396–399) in Table 4 provided some support for weak factorial invariance ($\Delta\chi^2_{SB}(8) = 15.81$, p = 0.045) but not for strong factorial invariance ($\Delta\chi^2_{SB}(8) = 41.53$, p < 0.001). Our study does not require the latter because we do not model latent level differences (Rönkkö, 2020).

[Insert Table 4 about here]

Results

We estimated all models using short- and long-form scales and all strain indicators (i.e., exhaustion, perceived stress, daily affective rumination). Because there were no substantial differences, we report the results using the short-form scales for all panels and refer to the three strain indicators as "strain." The results using the other scale versions and the different indicators are available in Additional Online Material 5 (https://osf.io/jef9w). Descriptive statistics and correlations of the study variables are shown in Table 5–Table 9. In all, the variables exhibit consistent variance and high correlation over time. The tables also contain results for family-work conflict (FWC), which we discuss in the post hoc analyses section. In what follows, we present the results by the four research aims.

[Insert Table 5, Table 6, Table 7, Table 8, and Table 9 about here]

Research Aim 1: Random Effects Assumption

[8] A purely empirical specification search eventually led to a non-significant chi-square ($\chi^2(69) = 86.94$, *p* = .07). However, these modifications did not lead to any substantial differences in the factor correlations (the most around 4%) as compared to the model without the ad hoc modifications. Thus, while our model does not yield the exact fit with the data, any changes to the model produce only trivial differences in the parameter estimates of interest. As such, we conclude that any remaining misspecifications were inconsequential.

We used CLPM and RI-CLPM analyses to assess the random effects assumption in WFC and strain dynamics. We started with CLPM models (Figure 3). Using the weak factorial invariance model as a starting point, we added (1) autoregressive paths for WFC and strain, (2) cross-lagged effects from WFC to strain and from strain to WFC, and (3) cross-sectional covariances between WFC and strain. The RI-CLPM models (Figure 4) extended the CLPMs by adding two latent variables as random intercepts (i.e., one for WFC and one for strain; RIWF and RIST in Figure 4). This splits the structural regression model into two parts: (1) cluster means random intercepts and (2) cluster mean-centered lambdas (Δ-factors in Figure 4). The autoregression and cross-lagged paths are within effects, and the covariance between the random intercepts quantifies the between effects.

Table 10 shows the estimates for CLPM models. The results show that WFC and strain persist strongly over time, and all cross-lagged effects are positive and statistically significant. The cross-lagged effects of WFC on strain vary between 0.067 and 0.147, with most being around 0.1, whereas the cross-lagged effects in the opposite direction are somewhat more substantial between 0.069 and 0.283. These effects are slightly larger than the meta-analytical estimate of 0.08 for both directions by Nohe et al. (2015). To rule out the possibility that the difference is due to scaling differences between Nohe et al.'s (2015) standardized estimates and our unstandardized ones, we repeated the comparison using standardized estimates with nearly identical results (not reported, results available in Additional Online Material 5 https://osf.io/jef9w). The consistent effects across various time lags hint at a possible unmodeled contextual influence rather than a within-person effect, which typically varies with the effect's duration. In summary, the CLPM reveals reciprocal WFC and strain effects over time, aligning with previous meta-analyses, albeit with marginally stronger effects on average.

[Table 10 about here]

A different picture emerges from RI-CLPM models. The nested model tests in Table 6 (the CLPM against the RI-CLPM, $\Delta\chi^2_{SB}(3)$ = 50.22–248.33, $p < 0.001$) show that the random effects assumption does not hold in any of the five panels, suggesting that stable between-person differences matter. Consequently, the RI-CLPM should be preferred over the CLPM, whose findings of reciprocal effects could be misleading.

Research Aim 2: Within-Person and Between-Person Effects in Comparison

To study how the within and contextual effects compare, we first calculated intraclass correlations (ICCs) to determine how much WFC and strain vary within and between persons. To do so, we transformed the data into long form and estimated a multilevel confirmatory factor analysis. The ICCs (Table 5–Table 9) are high, between .68 and .86, for all variables and all panels, providing initial evidence of the persistence of WFC and strain.

We also examined the RI-CLPM models in Table 11 to compare the within-person effects and the between-level covariance. In contrast to the mostly weak and non-significant autoregressive and cross-lagged effects, the between-level WFC-strain covariance is strong (0.448–0.608) and statistically significant across all panels and strain indicators. The cross-sectional error covariances are mostly statistically significant but much weaker (0.046–0.177), about one-fifth of the between covariances. The strong between-level covariance and weak within effects indicate that effects between WFC and strain are more stable and trait-like than varying from one time point to the next, that is, state-like. Returning to our alcohol drinking example, these results suggest that the WFC-strain effect is more like the relationship between alcohol consumption and liver damage than between alcohol consumption and being drunk.

[Table 11 about here]

Research Aim 3: Direction of Causality Between WFC and Strain

We examined the cross-lagged paths in the RI-CLPM models to infer the direction of causality. All but four cross-lagged effects are statistically non-significant. The significant effects are just below the 5% (p = .033, .049, .026, .028) threshold and become all non-significant if corrected for multiple comparisons (Bonferroni correction). These findings complicate causal claims, as the effect is primarily between-person, rooted in stable individual differences in WFC or strain. Thus, determining the sequence of causality is not possible.

Research Aim 4: The WFC-Strain Effect Over Different Lags and Timeframes

Finally, we inspected the RI-CLPM results across the five panels to understand how the WFC-strain relationship plays out over different timeframes. Drawing firm conclusions is difficult due to the mostly statistically non-significant cross-lagged and autoregressive effects, but we see a pattern: The autoregressive effects of strain (0.269–0.369) and the cross-lagged effects of strain on WFC (0.175–0.243), as well as the cross-sectional error covariances between WFC and strain (0.127–0.170), are generally larger in the daily panel than the others (on average 0.108, 0.041, and 0.092 for autoregressive, cross-lagged, and cross-sectional effects respectively), indicating that there may be short-term effects worth studying even if the contextual effect still dominates the daily panel. These effects drop dramatically to less than 0.1 when we move to the 1-week and 1-month panels[9]. In the longer 6-month and 1-year panels, the cross-sectional covariances appear to increase again (0.124-0.160). The results indicate that longer lags might reveal some within-individual effects, but the short-term effects uncovered by daily studies are much more apparent.

Posthoc Analyses

Family-Work Conflict (FWC)

For comparison, we also measured FWC in the 1-day, 1-week, and 1-month panels collected for this study. FWC was also available in the 6-month panel because it was needed for the larger research project from which this panel originated. FWC was measured using three or five items developed by Netemeyer et al. (1996). The sample FWC items included "The demands of my family or spouse/partner interfere with work-related activities"; "I have

[9] Note that even though there are two statistically significant autoregressive effects of WFC in the 1-week panel (0.336**, 0.318**), this is, in fact, just one WFC autoregression but estimated using two different strain indicators in the model.

to put off doing things at work because of demands on my time at home"; and "Things I want to do at work don't get done because of the demands of my family or spouse/partner."

We used the FWC data to estimate the same models we used for WFC. The CLPM results in Table 10 show significant crossed-lagged FWC-strain effects across all panels and strain indicators. However, like with the WFC data, the similarity of effect sizes between different lags suggests that the results may be driven by unmodeled contextual effects rather than the within effect. χ^2 difference tests (not reported) showed the random effects assumption to fail in all four panels, supporting this interpretation. The RI-CLPM results show a pattern similar to the corresponding WFC results, with most autoregressive and all cross-lagged effects being nonsignificant (see Table 11). The most substantial difference between the FWC and WFC results is that the between-level covariances between FWC and strain are smaller across the board than the WFC-strain covariances. The same pattern also holds for the standardized estimates in Additional Online Material 5 (https://osf.io/jef9w). Interestingly, in the daily study, the between-level covariances are about 80% of the corresponding WFC covariances and fall to about half in the other panels. These results suggest that FWC effects are less contextual than WFC effects in longer time frames.

Group Comparisons

Our finding that the effect of WFC on strain operates mainly as a stable difference raises the question of whether the stability of the effect varies between groups of individuals. To answer this question, we estimated our main models across different subsamples of sex, age (approximate median split at 45 years), and education groups (high school or university degree versus secondary school or lower vocational degree). We found a consistently strong covariance at the between-person level across subsamples. Whereas women show a larger between-level variance of WFC and strain in the daily study, gender differences in other panels are minor. The older age group (Age > 45 years) has higher between-person exhaustion variance and a stronger covariance with WFC, possibly due to prolonged exposure to unresolved WFC episodes leading to chronic exhaustion (Martz et al., 2019). The lower education group shows slightly higher between-person WFC variance and covariance, but these differences are often not statistically significant. See the Additional Online Material 7 (https://osf.io/bt3wa) for details on the subsample analyses.

Discussion

Prior WFC-strain research has almost entirely relied on statistical models that assume that trait-like differences between individuals play no role in the relationship. However, our systematic literature review found that this random effects assumption is rarely tested. Consequently, as our first research aim, we tested this assumption across five independent panels with five different lags (N = 26,133) and found it to fail (Research Aim 1). These results suggest that trait variance needs to be accounted for, and analysis methods that do so (e.g., RI-CLPM) should be preferred over those that do not (e.g., CLPM) because mismatched methods may lead to biased results (Gabriel et al., 2019, pp. 971–972, Ford et al. 2023).

A consistent pattern emerges across all five panels and all three strain indicators when comparing within-person effects against contextual effects (Research Aim 2). The contextual effects (trait-like differences) dominate the other effects across all lags. In contrast to prior research reporting reciprocal cross-lagged effects, we find that the WFC-strain effect is mostly non-significant and weak for the within effect (at around 0.1 or less, except for the daily WFC → strain relationship, Table 11) but always statistically significant and strong (at around 0.5–0.6, Table 11) for the between-level (contextual) effect. Lack of statistical power cannot explain the findings; our sample sizes are in the thousands. A power analysis over 28 representative conditions (Additional Online Material 10 [https://osf.io/pcjr5]) shows that the power of detecting a small cross-lagged effect of 0.1 (standardized) is in the 40-60% range with N = 2000. Therefore, we should detect much more cross-lagged effects if meaningful effects exist. Together with Ford et al. (2023), who found a similar result for the WFC → personal resources relationship, these results challenge the conclusion that “WFC is stable, adaptive, and reciprocal" (Allen & French, 2023, p. 456). Whereas these results do not explicitly answer how long WFC levels must be high for various types of strain to occur or vice versa, they clearly indicate that the effects are longer term and persistent rather than shorter term and transient.

Unfortunately, the combination of strong between-person effects and weak within-person effects makes inferences about the "chicken-and-egg" problem impossible with our research design, leaving our Research Aim 3 about causal direction open for future research.

Finally, regarding Research Aim 4, regardless of the lag used, the effect of WFC levels operates primarily at the between-person level, and fluctuations around WFC (or strain) baseline levels play a lesser role. These results do not mean that within-person effects do not exist but that their dynamics might be more complicated than simple cross-lagged effects from one time point to the next. Indeed, we found cross-sectional covariances between the error terms of WFC and strain in the daily data. These covariances disappeared with the weekly and monthly studies and started to increase again for the longer lags.

Interestingly, we find that affective rumination, which is mostly associated with short-term effects, functions like the other two strain indicators. This may be because, similar to exhaustion and perceived stress, work-related rumination "may extend work demands outside work because recurring cognitive representations of these demands prolong physiological activation" (Kinnunen et al., 2017; p. 2). In the long run, this activation may become chronic and lead to severe strain reactions later on (Brosschot et al., 2006). The three indicators may thus have much in common in that they are more likely to become evident when certain tasks or goals (e.g., managing the boundaries between work and private life) remain unresolved or unfinished (similar to the Zeigarnik's 1972 example of the waiter who can remember the order until it is delivered). They keep the memory occupied, trigger intrusive work-related thoughts, and prolong the exposure to demands (Syrek et al., 2017). In what follows, we outline these findings' theoretical, methodological, and practical implications.

Theoretical Implications

Overall, going back to our example of the effects of alcohol use, it seems that the WFC-strain effect is more like the alcohol consumption-liver damage relationship than the alcohol consumption-being-drunk relationship. Consequently, these results call for a between-person theory of the WFC-strain relationship.

Stability of WFC Effects

Smith et al.'s (2022) stability and change model (SCM) states that WFC levels can be expected to be primarily stable because they are driven by personality factors and the situation (Maertz et al., 2019). Both are relatively stable over short and mid-long periods, and as such, the WFC levels can change around major life events but remain otherwise stable. The idea of stability is further supported by the genetic study by Allen et al. (2022), which concludes that the WFC levels depend more on the person than the situation. However, the SCM model provides only a partial explanation of the stability of the WFC-strain relationship because it focuses exclusively on the stability of *WFC levels*, missing out on the stability of *WFC effects*.

Because both WFC and strain levels vary mainly between persons, we should ask what causes the variation in both and whether one is a cause of another one. Research on burnout might be a helpful starting point for addressing this question. While burnout is often theorized to be dynamic, empirical research shows that burnout levels are stable over time (Dunford et al., 2012). Dunford et al. (2012) propose to resolve this disconnect by arguing that burnout levels fluctuate during major life changes and stabilize later. They back their argument by demonstrating that while organizational insiders generally have consistent burnout levels, newcomers and job changers undergo more variability. Their burnout levels typically rise post-change but normalize within two years in their new role. This finding suggests that burnout development might operate similarly to the mechanisms proposed by the SCM model for WFC levels: both tend to be stable but can shift significantly due to major life events. Because we aimed to compare the short-lived within effect with the longer-term contextual one, we designed all five studies to target individuals in stable employment settings, specifically excluding significant job changes like shifts from full-time to part-time roles, transitions in or out of leadership positions, or employer changes. Studying the WFC-strain relationship around work-life changes is thus an attractive new research opportunity.

Our study provides indirect evidence against adaptation-level theory (Helson, 1948). If adaptation took place, the longitudinal effect between WFC and strain would not depend primarily on the longer-term WFC level but more on short-term changes to which adaptation has yet to occur. This does not mean that adaptation does not play a role in the longitudinal interplay between WFC and strain but that such processes might be linked more to situations of major work-life changes, such as starting a new job, that we did not study.

Accumulation of WFC Over Time

One key question in research on WFC levels is where they come from and how they relate to WFC episodes. Maertz et al. (2019) proposed that WFC levels are a product of the accumulation of unresolved WFC episodes. Our study indirectly supports this idea: If the WFC level accumulates from unresolved WFC episodes and causes strain, then an unresolved WFC episode should affect strain levels not only at the following time point but at all time points until the episode is resolved. In short panels, such as ours, an event that affects multiple time points would not show as a within-person effect but as a contextual effect.

Future research should study if and how specifically WFC levels accumulate, how they relate to strain, and whether this differs between strain outcomes. For instance, the impact of work-family conflict (WFC) might stay constant until WFC levels hit a critical threshold, at which point “things escalate”. In other words, beyond a certain point, even a minor rise in WFC levels could trigger a significant increase in strain, eventually leading to severe strain reactions or burnout over time. To predict the functional form or speed of WFC effects, work-family theory could build upon existing theories, such as the psychology of tipping points (O’Brien, 2020; O’Brien & Klein, 2017). For example, persons with a history of persistent and adverse WFC effects who are past the tipping point may not perceive a significant reduction in strain when the WFC level decreases. Conversely, should the adverse WFC effects slightly increase in magnitude, it would take far less time to infer that things changed for the worse (O'Brien & Klein, 2017).

Considering the WFC-strain relationship as one involving tipping points also provides a new interpretation to the recent result by Ford et al. (2023) that the reciprocal relationships between WFC and personal resources are either non-existent or so weak that they cannot produce the kind of loss cycles that COR theory (Hobfoll et al., 1989) proposes. If individuals react differently to WFC episodes based on their accumulated WFC levels, the loss cycle triggers only after a critical imbalance between WFC and resources has been reached. This also aligns with the central principle of COR theory that resource losses and gains depend on the current resource levels (Hobfoll et al., 2018). As such, it might be that loss cycles exist in the WFC context but are either not very common or are so heterogenous that they are not captured by studies that assume the effects of WFC and resources or strain to be linear and the same for every person, like we and Ford et al. (2023) did. Clearly, more research on the dynamics of WFC, personal resources, and strain is needed.

The FWC effect is less contextual than WFC. This aligns with Allen et al. (2022), who showed that genetic factors explain WFC more strongly than FWC. This might be because the two types of conflict take place in different contexts (work versus home), which, in turn, influences how they unfold over time. It might be, for example, that strain reactions induced by FWC incidents can be quickly dealt with (e.g., an unexpected call from a child asking where to find their lunch box) and are quickly forgotten because of the workload. This is consistent with Shockley and Allen's (2013) conclusion that FWC is perceived as more threatening than WFC because work partners are less forgiving than home partners. In this sense, FWC would lead to immediate strain reactions, while WFC would primarily be associated with delayed strain reactions.

Theorizing Longitudinal Effects in the Absence of a Clear Time-orientation

Our systematic review of the WFC-strain literature demonstrates that work-family theories are mostly vague regarding the temporal dynamism of the proposed effects (Tables 1 and 2). This ambiguity leads to the same theory being used to theorize short-term and long-term effects, such as those observed in daily studies vs. survey studies with longer lags. Longitudinal relationships can include more than one effect operating at different lags. Indeed, Allen et al. (2019; 2023) proposed that adaptation or loss-spirals may all occur but do so in different timeframes. For instance, in the case of the WFC-strain relationship, a shorter effect might be triggered by missing a family dinner because of working late, leading to a fight with a spouse, tiredness, and poor sleep that night. This differs from longer-term WFC (partner constantly missing family events because of work), which might lead to exhaustion or other severe consequences such as partnership problems. WFC can thus cause both short-term (within a day) stress reactions and longer-term strain, but the mechanisms that produce these outcomes are likely to be different.

Applying a theory to a particular (short-term or long-term) context requires justification. We know only one work-family framework (the work-home resources model; Ten Brummelhuis & Baker, 2012) that proposes different effects for different time frames. Work-home effects are either short- or long-term and, therefore, linked with different mediating processes (change in personal volatile vs. structural resources) and outcomes (daily vs. long-term). In the upcoming sections, we emphasize that researchers should thoughtfully decide on and discuss suitable lags, timeframes, and modeling techniques. These choices also impact the formulation of precise hypotheses, which we will detail further.

Methodological Implications

The Choice of Lags and Time Referents

Selecting the right time lag between measurements is crucial. A lag that is too short may not allow enough time for the effect to manifest, while one that is too long might miss the effect as it may have already vanished. When studies use different survey frequencies, a within effect in one study might be a contextual effect in another. For example, in a study with monthly surveys over a year, the contextual effect would represent the effects of that year's WFC (average of 12 months) on strain, which is very similar to the within effect in our 1-year lag panel. In other words, "the estimated lagged effects are specific to the time-interval used in study" (Kuiper & Ryan, 2018, p. 809), and this needs to be considered in the research design. For instance, if the effect of interest is a short-term within effect, it should be tested using a longitudinal study design with short time lags (e.g., daily diary studies). The lags should be equally spaced (Kuiper & Ryan, 2018; Mulder & Hamaker, 2021) unless special techniques for uneven time are used (Asparouhov et al., 2018).

It is also important to use time referents in survey questions (Podsakoff et al., 2019) and match these with the lag and research question. Consider, for example, the following question from Netemeyer et al.'s (1996) WFC scale that we used in our 1-year panel: "The demands of my work interfere with my home and family life." Because the item has no time referent, some informants might answer it based on their current feelings and others based on their memories of past experiences. The latter is more likely when asked to assess the situation over the last six months vs. today (e.g., Think about the last six months vs. Think about your day at work today) (Wang et al., 2017). Thus, the likelihood of finding mostly stable (contextual) effects may be higher when longer-term time referents are used. However, to our understanding, this effect remains to be studied in the WFC-strain context. Nevertheless, pairing the referent with the lag seems like a reasonable default option unless the research design dictates otherwise. For example, in our 1-year panel, the item "The demands of my work have interfered with my home and family life during the past year" would have been more appropriate.

The Choice of the Modeling Approach

Our findings and recommendations have consequences beyond the longitudinal WFC-strain research because the between-person differences and contextual effects have also been overlooked in other contexts (Antonakis et al., 2021; Hoffman & Walters, 2022). Our results demonstrate that the random effects assumption does not hold in any of the five independent panels. CLPM and RI-CLPM also produce substantially different results, highlighting the importance of testing the random effects assumption and choosing the appropriate model. We therefore urge scholars first to test the random effects assumption and use the RI-CLPM (as an extension of the CLPM) if this assumption fails.

RI-CLPM can also accommodate trends. Trends shared by all subjects are accommodated by allowing the item intercepts to vary over time (i.e., strong factorial invariance is not assumed). In multi-wave data (four waves of data or more), the RI-CLPM can also allow slopes to vary over time, which helps estimate within-person dynamics over time (Mulder & Hamaker, 2021, p. 641). Unfortunately, we cannot test for such time trends because it requires data from more than three waves of data. If such time trends and how they vary between individuals are the research interest, other dynamic models, such as the latent change score model (Usami et al., 2019), can be helpful.

Whereas RI-CLPM improves the CLPM that has dominated the WFC-strain research, it is hardly the endpoint of methodological development. Indeed, other analytical methods might be useful for WFC research and other areas. For instance, distributed lag models (Wooldridge, 2013, sec. 10.2) allow for analyzing both the impact propensity (effect to a specific time point) and long-run propensity (effects aggregated over time) by using multiple lags in the same model. Similar ideas where an "impulse" can affect multiple periods have been discussed in organizational research by Zyphur et al. (2019). However, the productive use of such models would require more extended panels than the three-wave design typical for WFC studies.

Earlier, we theorized that the WFC-strain effects vary between persons. Future research might attempt to uncover such effects by pairing RI-CLPM with a latent class analysis or using other models that allow combining profiles (latent classes) with random intercepts in panel data (Tseng, 2023). These techniques would allow relaxing the assumption that all individuals are affected by the same effect that the RI-CLPM makes. Indeed, while the RI-CLPM is currently best suited for analyzing dynamic effects over short panels, it may be a matter of time before more sophisticated methods that are already discussed in the literature (e.g., dynamic computational models, latent class modeling approach; Wang et al., 2017) become the “new defaults” for panel studies.

Finally, we may also consider cross-sectional designs, which dominated WFC research before the current trend toward panel studies (Casper et al., 2007). If the Stability and Change Model (Smith et al., 2022) holds for WFC, the WFC levels should exist around a stable equilibrium for most people. In this case, longitudinal studies provide little value over cross-sectional ones, which should be preferred for their simplicity and lower cost (Spector, 2019).

Closing the Theory-Method Gap Through Precise Hypotheses

More precise hypotheses would help close the gap between theory and method. Ford et al. (2023) provide an excellent example of this approach, which we demonstrate using a generic hypothesis involving WFC and strain in Table 12. The original hypothesis, “WFC has a positive cross-lagged relationship with strain,” would benefit from a further specification: Is it a within-person effect that immediately leads to a within-person change in the outcome and then quickly fades away or a contextual effect that takes a long time to unfold but tends to persist? The former would be compatible with theories in which single incidents of high WFC lead to an immediate stress response such as moodiness or fatigue but are unlikely to result in a severe strain. Conversely, the latter would be more consistent with theories proposing longer-term effects, such that strain develops because of prolonged exposure to high demands.

If the within effect was the most appropriate, the hypothesis could be made more explicit by adding the qualifier "within-person" in the hypothesis (see the second row of Table 12). However, as explained earlier and in prior studies (Kuiper & Ryan, 2018; Pitariu & Ployhart, 2010), the exact meaning of the within effect depends on the chosen lag between waves. Consequently, we recommend the time lag to be included, as in the third row of Table 12, to more explicitly show that the within effect is about short-lived effects where one variable affects another variable in one time point and one time point only.

Some theories call for testing longer-term effects. For example, adaptation-level theory states that because of adaptation, longer-term WFC levels should have little impact on the between-individual differences in strain, implying zero or very small contextual effects. In contrast, COR theory states that past values matter, implying a positive contextual effect. The fourth row of Table 12 shows how hypotheses about contextual effects could be written. Writing more precise hypotheses can also enable more rigorous tests of theories because a study can test multiple different effects for the same pair of variables. The final row shows an example of a hypothesis that could be used to test a theory that would predict that a long-term effect is more substantial than a short-term one.

[Insert Table 12 about here]

To summarize, hypotheses should explicitly state whether the effect is within or contextual and indicate a time frame for the effect to unfold. In doing so, we extend prior recommendations for more precise hypothesizing in the field of management (Edwards & Berry, 2010; Pitariu & Ployhart, 2010) by suggesting that, next to deciding upon such time dimensions as the rate of change or functional form, it is crucial to also theorize which effect is more dominant (within, contextual).

Limitations

Our five panel studies were modeled based on prior longitudinal WFC-strain research and thus share some limitations with these studies. First, to represent current research practice, we collected data on three waves, which limited how we could analyze the dynamic nature of the effect between WFC and strain. Future studies should gather data across additional waves to enhance insights into longer-term dynamics, where effects might not solely depend on the most recent WFC level but also on WFC levels from earlier periods, incorporating multiple distributed lags to capture more complex temporal patterns.

Second, our panel models include no control variables, following the current practice in longitudinal WFC studies. This does not seem too problematic as the random intercepts in the RI-CLPM offer a reasonable control for trait-like, stable differences such as gender, age, education, occupation, or personality. However, they do not help control for potential time-varying confounders such as job changes. Though we accomplished some degree of control by excluding observations with a significant change in their employment situation (i.e., change of the employer, moving in or out of a leadership position, switching from a full-time to a part-time job, or vice versa), there may be an important variable that was omitted from the analysis, leading to endogeneity (Antonakis et al., 2021).

Third, while our studies would allow us to show the direction of causality at the within level, our research design prevents us from establishing the direction of the contextual effect. This could be addressed using instrumental variables, quasi-experimental design, or collecting data over a much longer time frame (e.g., 10 years or more). While this may be challenging to do, such study designs are likely required to analyze the longer-term effects of WFC. Another approach would be to focus on how WFC and strain levels evolve around major life events. All our studies were designed to focus on individuals with stable work-life scenarios (i.e., no job changes). If the SCM model (Maertz et al., 2019) holds, our study subjects should expect relatively stable baseline WFC levels. Focusing on major life events that can affect the baselines could allow us to say something about the causal direction of the longer-term contextual effects of WFC and strain and their causal direction.

Fourth, psychological strain comes in many forms, of which we studied just three. Importantly, all three strain indicators we used (exhaustion, perceived stress, and affective rumination) are within the work domain, and exhaustion and perceived stress can be considered low-arousal strain indicators (Ford et al., 2014). Other types of strain that we did not study might have led to different results (e.g., family- and domain-unspecific strain, Allen et al., 2000; physical and high-arousal psychological strain, Ford et al., 2014). For example, the WFC effect might be (a) weaker for family- than work-related strain (e.g., parental distress vs. burnout, respectively; Nohe et al., 2015) and (b) stronger for high- than low-arousal strain (e.g., anxiety/irritation, and exhaustion, respectively; Ford et al., 2014).

Fifth, although we found very similar results across all five panels, the results may not hold for even longer panels. Moreover, our 1-year panel did not include matching time referents because the data originated from a larger project where time referents were not considered necessary. This raises the concern that our emotional exhaustion measure might measure more of a transient state than the average level over a year (Wang et al.'s 2017). Nevertheless, the fact that we observed very similar results over all panels suggests that this effect did not play a major role in our analysis.

Practical Implications

Our results lead to a few practical implications. First, our findings show that the random effects assumption does not hold and that there are significant differences in WFC and strain levels between persons. Consequently, employees and organizations seem well advised to investigate which specific traits and behaviors explain the different yet stable perceptions of WFC and strain in the workplace to develop effective interventions. Besides considering demographic and dispositional factors such as personality (Wayne et al., 2004), age (Matthews et al., 2010), or gender (Duxbury & Higgins, 1991), organizations should also take into account individual-level behaviors which could actively be influenced in order to improve employees' perceptions. For instance, trainings that target individual changes in boundary management (Allen et al., 2020), detachment behaviors (Sonnentag & Fritz, 2014), or mental resilience (Joyce et al., 2018) could be of value.

Second, we found the contextual effect to be particularly pronounced, implying that strain is a long-term consequence of prolonged exposure to WFC that is more trait-like than one time-point to the next. Thus, in addition to potential training targeting individual employees, organizations should invest in measures that prevent high WFC and strain from persisting over more extended periods. This might be difficult to accomplish, though, as it requires changes in steady characteristics at various organizational levels (leader, team, department) that are relatively resistant to change (e.g., leadership, team climate, organizational culture). However, various strategies seem possible. For example, organizations may strive to strengthen particular leadership and cultural climates (Boehm et al., 2015), which foster reasonable levels of job demands for employees (e.g., regarding reasonable deadlines, working hours) over longer periods. In fact, leaders might be in an ideal position to have a sustainable effect on their team members' long-term health and well-being through their actions and behaviors (e.g., role modeling behaviors, individualized consideration of needs, active management of work-home boundaries; Franke et al., 2014). This includes not only the prevention but also a constructive handling of WFC or family-work conflict episodes once they occur.

Moreover, teams and whole organizations could try to structure work demands over more extended periods, e.g., by allowing for time-outs or sabbaticals after intensive work periods or by using feedback systems to monitor employees' energy levels and intervene in cases of a lack of resources (Bruch & Vogel, 2011). Taken together, such initiatives could help to reduce WFC and strain levels over more extended periods, which is most likely needed to have a lasting effect on employees.

Conclusion

Work-family conflict and its relationship to strain have received much research attention, yet many important questions remain open. Whereas prior studies have documented that WFC levels tend to be chronic, the dynamic nature of its effects on strain remains unclear on both theoretical and empirical levels. Building on a systematic review of 95 studies and five independent, three-wave panel studies with varying time lags, we show that the WFC-strain effect might be more trait-like than previously thought; indeed, we find little evidence of the existence of one time point to the next within-person effects that most prior studies have analyzed. Our results challenge previous conclusions of a reciprocal effect or short effect cycles between WFC and strain. We use these findings to derive broader implications for the literature, urging scholars to (a) develop more explicit hypotheses, including anticipated effect types (within, contextual) as well as time horizons, (b) collect panel data with time lags that match these assumptions, and (c) employ analytical methods that take trait-like differences into account. Finally, we hope our results are appealing not only to researchers but also to practitioners who strive to help employees overcome the detrimental effects of WFC—a pertinent and widespread burden of the modern workplace.

References

Allen, T. D., & French, K. A. (2023). Work-family research: A review and next steps. *Personnel Psychology*, *76*(2), 437–471. https://doi.org/10.1111/peps.12573

Allen, T. D., French, K. A., Braun, M. T., & Fletcher, K. (2019). The passage of time in work-family research: Toward a more dynamic perspective. *Journal of Vocational Behavior*, *110*, 245–257. https://doi.org/10.1016/j.jvb.2018.11.013

Allen, T. D., Merlo, K., Lawrence, R. C., Slutsky, J., & Gray, C. E. (2021). Boundary management and work- nonwork balance while working from home. *Applied Psychology, 70*(1), 60-84.

Allen, T. D., Regina, J., Wiernik, B. M., & Waiwood, A. M. (2022). Toward a better understanding of the causal effects of role demands on work–family conflict: A genetic modeling approach. *Journal of Applied Psychology*. https://doi.org/10.1037/apl0001032

Antonakis, J., Bastardoz, N., & Rönkkö, M. (2021). On ignoring the random effects assumption in multilevel models: Review, critique, and recommendations. *Organizational Research Methods, 24*(2), 443-483. https://doi.org/10.1177/1094428119877457

Ashforth, B. E., Kreiner, G. E., & Fugate, M. (2000). All in a day's work: Boundaries and micro role transitions. *Academy of Management Review, 25*(3), 472-491. https://doi.org/10.5465/amr.2000.3363315

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 359–388. https://doi.org/10.1080/10705511.2017.1406803

Badawy, P., & Schieman, S. (2020). Control and the health effects of work–family conflict: a longitudinal test of generalized versus specific stress buffering. *Journal of Health and Social Behavior, 61*(3), 324-341.

Bakker, A. B., & Demerouti, E. (2013). The spillover-crossover model. In J.G. Grzywacz & E. Demerouti (Eds.), New frontiers in work and family. *New frontiers in work and family research* (pp. 54-70). Psychology Press. https://doi.org/10.4324/9780203586563

Bakker, A. B., Demerouti, E., Taris, T. W., Schaufeli, W. B., & Schreurs, P. J. (2003). A multigroup analysis of the job demands-resources model in four home care organizations. *International Journal of Stress Management, 10*(1), 16.

Beal, D. J., & Gabriel, A. S. (2019). Looking within: An examination, combination, and extension of within-person methods across multiple levels of analysis. In *The handbook of multilevel theory, measurement, and analysis* (pp. 305–327). American Psychological Association. https://doi.org/10.1037/0000115-014

Brosschot, J. F., Gerin, W., & Thayer, J. F. (2006). The perseverative cognition hypothesis: A review of worry, prolonged stress-related physiolo- gical activation and health. *Journal of Psychosomatic Research, 60,* 113– 124.

Bruch, H., & Vogel, B. (2011). Fully charged: How great leaders boost their organization's energy and ignite high performance. *Harvard Business Press*.

Brzykcy, A., & Boehm, S. (2022). No such thing as a free ride: The impact of disability labels on relationship building at work. *Human Relations, 75*(4), 734-763.

Boehm, S. A., Dwertmann, D. J. G., Bruch, H., & Shamir, B. (2015). The missing link? Investigating organizational identity strength and transformational leadership climate as mechanisms that connect CEO charisma with firm performance. *The Leadership Quarterly, 26*(2), 156-171. https://doi.org/10.1016/j.leaqua.2014.07.012

Borgmann, L. S., Kroll, L. E., Müters, S., Rattay, P., & Lampert, T. (2019). Work-family conflict, self-reported general health and work-family reconciliation policies in Europe: Results from the European Working Conditions Survey 2015. *SSM-population Health, 9,* 100465.

Casper, W. J., Eby, L. T., Bordeaux, C., Lockwood, A., & Lambert, D. (2007). A review of research methods in IO/OB work-family research. *Journal of Applied Psychology*, *92*(1), 28–43. https://doi.org/10.1037/0021-9010.92.1.28

Cho, E. (2016). Making reliability reliable: A systematic approach to reliability coefficients. *Organizational Research Methods, 19*(4), 651–682. https://doi.org/10.1177/1094428116656239

Clark, S. C. (2000). Work/family border theory: A new theory of work/family balance. *Human Relations, 53*(6), 747-770. https://doi.org/10.1177/001872670053600

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A Global Measure of Perceived Stress. *Journal of Health and Social Behavior*, *24*(4), 385–396. https://doi.org/10.2307/2136404

Cropley, M., Michalianou, G., Pravettoni, G., & Millward, L. J. (2012). The relation of post-work ruminative thinking with eating behaviour. *Stress and Health, 28*(1), 23-30.

Del Consuelo Medina, M., Calderon, A., Blunk, D. I., Mills, B. W., & Leiner, M. (2018). Organizational wellness program implementation and evaluation: a holistic approach to improve the wellbeing of middle managers. *Journal of Occupational and Environmental Medicine, 60*(6), 515-520.

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. https://doi.org/10.1037/0021-9010.86.3.499

Demerouti, E., Mostert, K., & Bakker, A. B. (2010). Burnout and work engagement: a thorough investigation of the independency of both constructs. *Journal of Occupational Health Psychology*, *15*(3), 209-222. https://doi.org/10.1037/a0019408

Dunford, B. B., Shipp, A. J., Boss, R. W., Angermeier, I., & Boss, A. D. (2012). Is burnout static or dynamic? A career transition perspective of employee burnout trajectories. *Journal of Applied Psychology, 97*(3), 637.

Duxbury, L. E., & Higgins, C. A. (1991). Gender differences in work-family conflict. *Journal of Applied Psychology, 76*(1), 60.

Edú-Valsania, S., Laguía, A., & Moriano, J. A. (2022). Burnout: A Review of Theory and Measurement. *International Journal of Environmental Research and Public Health*, *19*(3), Article 3. https://doi.org/10.3390/ijerph19031780

Edwards, J. R., & Berry, J. W. (2010). The presence of something or the absence of nothing: Increasing theoretical precision in management research. *Organizational Research Methods, 13*(4), 668-689. https://doi.org/10.1177/1094428110380467

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*(2), 121–138. http://psycnet.apa.org/journals/met/12/2/121/

Feng, G., Xu, X., & Lei, J. (2023). Tracking perceived stress, anxiety, and depression in daily life: a double-downward spiral process. *Frontiers in Psychology, 14,* 1114332.

Ford, M. T., Matthews, R. A., Wooldridge, J. D., Mishra, V., Kakar, U. M., & Strahan, S. R. (2014). How do occupational stressor-strain effects vary with time? A review and meta-analysis of the relevance of time lags in longitudinal studies. *Work & Stress, 28*(1), 9-30. https://doi.org/10.1080/02678373.2013.877096

Ford, M. T., Wang, Y.-R., Matthews, R. A., & Wayne, J. H. (2023). Energy, attentional resources and work–family conflict over the meso-term: Resource loss spirals revisited. *Journal of Organizational Behavior*. Advance online publication. https://doi.org/10.1002/job.2710

Franke, F., Felfe, J., & Pundt, A. (2014). The impact of health-oriented leadership on follower health: Development and test of a new instrument measuring health-promoting leadership. *German Journal of Human Resource Management, 28*(1-2), 139-161. https://doi.org/10.1177/239700221402800108

French, K. A., & Allen, T. D. (2020). Episodic work-family conflict and strain: A dynamic perspective. *Journal of Applied Psychology*, *105*(8), 863–888. https://doi.org/10.1037/apl0000470

Gabriel, A. S., Podsakoff, N. P., Beal, D. J., Scott, B. A., Sonnentag, S., Trougakos, J. P., & Butts, M. M. (2019). Experience sampling methods: A discussion of critical trends and considerations for scholarly advancement. *Organizational Research Methods*, *22*(4), 969–1006. https://doi.org/10.1177/1094428118802626

Goetz, T. M., & Boehm, S. A. (2020). Am I outdated? The role of strengths use support and friendship opportunities for coping with technological insecurity. *Computers in Human Behavior, 107,* 1-14.

Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development*, *58*(1), 80–92. https://doi.org/10.2307/1130293

Greenhaus, J. H., & Beutell, N. J. (1985). Sources of conflict between work and family roles. *Academy of Management Review, 10*(1). https://doi.org/10.5465/amr.1985.4277352

Guthier, C., Dormann, C., & Voelkle, M. C. (2020). Reciprocal effects between job stressors and burnout: A continuous time meta-analysis of longitudinal studies. *Psychological Bulletin, 146*(12), 1146-1173. https://doi.org/10.1037/bul0000304

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102–116. https://doi.org/10.1037/a0038889

Hamaker, E. L., & Muthén, B. (2020). The fixed versus random effects debate and how it relates to centering in multilevel modeling. *Psychological Methods, 25*(3), 365–379.

Halbesleben, J. R., Neveu, J. P., Paustian-Underdahl, S. C., & Westman, M. (2014). Getting to the "COR" understanding the role of resources in conservation of resources theory. *Journal of Management, 40*(5), 1334-1364.

Headey, B. (2010). The set point theory of well-being has serious flaws: on the eve of a scientific revolution?. *Social Indicators Research, 97*(1), 7-21.

Helson, H. (1948). Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychological Review, 55*(6), 297–313. https://doi.org/10.1037/h0056721

Hobfoll, S. E., Halbesleben, J., Neveu, J. P., & Westman, M. (2018). Conservation of resources in the organizational context: The reality of resources and their consequences. *Annual Review of Organizational Psychology and Organizational Behavior, 5,* 103-128. https://doi.org/10.1146/annurev-orgpsych-032117-104640

Hobfoll, S. E., & Lerman, M. (1989). Predicting receipt of social support: A longitudinal study of parents' reactions to their child's illness. *Health Psychology, 8*(1), 61-77. https://doi.org/10.1037/0278-6133.8.1.61

Hoffman, L., & Walters, R. W. (2022). Catching Up on Multilevel Modeling. *Annual Review of Psychology*, *73*(1), 659–689. https://doi.org/10.1146/annurev-psych-020821-103525

Huyghebaert-Zouaghi, T., Morin, A. J. S., Fernet, C., Austin, S., & Gillet, N. (2022). Longitudinal profiles of work-family interface: Their individual and organizational predictors, personal and work outcomes, and implications for onsite and remote workers. *Journal of Vocational Behavior*, *134*, 103695. https://doi.org/10.1016/j.jvb.2022.103695

Joyce, S., Shand, F., Tighe, J., Laurent, S. J., Bryant, R. A., & Harvey, S. B. (2018). Road to resilience: a systematic review and meta-analysis of resilience training programmes and interventions. *BMJ Open, 8*(6), e017858.

Junker, N. M., Baumeister, R. F., Straub, K., & Greenhaus, J. H. (2021). When forgetting what happened at work matters: The role of affective rumination, problem-solving pondering, and self-control in work–family conflict and enrichment. *Journal of Applied Psychology*, *106*(11), 1750–1766. https://doi.org/10.1037/apl0000847.supp

Kelloway, E. K., Gottlieb, B. H., & Barham, L. (1999). The source, nature, and direction of work and family conflict: A longitudinal investigation. *Journal of Occupational Health Psychology, 4,* 337–346.

Kenny, D. (1975). Cross-Lagged Panel Correlation: A Test for Spuriousness. *Psychological Bulletin*, *82*, 887–903. https://doi.org/10.1037/0033-2909.82.6.887

Kinnunen, U., Feldt, T., Sianoja, M., de Bloom, J., Korpela, K., & Geurts, S. (2017). Identifying long-term patterns of work-related rumination: Associations with job demands and well-being outcomes. *European Journal of Work and Organizational Psychology, 26*(4), 514-526.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (3rd ed.). Guilford Press.

Koeske, G. F., & Koeske, R. D. (1989). Construct Validity of the Maslach Burnout Inventory: A Critical Review and Reconceptualization. *The Journal of Applied Behavioral Science*, *25*(2), 131–144. https://doi.org/10.1177/0021886389252004

Kossek, E. E., Ruderman, M. N., Braddy, P. W., & Hannum, K. M. (2012). Work–nonwork boundary management profiles: A person-centered approach. *Journal of Vocational Behavior, 81*(1), 112-128. https ://doi.org/10.1016/j.jvb.2012.04.003

Kuiper, R. M., & Ryan, O. (2018). Drawing conclusions from cross-lagged relationships: Re-considering the role of the time-interval. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(5), 809-823.

Lapierre, L. M., & McMullan, A. D. (2016). A review of methodological and measurement approaches to the study of work and family. *The Oxford Handbook of Work and Family*, pp. 36-50.

Maertz, C. P., & Boyar, S. L. (2011). Work-family conflict, enrichment, and balance under "levels" and "episodes" approaches. *Journal of Management*, *37*(1), 68–98. https://doi.org/10.1177/0149206310382455

Maertz, C. P., Boyar, S. L., & Maloney, P. W. (2019). A theory of work-family conflict episode processing. *Journal of Vocational Behavior*, *115*, 103331. https://doi.org/10.1016/j.jvb.2019.103331

Matthews, R. A., Bulger, C. A., & Barnes-Farrell, J. L. (2010). Work social supports, role stressors, and work–family conflict: The moderating effect of age. *Journal of Vocational Behavior, 76*(1), 78-90.

Matthews, R. A., Wayne, J. H., & Ford, M. T. (2014). A work–family conflict/subjective well-being process model: A test of competing theories of longitudinal effects. *Journal of Applied Psychology*, *99*(6), 1173-1187. https://doi.org/10.1037/a0036674

Mulder, J. D., & Hamaker, E. L. (2021). Three extensions of the random intercept cross-lagged panel model. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(4) 638-648. https://doi.org/10.1080/10705511.2020.1784738

Netemeyer, R. G., Boles, J. S., & McMurrian, R. (1996). Development and validation of work–family conflict and family–work conflict scales. *Journal of Applied Psychology, 81*(4), 400–410. https://doi.org/10.1037/0021-9010.81.4.400

Nohe, C., Meier, L. L., Sonntag, K., & Michel, A. (2015). The chicken or the egg? A meta-analysis of panel studies of the relationship between work–family conflict and strain. *Journal of Applied Psychology*, *100*(2), 522-536. https://dx.doi.org/10.1037/a0038012

O'Brien, E. (2020). When small signs of change add up: The psychology of tipping points. *Current Directions in Psychological Science, 29*(1), 55-62. https ://doi.org/10.1177/0963721419884313

O'Brien, E., & Klein, N. (2017). The tipping point of perceived change: Asymmetric thresholds in diagnosing improvement versus decline. *Journal of Personality and Social Psychology, 112*(2), 161.

Oshio, T., Inoue, A., & Tsutsumi, A. (2017). Does work-to-family conflict really matter for health? Cross-sectional, prospective cohort and fixed-effects analyses. *Social Science & Medicine, 175,* 36-42.

Pauli, R., Gaum, P. M., Cropley, M., & Lang, J. (2023). Antecedents, outcomes and measurement of work related-cognition in non-work time: A multistudy report using the work-related rumination questionnaire in two languages. *Frontiers in Psychology, 14*.

Pitariu, A. H., & Ployhart, R. E. (2010). Explaining change: Theorizing and testing dynamic mediated longitudinal relationships. *Journal of Management, 36*(2), 405-429. https://doi.org/10.1177/0149206308331096

Podsakoff, N. P., Spoelma, T. M., Chawla, N., & Gabriel, A. S. (2019). What predicts within-person variance in applied psychology constructs? An empirical examination. *Journal of Applied Psychology*, *104*(6), 727–754. https://doi.org/10.1037/apl0000374

Rönkkö, M. (Director). (2020). *Measurement invariance—YouTube*. https://www.youtube.com/watch?v=G5PEgxoRVC4

Schertler, M., Glumann, N., & Boehm, S. A. (in press). How two megatrends affect each other: Studying the interplay of remote work and workplace inclusion with a random intercept cross-lagged panel model. *Academy of Management Discoveries*. https://doi.org/10.5465/amd.2022.0133

Shockley, K. M., & Allen, T. D. (2013). Episodic work–family conflict, cardiovascular indicators, and social support: An experience sampling approach. *Journal of Occupational Health Psychology, 18*(3), 262.

Shockley, K. M., Gabriel, A. S., & Yuan, Z. (2022). Profiles of attribution for work–family conflict episodes and their relation to negative emotions. *Journal of Organizational Behavior*, *43*(4), 643–661. https://doi.org/10.1002/job.2597

Smith, C. E., Wayne, J. H., Matthews, R. A., Lance, C. E., Griggs, T. L., & Pattie, M. W. (2022). Stability and change in levels of work–family conflict: A multi- study, longitudinal investigation. *Journal of Occupational and Organizational Psychology*, *95*(1), 1–35. https://doi.org/10.1111/joop.12372

Sonnentag, S., & Fritz, C. (2015). Recovery from job stress: The stressor- detachment model as an integrative framework. *Journal of Organizational Behavior, 36*, 72-103.

Spector, P. E. (2019). Do not cross me: Optimizing the use of cross-sectional designs. *Journal of Business and Psychology*, *34*, 125–137. https://doi.org/10.1007/s10869-018-09613-8

Syrek, C. J., Weigelt, O., Peifer, C., & Antoni, C. H. (2017). Zeigarnik's sleepless nights: How unfinished tasks at the end of the week impair employee sleep on the weekend through rumination. *Journal of Occupational Health Psychology, 22*(2), 225.

Taris, T. W., & Kompier, M. (2003). Challenges in longitudinal designs in occupational health psychology. *Scandinavian Journal of Work, Environment & Health*, *29*(1), 1–4. https://doi.org/10.5271/sjweh.697

Taris, T. W., & Kompier, M. A. J. (2014). Cause and effect: Optimizing the designs of longitudinal studies in occupational health psychology. *Work & Stress*, *28*(1), 1–8. https://doi.org/10.1080/02678373.2014.878494

Ten Brummelhuis, L. L., & Bakker, A. B. (2012). A resource perspective on the work–home interface: The work-home resources model. *American Psychologist*, *67*(7), 545-556. https://doi.org/10.1037/a0027974

Tseng, M.-C. (2023). Latent profile transition analysis with random intercepts (RI-LPTA). *Structural Equation Modeling: A Multidisciplinary Journal*, https://doi.org/10.1080/10705511.2023.2284671

Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods*, *24*(5), 637–657. https://doi.org/10.1037/met0000210

Wang, M., Beal, D. J., Chan, D., Newman, D. A., Vancouver, J. B., & Vandenberg, R. J. (2017). Longitudinal research: A panel discussion on conceptual issues, research design, and statistical techniques. *Work, Aging and Retirement, 3*(1), 1-24.

Ward, M. K., & Meade, A. W. (2023). Dealing with careless responding in survey data: Prevention, identification, and recommended best practices. *Annual Review of Psychology, 74*(1).

Wayne, J. H., Musisca, N., & Fleeson, W. (2004). Considering the role of personality in the work–family experience: Relationships of the big five to work–family conflict and facilitation. *Journal of Vocational Behavior, 64*(1), 108-130.

Westman, M. (2001). Stress and strain crossover. *Human relations*, *54*(6), 717-751.

Westman, M., Hobfoll, S. E., Chen, S., Davidson, O. B., & Laski, S. (2004). Organizational stress through the lens of conservation of resources (COR) theory. In Perrewe, P.L. & Ganster, D.C. (Eds.) *Exploring Interpersonal Dynamics* (*Research in Occupational Stress and Well Being, Vol. 4*) (pp. 167-229). Emerald Group Publishing Limited. https://doi.org/10.1016/S1479-3555(04)04005-3

Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th ed). South-Western Cengage Learning.

Xu, R., DeShon, R. P., & Dishop, C. R. (2020). Challenges and Opportunities in the Estimation of Dynamic Models. *Organizational Research Methods*, *23*(4), 595–619. https://doi.org/10.1177/1094428119842638

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

Zeigarnik, B. (1927). Das Behalten erledigter und unerledigter Handlungen [Remembering finished and unfinished tasks]. *Psychologische Forschung, 9,* 1–85.

Zyphur, M. J., Allison, P. D., Tay, L., Voelkle, M. C., Preacher, K. J., Zhang, Z., Hamaker, E. L., Shamsollahi, A., Pierides, D. C., Koval, P., & Diener, E. (2019). From data to causes I: Building a general cross-lagged panel model (gclm). *Organizational Research Methods, 23*(4). https://doi.org/10.1177/1094428119847278

Table

Table 1. Cross-Tabulation of Employed Work-Family Theories, Time Lags, Data Modeling Approaches, and Results

		Number of Waves				Length of the Time Lags								Estimation Strategy				Results					
																		WFC leads to Strain[b]			Strain leads to WFC		
	Total of studies	2 waves	3 waves	4 waves	>= 5 waves	<=1 day	< = 2 weeks	<= 1 month	<= 2 months	<= 6 months	<= 1 year	2 years	>= 3 years	Time-lagged regression[a]	CLPM	RI-CLPM	other	positive[c]	negative	inconclusive	positive	negative	inconclusive
COR Theory	31	14	12	–	5	3	3	2	1	4	12	4	2	10	16	–	5	18	2	3	19	1	–
JD-R Model	14	10	2	1	–	–	–	3	–	1	7	2	1	3	5	–	6	8	2	1	5	1	–
SSCM SCM	7	6	1	–	–	–	–	1	–	1	1	3	1	2	2	–	3	5	1	1	2	–	–
Boundary Theory Work-Family BT	4	3	1	–	–	–	–	–	1	–	2	–	1	2	2	–	–	3	1	–	1	–	–
W-HR Model	4	2	1	–	1	–	–	2	–	1	–	1	–	1	1	–	2	2	1	1	1	–	–
All studies	95	60	19	4	12	6	3	6	5	14	27	19	15	34	29	–	32	59	9	11	33	2	1

Note. [a] = includes hierarchical time-lagged regression analyses; [b] = includes also indicators of general health (e.g., perceived health, physical symptoms); [c] = Studies showing reciprocal relationships are double-coded (e.g., 1= WFC leads to strain: positive effect of WFC on strain; 2 = Strain leads to WFC: positive effect of strain on WFC); 36 of the 95 reviewed studies used a different theory or build upon relevant prior research and do not refer to a specific theory; COR = Conservation of Resources; JD-R Model = Job Demands-Resources Model; SSCM = Stress and Strain Crossover Model; BT = Border Theory; SCM = The Spillover-Crossover Model; W-HR = Work-Home Resources Model.

Table 2. Work-Family Theories and Related Examples of Stable Between-Person Differences, Short-Term Effects, and Long-Term Effects

Theory	Main Idea	Stable Between-Person Differences (Selected)	Proposed Effects and Their Time Orientation (Selected)	Time Orientation
Conservation of Resources Theory (Hobfoll, 1989; Hobfoll et al., 2018)	The COR theory is a "resource-oriented model […] based on the supposition that people strive to retain, protect, and build resources and that what is threatening to them is the potential or actual loss of these valued resources" (Hobfoll, 1989, p. 513).	"[R]esources are not distributed equally, and those people who lack resources are most vulnerable to additional losses." (Hobfoll, 1989, p. 519)	"Investment of resources may be observed in good marriages. In such marriages, both partners are constantly contributing from what they have to each other and to the relationship. There is a long-term expectation, however, that their investment will produce a payoff in terms of returned love, esteem, affection, and security (Clark, 1983)." (Hobfoll, 1989, p. 520)	Long-term
		"those who lack resources attempt to employ what resources they have, often producing self-defeating consequences." (Hobfoll, 1989, p. 519)	"COR theory posits that resource loss not only is more powerful than resource gain in magnitude but also tends to affect people more rapidly and at increasing speed over time." (Hobfoll et al., 2018, p. 105)	Not specified
			"[R] resource loss has a spiraling nature. Because resource loss is more powerful than resource gain, and because stress occurs when resources are lost, at each iteration of the stress spiral individuals and organizations have fewer resources to offset resource loss. This creates resource loss spirals whereby losses gain in both impact and momentum." (Hobfoll et al., 2018, p. 107)	Not specified
			"The role that time plays […] can take many forms, ranging from the amount of time over which resources are lost or gained (e.g., acute versus chronic stressors), to the length of recovery periods necessary to regain resources, to the specific timing that a resource becomes available relative to the timing of resource loss." (Hobfoll et al., 2018, p. 114)	Not specified
Job Demands-Resources Model (Demerouti et al., 2001; Bakker et al., 2003)	"[W]orking conditions can be categorized into two […] job demands and job resources (Demerouti et al., 2001, p. 499).	"[…] the relative contribution of specific job demands and resources to explaining burnout may vary across occupations because job demands as well as access to job resources may differ."	„Exhaustion is defined as a consequence of intensive physical, affective, and cognitive strain, for example as a long-term consequence of prolonged exposure to certain demands." (Demerouti et al., 2001, p. 500)	Long-term
	"Burnout develops […] when job demands are high and when job		"Demanding aspects of work (i.e., extreme job demands) lead to constant overtaxing and in the end, to exhaustion. [...] A lack of resources complicates the meeting of job demands, which further leads to withdrawal behavior. The long-term consequence of this withdrawal is disengagement from work." (Demerouti et al., 2001, p. 502)	Long-term

	resources are limited because such negative working conditions lead to energy depletion and undermine employees' motivation, respectively." (Demerouti et al., 2001, p. 499)	(Demerouti et al., 2001, p. 502)	"A lack of resources (e.g., autonomy, performance feedback) precludes actual goal accomplishment, which causes failure and frustration. The long-term consequence of this frustration is disengagement from work (cynism) and a reduced sense of professional efficacy" (Bakker et al., 2003, p. 22)	Long-term
Stress and Strain Crossover Model (Westman, 2001) Spillover-crossover model (SCM) (Bakker, & Demerouti, 2013)	"Stress and strain experienced by an individual cross over to others […] and influence their stress and strain […] through interpersonal factors (i.e., coping strategies, social support, social undermining)" (Westman, 2001, p. 732).	Persons differ from one another in terms of coping strategies (e.g., emotion-focused coping, problem-focused coping, avoidance-oriented coping) that they use in response to stress and strain (Powers et al., 2002; Nielsen, & Knardahl, 2014)	"In the long run, such [ineffective] coping strategies [that is, withdrawal from stressful situations] may also result in negative psychological outcomes for the coper". (Westman, 2001, p. 733)	Long-term
			"Stress experienced by one partner creates demands for support, and when unable to meet these expectations, the other is apt to feel anxious or guilty". "Job stress and strain influence social support by requiring social support from the donors, thus depleting their resources and enhancing their stress and strain". (Westman, 2001, p. 735)	Not specified
			"Social undermining mediates the process of crossover from the spouse to another. The strain of one spouse increases the process of social undermining between the partners. As one (Individual A) becomes more depressed he or she is more inclined to display undermining behaviors toward the partner. In turn, this undermining behavior of Individual A elevates the strain of the other partner (Individual B). Consequently, Individual B's strain increase social undermining behaviors toward Individual A". (Westman, 2001, p. 737)	Not specified
			"Job demands are hypothesized to evoke strain which can spill over into the home domain, and lead to work-family conflict. For instance, employees who are confronted with high emotional demands may feel fatigued after a day at work, and may continue to ruminate about work when at home. According to the SCM, this state of work-family conflict will have a negative impact on the interaction with the partner at home and indirectly on the partner's well-being. In contrast, job resources are hypothesized to foster engagement, which leads to work-family enrichment when these resources are high". (Bakker & Demerouti, 2013, p. 9)	Short-term
			"The SCM proposes that stress factors like job demands and burnout spill over to the home domain, and have an indirect negative impact on the	Long-term

			social support offered to the partner. This process holds for both partners, and the social support offered by men is positively related to the social support offered by women – in other words, support is reciprocated. This means that, in the long run, work-family conflict also reduces the social support one receives – work-family conflict undermines the quality of the relationship" (Bakker & Demerouti, 2013, p. 11)	
Boundary Theory (Ashforth et al., 2000) Work-Family Border Theory (Clark, 2000; Allen et al., 2014)	"Individuals create and maintain boundaries [around their roles] as a means of simplifying and ordering the environment. (Ashforth, & Kreiner, 2000, p. 474). Role boundaries can be described in terms of their flexibility and permeability (Ashforth et al., 2000, p. 474)	Persons differ in the degree to which they segment or integrate their work and home roles (Nippert-Eng, 1996) "The greater the role segmentation, the less difficult it tends to be to create and maintain role boundaries but the more difficult it tends to be to cross the role boundaries" (Ashforth et al., 2000, p. 477)	"In time, with repeated enactments of a transition script, both the cuing and the enactment of the transition process are likely to become relatively automatic or mindless". In plain words, the longer one switches between a pair of roles, the more automatic or "mindless" becomes the role transition/boundary crossing (Ashforth et al., 2000, p. 485).	Long-term
			"The cuing and enactment of a role schema tends to become more automatic over time so that one may enter a role and execute at least portions of it quite reflexively […] The teacher, interrupted by an emergency telephone from home, might resume his lecture with little difficulty or loss of continuity" (Ashforth et al., 2000, p. 486)	Long-term
			"Once the rites [of separation] and psychological disengagement have begun, it may be very difficult to fully re-engage an individual in workplace issues, even if he or she is still physically present in the workplace" (Ashforth, & Kreiner, 2000, p. 478)	Short-term
Work-home Resources Model (Ten Brummelhuis & Bakker, 2012)	"The [...] work–home resources (W-HR) model describes work–home conflict as a process whereby demands in one domain deplete personal resources and impede accomplishments in the other domain." (Ten Brummerlhuis & Bakker, 2012, p. 545)	"Work-home conflict is less likely among persons with key and macro resources because key and macro resources attenuate the negative relationship between contextual demands and personal resources" (Ten Brummelhuis, & Bakker, 2012, p. 551)	"Long-term work–home conflict and enrichment reflect durable processes between the work and home domains, whereby structural contextual demands and resources from one domain affect long-term outcomes in the other domain through a change in structural personal resources." (Ten Brummelhuis, & Bakker, 2012, p. 552)	Long-term
			"Short-term work–home conflict and enrichment reflect daily processes between the work and home domains, whereby volatile contextual demands and resources from one domain affect daily outcomes in the other domain through a change in volatile personal resources." (Ten Brummelhuis, & Bakker, 2012, p. 552)	Short-term

Note. The time orientation is determined based on a survey of 30 scholars on how they interpret these theories. See the additional online material (https://osf.io/brpwc) for a more detailed report of the survey study.

Table 3. Exclusion Process and the Final Study Sample

	Panel 1 1-day lag	**Panel 2** 1-week Lag	**Panel 3** 1-month Lag	**Panel 4** 6-month Lag	**Panel 5** 1-year Lag	**All Panels**	**Popu-lation**
Initial Sample	1,598	2,172	2,159	6,067	14,137	26,133	
Exclusion criteria							
Missingness in all study variables	1,598 (-0)	2,172 (-0)	2,159 (-0)	6,067 (-0)	14,121 (-16)[a]	16	
Change in sex over time	1,598 (-0)	2,172 (-0)	2,159 (-0)	6,067 (-0)	14,082 (-39)	39	
Lower education degree over time	1,598 (-0)	2,172 (-0)	2,159 (-0)	5,998 (-69)	13,693 (-389)	458	
Lower organizational tenure over time	1,598 (-0)	2,172 (-0)	2,159 (-0)	5,968 (-30)	13,142 (-551)	581	
Moving into or out of a leadership position	1,598 (-0)	2,172 (-0)	2,159 (-0)	5,797 (-171)	12,614 (-528)	699	
Change from a full-time into a part-time position or vice versa	1,598 (-0)	2,172 (-0)	2,159 (-0)	5,565 (-232)	12,241 (-373)	605	
Final study sample	1,598	2,172	2,159	5,565	12,241	23,735	
Women (sex = 2)	43.7%	45.5%	48.5%	52.3%	46.7%	47.8%	47%
Age (mean)	42.7	48.6	48.6	44.0	44.0	45.0	44
Asian	1%	0.7%	1%	-	-	-	- [b]
Arabic	0.7%	0.5%	0.5%	-	-	-	- [b]
Black	0,6%	0.1%	0.3%	-	-	-	- [b]
Latin American	0,4%	0.09%	0.3%	-	-	-	- [b]
White	95%	95.6%	95.3%	-	-	-	- [b]
Other	2,3%	3%	2,6%	-	-	-	- [b]
University degree (education = 1)	38.7%	32.3%	31.0%	37.7%	33.6%	34.5%	22%
High school diploma	24.9%	20.0%	19.7%	21.7%	20.7%	21.0%	38%
Secondary school degree	30.7%	40.0%	41.0%	36.7%	38.7%	38.1%	27%
Lower vocational degree	5.4%	7.0%	8.3%	3.7%	6.4%	6.0%	20%
No vocational degree (education = 5)	0.1%	0.2%	0.1%	0.0%	0.2%	0.1%	2,4%

Note. Initial sample includes only observations that passed attention checks; Data on the German workforce population were derived from the reports of the Federal Statistical Office of Germany. Gender non-conforming informants formed less than 1% of the sample and are not reported because this option was not included in all surveys. [a] = The data, part of a broader survey for various studies, includes observations irrelevant to this paper, as they lacked data on our variables but provided data on others. [b] = The Federal Ministry of Interior and Community has not collected data on ethnic origin since the Second World War. Breakdown by survey waves available via OSF (https://osf.io/hka7v)

Table 4. Model Fit Indices for Measurement Invariance Models and Main Models

Model	χ^2	df	p	$\Delta\chi^2_{SB}$	Δdf	p	CFI	TLI	RMSEA	SRMR
Panel 1 (1-day lag)										
Configural	157.86	99	<.001				0.99	0.99	0.02	0.02
Weak F.	168.35	107	<.001	11.19	8	.191	0.99	0.99	0.02	0.02
Strong F.	177.50	115	<.001	10.37	8	.240	0.90	0.99	0.02	0.02
CLPM	228.46	118	<.001				0.99	0.99	0.02	0.04
RI-CLPM	178.24	115	<.001	50.22	3	<.001	0.99	0.99	0.02	0.02
Panel 2 (1-week lag)										
Configural	226.74	99	<.001				0.99	0.99	0.02	0.02
Weak F.	242.79	107	<.001	16.76	8	.033	0.99	0.99	0.02	0.02
Strong F.	259.92	115	<.001	17.47	8	.026	0.99	0.99	0.02	0.02
CLPM	357.36	118	<.001				0.99	0.98	0.03	0.03
RI-CLPM	274.83	115	<.001	82.53	3	<.001	0.99	0.99	0.03	0.02
Panel 3 (1-month lag)										
Configural	213.84	99	<.001				0.99	0.99	0.02	0.02
Weak F.	223.51	107	<.001	11.88	8	.157	0.99	0.99	0.02	0.02
Strong F.	236.96	115	<.001	14.49	8	.070	0.99	0.99	0.02	0.02
CLPM	383.76	118	<.001				0.99	0.98	0.03	0.03
RI-CLPM	234.93	115	<.001	148.83	3	<.001	0.99	0.99	0.02	0.02
Panel 4 (6-month lag)										
Configural	243.04	99	<.001				1.00	0.99	0.02	0.02
Weak F.	257.07	107	<.001	15.57	8	.049	1.00	0.99	0.02	0.02
Strong F.	269.59	115	<.001	14.29	8	.075	0.99	0.99	0.02	0.02
CLPM	461.95	118	<.001				0.99	0.99	0.02	0.04
RI-CLPM	262.40	115	<.001	199.55	3	<.001	1.00	0.99	0.01	0.02
Panel 5 (1-year lag)										
Configural	353.42	96	<.001				0.99	0.99	0.01	0.01
Weak F.	366.12	104	<.001	15.81	8	0.045	0.99	0.99	0.01	0.01
Strong F.	411.38	112	<.001	41.53	8	<.001	0.99	0.99	0.01	0.02
CLPM	651.07	115	<.001				0.99	0.99	0.02	0.03
RI-CLPM	402.74	112	<.001	248.33	3	<.001	0.99	0.99	0.01	0.02

Note. Robust maximum likelihood estimation (MLR). $\Delta\chi^2_{SB}$ = Chi-square difference using Satorra-Bentler correction; Δdf = Degrees of freedom difference; χ^2 = Model chi-square; df = Degrees of freedom; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean squared residual; p = p value of nested chi-square model test.

Table 5. Means, Standard Deviations, and Correlations Between the Latent Study Variables in Panel 1 (1-Day Lag)

	Variables	Mean	SD	Rel	ICC	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	Age W1	44.13	12.11																			
2	Sex W1	1.48	0.51			-.19																
3	Education W1	2.07	0.96			.14	.02															
4	Perceived Stress W1	1.77	0.77	.87	.72	-.21	.11	-.01														
5	Perceived Stress W2	1.77	0.80	.88		-.15	.15	.02	.73													
6	Perceived Stress W3	1.77	0.81	.89		-.15	.10	.01	.69	.75												
7	Exhaustion W1	2.22	0.97	.86	.76	-.19	.19	.00	.81	.67	.64											
8	Exhaustion W2	2.18	0.95	.86		-.18	.18	.03	.70	.85	.73	.79										
9	Exhaustion W3	2.18	1.00	.89		-.18	.15	.02	.65	.72	.83	.71	.79									
10	Rumination W1	1.67	0.83	.91	.74	-.22	.10	-.03	.85	.66	.64	.75	.64	.62								
11	Rumination W2	1.67	0.83	.91		-.20	.09	-.00	.71	.86	.75	.60	.74	.69	.75							
12	Rumination W3	1.65	0.87	.92		-.17	.09	.01	.65	.69	.87	.56	.62	.76	.71	.78						
13	WFC W1	1.80	0.94	.92	.71	-.17	.09	-.07	.73	.59	.59	.68	.59	.61	.75	.62	.58					
14	WFC W2	1.78	0.96	.93		-.20	.09	-.01	.64	.75	.64	.61	.72	.64	.64	.76	.63	.71				
15	WFC W3	1.76	0.96	.94		-.14	.07	-.03	.57	.62	.73	.53	.60	.72	.62	.67	.72	.69	.73			
16	FWC W1	1.46	0.72	.93	.79	-.17	.02	-.06	.62	.52	.53	.50	.44	.46	.65	.55	.52	.68	.58	.60		
17	FWC W2	1.47	0.73	.93		-.17	-.02	-.04	.56	.61	.56	.46	.51	.52	.59	.62	.55	.58	.66	.61	.77	
18	FWC W3	1.43	0.71	.94		-.18	.00	-.01	.51	.56	.60	.43	.47	.53	.57	.59	.60	.54	.59	.68	.77	.84

Note. WFC = work-family conflict; FWC = family-work conflict; Rumination = Affective Rumination; W1 = wave 1, W2 = wave 2, W3 = wave 3; Rel = The congeneric reliability index (Cho, 2016). Latent variable correlations are from weak factorial invariance model. ICCs are calculated using long-form data. All correlations are significant at 5% significance level with p <.01. The reported means are indicator means.

Table 6. Means, Standard Deviations, and Correlations Between the Latent Study Variables in Panel 2 (1-Week Lag)

	Variables	Mean	SD	Rel	ICC	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Age W1	48.02	13.24																
2	Sex W1	1.47	0.50			-.15													
3	Education W1	3.76	0.99			-.14	-.03												
4	Perceived Stress W1	2.42	0.93	.88	.75	-.30	.15	.05											
5	Perceived Stress W2	2.33	0.97	.90		-.28	.15	.05	.77										
6	Perceived Stress W3	2.27	0.96	.90		-.27	.12	.01	.77	.80									
7	Exhaustion W1	2.69	0.96	.88	.81	-.26	.14	.01	.82	.70	.71								
8	Exhaustion W2	2.62	0.99	.89		-.26	.15	-.01	.75	.80	.74	.85							
9	Exhaustion W3	2.56	0.99	.90		-.27	.15	-.02	.75	.73	.81	.85	.89						
10	WFC W1	2.30	0.97	.91	.86	-.26	.03	.07	.66	.57	.61	.73	.69	.68					
11	WFC W2	2.24	0.98	.93		-.23	.04	.05	.59	.61	.61	.64	.71	.67	.82				
12	WFC W3	2.17	0.99	.93		-.24	.03	.05	.54	.54	.62	.59	.64	.69	.78	.84			
13	FWC W1	1.81	0.78	.92	.78	-.24	-.05	.09	.44	.41	.42	.40	.38	.38	.62	.54	.53		
14	FWC W2	1.78	0.78	.92		-.22	-.03	.08	.41	.43	.41	.37	.39	.37	.50	.60	.53	.75	
15	FWC W3	1.76	0.78	.93		-.23	-.05	.08	.37	.38	.44	.35	.39	.41	.49	.53	.62	.72	.78

Note. WFC = work-family conflict; FWC = family-work conflict; W1 = wave 1, W2 = wave 2, W3 = wave 3; Rel = The congeneric reliability index (Cho, 2016). Latent variable correlations are from weak factorial invariance model. ICCs are calculated using long-form data. All correlations are significant at 5% significance level with p <.01. The reported means are indicator means.

Table 7. Means, Standard Deviations, and Correlations Between the Latent Study Variables in Panel 3 (1-Month Lag)

	Variables	Mean	SD	Rel	ICC	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Age W1	47.94	13.27																
2	Sex W1	1.47	0.50			-.17													
3	Education W1	3.73	0.99			-.12	-.02												
4	Perceived Stress W1	2.49	0.96	.89	.77	-.33	.20	.02											
5	Perceived Stress W2	2.36	0.95	.89		-.33	.17	.04	.79										
6	Perceived Stress W3	2.29	0.92	.89		-.33	.16	.05	.73	.79									
7	Exhaustion W1	2.72	0.99	.88	.85	-.29	.16	-.02	.81	.75	.67								
8	Exhaustion W2	2.57	1.01	.89		-.30	.15	.01	.76	.84	.74	.86							
9	Exhaustion W3	2.54	1.01	.89		-.27	.13	.02	.70	.71	.79	.84	.86						
10	WFC W1	2.28	0.96	.91	.80	-.26	.04	.07	.65	.59	.55	.70	.64	.63					
11	WFC W2	2.22	0.97	.92		-.25	.06	.08	.59	.66	.58	.63	.71	.65	.78				
12	WFC W3	2.14	0.96	.93		-.25	.01	.07	.55	.56	.61	.61	.63	.69	.79	.83			
13	FWC W1	1.80	0.81	.92	.76	-.25	-.05	.10	.41	.41	.40	.39	.37	.36	.62	.51	.54		
14	FWC W2	1.75	0.82	.94		-.26	-.03	.09	.36	.47	.43	.37	.42	.39	.52	.63	.57	.73	
15	FWC W3	1.71	0.80	.93		-.22	-.06	.07	.38	.43	.47	.36	.39	.43	.51	.53	.65	.75	.79

Note. WFC = work-family conflict; FWC = family-work conflict; W1 = wave 1, W2 = wave 2, W3 = wave 3; Rel = The congeneric reliability index (Cho, 2016). Latent variable correlations are from weak factorial invariance model. ICCs are calculated using long-form data. All correlations are significant at 5% significance level with $p < .01$. The reported means are indicator means.

Table 8. Means, Standard Deviations, and Correlations Between the Latent Study Variables in Panel 4 (6-Month Lag)

	Variables	Mean	SD	Rel	ICC	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Age W1	44.29	11.21																
2	Sex W1	1.53	0.50			-.04													
3	Education W1	2.91	0.95			-.17	-.11												
4	Perceived Stress W1	2.57	0.83	.88	.68	-.15	.13	.01											
5	Perceived Stress W2	2.57	0.85	.88		-.18	.14	.02	.68										
6	Perceived Stress W3	2.55	0.86	.88		-.15	.09	-.01	.68	.67									
7	Exhaustion W1	2.86	0.97	.87	.76	-.13	.12	-.04	.81	.61	.63								
8	Exhaustion W2	2.83	0.99	.88		-.16	.13	-.02	.65	.82	.67	.76							
9	Exhaustion W3	2.85	1.01	.88		-.15	.09	-.02	.64	.63	.83	.74	.76						
10	WFC W1	2.46	1.02	.91	.73	-.13	.04	.05	.61	.52	.49	.67	.58	.53					
11	WFC W2	2.39	1.03	.92		-.14	.02	.05	.54	.63	.53	.56	.67	.55	.73				
12	WFC W3	2.48	1.05	.92		-.14	.01	.05	.52	.54	.63	.55	.57	.68	.71	.73			
13	FWC W1	1.91	0.86	.92	.68	-.18	-.08	.08	.40	.35	.35	.37	.33	.30	.57	.47	.46		
14	FWC W2	1.90	0.88	.93		-.22	-.06	.08	.36	.45	.38	.30	.41	.33	.43	.63	.49	.68	
15	FWC W3	1.94	0.89	.94		-.21	-.07	.07	.34	.38	.43	.31	.34	.39	.40	.46	.62	.66	.69

Note. WFC = work-family conflict; FWC = family-work conflict; W1 = wave 1, W2 = wave 2, W3 = wave 3; Rel = The congeneric reliability index (Cho, 2016). Latent variable correlations are from weak factorial invariance model. ICCs are calculated using long-form data. All correlations are significant at 5% significance level with p <.01. The reported means are indicator means.

Table 9. Means, Standard Deviations, and Correlations Between the Latent Study Variables in Panel 5 (1-Year Lag)

	Variables	Mean	SD	Rel	ICC	1	2	3	4	5	6	7	8
1	Age W1	44.21	11.43										
2	Sex W1	1.49	0.50			-.12							
3	Education W1	3.74	1.00			-.10	-.06						
4	Exhaustion W1	2.78	1.02	.86	.72	-.17	.07	-.07					
5	Exhaustion W2	3.02	0.98	.84		-.21	.09	-.05	.76				
6	Exhaustion W3	2.88	1.02	.84		-.21	.09	-.05	.71	.77			
7	WFC W1	2.65	1.04	.88	.69	-.16	-.06	.04	.59	.54	.49		
8	WFC W2	2.64	1.02	.89		-.22	-.03	.06	.51	.67	.54	.70	
9	WFC W3	2.55	1.04	.88		-.22	-.05	.07	.49	.55	.62	.69	.72

Note. W1 = wave 1, W2 = wave 2, W3 = wave 3; Rel = The congeneric reliability index (Cho, 2016). Latent variable correlations are from weak factorial invariance model. ICCs are calculated using long-form data. All correlations are significant at 5% significance level with p <.01. The reported means are indicator means.

Table 10. Summary of the CLPM Results Across Panels with Varying Lags

	Parameter					
	Autoregressive paths		Cross-lagged paths		Covariances	
	$Conf_t \rightarrow Conf_{t+1}$	$Strain_t \rightarrow Strain_{t+1}$	$Conf_t \rightarrow Strain_{t+1}$	$Strain_t \rightarrow Conf_{t+1}$	$Strain_1 \leftrightarrow Conf_1$	$Strain_{2,3} \leftrightarrow Conf_{2,3}$
Panel 1 (1-day lag)						
WFC						
Exhaustion	0.620*** (0.043) [<.001]	0.739*** (0.033) [<.001]	0.124*** (0.031) [<.001]	0.193*** (0.039) [<.001]	0.719*** (0.041) [<.001]	0.206*** (0.020) [<.001]
Rumination	0.560*** (0.044) [<.001]	0.716*** (0.036) [<.001]	0.094** (0.030) [0.002]	0.283*** (0.045) [<.001]	0.741*** (0.041) [<.001]	0.198*** (0.023) [<.001]
Stress	0.585*** (0.045) [<.001]	0.687*** (0.040) [<.001]	0.117*** (0.031) [<.001]	0.269*** (0.049) [<.001]	0.659*** (0.037) [<.001]	0.205*** (0.022) [<.001]
FWC						
Exhaustion	0.794*** (0.028) [<.001]	0.781*** (0.025) [<.001]	0.114*** (0.026) [<.001]	0.059** (0.019) [0.002]	0.446*** (0.036) [<.001]	0.075*** (0.015) [<.001]
Rumination	0.749*** (0.033) [<.001]	0.727*** (0.032) [<.001]	0.107** (0.031) [0.001]	0.114*** (0.027) [<.001]	0.548*** (0.035) [<.001]	0.089*** (0.015) [<.001]
Stress	0.766*** (0.033) [<.001]	0.706*** (0.033) [<.001]	0.129*** (0.029) [<.001]	0.097** (0.029) [0.001]	0.489*** (0.036) [<.001]	0.094*** (0.015) [<.001]
Panel 2 (1-week lag)						
WFC						
Exhaustion	0.776*** (0.027) [<.001]	0.842*** (0.021) [<.001]	0.087*** (0.019) [<.001]	0.094*** (0.025) [<.001]	0.668*** (0.026) [<.001]	0.090*** (0.009) [<.001]
Stress	0.803*** (0.023) [<.001]	0.727*** (0.024) [<.001]	0.147*** (0.020) [<.001]	0.069** (0.021) [0.001]	0.600*** (0.024) [<.001]	0.083*** (0.010) [<.001]
FWC						
Exhaustion	0.752*** (0.021) [<.001]	0.896*** (0.014) [<.001]	0.034* (0.017) [0.039]	0.062*** (0.012) [<.001]	0.285*** (0.019) [<.001]	0.037*** (0.008) [<.001]
Stress	0.749*** (0.022) [<.001]	0.790*** (0.019) [<.001]	0.103*** (0.022) [<.001]	0.060*** (0.014) [<.001]	0.327*** (0.021) [<.001]	0.049*** (0.010) [<.001]
Panel 3 (1-month lag)						
WFC						
Exhaustion	0.714*** (0.026) [<.001]	0.842*** (0.022) [<.001]	0.067** (0.020) [0.001]	0.141*** (0.023) [<.001]	0.668*** (0.026) [<.001]	0.109*** (0.013) [<.001]
Stress	0.763*** (0.023) [<.001]	0.731*** (0.023) [<.001]	0.096*** (0.021) [<.001]	0.085*** (0.020) [<.001]	0.621*** (0.026) [<.001]	0.108*** (0.012) [<.001]
FWC						

Exhaustion	0.748***	0.878***	0.036*	0.060***	0.299***	0.052***
	(0.021)	(0.015)	(0.018)	(0.012)	(0.020)	(0.010)
	[<.001]	[<.001]	[0.044]	[<.001]	[<.001]	[<.001]
Stress	0.743***	0.762***	0.099***	0.064***	0.325***	0.071***
	(0.022)	(0.017)	(0.020)	(0.013)	(0.021)	(0.010)
	[<.001]	[<.001]	[<.001]	[<.001]	[<.001]	[<.001]
Panel 4 (6-month lag)						
WFC						
Exhaustion	0.651***	0.747***	0.077***	0.160***	0.651***	0.207***
	(0.022)	(0.020)	(0.017)	(0.021)	(0.020)	(0.013)
	[<.001]	[<.001]	[<.001]	[<.001]	[<.001]	[<.001]
Stress	0.666***	0.607***	0.138***	0.173***	0.532***	0.167***
	(0.021)	(0.023)	(0.017)	(0.023)	(0.018)	(0.011)
	[<.001]	[<.001]	[<.001]	[<.001]	[<.001]	[<.001]
FWC						
Exhaustion	0.695***	0.791***	0.035*	0.058***	0.306***	0.105***
	(0.017)	(0.014)	(0.015)	(0.012)	(0.017)	(0.011)
	[<.001]	[<.001]	[0.023]	[<.001]	[<.001]	[<.001]
Stress	0.679***	0.673***	0.096***	0.094***	0.295***	0.100***
	(0.017)	(0.018)	(0.015)	(0.015)	(0.015)	(0.010)
	[<.001]	[<.001]	[<.001]	[<.001]	[<.001]	[<.001]
Panel 5 (1-year lag)						
WFC						
Exhaustion	0.657***	0.721***	0.096***	0.129***	0.628***	0.188***
	(0.016)	(0.016)	(0.014)	(0.015)	(0.016)	(0.010)
	[<.001]	[<.001]	[<.001]	[<.001]	[<.001]	[<.001]

Note. *** p<.001, ** p<.01, * p<.05 Standard errors in parentheses. p values in brackets. WFC = Work-family conflict, FWC = Family-work conflict, Rumination = Affective Rumination; $Conf_t$ = Conflict, either WFC or FWC, at time t. $Strain_t$ = Either Exhaustion, Perceived Stress, or Affective Rumination, at time t.

Table 11. Summary of the RI-CLPM Results Across Panels with Varying Lags

	Parameter						
	Autoregressive paths		Cross-lagged paths		Covariances		
	$Conf_t \rightarrow Conf_{t+1}$	$Strain_t \rightarrow Strain_{t+1}$	$Conf_t \rightarrow Strain_{t+1}$	$Strain_t \rightarrow Conf_{t+1}$	$Strain_1 \leftrightarrow Conf_1$	$Strain_{2,3} \leftrightarrow Conf_{2,3}$	Between-level
Panel 1 (1-day lag)							
WFC							
Exhaustion	0.023	0.369**	0.037	0.230*	0.134***	0.170***	0.596***
	(0.118)	(0.108)	(0.086)	(0.108)	(0.037)	(0.029)	(0.049)
	[0.843]	[0.001]	[0.668]	[0.033]	[<.001]	[<.001]	[<.001]
Rumination	0.007	0.298*	-0.065	0.175	0.153***	0.127***	0.608***
	(0.114)	(0.149)	(0.078)	(0.131)	(0.035)	(0.034)	(0.047)
	[0.952]	[0.045]	[0.402]	[0.181]	[<.001]	[<.001]	[<.001]
Stress	-0.013	0.269*	0.001	0.243*	0.131***	0.152***	0.543***
	(0.112)	(0.121)	(0.078)	(0.124)	(0.031)	(0.030)	(0.042)
	[0.907]	[0.026]	[0.992]	[0.049]	[<.001]	[<.001]	[<.001]
FWC							
Exhaustion	0.010	0.377**	0.036	0.078	0.090**	0.047	0.411***
	(0.128)	(0.143)	(0.129)	(0.093)	(0.026)	(0.028)	(0.048)
	[0.940]	[0.008]	[0.779]	[0.401]	[0.001]	[0.087]	[<.001]
Rumination	0.010	0.309*	-0.122	0.056	0.099***	0.035	0.480***
	(0.129)	(0.151)	(0.112)	(0.078)	(0.025)	(0.020)	(0.034)
	[0.941]	[0.041]	[0.279]	[0.470]	[<.001]	[0.081]	[<.001]
Stress	-0.043	0.255	-0.054	0.137	0.101***	0.050*	0.429***
	(0.111)	(0.130)	(0.109)	(0.086)	(0.024)	(0.023)	(0.038)
	[0.700]	[0.050]	[0.623]	[0.108]	[<.001]	[0.033]	[<.001]
Panel 2 (1-week lag)							
WFC							
Exhaustion	0.336**	0.042	0.078	-0.085	0.082***	0.051**	0.601***
	(0.109)	(0.137)	(0.078)	(0.119)	(0.016)	(0.015)	(0.027)
	[0.002]	[0.762]	[0.319]	[0.476]	[<.001]	[0.001]	[<.001]
Stress	0.318**	0.092	0.037	-0.024	0.079***	0.046**	0.536***
	(0.110)	(0.109)	(0.079)	(0.078)	(0.017)	(0.017)	(0.027)
	[0.004]	[0.398]	[0.637]	[0.763]	[<.001]	[0.008]	[<.001]
FWC							
Exhaustion	0.164	0.067	0.044	0.120	0.037**	0.030	0.255***
	(0.112)	(0.138)	(0.088)	(0.112)	(0.014)	(0.016)	(0.023)
	[0.144]	[0.629]	[0.621]	[0.285]	[0.009]	[0.069]	[<.001]
Stress	0.190	0.078	0.056	0.054	0.028	0.040**	0.293***
	(0.113)	(0.115)	(0.085)	(0.070)	(0.016)	(0.015)	(0.022)
	[0.093]	[0.495]	[0.508]	[0.439]	[0.075]	[0.005]	[<.001]
Panel 3 (1-month lag)							
WFC							
Exhaustion	-0.010	0.077	0.065	0.075	0.079***	0.071***	0.598***
	(0.075)	(0.111)	(0.061)	(0.095)	(0.017)	(0.018)	(0.026)
	[0.894]	[0.486]	[0.289]	[0.429]	[<.001]	[<.001]	[<.001]
Stress	-0.010	0.273**	-0.042	-0.009	0.089***	0.056***	0.532***
	(0.069)	(0.098)	(0.061)	(0.066)	(0.019)	(0.014)	(0.025)
	[0.880]	[0.005]	[0.488]	[0.888]	[<.001]	[<.001]	[<.001]

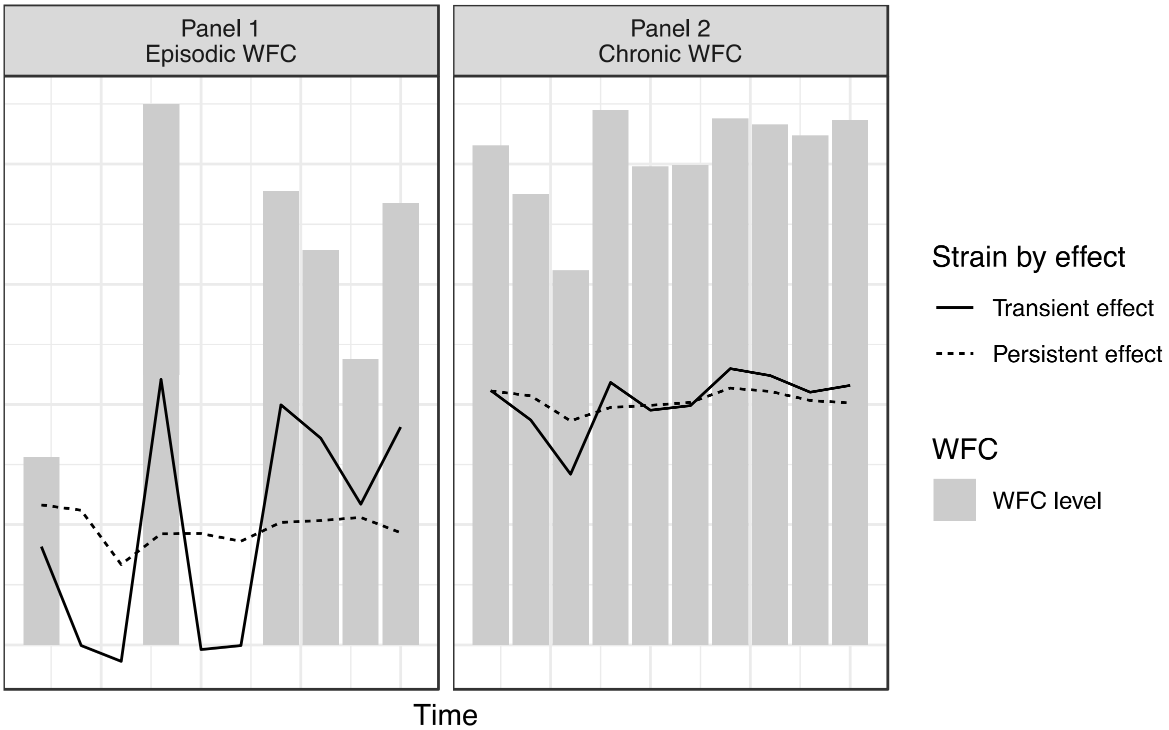
FWC							
Exhaustion	-0.041	0.107	0.061	0.073	0.033*	0.039**	0.274***
	(0.080)	(0.119)	(0.062)	(0.079)	(0.013)	(0.015)	(0.020)
	[0.611]	[0.365]	[0.326]	[0.354]	[0.012]	[0.009]	[<.001]
Stress	-0.026	0.240*	-0.008	-0.014	0.037*	0.039**	0.306***
	(0.081)	(0.096)	(0.065)	(0.056)	(0.015)	(0.012)	(0.021)
	[0.747]	[0.012]	[0.903]	[0.801]	[0.016]	[0.001]	[<.001]
Panel 4 (6-month lag)							
WFC							
Exhaustion	0.053	0.123	0.092	0.126	0.111***	0.160***	0.538***
	(0.068)	(0.088)	(0.055)	(0.067)	(0.018)	(0.016)	(0.022)
	[0.438]	[0.163]	[0.098]	[0.059]	[<.001]	[<.001]	[<.001]
Stress	0.053	-0.005	0.103*	0.136*	0.088***	0.124***	0.448***
	(0.067)	(0.072)	(0.046)	(0.057)	(0.015)	(0.016)	(0.019)
	[0.434]	[0.945]	[0.026]	[0.018]	[<.001]	[<.001]	[<.001]
FWC							
Exhaustion	0.122	0.163	0.044	0.036	0.054***	0.079***	0.255***
	(0.075)	(0.083)	(0.056)	(0.055)	(0.015)	(0.014)	(0.018)
	[0.104]	[0.050]	[0.434]	[0.518]	[<.001]	[<.001]	[<.001]
Stress	0.117	0.013	0.054	0.065	0.038**	0.074***	0.261***
	(0.075)	(0.070)	(0.050)	(0.051)	(0.013)	(0.013)	(0.015)
	[0.119]	[0.859]	[0.277]	[0.199]	[0.004]	[<.001]	[<.001]
Panel 5 (1-year lag)							
WFC							
Exhaustion	0.045	0.154*	0.069	0.068	0.120***	0.137***	0.517***
	(0.053)	(0.061)	(0.041)	(0.049)	(0.015)	(0.013)	(0.018)
	[0.395]	[0.011]	[0.094]	[0.160]	[<.001]	[<.001]	[<.001]

Note. *** p<.001, ** p<.01, * p<.05 Standard errors in parentheses. p values in brackets. WFC = Work-family conflict, FWC = Family-work conflict, Rumination = Affective Rumination; $Conf_t$ = Conflict, either WFC or FWC, at time t. $Strain_t$ = Either Exhaustion, Perceived Stress, or Rumination, at time t. Between-level = Between-level correlation between Conflict and Strain

Table 12. Examples of Hypotheses with Different Levels of Specificity

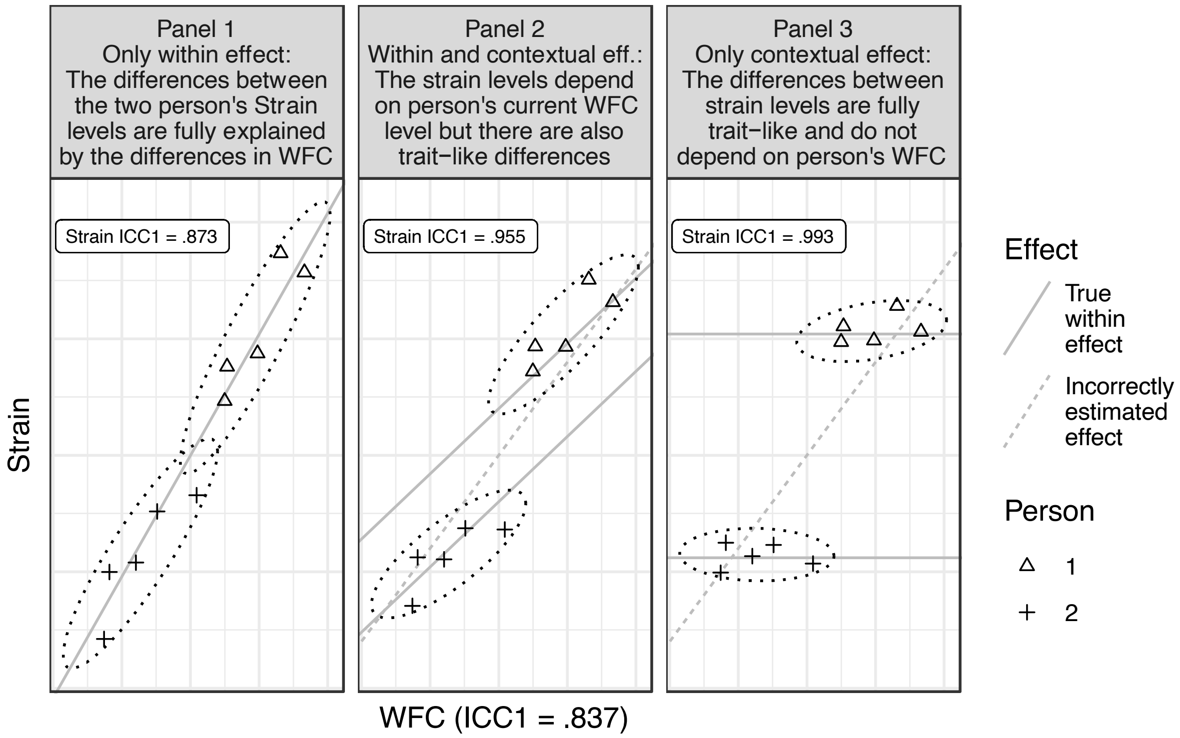
Effect and time orientation	**Hypothesis**
No effects or time orientation	WFC has a positive cross-lagged relationship with strain.
Within Effect	WFC has a positive *within-person* cross-lagged relationship with strain.
Within Effect with Explicit Time Lag	WFC has a positive *within-person* effect on strain *six months later and six months later only*.
Contextual Effect	WFC has a positive *long-term (contextual)* relationship with strain *such that the average WFC over all waves (eighteen months) has an effect even after controlling for lagged WFC.*
Comparison of Effects	WFC has a positive *long-term (contextual)* relationship with strain *that is stronger than the short-term effect of the WFC levels from last six months (within effect).*

Figure 1. Two Column Charts Demonstrating the Difference Between the Stability of Variable Levels and the Persistency of Effects



Note. Simulated data for two ten-wave panels. Columns show WFC levels during a particular wave. In Panel 1 the WFC levels strongly vary across the ten waves. In some waves the levels are equal to zero because there were no WFC episodes at these times. In Panel 2, the WFC levels are fairly stable and constantly high across the ten waves. Solid lines show the true transient WFC-strain effect. Dashed lines show the true persistent WFC-strain effect. In Panel 1 the true effect is transient, that is, it depends entirely on the current WFC level (e.g., WFC-strain effect wave 2 depends fully on WFC level in that particular wave). In Panel 2 the true effect is persistent, that is, it depends on the current WFC level plus the history of WFC levels throughout the entire study timeframe (ten waves). The history means here the rolling average of WFC levels over the ten waves. If the true effect is transient (Panel 1), these two effects can be easily distinguished. If the true effect is persistent (Panel 2), the two effects produce a similar pattern of results.

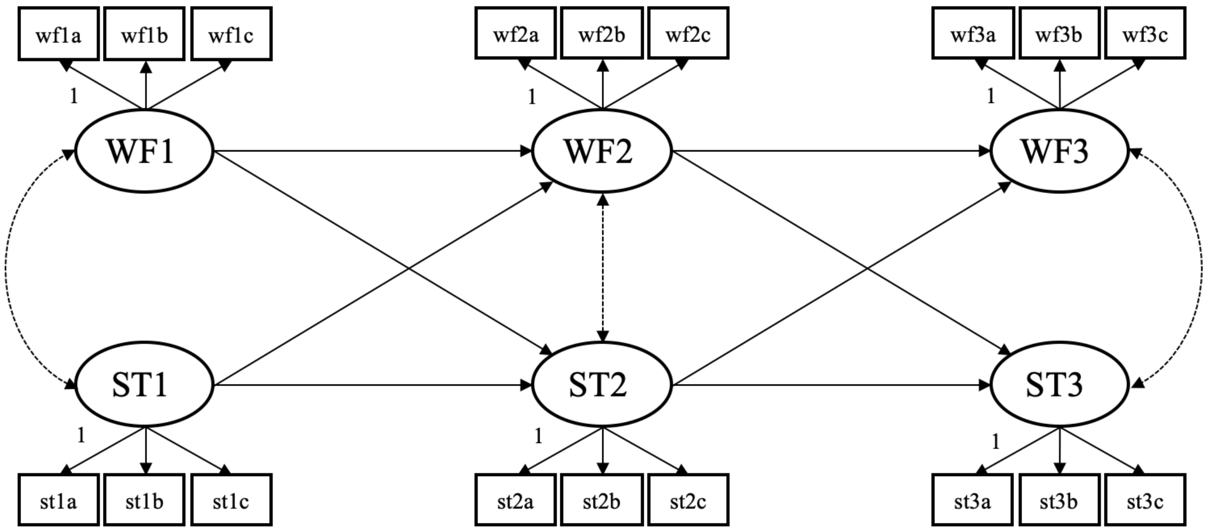
Figure 2. Three Scatterplots Demonstrating Within and Contextual Effects Using Five Repeated Measures of Two Hypothetical Persons



Note. Simulated data for three five-wave panels of two persons. All panels use the same work-family conflict (WFC) values, which are higher for person 1 (triangles) than person 2 (plus signs). Solid lines show the true within effect and dashed lines show incorrect estimates resulting from ignoring the contextual effect (e.g., if using CLPM). In panel 1 the two regression lines overlap because there is no contextual effect (the random effects assumption holds). The between effect (sum of within and contextual effects) is the same in all panels.

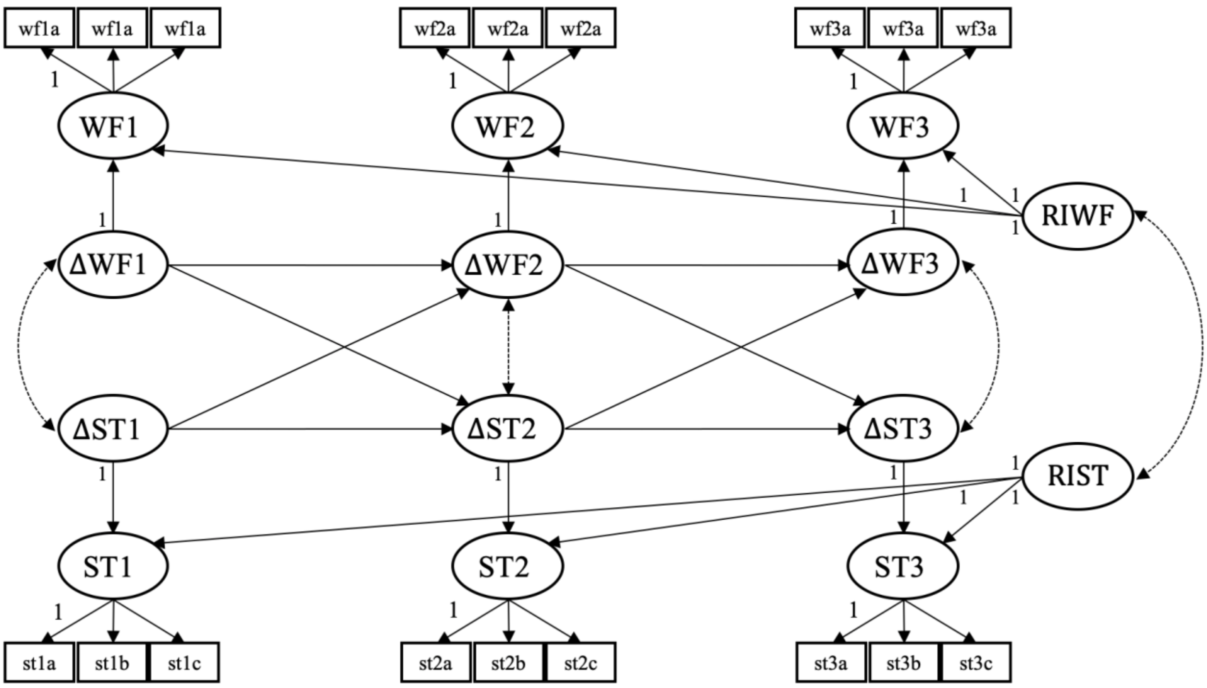
Panel 1 shows a purely within effect (with contextual effect equal to zero) where person 1's high WFC levels are caused by high current WFC levels. In this case, an intervention targeting the current WFC level could be used to bring the strain levels down to person 2's levels. Panel 2 presents a scenario where both effects are at play. The between effect, defined as the sum of the within and contextual effects, is approximately the same in the three panels. Panel 3 shows a pure contextual effect where the differences in strain are trait-like (e.g., because of prolonged exposure to high WFC levels). In this case an intervention targeting the current WFC level has no impact on the strain level.

Figure 3. Cross-Lagged Panel Model (CLPM)



Note. WF = work-family conflict; ST = strain (operationalized as exhaustion, perceived stress, daily affective rumination); Solid line = lagged paths; Dotted line = correlations of error terms. Circles illustrate latent variables. Correlations between indicator error terms are omitted for clarity.

Figure 4. Random-Intercept Cross-Lagged Panel Model (RI-CLPM)



Note. WF = work-family conflict; ST = strain (operationalized as exhaustion, perceived stress, daily affective rumination); Latent variables: WF = work-family conflict; ST = strain; Latent variables capturing within-person variance: ΔWF = work-family conflict; ΔST = strain; Latent variables capturing between-person variance: RIWF = Random intercept of work-family conflict; RIST = Random intercept of strain; Solid line = lagged paths; Dotted line = correlations of error terms or exogenous variables.