



A Personalised Approach to Identifying Important Determinants of Well-being

Joseph Ciarrochi¹ · Baljinder Sahdra¹ · Steven C. Hayes² · Stefan G. Hofmann³ · Brandon Sanford⁴ · Cory Stanton⁵ · Keong Yap⁶ · Madeleine I. Fraser⁶ · Kathleen Gates⁷ · Andrew T. Gloster⁸

Accepted: 1 April 2024 / Published online: 28 May 2024
© The Author(s) 2024

Abstract

Purpose To develop effective and personalized interventions, it is essential to identify the most critical processes or psychological drivers that impact an individual’s well-being. Some processes may be universally beneficial to well-being across many contexts and people, while others may only be beneficial to certain individuals in specific contexts.

Method We conducted three intensive daily diary studies, each with more than 50 within-person measurement occasions, across three data sets ($n_1 = 44$; $n_2 = 37$; $n_3 = 141$). We aimed to investigate individual differences in the strength of within-person associations between three distinct process measures and a variety of outcomes. We utilized a unique idiographic algorithm, known as i-ARIMAX (Autoregressive Integrated Moving Average), to determine the strength of the relationship (Beta) between each process and outcome within individuals (“i”). All of the computed betas were then subjected to meta-analyses, with individuals treated as the “study”.

Results The results revealed that the process–outcome links varied significantly between individuals, surpassing the homogeneity typically seen in meta-analyses of studies. Although several processes showed group-level effects, no process was found to be universally beneficial when considered individually. For instance, processes involving social behavior, like being assertive, did not demonstrate any group-level links to loneliness but still had significant individual-level effects that varied from positive to negative.

Discussion Using i-ARIMAX might help reduce the number of candidate variables for complex within-person analyses. Additionally, the size and pattern of i-ARIMAX betas could prove useful in guiding personalized interventions.

Keywords Well-being · Processes of change · Process-based therapy · Personalized interventions · Experience sampling · Daily diary · Meta-analysis · Process–outcome associations · Idiographic analysis · Autoregressive integrated moving average (ARIMA)

A client walks into the mental health practitioner’s (such as a psychologist’s) office seeking help. The practitioner’s job is to figure out, as quickly as possible, what they can do to help this specific individual with their specific

presenting problem in their specific context (Paul, 1969). For much of the last half-century, the general recommendation to accomplish that goal has been that the practitioner should generate a formulation understanding the client’s

✉ Joseph Ciarrochi
ciarrochj@gmail.com

¹ Institute for Positive Psychology and Education, Australian Catholic University, Sydney, Australia

² Behavior Analysis Program, Department of Psychology, University of Nevada, Reno, USA

³ Department of Clinical Psychology, Philipps-University Marburg, Germany, Marburg, Germany

⁴ Medical University of South Carolina, Charleston, SC, USA

⁵ Ralph H. Johnson VA Medical Center, Charleston, SC, USA

⁶ School of Behavior and Health Sciences, Australian Catholic University, Sydney, Australia

⁷ Quantitative Psychology Program, University of North Carolina, Chapel Hill, USA

⁸ Division of Clinical Psychology, Faculty of Behavioural Sciences and Psychology, University of Lucerne, Lucerne, Switzerland

presenting psychological problem, including a clinical diagnosis of a mental health disorder, and then administer an evidence-based treatment protocol that has been shown in randomized trials to improve outcomes for that disorder (e.g., Chambless & Ollendick, 2001; Hayes et al., 2019, 2022a). Practitioners often encounter several problems with this recommendation. First, not all components or processes targeted by a protocol are universally applicable to every person (Hayes et al., 2019; Sahdra et al., 2023, 2024). This variability is reflected in differing dropout and response rates from standardized protocols (Imel et al., 2013) and in some clients finding a single session satisfactory (Hoyt et al., 2020). Second, the client may show comorbidities or unique features that do not fit existing syndromal expectations. Indeed, the most common diagnostic category is “not otherwise specified” (Rajakannan et al., 2016). Third, the practitioner may find that the client responds well to some intervention components in the protocol but not others. If an Acceptance and Commitment Therapy (ACT; Hayes et al., 2012) protocol is being used, for example, one client may respond well to values and committed action interventions, and another to mindfulness and emotional acceptance interventions (Villatte et al., 2016).

Fourth, the formulation of treatment may change over time. As a patient progresses through treatment, new psychological symptoms may arise, requiring different approaches. For example, once a substance use disorder is successfully managed, painful childhood memories that were effectively avoided through substance use may surface. In cases where painful memories are at the root of addiction, a therapeutic approach such as Emotion Focused Therapy may prove more effective than Cognitive Behavioural Therapy in addressing and working through these underlying issues (Ehlers et al., 2014). Finally, practitioners may have to choose between different evidence-based treatment protocols, not knowing which would best suit this particular client. Although change processes may be similar across various protocols, practitioners must often undergo extensive training in individual protocols to ensure competent use. However, this can leave those who rely on evidence-based treatment recommendations to guide their clinical decision-making in a difficult situation when they try to tailor interventions to the unique needs of their clients.

Given the problems associated with complex, multifaceted protocols, there have been many models developed that focus on transtheoretical processes of change (Greenberg, 1986; Jones et al., 1988; Prochaska & DiClemente, 1983; Tedeschi & Moore, 2021). Understanding the problematic processes causing the client to feel “stuck” or distressed, and the processes for change, can help practitioners tailor evidence-based interventions to meet individual needs. In principle, a process focus should make it easier to personalize

interventions, as one can select the most relevant intervention kernel that bears on the most relevant process for a particular individual in a particular context (Hayes et al., 2019).

There has been an increasing call to identify the evidence-based intervention *kernels* (a fundamental component of interventions that effectively influences behavior) that comprise a package and the processes of change they affect (Embry & Biglan, 2008; Hayes et al., 2022a; Rosen & Davison, 2003). In broad terms, we will define a process of change as an evidence-based, theoretically coherent, contextually situated, modifiable biopsychosocial event or sequence of events that can lead to adaptive or maladaptive outcomes for a client (Hayes et al., 2020b). The Extended Evolutionary Meta Model (EEMM, see Ciarrochi et al., 2021; Hayes et al., 2020a) is a core theoretical model guiding the process-based therapy movement. The EEMM applies evolutionary concepts of context-appropriate variation, selection, and retention to key biopsychosocial dimensions and levels of organization related to human suffering, problems, and positive functioning (Hayes et al., 2022a). At the psychological level, commonly investigated processes include those focusing on cognition (e.g., functional beliefs), affect (e.g., low anxiety sensitivity), self (e.g., self-efficacy), motivation (e.g., values-based motivation), attention (e.g., mindfulness), and overt behavior (e.g., goal setting; see Hayes et al. (2022a). Processes at the social and biological level are also relevant. The core question of this paper is, how does one select the most relevant biopsychosocial process to target for a particular individual?

Researchers often attempt to answer this question by collecting data from a large group of participants, to examine the link between processes and outcomes for the group (e.g., via longitudinal or mediational analysis, perhaps as part of a randomized controlled trial), and then assume that these group level effects apply to each individual in the group (plus or minus some error; (Donald et al., 2022; Masuda et al., 2009). For example, suppose an ACT intervention improves a process of change, such as psychological flexibility for a group of participants, and flexibility correlates with or mediates outcomes. In that case, it is common to assume that ACT will likely improve that process and lead to better outcomes for the various individuals in that group (Wicksell et al., 2010).

Over two decades ago, Molenaar (2004) wrote a manifesto challenging group to individual generalizations. In subsequent years, theoreticians and researchers have further questioned the assumption that we can rely on group data to understand within-person development and change (e.g., Fisher et al., 2017; Hopwood et al., 2022; Molenaar, 2004; Rabinowitz & Fisher, 2020; Sanford et al., 2022; Wright et al., 2019). This approach relies on the assumption of *ergodicity*, which is the expectation that we can extrapolate

findings observed at the group level to individuals within the group. Ergodicity suggests that the statistical characteristics of a process, when averaged over time for the whole group,

represent each member's experiences. This means that if a system is ergodic, the behaviors and outcomes observed across the group as a whole would, on average, mirror those an individual would experience over time.

Ergodicity requires two things. First, a variable must be stationary; that is, the processes' statistical properties (mean, variance, autocorrelation structure) remain constant over time. However, individual development and improvement due to interventions imply non-stationarity (Molenaar, 2004), so intervention science focused on individuals is rarely interested in stationary variables. The second aspect of ergodicity requires that the same dynamic model applies to all individual elements. For example, it assumes that if there is a link between positive thinking and positive affect at the group level, positive thinking has the same positive effect on every individual in the group (Molenaar, 2004). Without these two properties, it is unknown to what extent group-level findings apply to individuals over time.

The violation of ergodicity is not a trivial matter. Research suggests that processes generally beneficial at the group level may be inert or even harmful to some individuals (Ferrari et al., 2022; Sahdra et al., 2023, 2024). For example, Sahdra et al. (2024) examined intensive longitudinal data and found that while valued action was associated with higher hedonic well-being (e.g., lower sadness, higher joy) at the group level, there was a subset of people labeled stoics, for whom it was not associated with higher hedonic well-being and indeed was associated with higher stress. In another intensive longitudinal study, Sahdra et al. (2023) found that compassion was associated at the group level with higher well-being. However, at the individual level, it was not associated with well-being if the person experienced conflict between self and other compassion.

In the present paper, we propose to examine how pervasive this issue is across several process measures. In three independent intensive daily diary studies, we will examine the ergodic assumption in the relationship between effects identified at the group and individual level. We also consider a variety of processes and positive and negative outcomes across these three studies to see how general these issues may be.

Identifying Key Processes of Change

What psychological processes significantly impact a particular individual's well-being? There are several ways the field has sought to address this question. Cross-sectional

studies, for instance, analyze data from one point in time and are used to understand the prevalence of health outcomes and determinants of health, and describe features of a population (Wang & Cheng, 2020). Cross-sectional analysis is inherently between-person and thus may not allow one to make inferences about within-person relationships or mechanisms of change (Robinson, 2009). For example, research shows that goal tenacity has a positive between-person link to student well-being (Sahdra et al., 2022), and thus suggests tenacity be promoted in student interventions. However, that may not hold true at the individual level. For example, providing an intervention that makes tenacious students even more tenacious may result in less well-being, even if tenacity is positive at the group level.

Longitudinal research involves the comparison of data collected from the same individuals across multiple time points to identify possible changes in outcomes due to interventions or natural development (van Weel, 2005). Longitudinal research is an improvement on cross-sectional research, especially as measurement frequency increases, and it allows one to examine within-person changes empirically (Donald et al., 2022; Hamaker et al., 2015). For example, longitudinal research shows that people with high self-esteem are more likely than others to improve their levels of social support (Marshall et al., 2014). Although this kind of research is more individually relevant in principle, the longitudinal link of a process predicting a changing outcome is commonly based on a group level or *fixed effect* (average effect across all individuals). Variation within individuals regarding the process-outcome relationship is frequently represented by *random slopes* and is considered error (Brockman et al., 2023). Further, even if within-person effects are examined, for example, by using multilevel models, these effects are estimated as individual deviations (in intercept and slope) from aggregated estimates (Fisher et al., 2018). This can yield parameter estimates that are biased if there are widely varying patterns of individual effects (Wright & Woods, 2020).

In addition, nomothetic modeling approaches, such as multilevel modeling of longitudinal data, tend to shrink individual-level estimates towards the group-level effect. In an experience sampling study, Sahdra et al. (2023) found that raw within-person associations between self-compassion and other-compassion were heterogeneous such that the two forms of compassion were linked positively for some individuals, negatively for others and were unrelated for yet others. A multilevel model linking the two forms of compassion showed a *fixed effect* that had a positive sign and model-implied individual estimates were all positive, suggesting that the nomothetic method was 'driving' individual-level estimates towards the group-level effect. Similarly, Sahdra et al. (2024) found high heterogeneity in

raw within-person associations of valued action and affect in daily life, but the multilevel model dramatically shrunk individual trajectories towards the nomothetic effect. While such shrinkage is not an issue if the goal is solely to make population-level inferences of the group-level effect, it becomes highly problematic when applying group means to predict the effects for specific individuals. Simply stated, group means fail to apply to many individuals.

Mediational analysis is a third, group-based approach focused on identifying the functionally important pathway of change in an intervention (Rijnhart et al., 2021). The typical mediational analysis estimates the intervention effect on the average process changes within a group (e.g., the intervention group improves in self-efficacy and the control group does not) and then estimates the extent to which that average process change predicts improvement in the average group well-being (e.g., reduces mean depression scores; for a systematic review of recent studies see Hayes et al., 2022a). Research across the three distinct literatures uniformly acknowledges individual differences in effects. Yet, these variations are usually treated as “error”. “Error,” in statistical language, indicates the gap between observed and model-predicted outcomes, capturing variability not explained by the model’s variables, which are usually group-level. Importantly, categorizing this variability as “error” does not imply it is random or beyond explanation.

Group-level findings can guide practitioners and scientists toward generally useful processes in the population (Hayes et al., 2022a). However, there is skepticism about the sufficiency of group averages in modeling individual processes (see, e.g. Hayes et al., 2019). An alternative, *idionomic* view considers each person as a system of interacting, dynamic processes shaping individual life trajectories (Fisher, 2015; Fisher et al., 2018; Molenaar, 2004, 2013). This approach argues that generalizations about populations, termed *nomothetic* conclusions, should result from individual system analyses rather than predetermine these analyses. It diverges from the traditional approach, which often generalizes from groups to individuals. By focusing on detailed studies of individuals, the idionomic method inverts this conventional hierarchy, highlighting the critical role of individual-level analysis in underpinning broader generalizations.

There has been a sharp recent increase in idionomic approaches to well-being. These include studies of within-person variation in process networks (Fisher et al., 2017; Rabinowitz & Fisher, 2020; Sanford et al., 2022; Wright et al., 2019), person-environment interactions (Hopwood et al., 2022), and within-person factor structures (Strohacker et al., 2021). However, despite this increase in idionomic research, the vast majority of psychological research on mental health and well-being still relies on top-down normative research,

implicitly or explicitly assuming that what is statistically good for the collective is also good for the individuals in the collective. The viability of that implicit, group-level assumption is being examined in the present study.

Within-Person Change: Mapping Processes to Outcomes

How can we identify which processes are the most important to an individual’s well-being? As a place to begin in this paper, we will focus on one simple relationship: the degree to which within-person changes in the process are associated with within-person changes in an outcome. Processes may relate in complex ways to an outcome, such as via an interaction with other variables. Complex statistical methods exist for clinically modeling networks of that kind (Beltz & Gates, 2017; Ong et al., 2022), but for the sake of this paper, we will focus only on modeling simple contemporaneous relationships between processes and outcomes. As will be seen, even that focus is not so simple. Our reason to begin an analysis of the model consistency feature of the ergodic assumption with simple within-person relationships is that this analysis requires much less power and sample size to analyze than statistics used to estimate more complex relationships, such as structural equation modeling (Donald et al., 2019), vector auto-regressive models (Bulteel et al., 2016), and network analysis (Beltz & Gates, 2017).

If we identify the processes most strongly associated with a specific outcome for a particular individual, this knowledge could be invaluable for both the client and the therapist. It could spotlight important processes to target in therapy, guiding the therapeutic focus. The key question in a process-based approach is *what treatment will most effectively target the key biopsychosocial processes of change for a specific person, given their current context, life history, and treatment goal* (Hayes et al., 2019). Within-person analyses can begin to identify those processes, person by person.

Current Study

The present study used three different process measures: the Process-Based Assessment Tool (PBAT; Ciarrochi et al., 2022), the Psy-Flex; (Gloster et al., 2021), and the Functional Analytic Assessment Template-Mobile (FIAT-M; Darrow et al., 2014). These measures all seek to identify processes that drive well-being, but are quite different in their focus, thus providing us with a broad sample of constructs. The PBAT focuses on concrete behavior (e.g., “I did something to hurt my relationship”), whereas the Psy-Flex uses more abstract and expansive language for processes (e.g., “I engage in

things that are important to me”). Both measures are primarily focused on the individual. In contrast, the FIAT is focused on social processes, such as asserting oneself and disclosing one’s feelings.

Each process measure is grounded in a theory that identifies the underlying causes of well-being and suffering. The PBAT seeks to measure adaptive and maladaptive forms of context-sensitive variation, selection, and retention across all six psychological dimensions and the bio-physiological and sociocultural levels of the Extended Evolutionary Meta Model (Ciarrochi et al., 2022). Selection items focus on the extent to which people engaged in value-consistent behavior in the areas of cognition, affect, attention, self, motivation, and overt behavior. Variation items focus on the extent people could change their behavior to be more value-consistent, and retention items focus on the extent people can persist in value-consistent behavior. The biophysiological level is assessed by two items related to health behaviors, and the sociocultural level by items assessing relationship behavior. Research has shown that the PBAT links in expected ways to clinically relevant outcomes and to need satisfaction; it also shows discriminant validity for positive and negative processes (Ciarrochi et al., 2022). For example, people can both hurt and help their relationships on the same day, indeed sometimes in the same five minutes.

We leveraged an intensive longitudinal dataset to build upon Sanford et al.’s (2022) work, which utilized network analysis to investigate complex, multivariate relationships among various PBAT processes and outcomes. Their findings revealed significant inter-individual differences in process-outcome networks. By employing multilevel modeling, they identified considerable within-person variability in these relationships, adhering to a nomothetic approach. The approach of this paper is idiomonic, concentrating on individual-level bivariate relationships between processes and outcomes, employing time series analysis, and using meta-analytic methods to evaluate the extent of heterogeneity.

The second process-based measure examined in this paper is the Psy-Flex, which focuses on key behaviors linked to psychological flexibility processes (Gloster et al., 2021). Its six individual items relate to attention (being present), affect (acceptance), cognition (non-reactivity to thoughts), self (“having a steady core inside me”), motivation (values awareness), and overt behavior (being engaged). Psychometric research indicates that the Psy-Flex exhibits a single-factor structure, reflecting overall psychological flexibility. It correlates as expected with well-being, distinguishes between clinical and non-clinical samples, and is responsive to clinical change (Benoy et al., 2019; Gloster et al., 2021).

The third process measure, the Functional Analytic Assessment Template-Mobile (FIAT-M; Stanton et al., in preparation), explicitly focuses on interpersonal behaviors

common to social repertoires, conceptualized as five non-orthogonal domains: Assertiveness, Bidirectional Communication, Conflict Resolution, Disclosures, and Emotional Expression. The FIAT-M is based conceptually on the original FIAT (Callaghan, 2006).

We utilized archival data from three different samples, each focusing on different clinically relevant measures of process, and positive and negative functioning. All three studies received full ethics review and approval. None of the data sets have been examined with idiographic time series analysis and meta-analytic estimates of heterogeneity. We aimed to uncover both group and individual-level connections between three distinct process measures and well-being outcomes, with a primary objective of determining the degree of heterogeneity in these links. Lower heterogeneity suggests that group averages more accurately reflect individual cases, while higher heterogeneity indicates a greater need for individual-focused analysis.

The Process-based Assessment Tool (PBAT)

Participants were recruited using Amazon’s Mechanical Turk (“mTurk”) service, both to maximize the potential pool of eligible participants and to secure a diverse sample in terms of age, gender, and nationality (Hauser & Schwarz, 2016). A total of 57 participants were recruited and completed at least one assessment. Participants who completed data collection (criteria are described below) ranged in age from 19 to 71 (average age = 38.5) and lived predominantly in the United States ($n=42$). Those living internationally were in Brazil ($n=8$), India ($n=4$), Italy ($n=2$), and Canada ($n=1$). Of the 57 original participants, seven were lost because of attrition, having missed over ten assessment periods in the first 35 days. These participants averaged 17.4 assessments out of the target of 60 and were not considered in any further analysis. Six of the 50 completers exhibited no variability on one or more assessment items or did not complete the measures used in the present study and were excluded. The analyzed sample was 44 (15 self-identified females, 24 self-identified males; 5 no answer for gender), with a mean age of 33.8 ($SD=13.03$).

Data was collected twice-daily across 35 days. To reward engagement in the study, a completion bonus was given to individuals who responded to at least 60 of the bi-daily assessment prompts. In total, participants were paid 5 dollars a day for their time and effort, including a completion bonus. An experience sampling app notified users via push notifications when to complete data. All items were completed using a 0–100 visual analog “finger swipe” scales to discourage anchoring.

The Process-Based Assessment Tool (PBAT; Ciarrochi et al., 2022) comprises 18 items focused on variation, selection,

and retention processes. The 14 selection items cover the domains of affect, cognitive processes, attention, social connection, motivation/autonomy, overt behavior/competence, and physical health, with one positive and one negatively valenced item for each. Two items assess the range of variation in behavior and two items assess behavioral retention across time; these item pairs also had one positively and one negatively valenced item. The stem for each item was “Over the past 12 hours” and the anchors were 0 = Strongly Disagree and 100 = Strongly Agree. Sample items include, “My thinking got in the way of things that are important to me” and “I felt stuck and unable to change by ineffective behavior.” The PBAT has been shown to link in theoretically expected ways to clinically relevant outcomes and to need satisfaction (Ciarrochi et al., 2022).

Concerning the outcomes, we assessed negative functioning using the Screening Tool for Psychological Distress (Stop-D; Young et al., 2007, 2015). This five-item scale asks “How much have you been bothered by”: Sadness - “Feeling sad, down, or uninterested in life?” Anxiety - “Feeling anxious or nervous?” Stress - “Feeling stressed?” Anger - “Feeling angry?” Perceived lack of social support - “Not having the social support you need?” (alpha = .90). To assess positive functioning, we utilized a single-Item Life Satisfaction Measure (Cheung & Lucas, 2014). The single item “In general, how satisfied are you with your life?” has good criterion validity because it produces similar observed correlations with a well-validated life satisfaction scale on self-reported happiness, physical health, and mental health.

The Functional Idiographic Assessment Template-Mobile (FIAT-M)

Data collected for the FIAT-M comes from a twice-daily diary study of social behaviors, loneliness, and mental health, which sought to evaluate the FIAT-M as a predictor of loneliness and other emotional health-related outcomes. Participants were non-treatment-seeking adults in the U.S. recruited from an American Mountain West university campus, its surrounding metropolitan area, and from the online survey panel service Prolific. Participant recruitment was equally split between college students and non-college attending working adults, between male and female and were majority non-white (White or European ancestry = 46.2%). Ages ranged from 18 to 55 years old ($M = 27.13$; $SD = 9.6$). Thirty-nine individuals comprise the total sample. Two participants showed no variability on measures and were excluded from further analysis, leaving 37 (18 male, 19 female) with a mean age of 26.54 ($SD = 9.4$).

Individuals in this sample completed twice-daily diary surveys for a minimum of 30 days and completed items related to social functioning. These items included the

FIAT-M described above, two items related to social support (alpha = 0.83; “I was supported by people in my life”), as well as a modified UCLA 3 Item Loneliness Scale (alpha = 0.85; Hughes et al., 2004; “I felt left out”, “I felt isolated from the world around me”, “I felt that I lacked a close relationship”).

Adapted from the Functional Idiographic Assessment Template (FIAT; Callaghan, 2006; Darrow et al., 2014), the FIAT-M measures interpersonal behaviours at a measurement interval suited for daily diary or event sampling research. The ten items on the FIAT-M are split into two categories of five items each, one category for discriminating opportunities for interpersonal interaction (SD) and one for acting on them (Bx). All items use a 0-100 scale to ensure sufficient variance. In a twice-a-day diary study attempting to validate the FIAT-M in a non-clinical sample, results showed that SD items were good predictors of Bx items, showing that these items functioned in the intended logical sequence for participants (Stanton et al., 2023).

Previous research using the FIAT questionnaire has found that while its items correlate with other constructs (i.e. quality of life, fear of negative evaluation, assertiveness, etc.) in expected directions, the underlying factor structure was more complex than initially considered. The authors speculated that a traditional psychometric framework might not be the ideal arena for the constructs that the FIAT measures (Darrow et al., 2014). Thus, Study 2 explored the FIAT categories through an *Experience Sampling Method (ESM)* study.”

The Psy-Flex

Participants were transdiagnostic patients who were a part of the Choose Change effectiveness trial for outpatients and inpatients chronically suffering from a range of mental disorders and psychological problems (Gloster et al., 2023). Following intake and informed consent procedures, patients completed a baseline assessment comprising a diagnostic interview and standardized questionnaires. Patients then engaged in a one-week ESM study using a smartphone and answered questions regarding their mood, cognitions, and behaviors. The ESM sampled six times daily for a total of 42 time points during the ESM week. For further details on the methodology, see Villanueva et al. (2019). There were 200 patients in total but not all participants completed all measures for this study. Psy-flex and positive and negative affect measures were available from 141 patients (66 males; 75 females) with age range from 18 to 64 years ($M = 35.86$, $SD = 11.40$).

We used the Psy-Flex to measure all six components of psychological flexibility, including indices of being present, being open to experience, leaving thoughts be/defusion,

having a steady self, having an awareness of values, and being engaged in life (Gloster et al., 2021). People respond on a five-point scale ranging from very often (5) to very seldom (1). Sample items include “I engage thoroughly in things that are important, useful, or meaningful to me” and “If need be, I can let unpleasant thoughts and experiences happen without having to get rid of them”. The items have been shown to reflect a higher-order psychological flexibility factor, to relate in expected ways to other measures of psychological flexibility and symptomatology, and to differentiate clinical and non-clinical samples (Gloster et al., 2021). To measure outcome, participants reported how they felt since the last scheduled prompt, in terms of negative affect (“how unhappy, without energy, distracted and distressed”; $\alpha = 0.88$) and positive affect (how optimistic, delighted, satisfied and grateful”: $\alpha = 0.87$). Ratings were made on a 100 point scale (0; not at all; 100; very much).

The i-ARIMAX Analytic Procedure

Our goal was to (1) identify the extent that within-person changes in clinically-relevant processes related to within-person changes in well-being, and (2) identify the extent to which the relationship varied from person to person. Idionomic analysis begins by focusing on individual-level relationships rather than on relationships based on the group average, and only makes group-level conclusions if they are consistent with the individual-level findings (Hayes et al., 2022a). This type of analysis does not assume that populations are homogenous and that each person in the population shares the same model structure and parameters. Rather, in idionomic analysis, model parameters, and structure can be specific to the individual (Molenaar, 2013).

Our analysis sought to establish the strength of relationship between each process and each outcome, within each individual. For example, we estimated the strength of within-person relationships and standard errors of that estimate for each of the six Psy-Flex processes for every person in the sample. These relationships then became the input for meta-analyses, with each person being treated as a separate “study”, allowing us to evaluate both the pooled effect across people and the variability in the effect.

Traditionally, one can estimate the strength of the relationship between processes and outcomes utilizing correlational or regression analysis. However, our time series data were expected to violate the assumptions of these traditional analyses in at least two ways. First, time series are often not stationary, as when the mean of the outcome changes. Second, the observations are often not independent, as earlier values often relate to later values (Chatfield & Xing, 2019). In scenarios of autocorrelation, the estimates obtained

through ordinary least squares (OLS) lose their efficiency, meaning they are not as precise as they could be. This lack of precision can lead to underestimated errors and exaggerated significance levels (t-scores), thereby undermining the trustworthiness of hypothesis tests and the accuracy of confidence intervals (Brockwell & Davis, 2013). Furthermore, neglecting to account for trends might distort the true nature of the relationship between variables (Bottomley et al., 2019).

To deal with these issues, we used an idionomic version of ARIMA (Autoregressive Integrated Moving Average; Chatfield & Xing, 2019). The AR (autoregression) component predicts values based on their past values, the I (integrated) component uses differencing to eliminate trends, and the MA (moving average) component captures the relationship between an observation and the residual errors from a moving average model of past observations. The AR and MA components necessitate stationarity in the dataset, implying that the time series’ mean, variance, and autocorrelation must remain constant over time (Ho & Xie, 1998; Jensen, 1990). While the differencing process (the I component) addresses mean stability by eliminating trends, we still must assume stability in variance and autocorrelation. This assumption mirrors that of other statistical methods, such as regression and multilevel modeling, which also presuppose constant relationships and variances across observations (Snijders et al., 1999).

I-Arimax is an extension of ARIMA. The ‘i’ in i-ARIMAX signifies individual-level analysis, while the ‘X’ represents the inclusion of an exogenous variable. I-ARIMAX enabled us to address mean trends and autocorrelation in time series data, thereby estimating the relationship strength between processes and outcomes and crafting customized models for each participant.

In ARIMAX models, the interplay of the parameters p , d , and q is crucial for enhancing forecast accuracy. The p parameter, focusing on autoregressive terms, emphasizes the importance of stability by leveraging past observations to predict future outcomes, suggesting that patterns or trends from the past are likely to persist. The

d parameter, which involves differencing, addresses the need to eliminate trends, thereby stabilizing the series over time and ensuring that predictions are based on momentary fluctuations rather than long-term trends. Finally, q , the moving average component, is key for integrating the effect of unexpected changes into the forecasting equation. This integration happens by adjusting forecasts based on the magnitude of past errors, specifically when these errors—stemming from unanticipated changes—demonstrate predictive value for future observations (Chatfield & Xing, 2019).

A simple way to think of ARIMA is as a filter that seeks to isolate meaningful patterns from the background noise in the temporal data (Nau, 2020). ARIMAX models add an exogenous variable (x), or variable that only predicts but is not predicted. The beta between x (process or exogenous variable) and Y (well-being or outcome) can be thought of as the strength of the relationship after controlling for the influence of trend, autoregressive effects, and moving average.

Manually fitting an ARIMA model and estimating the values for p , d , and q can be subjective and reliant on the skill of the analyst (Al-Qazzaz & Yousif, 2022). To solve this issue, the auto-Arima function in R seeks to automate the process of identifying the best ARIMA model by evaluating models with varying p , d , and q values and selecting the best fitting model (Hyndman & Khandakar, 2008). The function begins by conducting a Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test to determine if a time series is stationary or non-stationary (Kwiatkowski et al., 1992). If the time series is non-stationary, auto-arima will automatically apply a difference transformation to make the time series stationary. Next, auto-arima fits several models with different combinations of autoregressive (AR) and moving average (MA) terms. It chooses the model with the lowest corrected Akaike Information Criterion (AICc), the model that explains the greatest amount of variation using the fewest possible variables. The auto-Arima function allows the specification of an exogenous variable.

In the present paper, we developed an algorithm that applied auto-arima within each person, to estimate the link between every process and outcome pairing. These estimates then became the data for meta-analyses using the R package “*metafor*” (Viechtbauer, 2010). Each person’s estimate was treated like a study effect size with an estimate

of error. This allowed us to estimate pooled effects across participants, to estimate heterogeneity, and to present forest plots to illustrate that heterogeneity.

Results

Preliminary Analyses

We conducted i-ARIMAX analysis for every process-outcome pairing across all three datasets. Data were centered and scaled at the within-person level, focusing on within-person relationships. This approach improves interpretability by clearly distinguishing between-group and within-group variation (Paccagnella, 2006). The auto-arima component of i-ARIMAX identified a substantial variety of time series models for participants. Table 1 illustrates this variation in the numbers of parameters estimated for p (autoregressive components or stability), d (differencing components or trend), and q (moving average components or unexpected change). Not only did individuals differ in their ideal time series model as reflected by the relatively substantial percentages of participants requiring adjustments in these statistical variables, but samples and measures also differed. For example, for the PBAT sample, 50% of people experienced some change (reflected in differencing) in their negative affect time series, whereas 11% of Psy-Flex participants had differencing components added to their time series data to remove trends. This suggests that idiomonic statistical analysis can reveal differences in measures that may apply to their use as process variables when contextual sensitivity is key. Generally, a substantial minority of participants (between 8 and 19% depending on the variable) required one or more autoregressive components,

Table 1 Prevalence of autoregressive (P), Differencing (D), and Moving average (Q) Components in Auto-ARIMA analysis across individuals and three datasets

	Psy-Flex		PBAT		FIAT-M	
	Neg Af	Pos aff	Neg Af	Pos aff	Lonely	Supp
P (autoregression)						
No Autoregression	72%	66%	75%	80%	78%	76%
One Component	13%	19%	14%	5%	14%	14%
More than one Component	15%	15%	11%	16%	8%	11%
D (Differencing)						
No Autoregression	89%	88%	50%	68%	59%	68%
One Component	11%	12%	50%	12%	41%	32%
Over one Component	0%	0%	0%	0%	0%	0%
Q (Moving Average)						
No Autoregression	84%	85%	43%	57%	51%	43%
One Component	14%	11%	39%	23%	46%	27%
Over one Component	2%	4%	18%	20%	3%	30%

Note: The table presents a summary of auto-ARIMA analysis, detailing the frequency with which each component—correlation with past values (P), need for differencing to achieve stationarity (D), and impact of past errors on current predictions (Q)—appeared across models fitted for individuals in each dataset, categorized as none, one component, or multiple components

suggesting individuals differed in the stability of their outcomes (e.g., of mood). The moving average statistics (bottom Table 1) suggest that people differed in the extent they experienced unexpected changes in their outcome that were predictive of future changes (from 2 to 39% depending on the variable). The most common pattern in ARIMA models was “000”, which occurred 39% of the time for the PBAT, 51% of the time for the Psy-Flex, and 33% for the FIAT-M. This shows that a single statistical model would not have been adequate to describe all participants.

We argued in the analyses section that ordinary regression assumptions are often violated with time series when idiomatic analyses are applied to longitudinal data. However, although we anticipated differences in the regression and I-ARMAX coefficients, we also expected them to be closely related, as both methods use the same data to estimate the relationship’s strength. As a preliminary check to see if the I-ARIMAX approach was coherent with a simple regression approach, we conducted both analyses for each person across all processes and outcomes, comparing their results. Regression analysis was performed with the ARIMAX model by setting the p , d , and q parameters to 0. I-ARIMAX and regression yielded beta coefficients and standard errors for each individual within the samples. For each of the three samples, we then calculated the average and standard error of each coefficient. As can be seen in Table 2, the coefficients between regression and I-ARIMAX coefficients were high, having between 76 and 86% of variance in common. The average magnitude of the coefficients was also similar, being slightly smaller for i-ARIMAX. The level of error was smaller for i-ARIMAX compared to regression.

Main Analysis

In our next step, we utilized the *r* package, metafor, to conduct a meta-analytic examination of the within-person coefficients. This approach allows us to estimate the average effects and the heterogeneity of these effects across individuals using well-established meta-analytic tools. Table 3 presents the results for the Psy-Flex items. The pooled effects

suggest each process measured by the Psy-Flex generally has a moderate to strong link with negative affect and positive affect. I^2 represents the percentage of total variability across studies that is due to true heterogeneity rather than chance in a traditional meta-analysis (Higgins et al., 2003; Huedo-Medina et al., 2006). Rough guidelines for interpreting I^2 in the meta-analytic literature are that values less than 25% reflect low inconsistency, 25–50% reflect moderate inconsistency, 50 to 75% reflect high inconsistency, and over 75% show very high inconsistency (Higgins et al., 2003). While there are no absolute cutoffs, in the Cochrane library of meta-analyses, for example, the median I^2 is 21% (Ioannidis et al., 2007). If I^2 exceeds even 50%, it is common to search for subgroups or to avoid reporting pooled effects (Lo et al., 2019). In the present context “inconsistency” reflected the extent to which the strength of process-outcome links varied between people. To assess the significance of the value, the Q^2 statistic was employed. This metric calculates the sum of squared deviations between individual studies and the overall mean, normalized by the degrees of freedom, and serves to evaluate the statistical significance of heterogeneity (Huedo-Medina et al., 2006).

Both I^2 and Q^2 are important for determining if variation across studies (or in this case participants) can be attributed to heterogeneity beyond chance. However, these statistics have their limitations. If samples are small (e.g., $N < 7$), I^2 can be biased, reflecting an overestimation or underestimation of true heterogeneity (von Hippel, 2015). However, our samples included at least 37 people, minimizing bias. Another limitation of these statistics is that they are not an absolute measure of heterogeneity. To deal with this issue, we followed Borenstein’s et al. (2021) recommendation and reported the range of effects.

As can be seen in Table 3, most of the I^2 values in the present data sets are above 0.75, showing very high inconsistency. All Q^2 values are highly significant ($p < .0001$). Analogously to meta-analytic reporting, such a high level of heterogeneity suggests that the effects seen across different people are not easily comparable and thus that pooled reporting (as would de facto be the case when using classical

Table 2 Comparative metrics and correlation of beta coefficients from I-ARIMAX and standard regression across three samples and measures

	Psy-Flex		PBAT		FIAT	
	Neg Af	Pos aff	Neg Af	Pos aff	Lonely	Supp
Relationship strength: process & outcome						
Average Beta (Regression)	0.394	0.461	0.197	0.210	0.043	0.249
Average Beta (I-ARIMAX)	0.376	0.426	0.190	0.200	0.042	0.212
Beta correlation: Regression and I-ARIMAX	0.932	0.911	0.902	0.912	0.894	0.874
Standard error of beta						
Avg. SE (Regression)	0.183	0.173	0.120	0.120	0.121	0.117
Avg. SE (I-ARIMAX)	0.171	0.156	0.108	0.112	0.113	0.110

Note: The relationships are represented by absolute values. Averages are computed across participants in each sample, with each individual possessing a distinct beta coefficient for both regression and I-ARIMAX models. PBAT: Process-Based Assessment Tool; FIAT: Functional Idiographic Assessment Template

Table 3 Average (pooled) within-person relationships between each Psy-Flex process and outcomes, level of heterogeneity (Heter) of that relationship, and percentage of people showing different magnitudes of the relationship (beta)

Flex process	Pooled		Heter		Percentage of people within beta band						
	Beta	SE	I ²	Q ²	< -0.31	-0.30- -0.21	-0.20 -0.11	-0.10- 0.10	0.11- 0.20	0.21- 0.30	> 0.31
Link between process and negative affect											
FocusImpMoments	-0.39*	0.02	80	649	61%	11%	9%	14%	2%	1%	1%
AllowFeelings	-0.37*	0.03	80	653	56%	13%	12%	12%	4%	0%	3%
SteadySelf	-0.43*	0.02	75	646	67%	15%	6%	6%	1%	1%	2%
ChoseValue	-0.40*	0.02	71	461	61%	15%	8%	13%	3%	1%	0%
CommitAction	-0.38*	0.02	71	450	62%	11%	10%	13%	2%	1%	1%
ObsThoughtsDistance	-0.41*	0.03	83	949	65%	14%	4%	9%	5%	1%	2%
Link between process and positive affect											
FocusImpMoments	0.43*	0.02	89	2115	0%	1%	1%	14%	9%	11%	64%
AllowFeelings	0.40*	0.02	91	2361	1%	1%	3%	11%	12%	11%	62%
SteadySelf	0.47*	0.02	93	3517	1%	1%	1%	9%	8%	8%	73%
ChoseValue	0.45*	0.02	90	2184	1%	1%	1%	9%	11%	9%	69%
CommitAction	0.42*	0.02	87	2128	1%	0%	2%	10%	8%	16%	64%
ObsThoughtsDistance	0.48*	0.02	92	2546	1%	0%	1%	9%	7%	13%	69%

Note: * $p < .05$. All Q^2 tests of heterogeneity are highly significant, $p < .0001$

statistical methods) may not be appropriate. The right side of Table 3 presents the percentage and range of people with different magnitudes of beta.

We next examined the FIAT-M processes as they link to the outcomes of *loneliness* and *feeling supported*. Table 4 presents these results. Concerning *loneliness*, only one pooled effect was significant. *Experiencing interpersonal conflict* was generally linked to higher *loneliness*. However, it would be incorrect to conclude from this pooled effect that there were no other significant links to *loneliness*. The I^2 indicated high to very high heterogeneity in effects, suggesting that the “non-significant 0” effect simply does not describe all people well. For example, *expressing feelings* was associated with lower *loneliness* for about 14% of people ($\beta < -0.31$) but associated with higher *loneliness* for about 11% of people ($\beta > 0.31$).

In contrast to *loneliness*, pooled effects for predicting “*feeling supported*” tended to be significant. However, again these effects were highly heterogeneous. For example, *having the opportunity to express feelings* was strongly associated with *feeling supported* for 38% of people ($B > 0.31$), but was either not linked to *feeling supported* or negatively linked to *feeling supported* for 17% of people ($B < -0.11$).

Our final analysis focused on the PBAT. The results are presented in Table 5 (negative affect outcomes) and Table 6 (Positive Affect outcomes). Almost all processes showed a significant average effect with the outcomes in the expected direction, but once again all within-person effects were highly heterogeneous. Perhaps the strongest illustration of heterogeneity comes from three cases where there was no significant pooled effect: *Sticking to strategies* (negative affect only), *no outlet for feelings*, and *thinking got in the way* (positive affect only). For each process, the “average

effect of 0” poorly describes many people. For example, the process “sticking to strategies that have worked” was associated with less negative affect for 14% of people ($\beta < -0.31$) but more negative affect for 7% of the people ($\beta > 0.31$). Problematic thinking patterns were associated with lower life satisfaction for 23% of people ($\beta < -0.31$) but tended to have little effect or potentially a positive effect for 21% of people ($\beta > 0.11$).

To help provide an intuition about the significant heterogeneity of effects, Fig. 1 provides forest plots of one pair of process-outcome relationships across individuals for each of the three measures. For the FIAT-M, *asserting oneself* had a significant positive association with *loneliness* for seven people (confidence intervals don’t overlap with 0) and a negative association for four to five people. For the Psy-Flex, almost half of people showed a significant positive relationship between *allowing feelings* and *positive affect*, but the strength of that relationship varied substantially. A subset of people showed no association, and one person showed a significant negative link. Finally, for the PBAT, *problems with thinking* were significantly negatively associated with *positive affect* for 15 people, and significantly positively associated for 5 people.

The previous analysis reveals significant individual variations in the relationship between process and outcome across all process variables in the three datasets, demonstrating that the group average does not accurately represent many individuals. In contrast, the ergodic assumption posits that the group average reflects the experience of every group member. Thus, this aspect of the ergodic assumption was not supported in any of the analyses. Even before we face the stationarity requirements of ergodicity, these findings show why we need to look within individuals over time.

Table 4 Average (pooled) within-person relationships between each FIAT processes and outcomes, level of heterogeneity (Heter) of that relationship, and percentage of people showing different magnitudes of the relationship (beta)

Process	Pooled		Heter		Percentage of people within beta band						
	Beta	SE	I ²	Q ²	< -0.31	-0.30- -0.21	-0.20 -0.11	-0.10- 0.10	0.11- 0.20	0.21- 0.30	> 0.31
Opportunities for interpersonal action					Loneliness						
Assertive	0.03	0.04	73	139	5%	11%	5%	41%	22%	8%	8%
GiveRecFeedback	0.03	0.03	70	118	8%	8%	5%	46%	14%	14%	5%
InterperConflict	0.18*	0.04	81	190	0%	3%	14%	22%	14%	19%	30%
ChanceToBeClose	-0.02	0.04	77	155	16%	8%	8%	38%	14%	11%	5%
ExpressFeelings	0.04	0.04	75	142	11%	5%	11%	38%	11%	11%	14%
Behavioural processes											
AssertedNeeds	-0.04	0.04	80	200	14%	5%	11%	38%	24%	5%	3%
GaveFeedback	-0.04	0.04	73	133	14%	3%	16%	43%	14%	8%	3%
ResolvedConflict	0.02	0.03	62	95	0%	14%	11%	46%	14%	14%	3%
Disclosed	0	0.04	76	153	14%	0%	14%	35%	19%	11%	8%
ExpressedFeelings	-0.03	0.04	81	191	14%	14%	11%	30%	14%	8%	11%
Opportunities for interpersonal action					Feeling supported						
Assertive	0.22*	0.04	78	222	0%	0%	5%	30%	16%	19%	30%
GiveRecFeedback	0.20*	0.04	79	212	3%	3%	5%	16%	24%	19%	30%
InterperConflict	-0.02	0.04	76	148	5%	14%	19%	38%	8%	8%	8%
ChanceToBeClose	0.41*	0.04	82	341	0%	0%	0%	5%	11%	22%	62%
ExpressFeelings	0.21*	0.05	87	357	5%	3%	8%	16%	14%	16%	38%
Behavioural processes											
AssertedNeeds	0.26*	0.04	81	205	0%	5%	5%	14%	14%	19%	43%
GaveFeedback	0.25*	0.04	78	187	0%	5%	3%	14%	16%	22%	41%
ResolvedConflict	0.10*	0.03	71	140	3%	0%	3%	59%	16%	5%	14%
Disclosed	0.26*	0.03	62	97	0%	0%	0%	24%	14%	22%	41%
ExpressedFeelings	0.31*	0.03	74	146	0%	0%	3%	24%	5%	16%	51%

Note: * $p < .05$. All Q^2 tests of heterogeneity are highly significant, $p < .0001$

Table 5 Average (pooled) within-person relationships between each PBAT processes and *negative affect*, level of heterogeneity (Heter) of that relationship, and percentage of people showing different magnitudes of the relationship (beta)

Process	Pooled		Heter		Percentage of Betas within each band						
	Beta	SE	I ²	Q ²	< -0.31	-0.30- -0.21	-0.20 -0.11	-0.10- 0.10	0.11- 0.20	0.21- 0.30	> 0.31
Selection/Values selecting behavior											
ConnectToPeople	-0.21*	0.04	80	243	32%	16%	14%	32%	5%	0%	2%
PaidAttToImport	-0.22*	0.04	83	331	32%	11%	23%	27%	7%	0%	0%
PersonalImpor	-0.21*	0.03	80	249	30%	14%	25%	27%	5%	0%	0%
ExperienceRangeEmo	-0.12*	0.04	81	224	20%	9%	16%	39%	11%	2%	2%
ThinkingHelpedLife	-0.23*	0.04	82	258	41%	9%	9%	39%	2%	0%	0%
ImportantChallenge	-0.18*	0.03	79	217	25%	20%	14%	32%	9%	0%	0%
HurtConnect	0.19*	0.04	83	283	2%	0%	5%	34%	20%	7%	32%
StruggledConMoment	0.24*	0.04	88	366	2%	7%	0%	25%	14%	16%	36%
Complying	0.19*	0.04	81	274	0%	7%	2%	30%	14%	20%	27%
NoOutletForFeelings	0.25*	0.04	83	331	0%	0%	2%	30%	16%	16%	36%
ThinkingGotInWay	0.12*	0.05	89	432	5%	7%	2%	39%	11%	9%	27%
NoMeaningfulChall	0.14*	0.04	81	251	0%	5%	7%	52%	11%	5%	20%
Variation											
AbleToChangeBehavi	-0.19*	0.04	82	289	25%	16%	23%	30%	2%	2%	2%
StuckUnableChange	0.30*	0.03	79	223	0%	0%	2%	20%	9%	20%	48%
Retention											
StuckToStrategies	-0.06	0.04	88	424	14%	14%	14%	41%	5%	7%	7%
StruggledToKeepDoin	0.25	0.04	82	291	0%	0%	5%	27%	16%	14%	39%

Note: * $p < .05$. All Q^2 tests of heterogeneity are highly significant, $p < .0001$

Table 6 Average (pooled) within-person relationships between each PBAT processes and *life satisfaction*, level of heterogeneity (Heter) of that relationship, and percentage of people showing different magnitudes of the relationship (beta)

Process	Pooled		Heter		Percentage of Betas within each band						
	Beta	SE	I ²	Q ²	< -0.31	-0.30- -0.21	-0.20 -0.11	-0.10- 0.10	0.11- 0.20	0.21- 0.30	> 0.31
Selection/Behavior Building value											
ConnectToPeople	0.25*	0.03	73	156	0%	0%	2%	30%	16%	18%	34%
PaidAttToImport	0.27*	0.04	82	241	0%	0%	5%	27%	16%	11%	41%
PersonalImpor	0.27*	0.03	78	225	0%	0%	0%	27%	16%	7%	50%
ExperienceRangeEmot	0.18*	0.03	81	207	0%	5%	7%	36%	9%	14%	30%
ThinkingHelpedLife	0.29*	0.04	81	247	0%	0%	0%	36%	9%	7%	48%
ImportantChallenge	0.26*	0.03	76	189	0%	2%	2%	23%	11%	27%	34%
HurtConnect	-0.15*	0.03	73	172	18%	25%	5%	45%	5%	2%	0%
StruggledConMoment	-0.18*	0.04	83	245	30%	11%	14%	34%	7%	0%	5%
Complying	-0.15*	0.03	73	172	25%	14%	11%	41%	5%	5%	0%
NoOutletForFeelings	-0.02	0.04	79	226	30%	18%	16%	25%	7%	2%	2%
ThinkingGotInWay	-0.09	0.05	89	385	23%	11%	9%	36%	11%	5%	5%
NoMeaningfulChall	-0.17*	0.03	75	193	20%	11%	20%	41%	7%	0%	0%
Variation											
AbleToChangeBehavior	0.24*	0.04	81	253	0%	5%	5%	18%	18%	16%	39%
StuckUnableChange	-0.28*	0.04	82	245	39%	20%	18%	16%	7%	0%	0%
Retention											
StuckToStrategies	0.12*	0.04	85	332	7%	5%	7%	23%	18%	16%	25%
StruggledToKeepDoin	-0.21*	0.04	85	319	36%	18%	5%	30%	7%	2%	2%

Consider the simple forest plots of the within-person relationship for four individuals between negatively worded PBAT process items and negative affect, as shown in Fig. 2. These four people were chosen because they demonstrated contrasting profiles. The bottom triangle represents the pooled effects across all items within that person and shows that, generally, higher scores on the negative PBAT items were associated with higher negative affect, as might be expected. The patterns within a person were quite different, however. The item “*hurting health*” was significant for persons 2 and 4, but not for persons 1 and 3. “*Thinking got in the way*” seems a prominent problem for person 1 but not person 4. *Complying* is associated with less negative affect for person 2 but more negative affect for person 4.

Figure 3 similarly presents the relationships involving the Psy-Flex items and positive affect for four people (we picked participants to highlight different patterns). Although the pooled effects are similar (bottom triangle), the within-person patterns differ. Person 1 experiences positive affect when they have a *stable sense of self* and can *observe thoughts at a distance*. *Committed action* appears to be relatively unimportant for this person. In contrast, *committed action* appears to be the most important process for person 2. On days they commit to action, they experience the highest well-being; on days they are less committed, they experience lower well-being. For person 3, *focusing on the moment* appears to be central to well-being, and for person 4, all processes except for *values* and *committed action* appear to be important.

Finally, Fig. 4 presents the results for the FIAT-M and loneliness for four participants. Unlike the PBAT and Psy-Flex, there is little consistency in the pooled effects. Person 2 and 4 generally have a negative link between social processes and loneliness, person 3 has no significant pooled link, and person 1 has a significant positive pooled link. For person 1, almost every social context and behavior is associated with higher loneliness, whereas for person 2, the effects are largely reversed. When Person 1 is *assertive*, they feel more *lonely*, when Person 2 is *assertive*, they feel less *lonely*. For person 4, *expressing and disclosing* is associated with less *loneliness* and *conflict* with more *loneliness*. Person 3 shows an interesting pattern in which the *opportunity to be assertive* is associated with less *loneliness* (SD) but *asserting oneself* is not associated with less *loneliness* (BX).

Discussion

Across all three data sets and three measures of positive and three measures of negative functioning, the model consistency aspect of the ergodic assumption was always severely violated. I² was never below 0.61 and was typically above 0.75, suggesting that the strength of relationships between process and outcome differed substantially between people. Bornestein et al.’s (2021) conclusion about heterogeneity in meta-analysis appears to apply in the present instance: “When there is a great deal of heterogeneity, pooling the studies may not be appropriate. In such cases, it may be

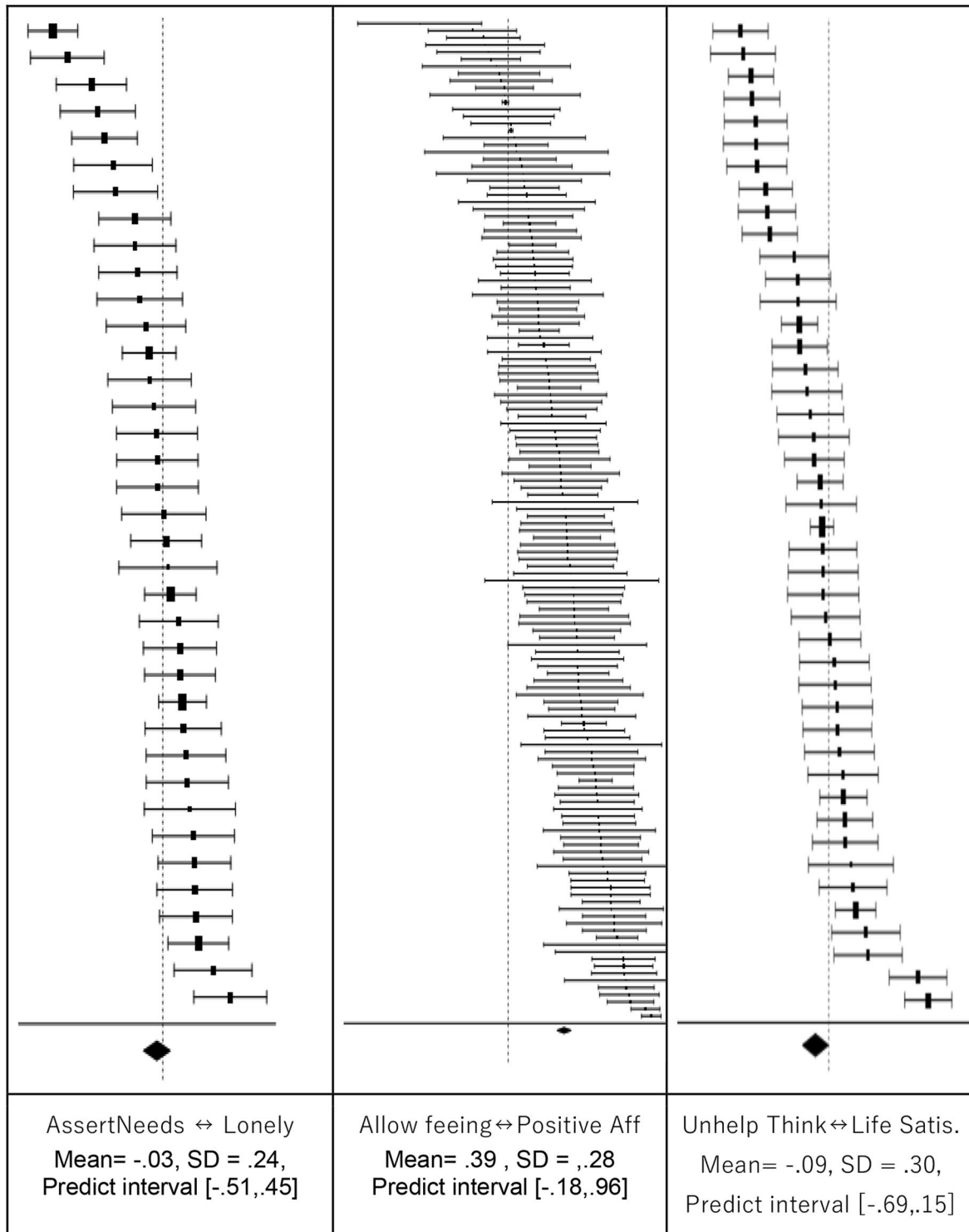


Fig. 1 Strength of process-outcome relationship for three behaviors: Asserting needs, Allowing/not controlling unpleasant feelings, and engaging in unhelpful thinking

Note: The middle line represents 0 relationship. Confidence intervals that don't overlap with this line to the left are negative relationships, and to the right, positive relationships

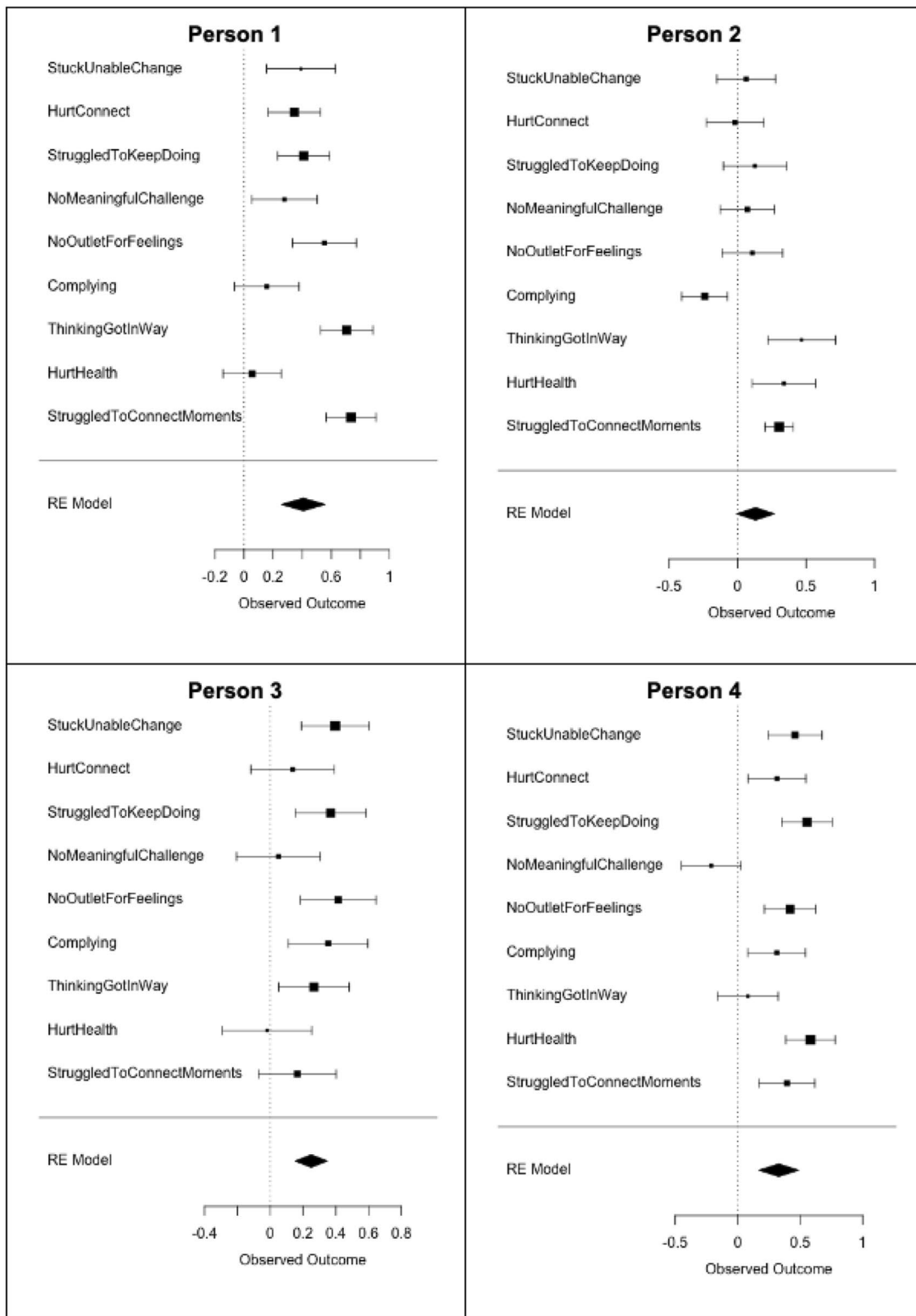


Fig. 2 Strength of the relationship between negative processes (Process-based assessment tool) and negative affect
 Note: *StuckUnableChange*: feeling stuck and unable to change ineffective behaviors; *HurtConnect*: actions that damaged connections with important people; *StruggledtoKeepDoing*: difficulty maintaining beneficial actions; *NoMeaningfulChallenge*: a lack of meaning-

ful self-challenges; *NoOutletForFeelings*: the absence of appropriate emotional outlets; *Complying*: actions taken solely to comply with others; *ThinkingGotInWay*: instances where thinking obstructed important activities; *HurtHealth*: behaviors detrimental to physical health; *StruggledToConnectMoments*: difficulties in engaging with daily moments

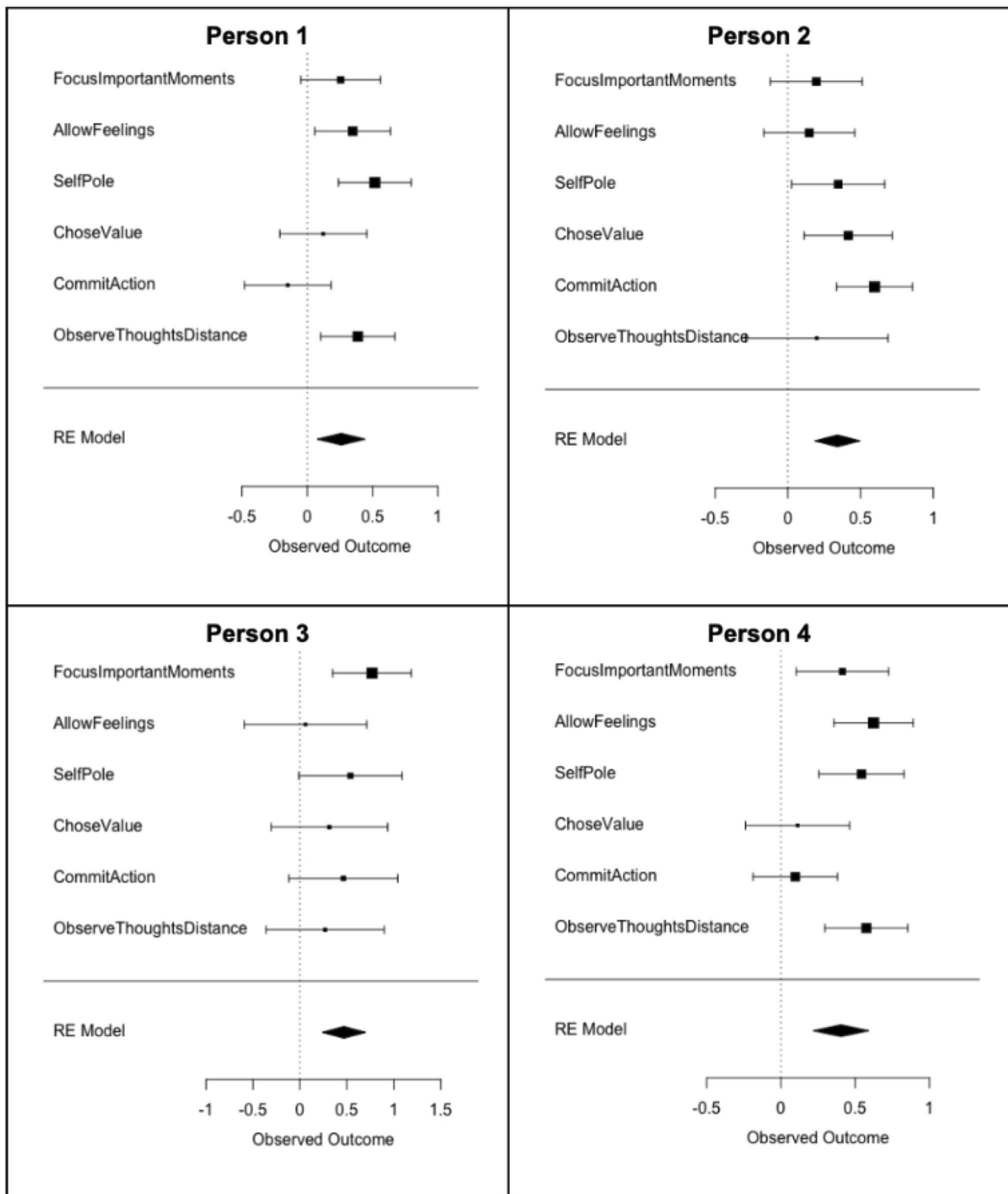


Fig. 3 Strength of the relationship between Psy-Flex processes and positive affect for four participants

Note: *FocusImportantMoments*: the ability to concentrate on present occurrences during significant moments; *AllowFeelings*: permitting unpleasant thoughts and experiences without immediate dismissal; *SelfPole*: noticing a stable core within oneself despite confusing

thoughts and experiences; *ChoseValue*: identifying and dedicating energy to personal priorities; *CommitAction*: engaging deeply in activities deemed important, useful, or meaningful; *ObserveThoughtsDistance*: viewing obstructive thoughts from afar without allowing them to dictate actions

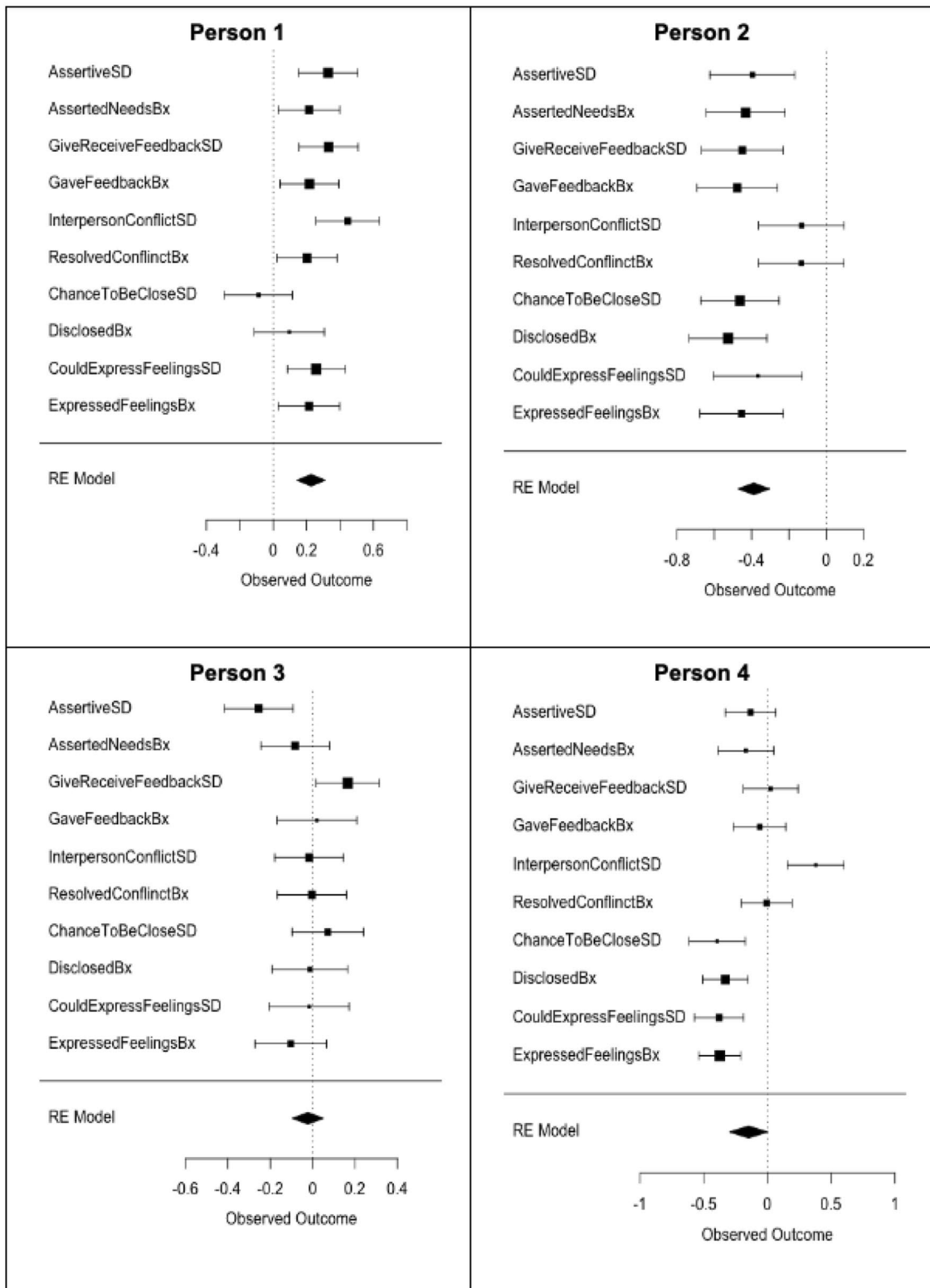


Fig. 4 Strength of relationship between loneliness and four processes, as measured by Functional Idiographic Assessment Template
 Note: SD indicates opportunities for action, Bx indicates taking of action

appropriate to report the results of the individual studies separately rather than trying to combine them” (p. 59). In this paper, the “individual studies” were individual persons and these comments suggest that combining their results into an average makes little analytic sense.

Whether these three datasets are exceptions or representative of psychological research remains unclear, raising questions about their typicality. Should these observations prove common, they would expose a fundamental shortcoming in the conventional analytical methods used to evaluate the effectiveness of evidence-based therapy. Violating ergodicity does not render classical statistical methods entirely ineffective for all purposes. However, it suggests that normative findings may not be reliably applicable to predicting and analyzing individual life trajectories. Consequently, idionomic methods should complement traditional statistical analyses such as randomized controlled trials, psychometric evaluations, and mediational analyses. This addition is crucial when applying results to specific individuals in psychotherapy, a universally adopted practice. It’s commonly assumed in psychological research that nomothetic generalizations serve as the “signal” for application to individuals, with variability often regarded as “noise.” As a statistical fact, the opposite may be true: Individual-level variability may be the key signal, and the collective average may be misleading.

If so, recognizing idiographic heterogeneity and violations in the ergodicity assumption is a first step in furthering clinical research and practice. Given the momentum provided by over 150 years of classical normative statistics as the source of individual prediction, only when we recognize that group averages cannot describe individual variation can we move to explain that clinically important variation. There are already a relatively small number of labs examining individual variation, although these labs are in the extreme minority compared to labs examining group-level effects. For example, Fisher and colleagues have used network analyses to model interindividual symptom dynamics (Fisher et al., 2017) and concussion symptomatology (Rabinowitz & Fisher, 2020). Wright and colleagues have used intensive time series data to show that people differ not only in the level of pathology but also in the range of symptoms, the temporal fluctuation of symptoms across days, and correlations between symptoms (Wright & Simms, 2016; Wright & Woods, 2020). Wright and colleagues have also shown that the structure of externalizing and internalizing behavior differs at the within compared to between-person level and is person-specific (Wright et al., 2015). Thus, there are clear methodologies for exploring individual-level networks of relationship when ergodicity is violated.

The present findings suggest that the link between clinically relevant processes and outcomes may almost always

violate the second model consistency aspect of the ergodicity assumption, namely, that the same dynamic models apply to all. In these datasets, what drives well-being for one person does not always drive well-being for another.

How do our results match theories that suggest certain processes should be of universal benefit? For instance, processes like observing thoughts from a distance and engaging in committed action, central to Acceptance and Commitment Therapy, seem to offer general benefits (Levin et al., 2012). However, we found that these processes were unrelated to or negatively associated with well-being among some individuals. We would suggest that these results do not invalidate ACT theory. Rather, they open the door for interesting questions about what moderates the link between processes and outcomes at the idiographic level. Whilst some processes may generally increase well-being, this won’t be true for everybody, in every context, at every time.

For instance, individuals might pursue actions aligned with their values, which, despite being meaningful, are challenging and stressful, thus not yielding hedonic well-being (Sahdra et al., 2024). People may also use the strategy of observing thoughts at a distance in a defensive way that is not linked to well-being. We acknowledge these hypotheses are speculative. However, acknowledging the variability in process-outcome links opens the door for exploring speculations like these in the future. Similarly, we have recently summarized the world’s literature on processes of change in mediational analyses in randomized trials (Hayes et al., 2022b). We do not suggest that the present result invalidates all the theories and measures identified there – but we suggest that they now need to be tested in an idionomic fashion.

Implications

The i-ARIMAX method described here focuses on bivariate relationships between a process and outcome and is likely to require less power than more complex multivariate analyses such as within-person structural equation modeling, network analyses, and factor analysis (Fisher et al., 2019; Sanford et al., 2022; Strohacker et al., 2021; Wright et al., 2015). We would suggest that i-ARIMAX might be useful for reducing the variables submitted to the more complex analysis. For example, if researchers were seeking to understand the within-person processes that predict relationship satisfaction, they might first use i-ARIMAX to identify the subset of processes that are most relevant to relationship satisfaction and then submit this subset to more complex, within-person structural equation modeling (Rush et al., 2019).

The results of the present study may also expand our notion of what it means for a measure to be valid. Typically, researchers present evidence of a scale’s validity by using

group-level statistics to show that the measure coheres across items and people and links to theoretically relevant criteria. In the present study, we showed that the pooled relationship between social behavior and an important criterion measure (loneliness) was often zero. Superficially, this implies that processes such as *asserting one's needs*, *expressing feelings*, or *resolving conflict* have no impact on *loneliness*. However, there were high levels of heterogeneity in the effects, suggesting that the zero effect estimate did not adequately describe the individual data. For some, *expressing feelings* was associated with more loneliness, for others, less loneliness. These findings raise the interesting possibility that a measure may lack criterion validity at the group level but still show practical utility at the individual level. Within the personalized intervention movement, we might prefer measures that discriminate between people over those with large average effects but cannot discriminate between people. In other words, what might be called “person-level” discriminant validity could be higher in measures with poor validity as measured by traditional normative psychometric analysis.

Similarly, whilst we may see heterogeneity of individual-level effects as a violation of ergodicity, we may also see it as a boon to personalized interventions. Heterogeneity of effects allows us to use measures to guide interventions and then evaluate if the measure has treatment utility, that is, improves outcomes (Ciarrochi et al., 2015). The findings in the present study may be useful in guiding future intervention research. Fisher et al. (2019) present an excellent example of this design. They had participants complete intensive daily surveys of symptoms, similar to the experience sampling methods utilized here. They then examined the idiosyncratic structure of the client's mood and anxiety pathology and used this information to construct personalized treatment plans for each individual. There was no control group in the design, but the authors could compare the effects of their personalized design to the effects observed in meta-analysis. The personalized design showed stronger effects. This encourages future research that compares personalized design based on intensive measures to standardized interventions. We hope i-ARIMAX can aid these designs.

To enhance their utility for clinicians, the algorithms from this study should ideally be integrated into clinical support tools (Lutz et al., 2022). These tools could streamline the assessment process, offering clinicians automated, straightforward insights into which processes might be most or least significant for a client's care. This could facilitate a more personalization and effective therapeutic approach by highlighting areas of potential focus or concern based on individual client profiles. There is meta-analytic evidence

that personalization can improve effect sizes (Lutz et al., 2022; Nye et al., 2023).

Limitations and Future Directions

The present paper focused on intensive self-report data. None of the methods presented in this paper are limited to self-report, however. Future research should evaluate i-ARIMAX using behavioral and physiological data, such as those collected passively from wearables and smartphones, or based on speech and text analysis. We still have much to learn about the within-person variation in the link between well-being and sleep, physical activity, heart rate variability, resting heart rate, diet, and other indices that link to well-being at the group level.

Our results show that there are substantial individual differences in the processes that drive well-being, but we do not yet know if this knowledge has treatment utility. Can experience-sampling measures and within-person analyses be used to improve treatment outcomes? What is the best way to convert within-person metrics to action? We might focus interventions on processes that are highly linked to outcomes for an individual. Should we also prioritize processes where the client typically scores below average? (e.g., Crutzen & Peters, 2023)? For example, if *having a meaningful challenge* is deeply important to person x (i.e., correlates strongly with well-being) and they are well below average in engaging in this process, then the process may be relatively influenceable. In contrast, if person X engages in many meaningful challenges in their life, then the practitioner may struggle to increase this process in their life: It may already be close to a ceiling. Other processes may be a better target for intervention. We don't yet know what the ideal algorithms are for personalizing interventions. We as a scientific community are only starting the process.

Ultimately, we must examine how normative or “group” statistics can be used with idiomonic statistics. If we know nothing about individual development, i.e., have no time series data on a client walking through the door seeking help, then group-level findings and one-off measures may be our best guess at what will work. But do we want to rely on guessing, especially when some processes, such as the social behaviors measured by FIAT and some behaviors in the PBAT, have little predictive value at the group level, even though they predict well-being for subsets of individuals?

Over the last fifty years, intervention science has invested billions of dollars in conducting thousands of trials on the efficacy of standardized treatment packages. Despite these efforts, effect sizes have not improved (Johnsen & Friberg, 2015; Jones et al., 2019; Ljótsson et al., 2017). We do not know if personalization metrics like those presented here can improve treatment outcomes, but we believe the time

has come to see if personalized interventions can do better than standardized interventions. We see no reason to believe that another fifty years of assessing complex, standardized packages in normative designs will lead to improvements.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10608-024-10486-w>.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

Declarations

Conflict of Interest: The authors declare that they have no conflict of interest. Dr. Hofmann receives financial support by the Alexander von Humboldt Foundation (as part of the Alexander von Humboldt Professor), the Hessische Ministerium für Wissenschaft und Kunst (as part of the LOEWE Spitzenprofessur), and the DYNAMIC center, which is funded by the LOEWE program of the Hessian Ministry of Science and Arts (Grant Number: LOEWE1/16/519/03/09.001(0009)/98).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Al-Qazzaz, R. A., & Yousif, S. A. (2022). High performance time series models using auto autoregressive integrated moving average. *Indonesian Journal of Electrical Engineering and Computer Science*, 27(1), 422–430. <https://doi.org/10.11591/ijeecs.v27.i1.pp422-430>
- Beltz, A. M., & Gates, K. M. (2017). Network mapping with GIMME. *Multivariate Behavioral Research*, 52(6), 789–804. <https://doi.org/10.1080/00273171.2017.1373014>
- Benoy, C., Knitter, B., Schumann, I., Bader, K., Walter, M., & Glöster, A. T. (2019). Treatment sensitivity: Its importance in the measurement of psychological flexibility. *Journal of Contextual Behavioral Science*, 13, 121–125. <https://doi.org/10.1016/j.jcbs.2019.07.005>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. Wiley. <https://play.google.com/store/books/details?id=2oYmEAAAQBAJ>
- Bottomley, C., Scott, J. A. G., & Isham, V. (2019). Analysing interrupted time series with a control. *Epidemiologic Methods*, 8(1). <https://doi.org/10.1515/em-2018-0010>
- Brockman, R., Ciarrochi, J., Parker, P., & Kashdan, T. B. (2023). Behaving versus thinking positively: When the benefits of cognitive reappraisal are contingent on satisfying basic psychological needs. *Journal of Contextual Behavioral Science*, 27, 120–125. <https://doi.org/10.1016/j.jcbs.2023.01.005>
- Brockwell, P. J., & Davis, R. A. (2013). *Introduction to time series and forecasting*. Springer Science & Business Media. <https://play.google.com/store/books/details?id=XtDkBwAAQBAJ>
- Bulteel, K., Tuerlinckx, F., Brose, A., & Ceulemans, E. (2016). Clustering vector autoregressive models: Capturing qualitative differences in within-person dynamics. *Frontiers in Psychology*, 7, 1540. <https://doi.org/10.3389/fpsyg.2016.01540>
- Callaghan, G. M. (2006). The functional idiographic assessment template (FIAT) system: For use with interpersonally-based interventions including functional analytic psychotherapy (FAP) and FAP-enhanced treatments. *The Behavior Analyst Today*, 7(3), 357–398. <https://doi.org/10.1037/h0100160>
- Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology*, 52, 685–716.
- Chatfield, C., & Xing, H. (2019). *The analysis of time series: An introduction with R*. CRC. <https://play.google.com/store/books/details?id=9tPOwQEACAAJ>
- Cheung, F., & Lucas, R. E. (2014). Assessing the validity of single-item life satisfaction measures: Results from three large samples. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment Care and Rehabilitation*, 23(10), 2809–2818. <https://doi.org/10.1007/s11136-014-0726-4>
- Ciarrochi, J., Zettle, R. D., Brockman, R., Duguid, J., Parker, P., Sahdra, B., & Kashdan, T. B. (2015). Measures that make a difference. In R. D. Zettle, S. C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds.), *The Wiley handbook of contextual behavioral science* (Vol. 15, pp. 320–346). Wiley. <https://doi.org/10.1002/9781118489857.ch16>
- Ciarrochi, J., Hayes, S. C., Oades, L. G., & Hofmann, S. G. (2021). Toward a unified framework for positive psychology interventions: Evidence-based processes of change in coaching, prevention, and training. *Frontiers in Psychology*, 12, 809362. <https://doi.org/10.3389/fpsyg.2021.809362>
- Ciarrochi, J., Sahdra, B., Hofmann, S. G., & Hayes, S. C. (2022). Developing an item pool to assess processes of change in psychological interventions: The process-based assessment tool (PBAT). *Journal of Contextual Behavioral Science*, 23, 200–213. <https://doi.org/10.1016/j.jcbs.2022.02.001>
- Crutzen, R., & Peters, G. J. Y. (2023). A lean method for selecting determinants when developing behavior change interventions. *Health Psychology and Behavioral Medicine*, 11(1), 2167719. <https://doi.org/10.1080/21642850.2023.2167719>
- Darrow, S. M., Callaghan, G. C., Bonow, J. T., & Follette, W. C. (2014). The functional idiographic assessment template-questionnaire (FIAT-Q): Initial psychometric properties. *Journal of Contextual Behavioral Science*, 3(2), 124–135. <https://doi.org/10.1016/j.jcbs.2014.02.002>
- Donald, J. N., Ciarrochi, J., Parker, P. D., & Sahdra, B. K. (2019). Compulsive internet use and the development of self-esteem and hope: A four-year longitudinal study. *Journal of Personality*, 87(5), 981–995. <https://doi.org/10.1111/jopy.12450>
- Donald, J. N., Ciarrochi, J., & Guo, J. (2022). Connected or cut-off? A 4-year longitudinal study of the links between adolescents' compulsive internet use and social support. *Personality & Social Psychology Bulletin*, 1461672221127802. <https://doi.org/10.1177/01461672221127802>
- Ehlers, A., Hackmann, A., Grey, N., Wild, J., Liness, S., Albert, I., Deale, A., Stott, R., & Clark, D. M. (2014). A randomized controlled trial of 7-day intensive and standard weekly cognitive therapy for PTSD and emotion-focused supportive therapy. *The American Journal of Psychiatry*, 171(3), 294–304. <https://doi.org/10.1176/appi.ajp.2013.13040552>

- Embry, D. D., & Biglan, A. (2008). Evidence-based kernels: Fundamental units of behavioral influence. *Clinical Child and Family Psychology Review*, 11(3), 75–113. <https://doi.org/10.1007/s10567-008-0036-x>
- Ferrari, M., Ciarrochi, J., Yap, K., Sahdra, B., & Hayes, S. C. (2022). Embracing the complexity of our inner worlds: Understanding the dynamics of self-compassion and self-criticism. In *Mindfulness*. <https://doi.org/10.1007/s12671-022-01897-5>
- Fisher, A. J. (2015). Toward a dynamic model of psychological assessment: Implications for personalized care. *Journal of Consulting and Clinical Psychology*, 83(4), 825–836. <https://doi.org/10.1037/ccp0000026>
- Fisher, A. J., Reeves, J. W., Lawyer, G., Medaglia, J. D., & Rubel, J. A. (2017). Exploring the idiographic dynamics of mood and anxiety via network analysis. *Journal of Abnormal Psychology*, 126(8), 1044–1056. <https://doi.org/10.1037/abn0000311>
- Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences of the United States of America*, 115(27), E6106–E6115. <https://doi.org/10.1073/pnas.1711978115>
- Fisher, A. J., Bosley, H. G., Fernandez, K. C., Reeves, J. W., Soyster, P. D., Diamond, A. E., & Barkin, J. (2019). Open trial of a personalized modular treatment for mood and anxiety. *Behaviour Research and Therapy*, 116, 69–79. <https://doi.org/10.1016/j.brat.2019.01.010>
- Gloster, A. T., Block, V. J., Klotsche, J., Villanueva, J., Rinner, M. T. B., Benoy, C., Walter, M., Karekla, M., & Bader, K. (2021). PsyFlex: A contextually sensitive measure of psychological flexibility. *Journal of Contextual Behavioral Science*, 22, 13–23. <https://doi.org/10.1016/j.jcbs.2021.09.001>
- Gloster, A. T., Haller, E., Villanueva, J., Block, V. J., Benoy, C., Meyer, A. H., Brogli, S., Kuhweide, V., Karekla, M., Bader, K., Walter, M., & Lang, U. (2023). *Transdiagnostic psychotherapy for non-responding in- and outpatients with common mental disorders: The choose change effectiveness trial*.
- Greenberg, L. S. (1986). Change process research. *Journal of Consulting and Clinical Psychology*, 54(1), 4–9. <https://doi.org/10.1037//0022-006x.54.1.4>
- Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, 20(1), 102–116. <https://doi.org/10.1037/a0038889>
- Hauser, D. J., & Schwarz, N. (2016). Attentive turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, 48(1), 400–407. <https://doi.org/10.3758/s13428-015-0578-z>
- Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (2012). *Acceptance and commitment therapy, second edition: The process and practice of mindful change*. 2nd edition. Guilford Publications. http://books.google.com.au/books?id=5_xAm0VF4X8C
- Hayes, S. C., Hofmann, S. G., Stanton, C. E., Carpenter, J. K., Sanford, B. T., Curtiss, J. E., & Ciarrochi, J. (2019). The role of the individual in the coming era of process-based therapy. *Behaviour Research and Therapy*, 117(September), 40–53. <https://doi.org/10.1016/j.brat.2018.10.005>
- Hayes, S. C., Hofmann, S. G., & Ciarrochi, J. (2020a). A process-based approach to psychological diagnosis and treatment: The conceptual and treatment utility of an extended evolutionary meta model. *Clinical Psychology Review*, 82, 101908. <https://doi.org/10.1016/j.cpr.2020.101908>
- Hayes, S. C., Hofmann, S. G., & Ciarrochi, J. (2020b). Building a process-based diagnostic system: An extended evolutionary approach. *Beyond the DSM: Toward a process-based alternative for diagnosis and mental health treatment*. Context Press/New Harbinger Publications, 251–278. <http://josephciarrochi.com/wp-content/uploads/2021/01/building-a-process-based-diagnostic-system-extended-evolutionary-approach-hayes-hofmann-ciarrochi-2020.pdf>
- Hayes, S. C., Ciarrochi, J., Hofmann, S. G., Chin, F., & Sahdra, B. (2022a). Evolving an idionomic approach to processes of change: Towards a unified personalized science of human improvement. *Behaviour Research and Therapy*, 156, 104155. <https://doi.org/10.1016/j.brat.2022.104155>
- Hayes, S. C., Ciarrochi, J., Hofmann, S. G., Chin, F., & Sahdra, B. (2022b). Evolving an idionomic approach to processes of change: Towards a unified personalized science of human improvement. *Behaviour Research and Therapy*, 156, 104155. <https://doi.org/10.1016/j.brat.2022.104155>
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *Bmj*, 327(7414), 557–560. <https://doi.org/10.1136/bmj.327.7414.557>
- Ho, S. L., & Xie, M. (1998). The use of ARIMA models for reliability forecasting and analysis. *Computers & Industrial Engineering*, 35(1), 213–216. [https://doi.org/10.1016/S0360-8352\(98\)00066-7](https://doi.org/10.1016/S0360-8352(98)00066-7)
- Hopwood, C. J., Wright, A. G. C., & Bleidorn, W. (2022). Person–environment transactions differentiate personality and psychopathology. *Nature Reviews Psychology*, 1(1), 55–63. <https://doi.org/10.1038/s44159-021-00004-0>
- Hoyt, M. F., Young, J., & Rycroft, P. (2020). Single session thinking 2020. *The Australian and New Zealand Journal of Family Therapy*, 41(3), 218–230. <https://doi.org/10.1002/anzf.1427>
- Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychological Methods*, 11(2), 193–206. <https://doi.org/10.1037/1082-989X.11.2.193>
- Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27, 1–22. <https://doi.org/10.18637/jss.v027.i03>
- Imel, Z. E., Laska, K., Jakupcak, M., & Simpson, T. L. (2013). Meta-analysis of dropout in treatments for posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology*, 81(3), 394–404. <https://doi.org/10.1037/a0031474>
- Ioannidis, J. P. A., Patsopoulos, N. A., & Evangelou, E. (2007). Uncertainty in heterogeneity estimates in meta-analyses. *Bmj*, 335(7626), 914–916. <https://doi.org/10.1136/bmj.39343.408449.80>
- Jensen, L. (1990). Guidelines for the application of ARIMA models in time series. *Research in Nursing & Health*, 13(6), 429–435. <https://doi.org/10.1002/nur.4770130611>
- Johnsen, T. J., & Friberg, O. (2015). The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: A meta-analysis. *Psychological Bulletin*, 141(4), 747–768. <https://doi.org/10.1037/bul0000015>
- Jones, E. E., Cumming, J. D., & Horowitz, M. J. (1988). Another look at the nonspecific hypothesis of therapeutic effectiveness. *Journal of Consulting and Clinical Psychology*, 56(1), 48–55. <https://doi.org/10.1037//0022-006x.56.1.48>
- Jones, P. J., Mair, P., Kuppens, S., & Weisz, J. R. (2019). An upper limit to youth psychotherapy benefit? A meta-analytic copula approach to psychotherapy outcomes. *Clinical Psychological Science*, 7(6), 1434–1449. <https://doi.org/10.1177/2167702619858424>
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. In *Journal of Econometrics* (Vol. 54, Issues 1–3, pp. 159–178). [https://doi.org/10.1016/0304-4076\(92\)90104-y](https://doi.org/10.1016/0304-4076(92)90104-y)
- Levin, M. E., Hildebrandt, M. J., Lillis, J., & Hayes, S. C. (2012). The impact of treatment components suggested by the psychological flexibility model: A meta-analysis of laboratory-based component studies. *Behavior Therapy*, 43(4), 741–756. <https://doi.org/10.1016/j.beth.2012.05.003>
- Ljótsson, B., Hedman, E., Mattsson, S., & Andersson, E. (2017). The effects of cognitive-behavioral therapy for depression are not falling: A re-analysis of Johnsen and Friberg (2015) [Review of

- the effects of cognitive-behavioral therapy for depression are not falling: A re-analysis of Johnsen and Friborg (2015)*. *Psychological Bulletin*, 143(3), 321–325. psycnet.apa.org.
- Lo, K., Stephenson, M., & Lockwood, C. (2019). Analysis of heterogeneity in a systematic review using meta-regression technique. *International Journal of Evidence-Based Healthcare*, 17(2), 131–142. <https://doi.org/10.1097/XEB.000000000000163>
- Lutz, W., Schwartz, B., & Delgado, J. (2022). Measurement-based and data-informed psychological therapy. *Annual Review of Clinical Psychology*, 18, 71–98. <https://doi.org/10.1146/annurev-clinpsy-071720-014821>
- Marshall, S. L., Parker, P. D., Ciarrochi, J., & Heaven, P. C. L. (2014). Is self-esteem a cause or consequence of social support? A 4-year longitudinal study. *Child Development*, 85(3), 1275–1291. <https://doi.org/10.1111/cdev.12176>
- Masuda, A., Hayes, S. C., Lillis, J., Bunting, K., Herbst, S. A., & Fletcher, L. B. (2009). The relation between psychological flexibility and mental health stigma in acceptance and commitment therapy: A preliminary process investigation. *Behavior and Social Issues*, 18(1), 25–40. <https://doi.org/10.5210/bsi.v18i1.2525>
- Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research and Perspectives*, 2(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1
- Molenaar, P. C. M. (2013). On the necessity to use person-specific data analysis approaches in psychology. *The European Journal of Developmental Psychology*, 10(1), 29–39. <https://doi.org/10.1080/17405629.2012.747435>
- Nau, R. (2020). *ARIMA models for time series forecasting*. *Duke Forecasting Home Page*. <https://people.duke.edu/~rnau/411arim.htm>
- Nye, A., Delgado, J., & Barkham, M. (2023). Efficacy of personalized psychological interventions: A systematic review and meta-analysis. *Journal of Consulting and Clinical Psychology*, 91(7), 389–397. <https://doi.org/10.1037/ccp0000820>
- Ong, C. W., Hayes, S. C., & Hofmann, S. G. (2022). A process-based approach to cognitive behavioral therapy: A theory-based case illustration. *Frontiers in Psychology*, 13, 1002849. <https://doi.org/10.3389/fpsyg.2022.1002849>
- Paccagnella, O. (2006). Centering or not centering in multilevel models? The role of the group mean and the assessment of group effects. *Evaluation Review*, 30(1), 66–85. <https://doi.org/10.1177/0193841X05275649>
- Paul, G. L. (1969). Behavior modification research: Design and tactics. In C. M. Franks (Ed.), *Behavior therapy: Appraisal and status* (pp. 29–62).
- Prochaska, J. O., & DiClemente, C. C. (1983). Stages and processes of self-change of smoking: Toward an integrative model of change. *Journal of Consulting and Clinical Psychology*, 51(3), 390–395. <https://doi.org/10.1037//0022-006x.51.3.390>
- Rabinowitz, A. R., & Fisher, A. J. (2020). Person-specific methods for characterizing the course and temporal dynamics of concussion symptomatology: A pilot study. *Scientific Reports*, 10(1), 1248. <https://doi.org/10.1038/s41598-019-57220-1>
- Rajakannan, T., Safer, D. J., Burcu, M., & Zito, J. M. (2016). National trends in psychiatric not otherwise specified (NOS) diagnosis and medication use among adults in outpatient treatment. *Psychiatric Services*, 67(3), 289–295. <https://doi.org/10.1176/appi.ps.201500045>
- Rijnhart, J. J. M., Lamp, S. J., Valente, M. J., MacKinnon, D. P., Twisk, J. W. R., & Heymans, M. W. (2021). Mediation analysis methods used in observational research: A scoping review and recommendations. *BMC Medical Research Methodology*, 21(1), 226. <https://doi.org/10.1186/s12874-021-01426-3>
- Robinson, W. S. (2009). Ecological correlations and the behavior of individuals. *International Journal of Epidemiology*, 38(2), 337–341. <https://doi.org/10.1093/ije/dyn357>
- Rosen, G. M., & Davison, G. C. (2003). Psychology should list empirically supported principles of change (ESPs) and not credential trademarked therapies or other treatment packages. *Behavior Modification*, 27(3), 300–312. <https://doi.org/10.1177/0145445503027003003>
- Rush, J., Rast, P., Almeida, D. M., & Hofer, S. M. (2019). Modeling long-term changes in daily within-person associations: An application of multilevel SEM. *Psychology and Aging*, 34(2), 163–176. <https://doi.org/10.1037/pag0000331>
- Sahdra, B. K., Ciarrochi, J., Basarkod, G., Dicke, T., Guo, J., Parker, P. D., & Marsh, H. W. (2022). High school students' tenacity and flexibility in goal pursuit linked to life satisfaction and achievement on competencies tests. *Journal of Educational Psychology*, 114(3), 622–636. <https://doi.org/10.1037/edu0000667>
- Sahdra, B. K., Ciarrochi, J., Fraser, M. I., Yap, K., Haller, E., Hayes, S. C., Hofmann, S. G., & Gloster, A. T. (2023). The compassion balance: Understanding the interrelation of self- and other-compassion for optimal well-being. *Mindfulness*, 14(8), 1997–2013. <https://doi.org/10.1007/s12671-023-02187-4>
- Sahdra, B. K., Ciarrochi, J., Klimczak, K. S., Krafft, J., Hayes, S. C., & Levin, M. (2024). Testing the applicability of idiomonic statistics in longitudinal studies: The example of doing what matters. *Journal of Contextual Behavioral Science*. <https://doi.org/10.1016/j.jcbs.2024.100728>
- Sanford, B. T., Ciarrochi, J., Hofmann, S. G., Chin, F., Gates, K. M., & Hayes, S. C. (2022). Toward empirical process-based case conceptualization: An idiomonic network examination of the process-based assessment tool. *Journal of Contextual Behavioral Science*, 25, 10–25. <https://doi.org/10.1016/j.jcbs.2022.05.006>
- Snijders, T. A. B., a., Tom, A. B., Snijders Roel, J., Bosker, & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage Publications (CA). <https://play.google.com/store/books/details?id=G8fhqL9xHOAC>
- Stanton, C. E., Follette, W. C., Hayes, S. C., Sanford, B. T., & Ciarrochi, J. (2023). *A complex network approach to loneliness, social connection and interpersonal functioning*.
- Strohacker, K., Keegan, R., Beaumont, C. T., & Zakrajsek, R. A. (2021). Applying P-technique factor analysis to explore person-specific models of readiness-to-exercise. *Frontiers in Sports and Active Living*, 3, 685813. <https://doi.org/10.3389/fspor.2021.685813>
- Tedeschi, R. G., & Moore, B. A. (2021). Posttraumatic growth as an integrative therapeutic philosophy. *Journal of Psychotherapy Integration*, 31(2), 180–194. <https://doi.org/10.1037/int0000250>
- van Weel, C. (2005). Longitudinal research and data collection in primary care. *Annals of Family Medicine*, 3 Suppl 1(Suppl 1), S46–S51. <https://doi.org/10.1370/afm.300>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
- Villanueva, J., Meyer, A. H., Rinner, M. T. B., Firsching, V. J., Benoy, C., Brogli, S., Walter, M., Bader, K., & Gloster, A. T. (2019). Choose change: Design and methods of an acceptance and commitment therapy effectiveness trial for transdiagnostic treatment-resistant patients. *Bmc Psychiatry*, 19(1), 173. <https://doi.org/10.1186/s12888-019-2109-4>
- Villatte, J. L., Vilardaga, R., Villatte, M., Plumb Vilardaga, J. C., Atkins, D. C., & Hayes, S. C. (2016). Acceptance and commitment therapy modules: Differential impact on treatment processes and outcomes. *Behaviour Research and Therapy*, 77, 52–61. <https://doi.org/10.1016/j.brat.2015.12.001>
- von Hippel, P. T. (2015). The heterogeneity statistic I(2) can be biased in small meta-analyses. *BMC Medical Research Methodology*, 15, 35. <https://doi.org/10.1186/s12874-015-0024-z>

- Wang, X., & Cheng, Z. (2020). Cross-sectional studies: Strengths, weaknesses, and recommendations. *Chest*, *158*(1S), S65–S71. <https://doi.org/10.1016/j.chest.2020.03.012>
- Wicksell, R. K., Olsson, G. L., & Hayes, S. C. (2010). Psychological flexibility as a mediator of improvement in acceptance and commitment therapy for patients with chronic pain following whiplash. *European Journal of Pain*, *14*(10), 1059.e1–e1059.e11.
- Wright, A. G. C., & Simms, L. J. (2016). Stability and fluctuation of personality disorder features in daily life. *Journal of Abnormal Psychology*, *125*(5), 641–656. <https://doi.org/10.1037/abn0000169>
- Wright, A. G. C., & Woods, W. C. (2020). Personalized models of psychopathology. *Annual Review of Clinical Psychology*, *16*, 49–74. <https://doi.org/10.1146/annurev-clinpsy-102419-125032>
- Wright, A. G. C., Beltz, A. M., Gates, K. M., Molenaar, P. C. M., & Simms, L. J. (2015). Examining the dynamic structure of daily internalizing and externalizing behavior at multiple levels of analysis. *Frontiers in Psychology*, *6*, 1914. <https://doi.org/10.3389/fpsyg.2015.01914>
- Wright, A. G. C., Gates, K. M., Arizmendi, C., Lane, S. T., Woods, W. C., & Edershile, E. A. (2019). Focusing personality assessment on the person: Modeling general, shared, and person specific processes in personality and psychopathology. *Psychological Assessment*, *31*(4), 502–515. <https://doi.org/10.1037/pas0000617>
- Young, Q. R., Ignaszewski, A., Fofonoff, D., & Kaan, A. (2007). Brief screen to identify 5 of the most common forms of psychosocial distress in cardiac patients: Validation of the screening tool for psychological distress. *The Journal of Cardiovascular Nursing*, *22*(6), 525–534. <https://doi.org/10.1097/01.JCN.0000297383.29250.14>
- Young, Q. R., Nguyen, M., Roth, S., Broadberry, A., & Mackay, M. H. (2015). Single-item measures for depression and anxiety: Validation of the screening tool for psychological distress in an inpatient cardiology setting. *European Journal of Cardiovascular Nursing: Journal of the Working Group on Cardiovascular Nursing of the European Society of Cardiology*, *14*(6), 544–551. <https://doi.org/10.1177/1474515114548649>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.