A personalised-approach to identifying important determinants of well-being

Joseph Ciarrochi[1], Baljinder Sahdra[1], Steven C. Hayes[2], Stefan G. Hofmann[3], Brandon Sanford[4], Cory Stanton[5], Keong Yap[6], Madeleine I. Fraser[6], Kathleen Gates[7], & Andrew T. Gloster[8]

[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, and Boston University, Boston, USA

[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] Department of Psychology, University of North Carolina at Chapel Hill, USA.

[8]Division of Clinical Psychology and Intervention Science, Faculty of Psychology, University of Basel, Basel, Switzerland.

Author Note

We have no conflict of interest to disclose. This work was supported, in part, by grants from the Swiss National Science Foundation awarded to ATG (SNSF Grants: PP00P1_190082 & PP00P1_163716/1).
We have no known conflict of interest to disclose.

Correspondence concerning this article should be addressed to Joseph Ciarrochi, Institute for Positive Psychology and Education, Australian Catholic University, PO Box 968, North Sydney, NSW 2059. E-mail:Joseph.Ciarrochi@acu.edu.au

Abstract

Background. Identifying the most important psychological drivers of well-being for a particular individual is critical to developing personalised interventions. **Methods.** We utilised three, intensive daily diary studies (within person measurement occasions N >50) across three data sets (n_1=44; n_2=37; n_3=141) to examine within-person associations between clinically-relevant processes and a variety of outcomes. We utilised a novel idiographic algorithm, ”i-ARIMAX,” to calculate the strength of relationship (beta) between every process and every outcome within individuals. We then submitted all betas to meta-analytic methods. **Results**. All process-outcome links were highly heterogeneous between individuals. Processes that were associated with positive outcomes for some people were often unrelated to outcomes for others or associated with negative outcomes. **Conclusion.** i-ARIMAX might be used to guide personalised interventions and to reduce the number of candidate variables for complicated within-person analysis.

A personalised-approach to identifying important determinants of well-being

A client walks through the door 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 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 identify the client's psychiatric disorder and then administer an evidence-based treatment protocol that has been shown in randomized trials to improve outcomes for that disorder (Chambless & Ollendick, 2001). Practitioners often encounter several problems with this recommendation. First, they may not have the time to administer the full treatment protocol. Perhaps they will only have one or two sessions with the client. Second, the client may show comorbidities or unique features that do not fit existing syndromal expectations. 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). Finally, practitioners may have to choose between different evidence-based treatment protocols, not knowing which one would be best suited for this particular client. Since extensive training in individual protocols is often claimed to be necessary for competent use, even though processes of change may be similar across various protocols, evidence-based clinicians are often left in an impossible situation when they try to fit interventions to the specific needs of individuals. .

There has been an increasing call to identify the evidence-based intervention kernels that comprise a package and the processes of change they affect (Hayes et al., 2022a; Rosen & Davison, 2003). In principle, a process focus should make it easier to personalise interventions, as one can select the most relevant intervention kernel that bears on a process for a particular individual in a particular context (Hayes et al., 2019). In broad terms, we will define a process of change here as an evidence-based, theoretically coherent,

contextually situated, modifiable biopsychosocial event or sequence of events that can lead to adaptive or maladaptive outcome for a client (Hayes et al., 2020a). 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 behaviour (e.g., goal setting; Hayes et al, 2022a). The core question of this paper is, how does one go about selecting the most relevant biopsychosocial process to target for a particular individual?

The most common general strategy for answering this question is to collect data from a large group of participants, to examine the link between processes and outcomes for the group (e.g., via correlation or mediational analysis, perhaps as part of a randomized controlled trial), and then to assume that these group level effects apply to each individual in the group (plus or minus some error). For example, if an ACT intervention improves a process of change such as psychological flexibility for a group of participants, and flexibility correlates with or mediates outcomes, it is common to assume that ACT will be likely to improve that process and lead to better outcomes for the various individuals in that group.

Recently, theoreticians and researchers have been questioning the assumption that we can rely on group data to understand how individuals develop and change (Fisher et al., 2017; Hopwood et al., 2022; Molenaar, 2004; Rabinowitz & Fisher, 2020; Sanford et al., 2022; Wright et al., 2019). One core of their concerns is that doing so requires that a key but rarely discussed assumption of classical statistics be met, namely that the predicted phenomenon is “ergodic”, the “ergodic assumption” (Molenaar, 2004).

To date, there has been no definitive rejection of Molenaar’s claim that ergodicity is an assumption of classical statistics as it applies to the probabilistic extension of normative results to individuals. It requires an understanding of complex mathematics of the ergodic theorem (Birkhoff, 1931; Neumann, 1932) to see why stationarity is a necessary aspect of the ergodic assumption. We will not address that aspect of the assumption here except to note in passing that its implications are grim since processes of change or changes in clinical outcomes cannot by definition be stationary. This would mean, for example, that

traditional mediational analyses or the usual analyses of results of randomized trials violate their own assumptions if the clinical purpose is to apply resulting knowledge to individuals, even probabilistically (Hofmann et al., 2020)

The second aspect of ergodicity–model consistency–is more easily understood without extensive mathematical knowledge, however, and it is more obviously relevant to a central question in clinical science and practice: To what extent do group level findings apply to individuals? Are processes of change that are identified as important at the group level also consistently important at the individual level (Molenaar, 2004). In the present paper we propose to examine this model consistency aspect of the ergodic assumption by considering the relationship between effects identified at the group level and at individual level in three intensive, daily diary studies, across a variety of processes and positive and negative outcomes.

Identifying key processes of change

For an intervention to be effective, it must target processes that have a functional link to the outcome of interest (Kok et al., 2016). Further, because practitioners often have limited time and resources and cannot target every possible process of interest, they need to focus on the most important ones such as those that are most modifiable and have the strongest link to outcome. Therefore the first step in developing an intervention often involves identifying the strength of relationship between hypothesised processes and outcomes, within a target population and context so that the most important processes can be targeted(Crutzen et al., 2017; Crutzen & Peters, 2022).

Researchers have used many tools to identify important processes in a context. The three most common approaches–cross-sectional, longitudinal, and mediational—usually estimate importance of processes at the group level, assuming that these estimates apply to the individuals in the group with some “unexplained” error. We briefly consider each of these approaches now.

Cross-sectional analysis is inherently between-person and thus if the ergodic assumption does not apply, these analyses may not apply within a person (Molenaar, 2004).

For example, research shows that goal tenacity is linked to student well-being (Sahdra, Ciarrochi, Basarkod, et al., 2022), and thus suggest tenacity be promoted in student interventions. That may not hold at the level of the individual. For example, providing an intervention that makes tenacious students even more tenacious may result in less life satisfaction even if tenacity is positive at the group level.

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. Within-person variation in the process-outcome link is often modelled as random slopes and treated as error. Second, 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 biassed if there are widely varying patterns of individual effects (Wright & Woods, 2020). In addition, multilevel model-implied estimates for individuals in a sample have been shown to be driven towards the group level effect (Sahdra, Ciarrochi, Fraser, et al., 2022).

Mediational analysis is a third, group-based approach focused on identifying the functionally important pathway of change in an intervention. 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 average group well-being (e.g., reduces mean depression scores) controlling for intervention (for a systematic review of recent studies see Hayes et al (2022a). The studies in all three different literatures commonly acknowledge individual variation in effects, but such between-person variation is statistically regarded as "error" in each approach. .

Recent theory and research questions the degree to which group averages provide an adequate model of individual-level processes (see, for example, Hayes et al., 2019). An alternative is to assume that person might be thought of as a system of interacting, dynamic processes that give rise to individual life trajectories (Fisher, 2015; Fisher et al., 2018; Molenaar, 2004, 2013), and that nomothetic generalizations should be explored only after these are modelled, and should be kept only if they improve idiographic fit. This idea has led to a sharp recent increase in “idionomic” approaches to well-being, in which individual-level analyses are taken to be the essential foundation of nomothetic implications, rather than vice versa (Sanford et al., 2022). 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. That implicit assumption is being examined in the present study.

Identifying important within-person processes

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 index: the degree to which within-person changes in the process are associated with within-person changes in an indexed outcome. Processes may relate in complex ways to an outcome, such as via an interaction with other variables. Complex statistical methods exist for clinically modelling networks of that kind (Beltz & Gates, 2017; Ong et al., 2022) , but for the sake of this paper, we will focus only on modelling simple contemporaneous relationships between processes and outcomes. As will be seen, even that focus is not that 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 modelling (Donald et al., 2019), vector auto-regressive models (Bulteel et al., 2016), and network analysis (Beltz & Gates, 2017). Even the relatively simple analyses explored here might identify potentially important variables and eliminate unimportant ones that could enter relevant functional analyses. 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.

A meta-framework for understanding processes

We turn our attention to the processes that we focus on in the present set of studies. To put these processes into a wider context, we first nest them within a theoretical framework, the Extended Evolutionary Meta-Model (EEMM; Ciarrochi et al., 2021; Hayes et al., 2020b). The EEMM provides a coherent way, grounded in multi-dimensional, multi-level evolutionary science, of examining the wide variety of specific models of processes of change that currently exist.

Within the EEMM, processes differ in the psychological dimensions they may target–(cognitive, affective, attentional, self, motivational, and overt-behavior)–and in the levels of organization they may involve – bio-physiological, individual psychological, and sociocultural. Finally, processes can also be understood in terms of the extent they promote the four foundational drivers of evolutionary change (Gloster & Haller, 2022; Hayes et al., 2022b) : variation (did you get the client to try something new?), selection (how do you know the new behavior building value or producing success?), retention (do clients maintain healthy variations across time and circumstance?) and context (do clients know when to deploy new skills then they are needed). The EEMM has been shown to successfully summarise a wide variety of processes of change (Ciarrochi et al., 2021; Hayes et al., 2022a).

The present study used three different process measures, which can be nested with the EEMM. The process-based assessment tool (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 EEMM. Selection items focus on the extent 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 behavior to be more value consistent, and retention items focus on the extent people can persist at value-consistent behavior. The biophysiological level is assessed by two items related to health behaviors, and the sociocultural level by items assessing relationship health. Research has shown that the PBAT links in expected ways to clinically relevant outcomes and to need satisfaction; and it shows discriminate 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.

Research has been conducted with the PBAT in an intensive longitudinal design and has shown that it links to common clinical outcomes for virtually all participants in individual network analysis (Sanford et al., 2022). Further, Sanford et al (2022) showed that relationships needed to be studied at the individual level because participants' responses rarely conformed to the group averages. Not all processes were beneficial to all people.

The second process-based measure examined in this paper is the Psy-Flex, which focuses on key behaviours 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). Research has shown that the Psy-Flex has a single factor structure, which is consistent with it reflecting psychological flexibility overall and is linked in expected ways to well-being, discriminates between clinical and non-clinical samples, and is sensitive 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 behaviours common to social repertoires, conceptualised 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).

Current study

We utilised archival data from three different samples, each focusing on a different clinically-relevant process measure and on an indices of positive and negative functioning. All three studies received full ethics review and approval .

Sample 1: 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. A total of 57 participants were recruited and completed at least one assessment. Participants who completed data collection (criteria are described below) were evenly represented regarding gender (nfemale = 24), ranged in age from 19 to 71 (mage = 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, 7 were lost because of attrition, having missed over 10 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 female, 24 male; 5 no answer for gender), with a mean age of 33.8 (SD=13.03).

Data was collected twice-daily, across 35 days. In order 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. While formal power recommendations have yet to be established (Fried et al., 2017), 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 in order 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 valanced item for each. Two items assess range of variation in behavior and two items assess behavioral retention across time; these item pairs also had one positively and one negatively valanced item. The stem for each item was “Over the past 12 h” 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 utilised 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.

Sample 2: the Functional Idiographic Assessment Template-Mobile (FIAT-M).

Sample 2 comes from a twice-daily diary study of social behaviours, 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 18-55 years old (*M* = 27.13; *SD* = 9.6). 39 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 Sample 2 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 = .83; “I was supported by people in my life”), as well as a modified UCLA 3 Item Loneliness Scale (alpha = .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 in order 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, in preparation).

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 investigated the FIAT categories as inspiration for EMA items instead.

Sample 3. 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, 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). 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 experience sampling methodology (ESM) using a study-issued smartphone and answering questions regarding their mood, cognitions, and behaviours. 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 male; 75 female) 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, and 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 = .88) and positive affect (how optimistic, delighted, satisfied and grateful”: alpha - .87). Ratings were made on a 100 point scale (0;not at all; 100; very much)

The i-ARMAX 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 with the individual, rather than with aggregated estimates based on the group, and is only then examined for nomothetic patterns. This type of analysis makes no assumption 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 relationship and standard error 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 relationship between processes and outcomes utilising correlational or regression analysis. However, our daily diary, time series data was expected to violate the assumptions of these traditional analysis in at least two ways. First, time series are often not stationary, as when the mean and variance of the outcome changes. Second, the observations are often not independent, as earlier values often relate to later values (Chatfield & Xing, 2019).

To deal with these issues, we used an idiographic version of ARIMAX (Autoregressive Integrated Moving Average) to predict outcomes with an x exogenous process variable. The AR (autoregression) component of the model involves values being regressed on their own lagged values, the I (integrated) component involves differencing to remove trends, and the MA (moving average) component involves modelling the dependency between an observed value and residual error from a moving average model applied to previous observations. We used this analysis to establish the strength of

relationship between each process and each outcome, fitting a unique model for each individual (the “i” in **i-**ARIMAX stands for individual).

ARIMAX models use the standard notion of p (number of autoregressive terms), d (number of times that the raw observations are differences for stationarity), and q (number of lagged forecast errors in the prediction equation, or the size of the moving average) (Chatfield & Xing, 2019). High, positive AR estimates suggest getting "stuck", whereas negative ones suggest oscillation, and those close to zero can be considered to be variables that have the potential to be more highly influenced by other variables. High MA estimates suggest that there are sudden changes in the time series.

The simple way to think of ARIMAX is as a filter that seeks to separate the signal from the noise for one variable (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 relationship, after removing 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. 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 using unit root tests 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.

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. 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 sudden 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 idionomic statistical analysis can reveal differences in measure 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, suggesting individuals differed in the stability of their outcomes (e.g., of mood). The moving average statistics (bottom Table 1) suggests that people differed in the extent they experienced sudden changes in their outcome (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.

Table 1. Frequency percentage of P (order of autoregressive component), D (order of differencing component), Q (order of the moving average) identified by auto-arima in predicting negative outcomes across three different 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%

We argued in the introduction that ordinary regression assumptions are often violated with time series when idionomic analyses are applied to longitudinal data. This possibility can be examined empirically by comparing within-person regression results with i-ARIMAX results. If i-ARIMAX is a “filter for noise” that enters more traditional statistical analyses, then we should see less unexplained error in the iARIMAX analyses compared to multiple regression models. We expected the coefficients of regression and i-ArIMAX to differ but still be highly related because both procedures are using the same data to estimate the strength of the relationship (using absolute values to focus on magnitude). We conducted both an i-ARIMAX and regression analysis within every person for every process and outcome and then compared the results of the two approaches. As can be seen in Table

2, the coefficients between regression and I-ARIMAX coefficients were high, having between 76% to 86% of variance in common. The average magnitude of the coefficients were also similar, being slightly smaller for i-ARIMAX. Consistent with expectations, the level of error was smaller for i-ARIMAX compared to regression.

Table 2. A comparison of results using i-ARIMAX and standard regression in predicting positive and negative outcome

	Psy-Flex		**PBAT**		**FIAT-M**	
	Neg Af	**Pos aff**	**Neg Af**	**Pos aff**	**Lonely**	**Supp**
Strength of relationship: process & outcome						
Regression average B	.394	.461	.197	.210	.043	.249
ARIMAX average B	.376	.426	.190	.200	.042	.212
Correlation between regression & i-ARIMAX B	.932	.911	.902	.912	.894	.874
Level of Error						
Regression average SE	.183	.173	.120	.120	.121	.117
I-ARIMAX Average SE	.171	.156	.108	.112	.113	.110

Note: Strength of relationship is average of absolute value of relationships between processes and outcomes

Main analysis

We next submitted the within person coefficients to meta-analytic examination using the r package, *metafor*. The logic of this approach is that we can estimate pooled (average) effects, as well as the heterogeneity of these effects across individuals, by using widely accepted 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 traditional meta-analysis. Rough guidelines for interpreting I^2 in the meta-analytic literature are that

values less than 25% reflects low inconsistency, 25% to 50% reflects moderate inconsistency, 50 to 75% reflects high inconsistency, and over 75% shows very high inconsistency. 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. As can be seen in Table 3, most of the I^2 values in the present data sets are above .75, showing very high inconsistency. The Q^2 is the sum squared differences between studies and the overall mean, divided by degrees of freedom and is used to test for statistical significance of the heterogeneity (Huedo-Medina et al., 2006). 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 statistical methods) may not be appropriate. The right side of Table 3 presents the percentage of people with different magnitudes of beta.

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	< -.31	-.30- -.21	-.20 -.11	-.10- .10	.11- .20	.21- .30	> .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.

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 < -.31) but associated with higher *loneliness* for about 11% of people (beta > .31).

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	< -.31	-.30- -.21	-.20 -.11	-.10- .10	.11- .20	.21- .30	> .31
Opportunities for interpersonal action									**Loneliness**		
Assertive	0.03	0.04	73	139	5%	11%	5%	41%	22%	8%	8%
GiveRecFeedbak	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.

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> .31), but was either not linked to *feeling supported* or negatively linked to *feeling supported* for 17% of people (B < -.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 not a significant pooled effect: *Sticking to strategies* (negative affect only), *no outlet for feelings*, and *thinking got in the way* (positive affect only). For each of these processes, the "average effect of 0" poorly describes a substantial number of people. For example, the process "sticking to strategies that have worked" was associated with less negative affect for 14% of people (beta < -.31) but more negative affect for 7% of the people (beta > .31). Problematic thinking patterns were associated with lower life satisfaction for 23 % of people (Beta <-.31) but tended to have little effect or potentially a positive effect for 21% of people (Beta > .11).

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	< -.31	-.30- -.21	-.20 -.11	-.10- .10	.11- .20	.21- .30	> .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%

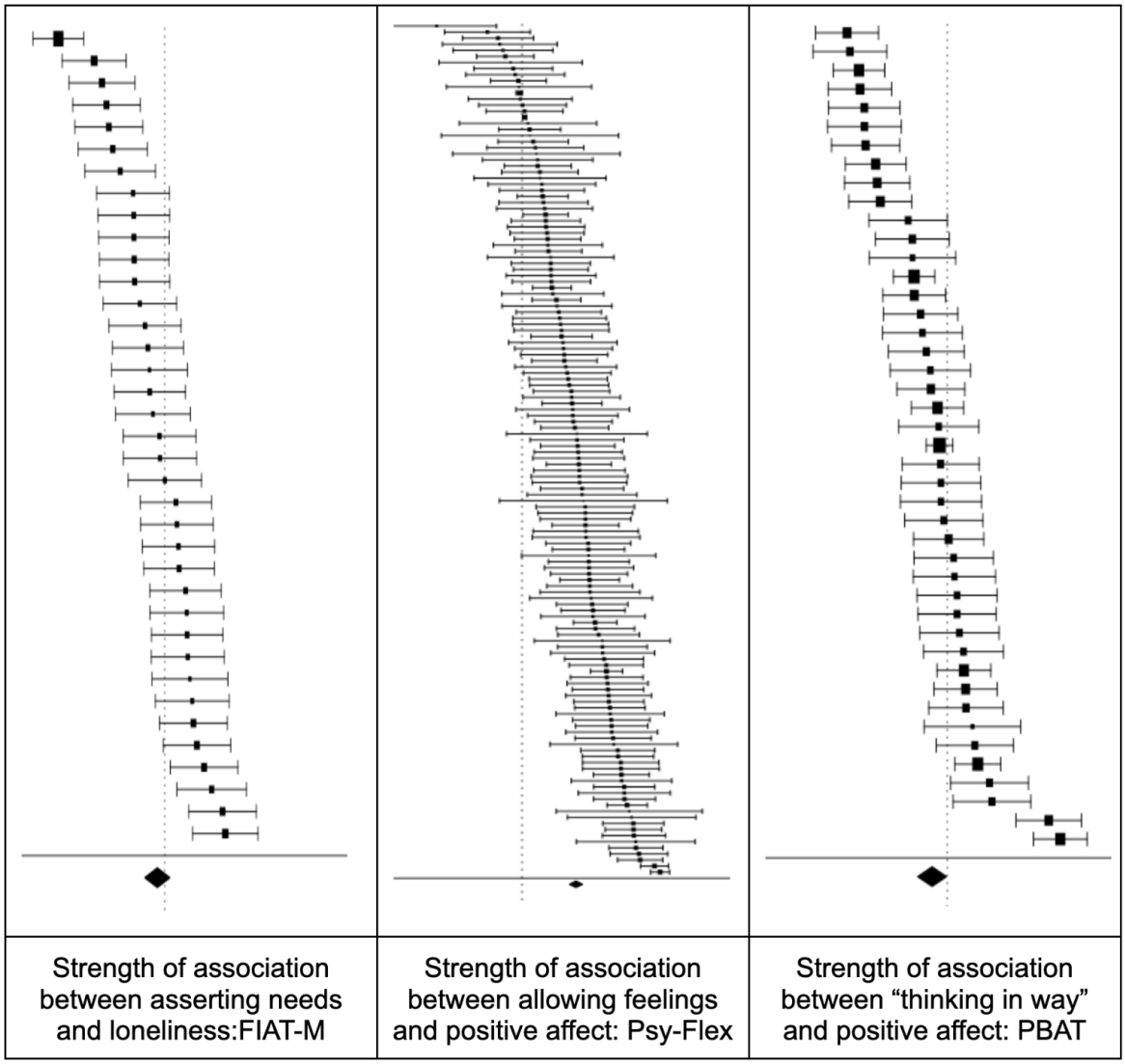
Note: * p< .05. All Q^2 tests of heterogeneity are highly significant, *p*<.0001.

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	< -.31	-.30- -.21	-.20 -.11	-.10- .10	.11- .20	.21- .30	> .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											
AbleToChangeBehavio	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%
StruggledToKeepDoing	-0.21*	0.04	85	319	36%	18%	5%	30%	7%	2%	2%

To help provide an intuition about the significant heterogeneity of effects, Figure 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 of the 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.

Figure 1: Strength of relationship for each person for items in the FIAT-M, Psy-Flex, and PBAT



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

The previous analysis shows that there are notable individual differences in the link between process and outcome for every process variable across all three datasets. Thus, the second aspect of the ergodic assumption, that the same dynamic model applies to all, 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, rather

than across individuals, in order to make good clinical decisions. Consider the simple forest plots of the within-person relationship for four individuals between negatively worded PBAT processes items and negative affect as is shown in Figure 2. 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-person were quite different, however. The item "*hurting health*" was significant for person 2 and 4, but not for person 1 and 3. "*Thinking got in the way*" seems to be 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 2: Strength of relationship between negative PBAT processes and negative affect

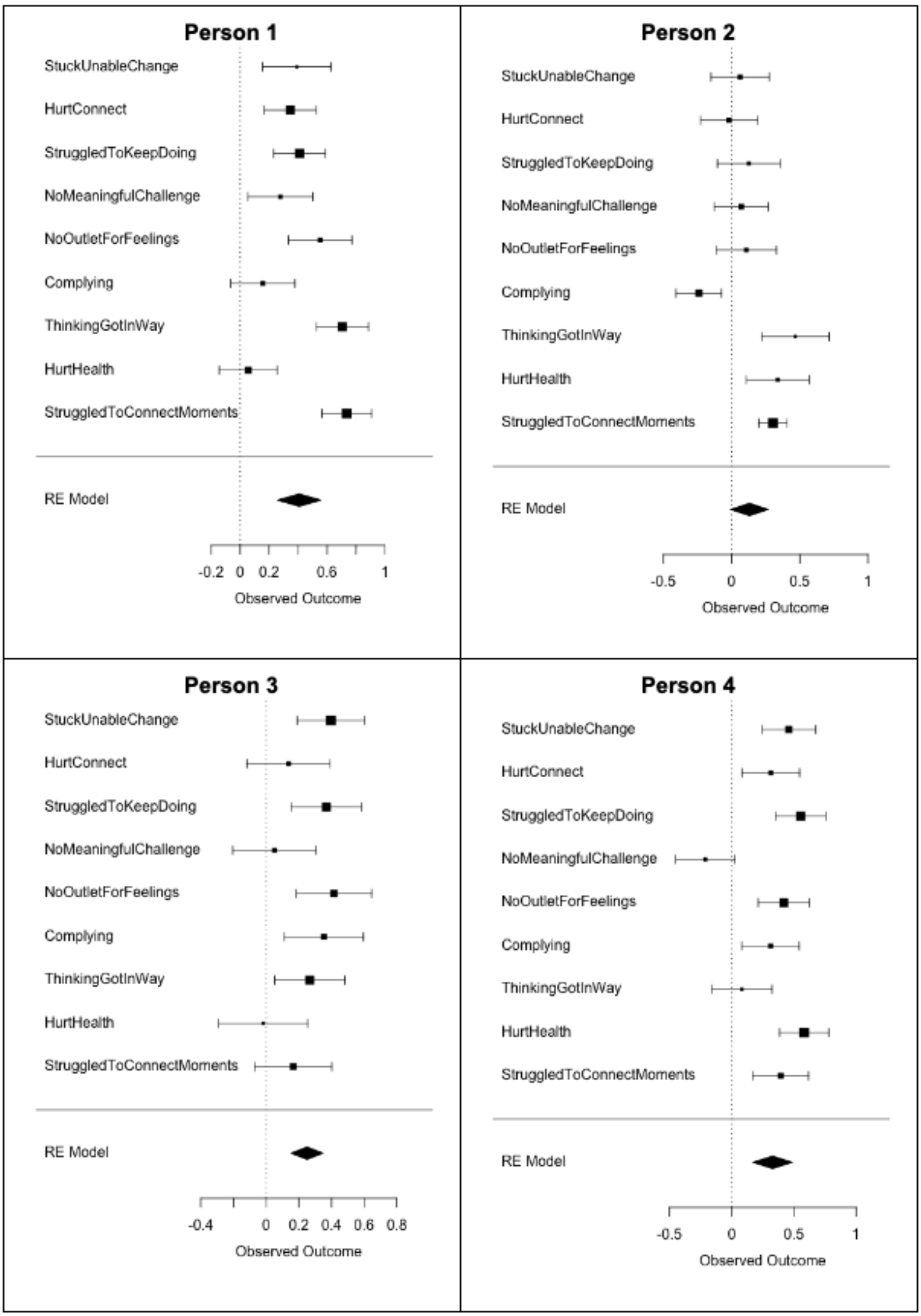
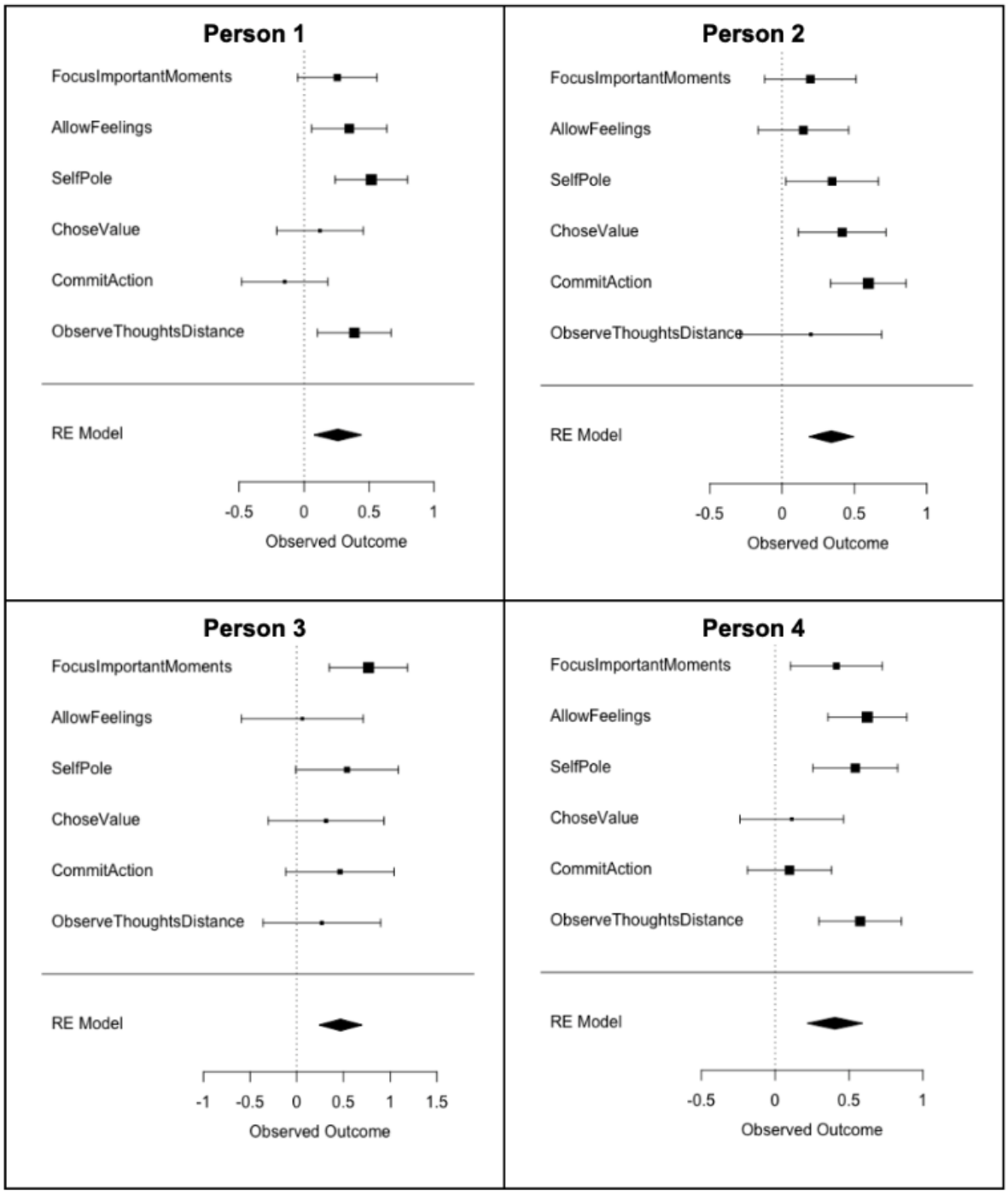

Person 1
StuckUnableChange
HurtConnect
StruggledToKeepDoing
NoMeaningfulChallenge
NoOutletForFeelings
Complying
ThinkingGotInWay
HurtHealth
StruggledToConnectMoments
RE Model
-0.2 0 0.2 0.6 1
Observed Outcome
Person 2
StuckUnableChange
HurtConnect
StruggledToKeepDoing
NoMeaningfulChallenge
NoOutletForFeelings
Complying
ThinkingGotInWay
HurtHealth
StruggledToConnectMoments
RE Model
-0.5 0 0.5 1
Observed Outcome
Person 3
StuckUnableChange
HurtConnect
StruggledToKeepDoing
NoMeaningfulChallenge
NoOutletForFeelings
Complying
ThinkingGotInWay
HurtHealth
StruggledToConnectMoments
RE Model
-0.4 0 0.2 0.4 0.6 0.8
Observed Outcome
Person 4
StuckUnableChange
HurtConnect
StruggledToKeepDoing
NoMeaningfulChallenge
NoOutletForFeelings
Complying
ThinkingGotInWay
HurtHealth
StruggledToConnectMoments
RE Model
-0.5 0 0.5 1
Observed Outcome

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 are different. 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, and 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.

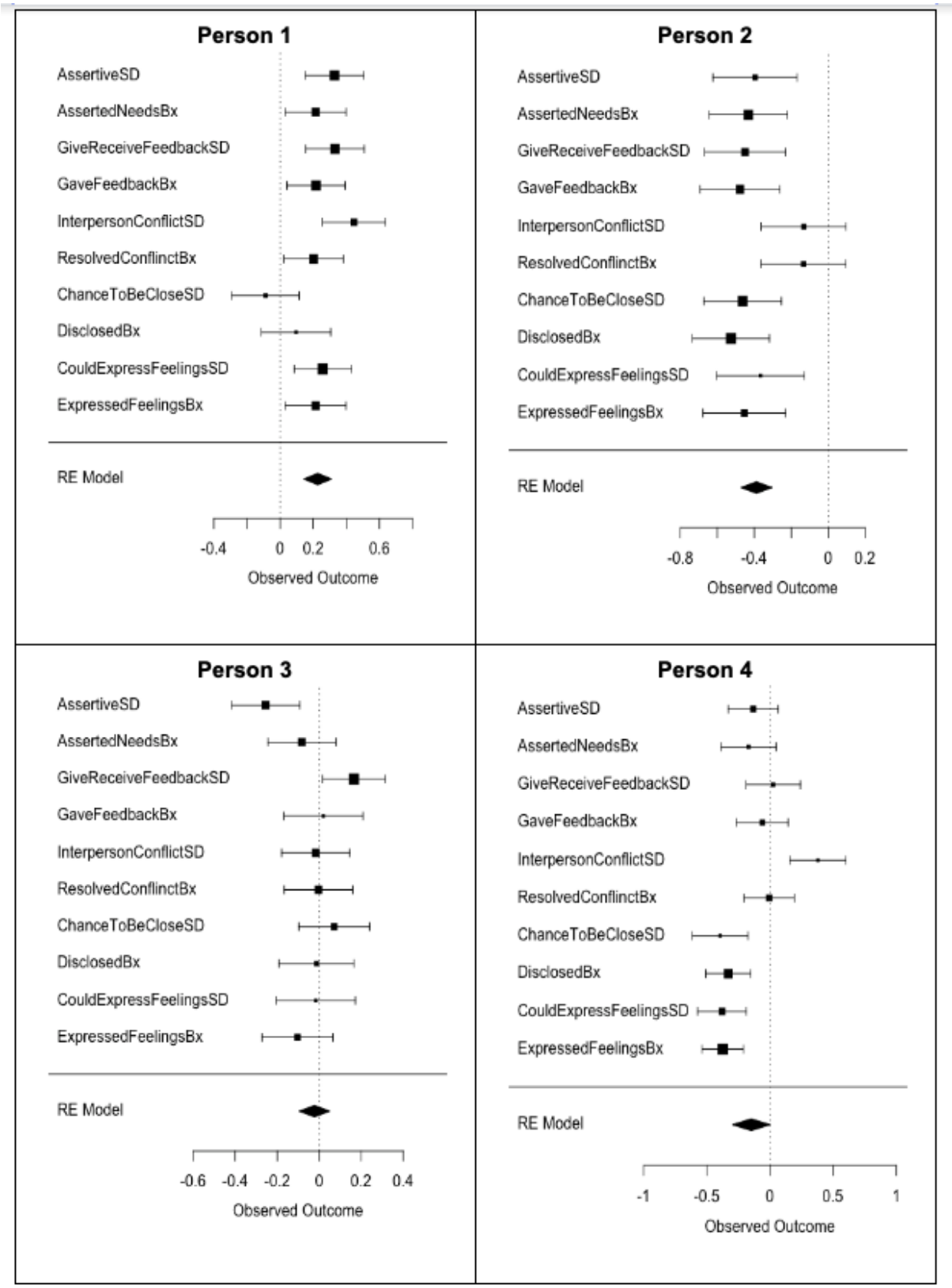
Figure 3: Strength of relationship between Psy-Flex processes and positive affect for four participants



Finally, Figure 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 the 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 actually *asserting oneself* is not associated with less *loneliness* (BX)

Figure 4: Strength of relationship between FIAT-M processes and loneliness for four participants



Note: SD indicates opportunities for action, Bx indicates taking of action.

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. I^2 was never below .61 and was typically above .75, suggesting that the strength of relationships between process and outcome differed substantially between people. Bornestein et al's (2009) conclusion about heterogeneity in meta-analysis appears to apply in the present instance: ".When there is a great deal of heterogeneity, it may not be appropriate to pool the studies. In such cases, it may be appropriate to report the results of the individual studies separately rather than trying to combine them. (p59)." In this paper, the "individual studies" were individual persons and these comments suggest that combining their results into an average makes little analytic sense.

We do not yet know if these three datasets are outliers or whether they are typical of psychological research but that question is now on the table. If the findings reported here are common, a key flaw in the logic of traditional analytic approaches to evidence-based therapy has been revealed. Violating ergodicity does not mean that classical statistical methods are fatally flawed for all purposes, but it appears to mean that normative findings cannot be reliably applied to the prediction and analysis of individual life trajectories, and thus that idionomic methods need to be added to such methods as randomized controlled trials, psychometric analyses, mediational analyses, and other traditional statistical analyses if their results will be applied even probabilistically to specific people. Most behavioral scientists appear to believe that nomothetic generalizations are the "signal" of psychological research should be applied to individuals because idiographic variability is "noise." As a statistical fact, the opposite may be true: Idiographic variability may be the key signal.

If so, then recognizing idiographic heterogeneity and the violation of 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 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 level of pathology but also 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 showed that the structure of externalising and internalising behaviour differs at the within compared to between-person level, and is person specific and highly idiographic (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 seem to drive well-being for another.

How do our results match up with theories that suggest certain processes should be of universal benefit? Consider a nomothetic theory that is arguably among the best supported in all of behavioral science, Self-Determination Theory (SDT; Bradshaw et al., 2021; Donald et al., 2020; Ryan & Deci, 2017). SDT suggests all people experience well-being when they get their needs for connection, competence, and autonomy met (Bradshaw et al., 2021; Donald et al., 2020; Ryan & Deci, 2017). Yet in the present study we showed, for example, that *hurting one's connection with others* is sometimes not associated with lower well-being, *complying* (e.g., low autonomy) is sometimes associated with higher, not lower, well-being, and *having a personally important challenge* (competence needs) is not always associated with well-being.

We would suggest that these results do not invalidate SDT theory. Rather, they open the door for a series of interesting questions about what moderates the link between processes and need satisfaction at the idiographic level. Whilst some processes may

generally increase one's chances of getting needs met, this won't be true for everybody, in every context, at every time. For example, people in an extremely supportive social environment may experience little fluctuations in their felt support, even when their social behaviour is sometimes unpleasant. Other people may feel they comply substantially in one unimportant domain (low autonomy), but do what they want when it counts (high autonomy). Some people may have a satisfying level of challenge in their life and fluctuations in day to day challenge may not affect their 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.

In a similar way, 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.

Applied implications

The i-ARIMAX method described here focuses on bi-variate relationships between a process and outcome and is likely to require less power than more complex multivariate analysis such as within-person structural equation modelling, network analyses, and factor analysis (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 modelling.

The results of the present study raises serious questions about 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 the FIAT-M 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 that have 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.

In a similar way, whilst we may see heterogeneity of idiographic 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 such 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 utilised here. They then examined the idiosyncratic structure of client mood and anxiety pathology and used this information to construct personalised treatment plans for each individual. There was no control group in the design, but the authors could compare the effects of their personalised design to effects observed in meta-analysis. The personalised design showed stronger effects. This encourages future research that compares personalised design based on intensive measures to standardised interventions. We hope i-ARIMAX can aid these designs.

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 behavioural and physiological data, such as those collected passively from wearables and smart phones, 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 known to link to well-being at the group level. The EEMM seems well suited to serving as a kind of meeting place, an intellectual Agora, for the marketplace of ideas that will be needed for such a comprehensive turn in the mainstream methodology of behavioral science research (Ciarrochi et al., 2021)

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. Or should we also prioritise processes with the client is typically below the average functioning (Crutzen & Peters, 2022).? 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 a lot of 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 personalising interventions. We as a scientific community are only starting the process.

Ultimately, we need to examine how more normative or "group" statistics can be used with idionomic 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 behaviours measured by FIAT and some behaviours 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, we as a society have invested billions of dollars in conducting thousands of trials on the efficacy of standardised treatment packages. Despite these efforts, effect sizes have not improved (Johnsen & Friborg, 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 personalised interventions can do better than standardised interventions. We do not believe another fifty years of assessing complex, standardised packages will lead to improvements.

References

Asher, S. R., & Paquette, J. A. (2003). Loneliness and peer relations in childhood. *Current directions in psychological science*, 12(3), 75-78.

Beltz, A. M., & Gates, K. M. (2017). Network Mapping with GIMME. *Multivariate Behavioral Research*, *52*(6), 789–804.

Benoy, C., Knitter, B., Schumann, I., Bader, K., Walter, M., & Gloster, A. T. (2019). Treatment sensitivity: Its importance in the measurement of psychological flexibility. *Journal of Contextual Behavioral Science*, *13*, 121–125.

Birkhoff, G. D. (1931). Proof of the Ergodic Theorem. *Proceedings of the National Academy of Sciences of the United States of America*, *17*(12), 656–660.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. John Wiley & Sons.

Bradshaw, E. L., Sahdra, B. K., Ciarrochi, J., Parker, P. D., Martos, T., & Ryan, R. M. (2021). A configural approach to aspirations: The social breadth of aspiration profiles predicts well-being over and above the intrinsic and extrinsic aspirations that comprise the profiles. *Journal of Personality and Social Psychology*, *120*(1), 226–256.

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.

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.

Chambless, D. L., & Ollendick, T. H. (2001). Empircally 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 Press, Taylor & Francis Group.

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.

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.

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.

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). John Wiley & Sons, Ltd.

Crutzen, R., & Peters, G.-. J. Y. (2022). A lean method for selecting determinants when developing behavior change interventions. In *psyarxiv.com*. https://doi.org/10.31234/osf.io/e467q

Crutzen, R., Peters, G.-J. Y., & Noijen, J. (2017). Using Confidence Interval-Based Estimation of Relevance to Select Social-Cognitive Determinants for Behavior Change Interventions. *Frontiers in Public Health*, *5*, 165.

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.

Donald, J. N., Bradshaw, E. L., Ryan, R. M., Basarkod, G., Ciarrochi, J., Duineveld, J. J., Guo, J., & Sahdra, B. K. (2020). Mindfulness and Its Association With Varied Types of Motivation: A Systematic Review and Meta-Analysis Using Self-Determination Theory. *Personality & Social Psychology Bulletin*, *46*(7), 1121–1138.

Donald, J. N., Ciarrochi, J., & Guo, J. (2022). Connected or Cutoff? A 4-Year Longitudinal Study of the Links Between Adolescents' Compulsive Internet Use and Social Support. *Personality & Social Psychology Bulletin*, 1461672221127802.

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.

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.

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.

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.

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.

Gloster, A. T., Block, V. J., Klotsche, J., Villanueva, J., Rinner, M. T. B., Benoy, C., Walter, M., Karekla, M., & Bader, K. (2021). Psy-Flex: A contextually sensitive measure of psychological flexibility. *Journal of Contextual Behavioral Science*, *22*, 13–23.

Gloster, A. T., & Haller, E. (2022). Meaningful and Lasting Change - Psychotherapy in the Light of Evolutionary Processes. *Clinical Psychology in Europe*, *4*(3), e9859.

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*.

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.

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*. https://josephciarrochi.com/wp-content/uploads/2022/06/An-Idionomic-Approach-Hayes-et-al-preprint-BRAT.pdf

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.

Hayes, S. C., Hofmann, S. G., & Ciarrochi, J. (2020a). 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.

Hayes, S. C., Hofmann, S. G., & Ciarrochi, J. (2020b). 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.

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.

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.

Hofmann, S. G., Curtiss, J. E., & Hayes, S. C. (2020). Beyond linear mediation: Toward a dynamic network approach to study treatment processes. *Clinical Psychology Review*, *76*, 101824.

Hopwood, C. J., Wright, A. G. C., & Bleidorn, W. (2022). Person–environment transactions differentiate personality and psychopathology. *Nature Reviews Psychology*, *1*(1), 55–63.

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.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, *27*, 1–22.

Ioannidis, J. P. A., Patsopoulos, N. A., & Evangelou, E. (2007). Uncertainty in heterogeneity estimates in meta-analyses. *BMJ* , *335*(7626), 914–916.

Johnsen, T. J., & Friborg, O. (2015). The effects of cognitive behavioral therapy as an anti-

depressive treatment is falling: A meta-analysis. *Psychological Bulletin*, *141*(4), 747–768.

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.

Kok, G., Gottlieb, N. H., Peters, G.-J. Y., Mullen, P. D., Parcel, G. S., Ruiter, R. A. C., Fernández, M. E., Markham, C., & Bartholomew, L. K. (2016). A taxonomy of behaviour change methods: an Intervention Mapping approach. *Health Psychology Review*, *10*(3), 297–312.

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

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 Friborg (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.

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.

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.

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.

Nau, R. (2020). *ARIMA models for time series forecasting*. Duke Forecasting Home Page. https://people.duke.edu/~rnau/411arim.htm

Neumann, J. V. (1932). Physical Applications of the Ergodic Hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, *18*(3), 263–266.

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.

Paul, G. L. (1969). Behavior modification research: Design and tactics. In C. M. Franks (Ed.), *Behavior therapy: Appraisal and status* (pp. 29–62).

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.

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.

Ryan, R. M., & Deci, E. L. (2017). *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. Guilford Publications.

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.

Sahdra, B. K., Ciarrochi, J., Fraser, M. I., Yap, K., Haller, E., Hayes, S. C., Hofmann, S., & Gloster, A. T. (2022). *There is no self without the other: An experience sampling study of the interconnectedness of self-and other-compassion and its link to well-being*. https://psyarxiv.com/8zw4d/download/?format=pdf

Sanford, B. T., Ciarrochi, J., Hofmann, S. G., Chin, F., Gates, K. M., & Hayes, S. C. (2022). Toward empirical process-based case conceptualization: An idionomic network

examination of the process-based assessment tool. *Journal of Contextual Behavioral Science*, *25*, 10–25.

Stanton, C.E., Follette, W.C., Hayes, S.C., Sandford, B.T., Ciarrochi, J. (in preparation). *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.

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, *36*(3), 1–48.

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.

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.

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.

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.

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.

Wright, A. G. C., & Woods, W. C. (2020). Personalized Models of Psychopathology. *Annual Review of Clinical Psychology*, *16*, 49–74.

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.

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.