Behaviour Research and Therapy 118 (2019) 43–53

Contents lists available at ScienceDirect

Behaviour Research and Therapy

journal homepage: www.elsevier.com/locate/brat

An idiographic statistical approach to clinical hypothesis testing for routine psychotherapy: A case study

Casey L. Brown[a,*], Hannah G. Bosley[a], Alan D. Kenyon[b], Kuan-Hua Chen[a], Robert W. Levenson[a]

[a] Department of Psychology, University of California, Berkeley, CA, USA
[b] Independent Engineer, Berkeley, CA, USA

ARTICLE INFO

Keywords:
Cognitive behavior therapy
Person specific
Single case
Case-formulation

ABSTRACT

In order to develop more targeted, efficient, and effective psychotherapeutic interventions, calls have been made in the literature for greater use of idiographic hypothesis testing. Idiographic analyses can provide useful information regarding mechanisms of change within individuals over time during treatment. However, it remains unclear how clinicians might utilize idiographic statistical analyses during routine treatment to test clinical hypotheses, and in turn, guide treatment. We present an idiographic statistical framework for clinical hypothesis testing with routine treatment data that enables clinicians to examine a) whether the client's symptoms and hypothesized mechanisms change over time, b) whether trajectories of change reflect the timing of interventions, c) whether mechanisms predict subsequent symptoms, and d) whether relationships exist between *multiple* mechanisms, symptoms, or other treatment-related constructs over time. We demonstrate the utility of the approach for clinical hypothesis testing by applying it to routine treatment data collected from a 56 year-old male who presented with a combination of anger problems, anxiety, and depressive symptoms. We discuss how results from analyses can inform the case-formulation and guide clinical decision-making. We aim to make these methods more accessible by providing an online platform where clinicians can enter client data, test their clinical hypotheses using idiographic analyses, and utilize the results to disseminate their findings.

1. Introduction

Cognitive Behavioral Therapy (CBT) and other empirically supported treatments (ESTs) remain ineffective for many individuals (Hofmann, Asnaani, Vonk, Sawyer, & Fang, 2012). Researchers have argued that a greater emphasis on idiographic analysis of clinical data (i.e., analyzing change processes within individuals over time) is necessary for more effective interventions (Barlow & Nock, 2009; Molenaar, 2004). Idiographic analyses can shed light on the psychological mechanisms that promote symptom change during treatment, improving our understanding of how, when, and for whom change occurs during psychotherapy (Boswell, Anderson, & Barlow, 2014; Boswell & Bugatti, 2016). As a result, idiographic analyses are becoming more common in research (e.g., Fisher, Reeves, Lawyer, Medaglia, & Rubel, 2017; Wright et al., 2016). Idiographic analyses may also prove useful in clinical practice settings to improve clinical hypothesis testing and decision-making. In fact, many clinicians have begun to rely on idiographic hypothesis testing, using a case-formulation approach to improve clinical decisions (Persons & Hong, 2016).

In the case-formulation approach to CBT, a patient's symptoms and the potential mechanisms underlying those symptoms are assessed in order to develop a personalized formulation of the client's problems (Persons & Tompkins, 1997). For example, a clinician may develop a formulation that repetitive negative thoughts and poor sleep hygiene are two mechanisms underlying a client's depressive symptoms. A clinician then selects interventions from a variety of ESTs to target the hypothesized mechanisms in the formulation in order to reduce symptoms (Persons, 2012). Changes in mechanisms and symptoms are monitored as treatment progresses, and the formulation and treatment are adjusted as necessary based on data from the patient (Persons, 2012; Persons, Brown, & Diamond, 2019).

Because the case formulation approach to CBT and other ESTs encourage the use of client's self-report data to test clinical hypotheses and guide treatment, clinicians are increasingly collecting quantitative data from their clients to make inferences about change processes by visually inspecting client data. In addition to visual inspection, other useful methods have been developed to provide therapists with feedback on treatment effectiveness (e.g., Lutz, Zimmermann, Müller, Deisenhofer, & Rubel, 2017; Lyon, Lewis, Boyd, Hendrix, & Liu, 2016; Shimokawa, Lambert, & Smart, 2010). However, these methods typically do not

* Corresponding author. UC Berkeley Psychology Clinic, Department of Psychology, University of California, 2121 Berkeley Way, Berkeley, CA, 94704, USA.
E-mail address: CaseyLBrown@Berkeley.edu (C.L. Brown).

https://doi.org/10.1016/j.brat.2019.03.014
Received 26 April 2018; Received in revised form 7 March 2019; Accepted 19 March 2019
Available online 27 March 2019
0005-7967/

offer therapists optimal flexibility in measure selection (Lyon et al., 2016). For example, therapists must use a specific measure or a set of measures that are not personalized to their client's unique problems. More importantly, existing approaches often do not answer questions of clinical relevance beyond the basic issue of, "Is treatment working for this patient?"

In addition to understanding whether symptoms and hypothesized mechanisms are changing during treatment, it is also useful to know *how* they're changing. Therapists using a case-formulation approach typically want to know whether they're targeting the right mechanisms to reduce symptoms. They also want to understand the relationships between mechanisms, symptoms, and other treatment constructs in order to tailor the formulation, prioritize treatment targets, and improve their clinical decisions. Researchers have used impressive idiographic methods to examine these kinds of clinical questions in research settings (e.g., Fisher, 2015). However, such methods have not been applied to routine treatment data, despite the valuable information they could provide. Even simpler idiographic approaches (e.g., person-specific regression) that can tackle important questions of clinical interest are also not typically utilized by clinicians.

There are many possible reasons practicing clinicians do not take advantage of idiographic statistics to test clinical hypotheses. Clinicians may believe that idiographic statistical procedures require treatments to be delayed or halted for the purpose of data collection. They may perceive analyses to engender atypical time demands for themselves (e.g., learning to write code in modern statistical software) or for their clients (e.g., filling out questionnaires multiple times per day). Moreover, clinicians may assume that statistics cannot be applied to "messy" clinical data, which may include sporadic missing data and measures added partway through treatment. As a result, potentially valuable data already collected during treatment remain underutilized. There is a need for systematic approaches and tools that enable clinicians to readily incorporate existing idiographic statistics into routine clinical practice.

1.1. The current study

In the current study, we describe and apply an idiographic statistical approach for clinical hypothesis testing in routine treatment to assess a) whether symptoms and hypothesized mechanisms change over time (using within-person linear regression analyses), b) whether trajectories of change reflect the timing of interventions (adding quadratic time parameters to within-person linear regression models), c) whether hypothesized mechanisms predict subsequent symptom levels (using within-person time-lagged regressions), and d) whether relationships exist between *multiple* mechanisms, symptoms, or other treatment-related constructs over time (using p-technique exploratory and confirmatory factor analyses and dynamic factor modeling). In order to assess the utility of the approach for clinical hypothesis testing with routine treatment data, we apply the analyses to self-report data collected from a single client treated with the case formulation approach to CBT. We discuss how findings can inform case-formulation and guide clinical decision-making. Finally, we present a web-based platform that enables clinicians to readily apply idiographic analyses to routine treatment data.

2. Method

2.1. The approach: idiographic statistical analyses for routine treatment data

To test clinical hypotheses using idiographic statistics, the case-formulation therapist must start, as usual, by developing a case-formulation. Next, the therapist selects and implements interventions hypothesized to be effective for the client's specific case-formulation. The therapist monitors the symptoms and mechanisms hypothesized in the client's formulation over the course of multiple sessions using quantitative measures the clinician prefers (for more detailed descriptions of the case-formulation approach to therapy, see Frank & Davidson, 2014; Persons, 2012). With these kinds of quantitative time-series data, the therapist can begin testing hypotheses using the idiographic statistical approach described below. The statistical methods we describe may seem daunting to therapists who lack familiarity with statistics, thus, we also provide more simplified explanations in the web-based platform where analyses can be conducted easily.

Examining linear changes in symptoms and mechanisms. Typically the therapist's first question is whether a patient's symptoms and hypothesized mechanisms are changing over time. Linear changes in patient's symptoms and mechanisms can be examined using person-specific ordinary least squares regression. Regression models can test the relationship between time (coded in days; as an independent variable) and symptoms or mechanisms (dependent variables). As with any regression, p-values can be examined to assess significance. However, clinicians should pay special attention to effect sizes for each variable because in routine treatment clinicians likely have fewer time-points of data, resulting in lower power to detect significant effects. For example, Cohen's D can be computed as a measure of effect size, where $D = \mathrm{t} * \sqrt{2/n}$, with n referring to the number of observations of the symptom or mechanism of interest.

Examining quadratic changes in symptoms and mechanisms. Symptoms and mechanisms may change at different rates during treatment leading to different trajectories of change over time (Stulz, Lutz, Leach, Lucock, & Barkham, 2007). Examining trajectories of change can reveal whether symptoms and hypothesized mechanisms are changing more rapidly in certain phases of treatment, which may help clinicians link patient change to particular aspects of treatment. Clinicians can assess the trajectory of change for a mechanism or symptom by adding a quadratic time parameter (reflecting curvilinear change of a symptom or mechanism over time) to the existing linear model. Clinicians can then determine whether the linear or quadratic effect of time provides the best model fit. If the quadratic parameter results in significant decreases in the deviance statistic (indicative of good model fit) relative to the linear model alone, the quadratic parameter is retained and the shape of change for that variable is considered to be curvilinear.

When change is curvilinear, the particular shape of the curve has implications for the rate of change during different phases of psychotherapy. The sign of the beta coefficient of the quadratic time parameter indicates the shape of the curve. A positive coefficient indicates a convex curve, whereas a negative coefficient indicates a concave curve. When linear change is significant and the quadratic parameter improves model fit, a positive coefficient likely indicates quick change at first that tapered off in later sessions. In contrast, a negative coefficient likely indicates slow change at first that became more pronounced later in treatment. The shape of the curve may alert the therapist to other important details: for example, if curvilinear change is found in the absence of linear change, one possibility is that the client is regressing (e.g., symptoms lowered initially, but then began to rise). Such findings could encourage the therapist to develop formulation hypotheses regarding why the client is unable to maintain change.

Examining whether mechanisms predict subsequent symptoms. Clinicians often wonder about the accuracy of their case-formulations and whether they are targeting the right mechanisms during treatment in order to reduce symptoms. Time-lagged linear regressions can assess the effect that a mechanism has on a symptom across time. Clinicians can examine whether a mechanism at Time 1 predicts a symptom at Time 2 when controlling for the effect that the symptom at Time1 has on itself at Time 2 (i.e., controlling for the autoregression of that symptom, often referred to as "Granger causality"; Granger, 1969). While current thinking rejects regression-based methods of causality testing (c.f. Sekhon, 2009), clinicians can interpret these data with caution as a preliminary step toward understanding the direction and strength of time-lagged connections from mechanisms to symptoms.

To conduct such an analysis, symptom and mechanism variables are reproduced and "lagged" by 1 observation. Separate regression models are then constructed for each mechanism in which time-forward symptoms are modeled as the dependent variable, with the time-lagged mechanism variable and the time-lagged symptom as predictors. Missing data can be excluded as a function of listwise deletion. To control for uneven spacing of observations and missingness that can occur with routine treatment data, an additional variable can be created for each mechanism that represents the elapsed time between each observation and the one preceding it. For each model, the elapsed-time variable can be modeled as an interaction term with the time-lagged predictor variable and time-lagged symptom. If the interaction effect is not significant, it can be removed and elapsed time is modeled as a covariate. This procedure helps to address the problem of missing data because missingness is treated as a variant of uneven sampling and regressed out in the interaction term (consistent with Clasen, Fisher, & Beevers, 2015).

Examining associations between multiple mechanisms, symptoms, or other treatment related constructs over time. Therapists may also want to understand the nuanced relationships between *multiple* mechanisms, symptoms, or other treatment constructs over time in order to further tailor the formulation and prioritize treatment targets. Structural equation modeling can be used to examine the structural and temporal dynamics of multiple treatment constructs, consistent with methods used by Fisher (2015). First, time-series data can be subjected to exploratory factor analysis (EFA) to determine the latent structure of the data (i.e., how symptoms, mechanisms, or other treatment constructs cluster together within the individual across time). An EFA can be conducted using the Psych package (Revelle, 2013) in R Version 3.2.1, using maximum likelihood estimation. Consistent with Tabachnick and Fidell (2007) we suggest starting with oblique (oblimin) rotation and then assessing whether factors are correlated rather than orthogonal (correlations > ∼0.32 are said to warrant oblique rotation). To determine the number of factors, an iterative approach can be used. First, a one-factor model is assessed, then a two-factor model, and factors continue to be added until the standardized root mean square residual (SRMR) falls below 0.08 indicating acceptable model fit (Lo, Molenaar, & Rovine, 2017).

Next, a confirmatory factor analysis (CFA) can be run using LISREL Version 8.8 to assess whether the factor structure from the EFA provides a good data fit. A pattern matrix is created to represent the factor structure indicated by the EFA as an array of 0s and 1s. An a priori decision rule is used to omit factor loadings that fall below a certain threshold (we specify a threshold of 0.30, but this threshold can be adjusted). Non-significant factor loadings are omitted in an iterative fashion (smallest to largest) and cross-loadings (as revealed by modification indices) are iteratively added in order from largest to smallest. The final model is selected when fit index values indicate good fit (with recommended cutoff values near 0.95 for CFI and TFI, 0.06 for RMSEA, and 0.08 for SRMR; Hu & Bentler, 1999). The raw time-series data are then multiplied by a weighting matrix generated from the factor loadings of the confirmatory factor solution using the "components" option of the factor scores function in the Psych package resulting in a factor-scored time series for each factor.

Then another set of structural models using a lag-1 vector autoregressive framework are run to assess how the resulting factors relate to one another within and across time. This method, called dynamic factor modeling (DFM; Molenaar, 1985), allows clinicians to assess contemporaneous and time-lagged relationships between factors. First, a linear de-trending procedure is used in which factor scores are regressed on a linear time parameter, and the residuals of that model (with the linear trend regressed out) are used in subsequent analyses. Each de-trended factor-scored time series is duplicated and lagged by one observation, resulting in one time-lagged and one time-forward time-series for each factor. Setting the factor loading matrix to "identity" and the measurement error matrix to "0", an initial model can be run including only contemporaneous correlations and autoregressions. The Lagrange multiplier test can be used iteratively to identify cross-lagged relationships between factors. We use a cutoff threshold of 4 for the modification index because the minimum chi-square value for significance with 1 degree of freedom is 3.84 (conservatively, we rounded up to 4). Thus, modification indices lower than 4 indicate that the addition of that parameter would not result in statistically significant improvement to model fit (Jöreskog, 1993; Lei & Wu, 2007).

2.2. Applying the approach to routine treatment data: participant

"Arnold", a 56-year-old divorced white heterosexual male, working at a low-income job, was referred to the graduate training clinic at the University of California, Berkeley because of its affordable sliding scale fee. In an initial phone screening, the client's chief complaint was high levels of anger. The client was previously fired because of his angry outbursts. He had experienced homelessness, and had lived in his car for significant periods of time. He feared losing control of his anger, losing his current job, and becoming homeless again. The client also described experiencing symptoms of anxiety and depression. As is typical with treatment in the Berkeley training clinic, no attempt was made to have the patient undergo a structured diagnostic interview. The client was not receiving any other adjunctive treatment or medication. He reported that he had received 12 sessions of weekly individual psychotherapy five years prior that he found "largely unhelpful." The client consented to the use of his clinical record for research purposes, and procedures for examining clinical records were approved by Berkeley's Committee for the Protection of Human Subjects.

2.3. Procedure

Case formulation. During the initial phone screen and sessions, the therapist worked collaboratively with Arnold to develop a case formulation (See Fig. 1) based on data collected through clinical interview and scores on standardized questionnaires (See Measures section). Several primary symptoms were identified, including excessive *anger*, symptoms of *anxiety*, and *depressive symptoms*. The therapist hypothesized several mechanisms thought to bring about and maintain these symptoms. The client's obsessive beliefs and perfectionistic tendencies were likely contributing to his symptoms. For example, the therapist hypothesized that the client's intense desire to control his thoughts and angry feelings (*importance of controlling thoughts*) promoted his excessive anger and his likelihood of future angry outbursts. The client's perfectionism also seemed to promote anger. He had high standards for

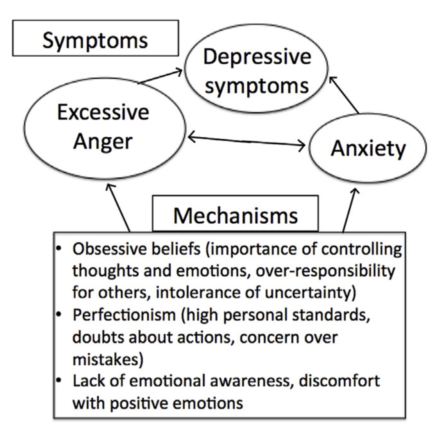

Fig. 1. The case formulation used to guide treatment, depicts relationships between symptoms and hypothesized mechanisms.

himself and others. For example, he frequently became frustrated when he or others were inefficient or made mistakes in the workplace (*high personal standards*). He felt a strong sense of responsibility to avoid errors, because in his industry, mistakes could result in physical injury (*high responsibility and threat estimation*). As a result, when he made a minor mistake at work, he would jump to extreme conclusions that he might lose his job (*high concern over mistakes*). Understandably, the client was extremely concerned when he lost control and inappropriately expressed his anger (e.g., telling off his boss). However, when the client managed his anger appropriately (e.g., politely explaining to a person over the phone that they had dialed the wrong department), he frequently doubted whether he had expressed his emotions appropriately (*doubts about actions*). Arnold struggled to deal with the uncertainty surrounding whether he had expressed inappropriate anger, and would ruminate over these incidents, wanting certainty that he handled every bit of anger perfectly (*perfectionism and intolerance of uncertainty*). The client did not seem to notice or differentiate low versus high intensity emotions (*lack of emotional awareness*), and the client reported discomfort with positive emotions, suppressing or avoiding positive emotions when he noticed them. Arnold was in an on-again off-again romantic relationship and had tremendous difficulty expressing his personal desires and feelings of love and affection in his relationship. These emotional difficulties were hypothesized to interfere with his relationships and promote anger (Levenson et al., 2017). Arnold agreed with all of the aforementioned aspects of the formulation.

In addition, Arnold described scenarios where he would behave in a condescending or insubordinate manner towards superiors (e.g., pointing out his boss's shortcomings in front of his department). Arnold described these scenarios as if they were accomplishments, where he asserted his intelligence. Arnold did not recognize these behaviors as problematic, and they were not included in the formulation. Nonetheless, the therapist worked to help the client weigh the consequences of such behavior.

Data collection. Data were collected during treatment to develop a case-formulation and monitor progress (See Table 1 for data summary). Arnold completed a small battery of standardized questionnaires on days when he arrived early to the clinic before treatment (usually 10 min). Occasionally, the client arrived later, and the therapy session was prioritized over self-report data collection. He also provided data on his emotions as a part of treatment homework assignments. Measures were selected to conform to the client's specific case-formulation (described below) and were selected by the therapist (the first author) and her clinical supervisor.

2.4. Symptom measures

Anger. Beginning on the 10th session, as a homework assignment, the client completed retrospective daily ratings of his anger levels on a scale of 1–5 with higher scores indicating more anger. Anger ratings were averaged each week following the day of his weekly therapy sessions. Weekly therapy appointment served as a natural point of division between weeks, and the rating he gave the night following his therapy session was applied to the following week.

Depressive symptoms. The Beck Depression Inventory-II (BDI; Dozois, Dobson, & Ahnberg, 1998) was used to measure symptoms of depression (e.g., "I'm so sad or unhappy that I can't stand it"), with higher scores indicating greater depressive symptoms. The client's initial BDI was 30 and his last score was 0.

Anxiety. The Beck Anxiety Inventory (BAI; Steer & Beck, 1997) was used to measure symptoms of anxiety (e.g., "unable to relax") with higher scores indicating greater anxiety symptoms. The client's initial BAI score was 13 and his last score was 0.

Table 1
Data types and structure.

Measure	Scale	Observations	Session number 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Symptom Measures																														
Anger	Daily homework	19										✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Depressive symptoms	BDI-II	16	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓						✓		✓			
Anxiety	BAI	15	✓		✓	✓		✓		✓	✓	✓	✓	✓	✓	✓		✓	✓						✓		✓			
Mechanism Measures																														
Obsessive Beliefs	OBQ-44 Total	12				✓						✓	✓		✓		✓		✓	✓			✓		✓		✓		✓	✓
Importance/control of thoughts	OBQ-44 subscale	12				✓						✓	✓		✓		✓		✓	✓			✓		✓		✓		✓	✓
Perfectionism/intolerance of uncert.	OBQ-44 subscale	12				✓						✓	✓		✓		✓		✓	✓			✓		✓		✓		✓	✓
Responsibility/threat estimation	OBQ-44 subscale	12				✓						✓	✓		✓		✓		✓	✓			✓		✓		✓		✓	✓
Doubts about actions	FMPS subscale	12		✓								✓	✓		✓		✓		✓	✓			✓		✓		✓		✓	✓
Concern over mistakes	FMPS subscale	12		✓								✓	✓		✓		✓		✓	✓			✓		✓		✓		✓	✓
Personal standards	FMPS subscale	12		✓								✓	✓		✓		✓		✓	✓			✓		✓		✓		✓	✓
Emotional Awareness	TAS-20	9		✓								✓			✓		✓						✓		✓		✓		✓	✓
Emotion Log Data	Daily homework	67																			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note. Check marks indicate that Arnold completed the measure in the corresponding session. The complete raw data can be accessed at www.changestat.org.

Table 2
Results for person specific linear and quadratic regressions.

Linear Models

Variable	β	p	t	Cohen's D	DF
Anger	−0.87	.016	−2.67	−0.87	17
Depressive symptoms (BDI)	−0.74	.006	−3.21	−1.13	14
Anxiety symptoms (BAI)	−0.73	.011	−2.96	−1.08	13
Total obsessive beliefs (OBQ-44)	−0.95	.001	−4.57	−1.87	10
Importance/Control of Thoughts (OBQ-44)	−0.92	.002	−4.2	−1.71	10
Perfectionism/Intolerance of Uncertainty (OBQ-44)	−0.94	.001	−4.4	−1.80	10
Responsibility/Threat Estimation (OBQ-44)	−0.86	.005	−3.57	−1.46	10
Concern over Mistakes (FMPS)	−0.87	.003	−3.99	−1.63	10
Personal Standards (FMPS)	−0.01	.976	−0.03	−0.01	10
Doubts about Actions (FMPS)	−0.48	.162	−1.51	−0.62	10
Emotional Awareness (TAS-20)	−0.81	.005	−4.05	−1.91	7

Quadratic Models

Variable	Parameter	β	p	t	Cohen's D	DF
Depressive symptoms (BDI)	Linear time	−2.89	< .001	−5.73	−2.03	13
	Quadratic time	2.55	< .001	4.47	1.58	13
Anxiety symptoms (BAI)	Linear time	−2.26	.008	−3.17	−1.16	12
	Quadratic time	1.79	.044	2.26	0.83	12
Total obsessive beliefs (OBQ-44)	Linear time	−3.41	< .001	−6.21	−2.54	9
	Quadratic time	2.18	.001	4.59	1.87	9
Importance/Control of Thoughts (OBQ-44)	Linear time	−3.04	.004	−3.89	−1.59	9
	Quadratic time	1.87	.021	2.79	1.14	9
Perfectionism/Intolerance of Uncertainty (OBQ-44)	Linear time	−3.41	< .001	−5.78	−2.36	9
	Quadratic time	2.18	.002	4.29	1.75	9
Responsibility/Threat Estimation (OBQ-44)	Linear time	−3.49	.001	−4.66	−1.90	9
	Quadratic time	2.32	.005	3.59	1.47	9
Personal Standards (FMPS)	Linear time	−2.41	.078	−1.99	−0.81	9
	Quadratic time	2.32	.071	2.04	0.83	9
Emotional Awareness (TAS-20)	Linear time	0.84	.221	1.37	0.65	6
	Quadratic time	−1.55	.033	−2.77	−1.31	6

2.5. Mechanism measures

Obsessive beliefs. The Obsessive Beliefs Questionnaire (*OBQ-44;* Bhar et al., 2005) measures beliefs and appraisals involved in the development of obsessions. We examined the OBQ-44 total score and three subscales including a) *importance and control of thoughts* (e.g., "Having intrusive thoughts means I'm out of control") with higher scores indicating higher distress over unwanted thoughts and the need to control those thoughts; b) *perfectionism and intolerance of uncertainty* (e.g., "For me, things are not right if they are not perfect), with higher scores indicating higher rigidity and distress over feeling uncertain); and c) *responsibility and threat estimation* (e.g., Avoiding serious problems … requires constant effort on my part"), with higher scores indicating greater desire to prevent harm and responsibility for bad things that happen. His initial OBQ-44 total score was 197 and his last score was 106.

Perfectionism. The Frost Multidimensional Perfectionism Scale (FMPS; Stober·, 1998) assesses aspects of perfectionism. We examined three subscales a) *doubts about actions* (e.g., I usually have doubts about the simple everyday things I do") with higher scores indicating more doubt; b) *concern over* mistakes (e.g., "If I do not do well all the time, people will not respect me") with higher scores indicating greater concern; and c) *personal standards (*e.g., "I set higher goals than most people") with higher scores indicating higher standards. We did not examine subscales (e.g., parental criticism) unrelated to the formulation for this client. His initial and last scores on each subscale were as follows: doubts about actions 7 and 4, concern over mistakes 36 and 12, and personal standards 33 and 29.

Emotional awareness. The Toronto Alexithymia Scale (TAS-20; Bagby, Parker, & Taylor, 1994) measures overall dysfunctions in emotional awareness (e.g., "I often don't know why I'm angry"). Higher scores indicate lower emotional awareness. His initial score was 50 and his last score was 28.

Emotion log data. Beginning on the 19th session as a homework assignment, the client began retrospective daily ratings of a variety of emotions on a scale of 1–5 (5 being more intense) including: happy, interested, excited, caring, affection, love, loved, compassion, grateful, proud, confident, hurt, sad, envious, jealous, afraid, regret, irritated, angry, resentment, disgust, contempt, ashamed, guilty, and anxious. This resulted in 67 daily surveys in which the client rated his experience of 25 emotions.

2.6. Treatment

In total, the client underwent 29 sessions of treatment (50 min durations) using the case-formulation approach to CBT. Arnold and the therapist collaboratively developed three main treatment goals: 1) reduce anger and the fear of losing control of anger, 2) reduce symptoms of depression and anxiety, and 3) increase positive emotions and comfort expressing positive feelings in relationships. In the initial phase of treatment, the therapist focused on the first goal of reducing anger. During this phase of treatment, the client began to monitor his anger as homework, each day rating his anger on a 1–5 scale in an anger log. He was asked to write down scenarios causing anger each day so he could discuss them later in therapy. To reduce anger, the therapist targeted the client's obsessive beliefs and perfectionism, using thought records, socratic dialogue, behavioral chains, pros and cons lists, and behavioral experiments.

The therapist noticed that depressive and anxiety symptoms seemed to decline quickly in the early part of treatment. When the therapist believed (based on her and her supervisor's clinical judgments and after visually inspecting the data) that obsessive beliefs and some aspects of perfectionism had improved, the therapist shifted treatment priorities. In this second part of treatment, the therapist prioritized the third goal of increasing positive emotions, but continued to work toward reducing personal standards and anger. The therapist primarily targeted the mechanism of emotional awareness. Starting at session 19, the therapist asked Arnold to monitor a variety of emotions on a daily emotion log as a homework assignment to increase his emotional awareness and comfort with positive emotions. The therapist used behavioral experiments and thought records to address Arnold's maladaptive emotion beliefs (e.g., expressing love will backfire). Treatment ended when the UC Berkeley training clinic closed for the summer, and the client was provided referrals for continued therapy elsewhere. The clinic administered an end of treatment feedback form where the client wrote, "Without [treatment] I would not have made significant improvement in how I handle anger."

3. Results

3.1. Linear changes in symptoms and mechanisms

For Arnold, a linear regression model was run for each symptom measure (i.e., BDI, BAI, anger) and each mechanism (i.e., OBQ-44 total, OBQ-44 subscales, FMPS subscales, TAS-20). Each model tested the relationship between time (coded in days; independent variable) and changes in one symptom or mechanism (dependent variable). Table 2 presents results for linear change for all symptom and mechanism variables. All symptoms changed significantly over time including depressive symptoms ($d = -1.13$), anxiety symptoms ($d = -1.08$), and anger ($d = -0.87$), producing large effects. Most mechanisms changed significantly over time including the OBQ-44 total score ($d = -1.87$), OBQ-44 subscales (importance/control of thoughts [$d = -1.71$], perfectionism/intolerance of uncertainty [$d = -1.80$], and responsibility/threat estimation [$d = -1.46$]), FMPS concern over mistakes ($d = -1.63$), and TAS-20 total score ($d = -1.91$). We did not see significant change over time for personal standards or doubts about actions. The negligible Cohen's D effect size for changes in personal standards over time reflects a lack of linear change over time ($d = -0.01$), however, the effect size for doubts about actions was medium in size ($d = -0.62$).

3.2. Quadratic changes in symptoms and mechanisms

Table 2 presents results for all models for which a quadratic time parameter was retained. Fitting the quadratic parameter resulted in significant decreases in deviance relative to the linear model for depression ($x^2 = 5.24$, df = 1, $p < .001$), anxiety ($x^2 = 2.49$, df = 1, $p = .024$), OBQ-44 total ($x^2 = 2.51$, df = 1, $p < .001$), FMPS personal standards ($x^2 = 3.49$, df = 1, $p = .041$) and the TAS-20 total score ($x^2 = 1.33$, df = 1, $p = .006$), indicating these variables showed curvilinear change over time during treatment.

Arnold's treatment timeline predicted rapid early change in obsessive beliefs (targeted at the beginning of treatment), and rapid late change in emotional awareness (targeted towards the latter part of treatment). As hypothesized, the coefficient for the quadratic parameter in the OBQ-total model was positive, indicating that obsessive beliefs showed rapid change early in treatment, which tapered off in later sessions. Conversely, in the model for TAS-20, the coefficient for the quadratic time parameter was negative, indicating that change over time in emotional awareness was slow at the beginning of treatment, and more rapid in the second part of treatment. Fig. 2 depicts these two curvilinear shapes of change.

3.3. Mechanisms predicting subsequent symptoms

Because Arnold's primary symptom was anger, we examined the formulation hypothesis that lower scores in problematic mechanisms (OBQ-Total, all OBQ-44 and FMPS subscales, and the TAS-20 total score) would be associated with subsequently lower scores in anger. For example, the formulation hypothesized lower personal standards at one session would predict less anger at the following session. To examine

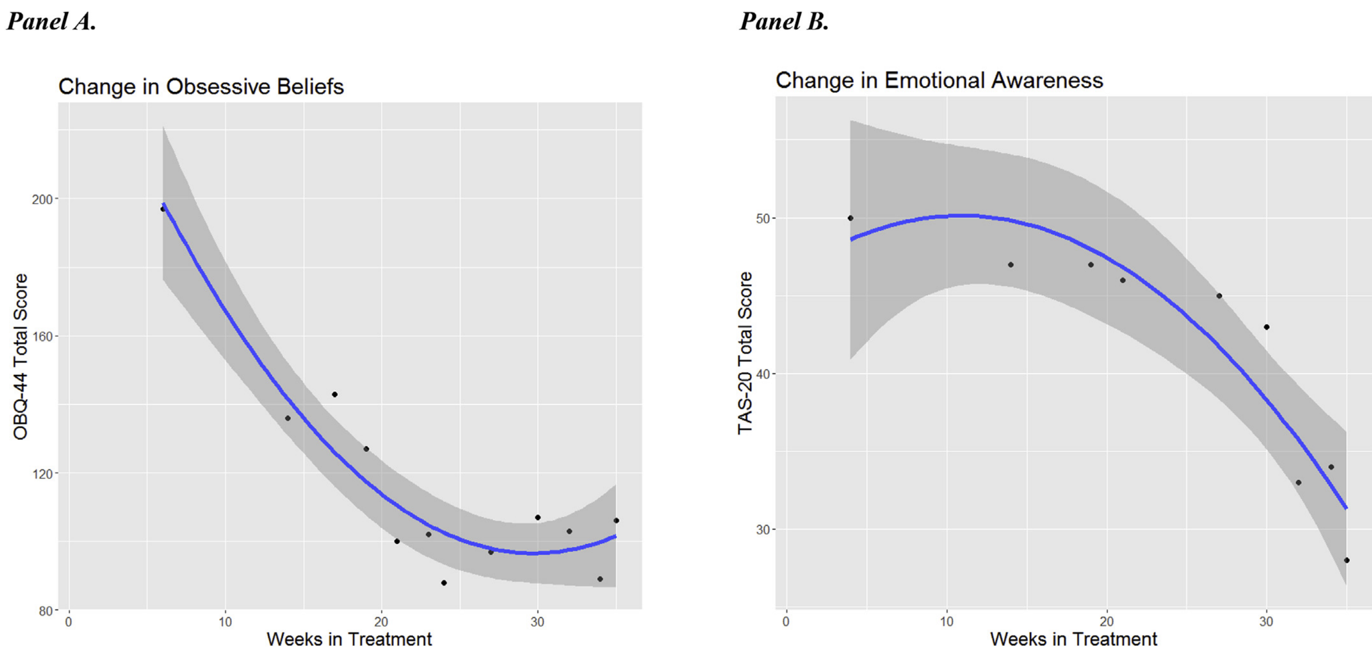

Fig. 2. Panel A depicts curvilinear change in obsessive beliefs over time. Panel B depicts curvilinear change in emotional awareness over time (higher values indicate lower emotional awareness). The therapist targeted obsessive beliefs early in treatment, whereas emotional awareness was targeted later in treatment. In line with hypotheses, obsessive beliefs changed more rapidly in the early part of treatment, whereas emotional awareness changed more rapidly in the later part of treatment.

the time-lagged effects of each mechanism variable on anger, separate regression models were constructed for each mechanism in which time-forward anger was modeled as the dependent variable, predicted by the time-lagged mechanism variable, as well as time-lagged anger. As described previously, an interaction term with elapsed time was modeled for each predictor, to correct for uneven spacing of time-points.

Table 3 presents the final time-lagged regression models. As hypothesized, the OBQ total score (d = 2.57), as well as two subscales (importance/control of thoughts [d = 2.35]; perfectionism/intolerance of uncertainty [d = 3.23]) demonstrated significant lagged effects on anger, producing large effects. These effects were in the expected direction, suggesting that when the client had lower scores in these problematic obsessive beliefs and perfectionistic tendencies, he reported higher anger at the following time point. Contrary to hypotheses, personal standards demonstrated a significant *negative* lagged effect on anger at the subsequent time-point, producing a large effect (d = 1.19). This finding indicates that when the client reported lower personal standards, he reported higher anger at the following time-point. We did not find significant lagged effects of the OBQ responsibility for harm subscale (d = 0.38) or emotional awareness (d = 0.02) on subsequent anger.

3.4. Associations between multiple treatment constructs over time

In line with methods described above, we applied EFA, CFA, and DFM to Arnold's emotion log data. First, using EFA (with oblique rotation), we examined which of Arnold's emotion log items clustered together across time. For example, we hypothesized that on days when Arnold was feeling greater affection, he was also feeling more loved and caring. We expected these positive emotions to make up one factor, distinct from negative emotions. Further, using DFM, we tested relationships *between* emotion factors over time. For example, Arnold's formulation hypothesized that his discomfort with positive emotions promoted anger. Thus, for Arnold, experiencing a cluster of positive emotions may be followed on a subsequent day by a cluster of emotions including anger.

A four-factor solution resulted from the EFA (SRMR = 0.06); the three-factor solution yielded a SRMR of 0.09. Next, a CFA was run using standardized factor loadings from the EFA. After one iteration of the confirmatory model, the item "grateful" was removed from the third factor due to a factor loading of 0.15 (below the threshold of 0.30). Three items (*envious*, *jealous*, and *afraid*) were removed entirely because they did not meet the loading threshold of 0.30 on any factor. The final confirmatory model demonstrated good fit according to absolute fit indices (RMSEA = 0.067, SRMR = 0.067, CFI = 0.85, TFI = 0.83). The completely standardized solution and all factor loadings and correlations among the factors are presented in Table 4. The first three factors suggest that 1) negative affects, 2) positive interpersonal affects, and 3) positive intrapersonal affects cluster together over time. The fourth factor, which we interpret as a "superiority" cluster, suggests that on days when Arnold experienced greater contempt and disgust, he also felt more pride, confidence, and less hurt.

Next, we examined relationships between these four factors across time using DFM. The model initially specified only autoregressive beta paths (that is, each of the four factors at time $t-1$ predicting only itself at time t). Lagrange multiplier tests suggested the presence of a cross-lagged path from Factor 2 at time $t-1$ to Factor 4 at time t. This path was opened, and the model was then re-run. The resulting model provided excellent fit across all fit indices (SRMR = 0.062, RMSEA $<$ 0.001, CFI = 1.00, TFI = 1.07). No further modification indices emerged. Fig. 3 presents the completely standardized solution, containing contemporaneous correlations between factors, and autoregressive and cross-lagged beta paths, indicating how factors relate to one another across time. Within the same day, the two positive emotion factors were highly correlated, and had negative associations with the negative affect factor, and positive associations with the "superiority" factor. Results suggest that when the client reported higher interpersonal positive affects, he tended to report higher levels of the "superiority" cluster of affects on the subsequent day.

4. Discussion

We described an approach that leverages a combination of idiographic statistical analyses to systematically test clinical hypotheses during routine treatment. We implemented the approach using data collected from an adult male with a mix of excessive anger, depressive

Table 3
Results from time-lagged regression models predicting anger.

Variable/Mechanism Predicting Anger	Predictors in Model	β	p	t	Cohen's D
Total obsessive beliefs (OBQ-44)	Lagged Mechanism	0.70	.002	5.74	2.57
	Anger Autoregression	0.10	< .001	8.17	3.65
	Elapsed Time	0.04	.713	0.39	0.17
	Anger*Elapsed Time	−2.92	.001	−6.54	−2.92
Importance/Control of Thoughts (OBQ-44)	Lagged Mechanism	0.70	.003	5.25	2.35
	Anger Autoregression	0.84	.001	6.49	2.90
	Elapsed Time	0.28	.085	2.14	0.96
	Anger*Elapsed Time	−2.26	.003	−5.25	−2.35
Perfectionism/Intolerance of Uncertainty (OBQ-44)	Lagged Mechanism	0.81	< .001	7.23	3.23
	Anger Autoregression	1.07	< .001	10.67	4.77
	Elapsed Time	−0.10	.279	−1.21	−0.54
	Anger*Elapsed Time	−3.53	< .001	−8.6	−3.85
Responsibility/Threat Estimation (OBQ-44)	Lagged Mechanism	0.25	.43	0.86	0.38
	Anger Autoregression	0.66	.068	2.22	0.99
	Elapsed Time	−0.07	.820	−0.24	−0.11
Personal Standards (FMPS)	Lagged Mechanism	−0.64	.048	−2.48	−1.11
	Anger Autoregression	0.61	.034	2.74	1.23
	Elapsed Time	0.28	.330	1.06	0.47
Doubts about Actions (FMPS)	Lagged Mechanism	0.46	.147	1.66	0.75
	Anger Autoregression	0.62	.057	2.35	1.05
	Elapsed Time	0.10	.739	0.35	0.16
Concern over Mistakes (FMPS)	Lagged Mechanism	0.33	.305	1.12	0.50
	Anger Autoregression	0.63	.072	2.18	0.97
	Elapsed Time	0.04	.905	0.13	0.06
Emotional Awareness (TAS-20)	Lagged Mechanism	0.02	.97	0.04	0.02
	Anger Autoregression	0.75	.201	1.63	0.87
	Elapsed Time	0.32	.436	0.90	0.48

Note. Time-lagged regression models with anger as an outcome variable, predicted by lagged anger (autoregression), elapsed time, and each lagged mechanism. If the interaction between lagged anger and elapsed time or the interaction between the lagged mechanism and elapsed time was significant than the interaction was retained in the model.

and anxiety symptoms. Below, we describe the results of our analyses and how they can inform the case-formulation and guide clinical decision-making. In addition, we discuss an online platform we've developed to easily implement idiographic statistical analyses with routine treatment data.

4.1. Testing linear and quadratic changes in symptoms and mechanisms

In line with clinical hypotheses, all symptoms and most mechanisms declined significantly over time. Contrary to hypotheses, there was no evidence for linear change in personal standards. However, personal standards demonstrated marginally significant curvilinear change over time, suggesting that interventions to reduce personal standards may have been effective, but change was short-lived. As hypothesized, mechanisms of obsessive beliefs decreased and emotional awareness increased more rapidly during the times they were targeted by the therapist during treatment, demonstrating the potential for analyses to link changes in mechanisms to specific periods of treatment. These findings provide useful information regarding the effectiveness of treatment for altering specific symptoms and mechanisms in the formulation.

4.2. Testing whether mechanisms predict symptoms

Lower levels of several hypothesized mechanisms in the formulation predicted less anger at subsequent sessions (e.g., obsessive beliefs). However, contrary to formulation hypotheses, there was no evidence linking emotional awareness to subsequent anger. Moreover, lower personal standards were significantly associated with subsequent *increases* in anger (opposite the expected direction). These aspects of the formulation were hypothesized based on nomothetic research literature, for example, suggesting that higher personal standards are associated with higher anger, especially when other domains of perfectionism are high (Dunn, Gotwals, Causgrove Dunn, & Syrotuik, 2006; Rice & Lapsley, 2001). The fact that formulation hypotheses based on research literature were unsupported within this individual highlights the importance of examining person specific data to identify change processes unique to a specific client. When a therapist conducts interventions that are not well suited to a specific client's needs, this can have negative consequences, such as wasted time, and potentially worsened client symptoms. If Arnold's therapist or supervisor were aware of the negative association between personal standards and subsequent anger during treatment, they might have stopped intervening to alter personal standards, avoided subsequent increases in Arnold's anger, and potentially created a more efficient and effective treatment. Thus, findings from idiographic analyses can help therapists to substantiate or rule out mechanistic hypotheses in the case formulation to guide treatment choices.

4.3. Testing associations between multiple treatment constructs over time

We used Arnold's emotion log data to examine the formulation hypothesis that his experience of positive emotions and the associated discomfort led to subsequent anger. Across published studies on the structure of emotion, it is generally found that a two-factor structure (positive and negative affect) is well-fit to group-aggregated data (e.g. Watson, Clark, & Tellegen, 1988). For Arnold, a four-factor model had the best fit. Factors included a) negative affects, b) positive

Table 4
Results from confirmatory factor analysis.

	Factor 1: Negative Emotion	Factor 2: Interpersonal Positive Emotion	Factor 3: Intrapersonal Positive Emotion	Factor 4: "Superiority"
Happy	–	–	0.61	–
Interested	–	–	0.55	–
Excited	–	–	0.68	–
Caring	–	0.52	–	–
Affection	–	0.58	–	–
Love	–	0.71	–	–
Loved	–	0.63	–	–
Compassion	–	0.40	–	–
Grateful	–	0.43	–	–
Proud	–	–	0.56	0.42
Confident	–	–	0.46	0.40
Hurt	0.62	–	–	−0.42
Sad	0.51	–	–	–
Regret	0.43	–	–	–
Irritated	0.56	–	–	–
Angry	0.56	–	–	–
Resentment	0.61	–	–	–
Disgust	0.43	–	–	0.40
Contempt	–	–	–	0.67
Ashamed	0.46	–	–	–
Guilty	0.54	–	–	–
Anxious	0.41	–	–	–
Correlations Between Factors				
Factor 1	1	-.44	-.35	.34
Factor 2	-.44	1	.67	.06
Factor 3	-.35	.67	1	-.31
Factor 4	.34	.06	-.31	1

interpersonal affects, and c) positive intrapersonal affects. The fourth (and arguably most interesting) factor revealed that on days when the client was feeling more contempt and disgust, he felt more pride, confidence, and less hurt. We interpreted this cluster as a "superiority" cluster because it seems closely related to the client's problematic superiority behavior in the workplace (e.g, he would patronize and undermine his superiors with disgust and contempt). The "superiority" cluster advances a novel formulation hypothesis: that Arnold may engender contempt and disgust towards superiors in order to feel better about himself and less hurt. Thus, using a data driven idiographic approach to test hypotheses has the added benefit of elucidating novel mechanism hypotheses for the formulation.

Contemporaneously (i.e., within the same day), the intrapersonal and interpersonal positive emotion factors were highly correlated. These two positive emotion factors had contemporaneous negative associations with the negative affect factor, and positive associations with the "superiority" factor, suggesting the superiority cluster is activated on days when the client is feeling positive emotions. Additionally, examining associations among these four factors *across* time, we found that when the client reported higher interpersonal positive affect, he tended to report higher levels of the "superiority cluster" emotions on the subsequent day. These results do not support the formulation hypothesis that positive emotional discomfort relates to the subsequent symptom of increased anger, but instead suggest the client's self-reported discomfort and vulnerability surrounding positive interpersonal emotions promoted the potentially problematic "superiority" cluster. These results help link the problematic superiority behavior to the case formulation through the mechanism of positive emotions. By working to increase positive emotions, the therapist may have been doing the client a disservice, unknowingly activating problematic superiority behaviors.

4.4. Online platform for idiographic analysis of routine treatment data

In the future, idiographic analyses will have the greatest potential to help guide treatment choices if the methods are made easily accessible and user friendly. To address this, we've developed a website that offers the ability to enter clinical data online and receive statistical output {www.changestat.org}. In its current form, our website enables users to enter/upload and store multiple client datasets. Plots are provided for each variable across time, along with results from person specific regressions with linear and quadratic parameters. Time-lagged regression results are also provided between pairs of variables. The website offers the capability of conducting p-technique EFA and examining time-lagged relationships between pairs of factors.

How might clinicians utilize the approach during treatment? After several sessions, clinicians can upload quantitative data onto the website (sampled as frequently as once a day). Data can be updated as treatment progresses in order to monitor changes in key mechanisms and symptoms. Plots of each variable allow for visual inspection, and statistical values enable clinicians to infer significant change in a variable over time, as well as differences in the rates of change between different parts of treatment. As clinicians work to alter one problematic mechanism, they can gain information regarding whether that mechanism is effectively being altered, and whether it leads to reduction of symptoms. Simultaneously, clinicians can gather information suggesting the next best mechanisms to target. Importantly, having statistics to support clinical hypothesis testing may enable clinicians to more easily prepare quantitative case reports or case series for the scientific literature, helping to bridge the gap between science and practice. Although there will always be obstacles to collecting data in clinical settings (Boswell, Kraus, Miller, & Lambert, 2015), we plan to

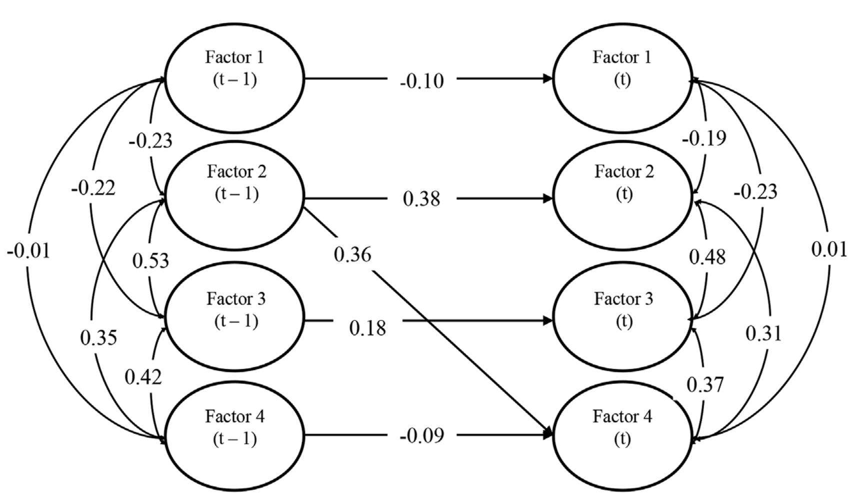

Fig. 3. The completely standardized factor solution for the client's emotion log data is presented below. The four factors are represented as circles on the left. Circles on the right represent the four factors at the subsequent time-point. The four factors are 1) negative emotion items, 2) interpersonal positive emotion item, 3) intrapersonal positive emotion items, and 4) emotion items thought to relate to the client's superiority behavior. The solution contains autoregressive (horizontal arrows) and cross-lagged (diagonal arrow) beta paths as well as contemporaneous correlations between the factors (curved lines).

continue to improve the accessibility, usability, and convenience of these methods. We hope that the valuable information obtained with these approaches will incentivize therapists to collect more quantitative data during treatment.

4.5. Strengths, limitations, and future directions

These findings contribute to our basic understanding of how idiographic analytic methods can be applied to routine treatment data to inform case formulations and clinical decisions. There is strong empirical support for the benefits of monitoring treatment effectiveness (e.g., Shimokawa et al., 2010), and the analytic approaches outlined here provide useful information beyond treatment effectiveness that could similarly improve treatments when used in ongoing psychotherapy. The current study takes a necessary first step in applying this approach to routine treatment data to illustrate its utility for clinical hypothesis testing. However, the study and approach are not without limitations.

First, as will often be the case with data collected during routine treatment, the number of time-points was limited. Thus, we were statistically underpowered for our analyses with the fewest observations. Additionally, analyses were conducted after treatment ended. Future research should assess the utility of implementing idiographic hypothesis testing *during* treatment. There are also limitations to the approach. Although the approach does not require that treatment be delayed or halted, there is a necessary delay between the time a therapist develops a clinical hypothesis and the time the therapist has collected the amount of quantitative data necessary to reliably test that hypothesis.

Concerning the models themselves, models risk over-fitting the data or interpreting noise and error. Models also may reflect idiosyncratic interpretations of items as opposed to substantive factors. We made several choices with our approach, based upon previously published applications of p-technique factor analysis and DFM (see Fisher, 2015). However, clinicians may have compelling rationales for alternative choices. For example, we suggest starting with oblique rotation in EFA, and then assessing whether factors are correlated rather than orthogonal. However, if factors are not correlated and theoretically orthogonal, clinicians may want to use orthogonal rotation. Additionally, we included a CFA. Although CFA can provide greater confidence that the factor structure from the EFA optimally reflects the latent structure of the data, CFA may also inflate correlations between factors. Although it is useful to gain confidence in the putative factor structure, the CFA can also be omitted, and factor scores could instead be generated from the EFA solution.

Our website takes an important step in making idiographic analyses more accessible to clinicians, however, it has tremendous room for growth and improvement. While it is not a large time-cost for clinicians to upload or input session scores, our website would be maximally convenient if measures could be completed and scored within the web-platform itself. Additionally, by using a set of measures tailored to the individual, comparison to other patients is difficult. These models will become even more useful when comparisons can be made to normative distributions of effect sizes, in order to assess where a patient falls in comparison to other individuals during treatment. As such, researchers have argued for the importance of bridging idiographic statistical approaches with group aggregated approaches (e.g., Beltz, Wright, Sprague, & Molenaar, 2016) to enable such comparisons. It is our hope that as idiographic analyses in routine treatment become more common, our platform could aggregate and synthesize data across individuals to provide personalized normative information to clinicians.

5. Conclusions

The current study demonstrates the utility of applying idiographic data analyses to the kinds of self-report data that can be readily collected during routine treatment. As illustrated with a complex case characterized by excessive anger, depressive symptoms, and anxiety, this approach (a) provided personalized information about the effectiveness of treatment, (b) helped to confirm and disconfirm formulation hypotheses, (c) expanded the case formulation, and d) provided information to guide treatment choices. Practicing clinicians routinely think about change processes in complex and challenging cases and many collect data to monitor progress. Our approach offers methods that enable clinicians to deliver more personalized empirically based treatment. It further empowers them to make valuable contributions to the scientific understanding of psychotherapeutic change processes. As idiographic analytic approaches become common in routine treatment and in research settings, analyses can be aggregated or replicated across individual cases (Fisher, Newman, & Molenaar, 2011) and across practitioners, to shed light on mechanisms of change, improve treatments, and help the many individuals suffering from mental health problems.

Conflicts of interest

For all authors, none declared.

Funding

Preparation of this manuscript was supported by a National Institute of Mental Health pre-doctoral fellowship awarded to Casey L. Brown (T32MH020006) and a National Institute on Aging grant awarded to Robert W. Levenson (R01AG041762).

Abbreviations

CBT (Cognitive Behavioral Therapy); ESTs (empirically supported treatments); BDI (Beck Depression Inventory-II); BAI (Beck Anxiety Inventory); OBQ-44 (Obsessive Beliefs Questionnaire); FMPS (Frost Multidimensional Perfectionism Scale); TAS-20 (Toronto Alexithymia Scale).

Acknowledgements

The authors thank Jacqueline B. Persons for her exceptional clinical supervision and helpful comments on the manuscript. The authors also thank Aaron J. Fisher, Nancy H. Liu, and Amy H. Sanchez for their helpful feedback.

References

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia scale—I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38*(1), 23–32. http://doi.org/10.1016/0022-3999(94)90005-1.

Barlow, D. H., & Nock, M. K. (2009). Why can't we be more idiographic in our research? *Perspectives on Psychological Science, 4*(1), 19–21. http://doi.org/10.1111/j.1745-6924.2009.01088.x.

Beltz, A. M., Wright, A. G. C., Sprague, B. N., & Molenaar, P. C. M. (2016). Bridging the nomothetic and idiographic approaches to the analysis of clinical data. *Assessment, 23*(4), 447–458. http://doi.org/10.1177/1073191116648209.

Bhar, S., Bouvard, M., Calamari, J., Carmin, C., Clark, D. A., Cottraux, J., et al. (2005). Psychometric validation of the obsessive belief questionnaire and interpretation of intrusions inventory—Part 2. *Behaviour Research and Therapy, 43*, 1527–1542. http://doi.org/10.1016/j.brat.2004.07.010.

Boswell, J. F., Anderson, L. M., & Barlow, D. H. (2014). An idiographic analysis of change processes in the unified transdiagnostic treatment of depression. *Journal of Consulting and Clinical Psychology, 82*(6), 1060–1071. http://doi.org/10.1037/a0037403.

Boswell, J., & Bugatti, M. (2016). An exploratory analysis of the impact of specific interventions: Some clients reveal more than others. *Journal of Counseling Psychology, 63*(6), 710.

Clasen, P. C., Fisher, A. J., & Beevers, C. G. (2015). Mood-reactive self-esteem and depression vulnerability: Person-specific symptom dynamics via smart phone assessment. *PLoS One, 10*(7), e0129774. http://doi.org/10.1371/journal.pone.0129774.

Dozois, D. J. A., Dobson, K. S., & Ahnberg, J. L. (1998). A psychometric evaluation of the Beck depression inventory-II. *Psychological Assessment, 10*(2), 83–89. http://doi.org/10.1037/1040-3590.10.2.83.

Dunn, J. G. H., Gotwals, J. K., Causgrove Dunn, J., & Syrotuik, D. G. (2006). *Examining the*

relationship between perfectionism and trait anger in competitive sport. Vol. 4, 7–24.
Fisher, A. J., Newman, M. G., & Molenaar, P. C. M. (2011). A quantitative method for the analysis of nomothetic relationships between idiographic structures: Dynamic patterns create attractor states for sustained posttreatment change. *Journal of Consulting and Clinical Psychology, 79*(4), 552–563. http://doi.org/10.1037/a0024069.
Fisher, A. J. (2015). Toward a dynamic model of psychological assessment: Implications for personalized care. *Journal of consulting and clinical psychology, 83*(4), 825.
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. http://doi.org/10.1037/abn0000311.
Frank, R., & Davidson, J. (2014). *The transdiagnostic road map to case formulation and treatment planning: Practical guidance for clinical decision making.* New Harbinger Publications.
Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Source: Econometrica Econometrica, 37*(3), 424–438.
Hofmann, S. G., Asnaani, A., Vonk, I. J. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research, 36*(5), 427–440. http://doi.org/10.1007/s10608-012-9476-1.
Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidiscip. J. 6*(1), 1–55.
Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen, & J. S. Long (Eds.). *Sage focus editions* (pp. 294). SAGE Publications.
Lei, P. W., & Wu, W. (2007). Introduction to structural equation modeling: Issues and practical considerations. *Educational Measurement: Issues and Practice, 26*(3), 33–43.
Levenson, R. W., Lwi, S. J., Brown, C. L., Ford, B. Q., Otero, M. C., & Verstaen, A. (2017). Emotion. In G. G. B. J. T. Cacioppo, & L. G. Tassinary (Eds.). *Handbook of Psychophysiology* (pp. 444–464). Cambridge University Press. https://doi.org/10.1017/9781107415782.020.
Lo, L. L., Molenaar, P. C. M., & Rovine, M. (2017). Determining the number of factors in P-technique factor analysis. *Applied Developmental Science, 21*(2), 94–105. http://doi.org/10.1080/10888691.2016.1173549.
Lutz, W., Zimmermann, D., Müller, V. N. L. S., Deisenhofer, A.-K., & Rubel, J. A. (2017). Randomized controlled trial to evaluate the effects of personalized prediction and adaptation tools on treatment outcome in outpatient psychotherapy: Study protocol. *BMC Psychiatry, 17*(1), 306. http://doi.org/10.1186/s12888-017-1464-2.
Lyon, A. R., Lewis, C. C., Boyd, M. R., Hendrix, E., & Liu, F. (2016). Capabilities and characteristics of digital measurement feedback systems: Results from a comprehensive review. *Administration and Policy in Mental Health and Mental Health Services Research, 43*(3), 441–466. http://doi.org/10.1007/s10488-016-0719-4.
Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika, 50*(2), 181–202. http://doi.org/10.1007/BF02294246.
Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdiscip.Res. Perspect. 2*(4), 201–218. http://doi.org/10.1207/s15366359mea0204_1.
Persons, J. (2012). *The case formulation approach to cognitive-behavior therapy.* New York: Guilford Press.
Persons, J. B., Brown, C. L., & Diamond, A. (2019). Case formulation-driven cognitive behavioral therapy. In D. K.S., & D. D. (Eds.). *Handbook of cognitive behavioral therapies*(4th ed.). New York: Guilford Press.
Persons, & Hong, J. (2016). Case formulation and the outcome of cognitive behaviour therapy. In N. Tarrier, & J. Johnson (Eds.). *Case formulation in cognitive behaviour therapy: The treatment of challenging and complex cases* (pp. 14–37). (2nd ed.). New York: Routledge.
Persons, J. B., & Tompkins, M. A. (1997). Cognitive-behavioral case formulation. In T. D. Eells (Ed.). *Handbook of psychotherapy case formulation* (pp. 314–349). (2nd ed.). Guilford Press.
Revelle, W. (2013). *Psych: Procedures for psychological, psychometric, and personality research. R package version 1.3. 10.* Evanston, IL: Northwestern University.
Rice, K. G., & Lapsley, D. K. (2001). Perfectionism, coping, and emotional adjustment. *Journal of College Student Development, 42*(2).
Sekhon, J. S. (2009). Opiates for the matches: Matching methods for causal inference. *Annual Review of Political Science, 12*(1), 487–508. http://doi.org/10.1146/annurev.polisci.11.060606.135444.
Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology, 78*(3), 298–311.
Steer, R. A., & Beck, A. T. (1997). Beck anxiety inventory. In C. P. Zalaquett, & R. J. Wood (Eds.). *Evaluating stress: A book of resources* (pp. 23–40). Lanham, MD: Scarecrow Education. Retrieved from http://psycnet.apa.org/record/1997-09146-002.
Stober, J. (1998). The frost multidimensional perfectionism scale revisited: More perfect with four (instead of six) dimensions. *24*(4), 481–491.
Stulz, N., Lutz, W., Leach, C., Lucock, M., & Barkham, M. (2007). Shapes of early change in psychotherapy under routine outpatient conditions. *Journal of Consulting and Clinical Psychology, 75*(6), 864–874. http://doi.org/10.1037/0022-006X.75.6.864.
Tabachnick, B., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon/Pearson Education.
Wright, A. G. C., Hallquist, M. N., Dtepp, S. D., Scott, L. N., Beeney, J. E., Lazarus, S. A., et al. (2016). Modeling heterogeneity in momentary interpersonal and affective dynamic processes in borderline personality disorder. *Assessment, 23*(4), 484–495. http://doi.org/10.1177/1073191116653829.