Available online at www.sciencedirect.com
ScienceDirect
Behavior Therapy 50 (2019) 938–951
Behavior Therapy
www.elsevier.com/locate/bt

A Clinician's Primer for Idiographic Research: Considerations and Recommendations

Marilyn L. Piccirillo
Emorie D. Beck
Thomas L. Rodebaugh
Washington University in St. Louis

Theorists and clinicians have long noted the need for idiographic (i.e., individual-level) designs within clinical psychology. Results from idiographic work may provide a possible resolution of the *therapist's dilemma*—the problem of treating an individual using information gathered via group-level research. Due to advances in data collection and time series methodology, there has been increasing interest in using idiographic designs to answer clinical questions. Although time series methods have been well-studied outside the field of clinical psychology, there is limited direction on how clinicians can use such models to inform their clinical practice. In this primer, we collate decades of published and word-of-mouth information on idiographic designs, measurement, and modeling. We aim to provide an initial guide on the theoretical and practical considerations that we urge interested clinicians to consider before conducting idiographic work of their own.

Keywords: idiographic; single subject; within-person methodology; time-series methodology

Clinicians who provide treatment typically focus directly on an individual. Clinical scientists who conduct and disseminate treatment research, on the other hand, typically focus on a group. This situation results in a tension known as *the therapist's dilemma*—a phrase describing the situation of being trained primarily in nomothetic, group-based research methodology, yet being tasked with treating a single individual (Levine, Sandeen, & Murphy, 1992). This dilemma results in serious consequences when one considers that theorists have increasingly argued that group-level methods cannot properly model intraindividual variability (Borsboom, Mellenbergh, & van Heerden, 2003; Fisher, Medaglia, & Jeronimus, 2018; Molenaar, 2004). The existing theoretical and empirical work suggests that using findings from nomothetic research to inform individual-level care may result in misguided interventions.

Clinicians have long sensed the implications of the therapist's dilemma and have argued that, for an intervention to be successful, it must produce beneficial change not only in experimental settings, but also both in general practice settings and in settings for a specific, *individual* patient (Howard, Moras, Brill, Martinovich, & Lutz, 1996). The desire to evaluate whether interventions work for a single individual has led both clinicians and researchers to advocate for an increased idiographic focus in psychotherapy and psychotherapy research (Howard et al., 1996). Indeed, idiographic methods have been used for decades in psychotherapy process research (see e.g., Hoenders, Bos, De Jong, & De Jonge, 2012; Jones & Nesselroade, 1990; Jones, Ghannam, Nigg, & Dyer, 1993; Luborsky, 1953; Luborsky & Mintz, 1972; Tschacher & Ramseyer, 2009; Tschacher, Zorn, & Ramseyer, 2012; see also Russell, Jones, & Miller, 2007, for a review of idiographic methods used to study psychotherapy process). However, due to limitations in data collection technology and the complexity of data analytic techniques, idiographic methods are not routinely used within clinical settings. Instead,

* Address correspondence to Marilyn Piccirillo, Department of Psychology, Washington University in St. Louis, Campus Box 1125, St. Louis, MO 63130; e-mail: piccirillom@wustl.edu.

clinicians who use empirically based therapies continue to rely heavily on findings demonstrated at the group level (e.g., through effectiveness or efficacy studies). Therefore, integration of idiographic methods within the field of clinical psychology is sorely needed to ensure that psychopathological models originally developed from group-level research also apply in the field of personalized care.

Although idiographic methodology has been used in the past to study psychotherapy process, clinicians have not systematically implemented these methods in their work. In part, we believe this lack of implementation may reflect previous limitations in the accessibility and feasibility of idiographic data collection methods and updated statistical methods. However, we see clear signs that idiographic methods are becoming increasingly more accessible and feasible. One such sign is the increasing availability of smartphone technology and dynamic data collection methods, such as experience sampling methodology (ESM) or ecological momentary assessment (EMA), which are becoming increasingly accessible to clinicians (Myin-Germeys et al., 2009) and are a necessary part of modeling individual processes (Fisher, 2015).

A second auspicious sign is that innovative statistical methods are better able to model individual processes (e.g., developmental packages in R; dynamic structural equation modeling in Mplus; Muthén & Muthén, 1998-2017). Furthermore, the paywall previously blocking many researchers and clinicians from implementing such models is breaking down. Most of the statistical analyses needed for idiographic modeling can be conducted using the R language and environment, which is a free and open-source software program (R Core Team, 2017). Moreover, specific R packages, which include a number of idiographic-specific packages, are generally accompanied by user-friendly guides (graphicalVAR, Epskamp, 2018; gimme, Gates & Molenaar, 2012; psych, Revelle, 2018). Further, the clinicians who are more likely to be interested in idiographic work are also more likely to be trained to use R during graduate school, as opposed to past trends of clinicians being trained to use software packages that are difficult to access outside of an academic setting. Thus, we believe that idiographic modeling is becoming more and more feasible for clinicians because data collection methods, analytical tools, and (in the case of newer clinicians) the clinicians themselves are increasingly up to the task. Full discussion and review of available idiographic methods is outside the scope of this primer and available elsewhere (Piccirillo & Rodebaugh, in press); however, we invite interested readers to review the current methods for idiographic modeling provided in Table 1 to decide which technique is most relevant to their study design.

Our sense is that many clinical scientists, and clinicians more generally, see idiographic modeling as an important scientific advance for the field of psychotherapy. It is easy to understand why such models would seem promising. Idiographic models might allow these clinicians to (a) develop hypotheses for intervention regarding individual patients (Epskamp et al., 2018) and (b) make claims about individual-level causal processes (Borsboom et al., 2003; Kroeze et al., 2017). Eventually, these models may be able to identify processes that hinder an individual's treatment or represent new avenues for therapeutic intervention, even if that new avenue

Table 1
Examples of Commonly Used Idiographic Methods in Clinical Psychology

Technique	Description of Technique	Exemplar Paper
P-Technique Factor Analysis / Principle Components Analysis	An exploratory factor analysis for one individual.	Borkenau et al., 1998; Molenaar, 2004
P-Technique Dynamic Factor Analysis	An exploratory factor analysis for one individual that accounts for time trends.	Molenaar, 1985; Molenaar, De Goojier, & Schmitz, 1992
Association Networks	Shows bivariate correlations between observed variables.	Epskamp, Kruis, & Marsman, 2017
Graphical VAR Models	Shows lagged partial correlations between observed variables.	Epskamp, Waldorp, Mõttus, & Borsmoom, 2018
Within-Person SEM	Uses unified SEM to simultaneously estimate lagged and contemporaneous relationships between observed variables.	Beltz & Gates, 2017; Fisher & Boswell, 2016
Dynamic Bayesian Multilevel Model	Shows multivariate lagged relationships and other dynamic estimates of time series variables (e.g. variability, intensity, etc.).	Krone, Albers, & Timmerman, 2016
Dynamic SEM	Shows lagged and contemporaneous relationships between observed and latent variables. A multilevel extension is also available.	Hamaker & Wichers, 2017; Muthén & Muthén, 1998-2017

only applies to single patient for whom the model was constructed. That is, the therapist's dilemma might be completely dissolved by these new methods. However, the current state of idiographic research is far from dilemma-free. In large part, this is because of the complexity surrounding idiographic measurement, assessment, and modeling. Idiographic research within clinical psychology and psychotherapy represents an exciting frontier, but guides to this frontier are few.

To date, there have been few published (practical) clinician guidelines for designing idiographic studies, collecting data, and constructing individual-level models. Although some researchers have published theory-based papers that can assist interested researchers (Kroeze et al., 2017), to our knowledge, the most relevant guide for *clinicians* was published over a decade ago by Borckardt and colleagues (2008). Although this guide filled a much-needed gap in the literature, idiographic methodology has changed considerably, in terms of both methods (e.g. advent of ecological momentary assessment, Borckardt et al., 2008) and modeling (e.g., the development of time-varying modeling techniques, Bringmann, Hamaker, et al., 2016).

Although time series methods have been well-studied in other fields (e.g., economics), our read of the field—at least in terms of how idiographic methodology has been applied to clinical psychology—is that it is far from settled at present. Indeed, many of the updated resources that we (i.e., the authors) used recently when collecting and modeling our own individual-level data were shared during academic conferences (e.g., Association for Behavioral and Cognitive Therapies), across social media sources (e.g., academic Twitter), or via personal correspondence. Unfortunately, to our knowledge, none of these current resources have been collated in a comprehensive paper, contributing to the limitations of conducting idiographic research in clinical psychology. Thus, we aim to provide an update and extension of the considerations and recommendations discussed in previous work.

The following is a review of the nuanced challenges that we think clinicians would be well-advised to consider *prior* to collecting idiographic data. Of note, by referring to *clinicians*, we do not mean to imply that all or even most clinicians will be interested in collecting the data and conducting the analyses we describe here. We do, however, believe that there is a significant group of clinicians who are interested in doing so, and that those clinicians in that group are likely to read this journal. Accordingly, we have designed this review to be an introductory guide for clinical scientists and clinicians alike who are familiar with the rationale for pursuing idiographic work and, most important, who sustain an interest in integrating idiographic work into their clinical practice. We begin by presenting an example of how idiographic methods may be integrated into clinical practice. We then advance into describing different topics related to idiographic work, beginning with the design and data collection stage.

A Practical Hypothetical Example

Imagine a man is seeking treatment for social anxiety disorder and depression. Upon intake, he reports symptoms consistent with a depressive episode and social anxiety disorder. He also reports problematic alcohol use in social situations, in an attempt to manage his social anxiety. Due to a wait list at the clinic, the therapist suggests that the client complete daily assessment of symptoms and behavior to provide an initial guide for therapy. Using the client's report, the therapist chooses specific items to assess this client's emotional experience. These items include symptoms, such as depressed mood, anhedonia, and social anxiety, as well as behavior since the last assessment, such as alcohol use. It is important to note that such retrospective assessment introduces reporting biases, but is sometimes the only option for gathering data on behaviors that may occur with lower frequency over the course of the day (i.e., situational alcohol use). We will discuss this further later on. The client completes the assessment multiple times a day over a period of time.

After collecting enough data, the interested clinician uses an idiographic analysis to construct a series of individual-specific models for the client. These models demonstrate both *contemporaneous* and *directed* paths between items, which are represented as a network of inter-relations among the measured variables. *Contemporaneous* paths show how items measured at the same time point are correlated across time. Positive relationships suggest that two variables are likely to co-occur over time. Negative relationships suggest that when one variable has a lower level, the other is likely to have a higher level at the same point in time. The *directed* paths show associations between items across measurement occasions, such that a positive directed path indicates that higher values of one variable prospectively predicts higher levels of another variable. Likewise, a negative directed path indicates that higher levels of the variable prospectively predicts lower levels of the second variable.

Examining this specific client's model more closely, the contemporaneous paths may demonstrate the correlations seen in Figure 1A, with positive paths between depressed mood, anhedonia, and alcohol use, but a negative path between social anxiety and alcohol use. The directed paths may look similar to

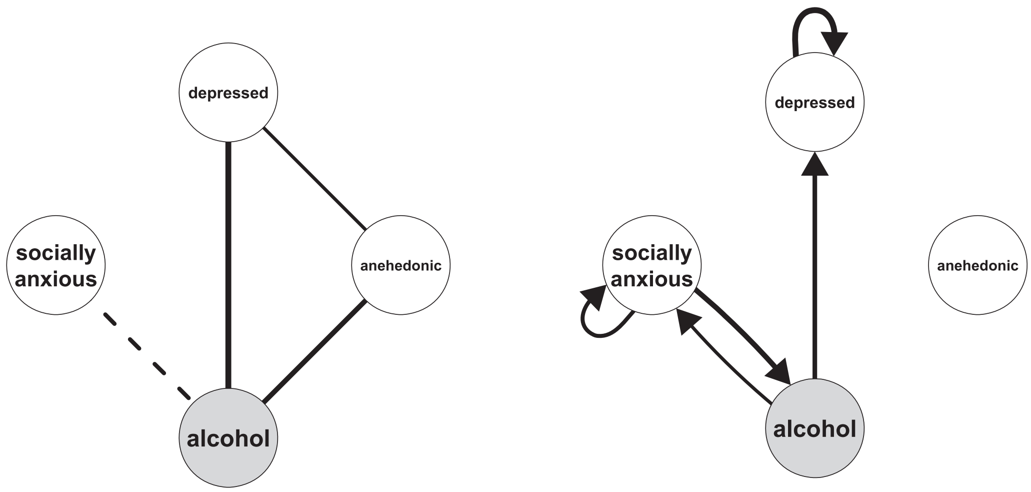

FIGURE 1 (A) Network representation of contemporaneous undirected relationships among anxiety / depression symptoms and alcohol use. Solid edges (lines) represent positive relationships among nodes (circles), while dashed edges represent negative associations. (B) Network representation of directed (lagged) relationships among the same symptoms and behavior as (A). Solid edges and dashed edges represent positive and negative associations, respectively. The width of the edge represents the relative strength of the relationship.

the model presented in Figure 1B. Here we see increased alcohol use predicting more social anxiety and depressed mood at a future time point, and social anxiety predicting future alcohol use. In addition, social anxiety and depression predict themselves, suggesting that once they onset, they persist over time. These results could be very useful to the clinician in gently introducing the notion that alcohol use is causing the client problems. The client may not be aware that alcohol use predicts future social anxiety and depressed mood and may instead only be aware that his social anxiety decreases around the time he is using alcohol. Thus, limiting alcohol use may reduce both the onset and duration of social anxiety and depressive symptoms. Presenting results from such a model might increase the client's interest in changing his behavior.

Over the course of therapy, one might expect that the positive, directed relationship between social anxiety and alcohol use would decrease as the client learns strategies for managing anxiety in social situations and reduces the use of alcohol in these anxiety-provoking situations. The clinician could also collect data at a second time point, perhaps during a break in therapy during which the client tries out using the skills he has learned. A model derived from this time period may demonstrate that the maladaptive directed relationships are no longer present (or greatly reduced), which may suggest the end of therapy. Moreover, these results could provide additional evidence of the positive changes that the client has experienced regarding his social anxiety and mood.

Design and Data Collection

Now that we have presented an example of how these methods could be integrated into a clinical setting, we turn to our discussion of how to begin idiographic work. There are a number of important considerations, which we discuss below. We briefly summarize the existent literature and provide guidelines and recommendations for how to address each issue.

ASSUMPTIONS OF IDIOGRAPHIC METHODS: STATIONARITY

One such assumption behind most methods of idiographic analysis is stationarity. Stationarity requires that each variable demonstrate a similar mean, variance, and relationship over time with other variables and itself (Bringmann et al., 2016). This means that developmental trends (e.g., levels of autonomy increasing during adolescence), cyclic trends (e.g., seasonal affective symptoms), or autoregressive trends (e.g., an increasing tendency for depressed mood to persist and therefore predict itself more strongly as a major depressive episode begins; van de Leemput et al., 2014) may violate stationarity.

We provide an example of stationarity to help to make this assumption more intuitive. Consider a person whose anxiety leads to feelings of excitement (via increased activity) during the first part of a time series, followed by the anxiety leading to depressive mood (due to decreased activity) during the latter half of a time series. It is easy to intuit how such a thing could happen: The individual is, at first, motivated by his or her anxiety, but as continued efforts fail to alleviate the anxiety, the anxiety leads to increased depression. What will happen when such data are modeled? The negative association between anxiety and depression in the first part of the time series may be partially or wholly negated by the positive association in the second part of the time series. Examining the resulting model, the clinician might be surprised to find no relationship between anxiety and depression *on average* for this individual, but that finding would be completely attributable to the fact that the process modeled violates stationarity.

It is important to note that stationarity does not at all imply that symptoms have plateaued or reached a steady state. It is not a problem for statistical models for symptoms to vary across time; indeed ideally they should vary highly. The problem occurs when the variables change systematically in response to another variable that is not in the model, or when the relationships between variables change over time, as depicted in the example we give above.

Researchers across the social sciences have developed and used several methods to test for (e.g., Kwiatkowski-Phillips-Schmidt-Shin test; Kwiatkowski, Phillips, Schmidt, & Shin, 1992) and esolve violations to stationarity (e.g., use of residuals from ordinary least squares regression models, Fisher et al., 2017; detrending, Beltz & Gates, 2017, etc.). These methods largely address violations at the mean and variance levels of stationarity. Other, more complex, time-varying methods can accommodate autoregressive violations to stationarity. These methods will be discussed in greater detail below, in the section on idiographic modeling and analysis. Suffice it to say that, for now, such methods are less accessible in that they require more data (in terms of time points) and can only handle a limited number of variables effectively.

Despite the existence of methods to correct for violations of stationarity, we suggest that one does the utmost to *start* with stationary variables and processes. Doing so requires the clinician to consider whether the psychological process being studied (e.g., depressed mood as caused by decreased activity and levels of anxiety) is plausibly stationary. Consider a situation in which a patient is undergoing active treatment in which there is a systematic reduction in symptoms (i.e., a violation to stationarity). In this case, removing the (linear) trend though a detrending process may be enough; alternatively, perhaps it would be better to plan to collect data once symptoms have stabilized (e.g., once response to treatment has stalled).

Both clinicians and researchers, upon learning about stationarity, have expressed disappointment that idiographic methods are not currently well-suited to examining change over time. However, we believe that this concern represents a misunderstanding of stationarity. These methods were developed specifically to model change over time in the level or *value* of variables. That is, these methods are entirely in keeping with attempts to discover what factors rule over a client's level of depression or anxiety at a given point in time. What these methods were not developed to handle are (a) changes in the *processes* that produce those values or (b) changes in values that arise from outside of the system being modeled. As a result, we contend that these techniques are likely to be useful before active treatment has begun, as well as when treatment has stalled—and thus the processes at play are expected to be consistent for some period of time. When a clinician wants to measure change in symptoms across time, other methods may be more appropriate (e.g., examining clinically significant and reliable change; Jacobson & Truax, 1991).

Furthermore, if there are reasons to believe that relationships between variables may not be stationary and that use of a time-varying method is warranted, the researcher should be aware that it is unlikely that more than two variables could be analyzed effectively at present. Moreover, as time-varying methods are still a relatively new addition to the field, they are limited in their utility and almost certainly increase the complexity of the data analytic approach. For example, time-varying methods often require a considerable amount of observations (based on experience and discussion with researchers, our best guess is that over 100 time points may often be needed; L.F. Bringmann, personal communication, July 6, 2018). The need for increased observations likely increases patient and clinician burden. As these techniques are further developed, this concern may no longer be an issue in the future, but they are certainly an important concern at present (Bringmann, Hamaker, et al., 2016).

Until time varying methods have advanced further, we recommend that clinicians interested in beginning idiographic work consider collecting data during a time period that is more likely to result in an approximately stationary set of variables. For example, there may be natural delays before beginning therapy (e.g., scheduling limitations, clinic waitlists, etc.), especially within larger group or program practices. Thus, interested clinicians may consider collecting data during this waiting period when the process is more plausibly approximately stationary. Alternatively, given that rapid therapeutic change is mostly likely to occur during the initial sessions (Howard, Kopta, Krause, & Orlinsky, 1986), data collection could also take place after the clinician has observed a slowing down or lack of progress in treatment as evidenced by similar scores on progress measures or a clinician's subjective sense that a client is no longer making substantial gains. This may suggest a suitable time to collect additional idiographic assessments as the process may be more plausibly stationary. We end this section with a word of warning: the fact is that little is known about whether psychological processes (i.e., relationships between pairs of any two psychological constructs over time) are *ever* plausibly stationary. Much work stands to be done in this area.

VARIABLE AND ITEM SELECTION

As in every psychological study, the items and measures used are only as useful as the constructs that the measures operationalize. Some of our suggestions here are more general and some are specific to EMA and idiographic assessment. Our first recommendation regarding item selection seems simple: The items should produce variables that are continuous (and not ordinal or binary) whenever possible. We say it *seems* simple because clinicians are likely used to dealing with variables that are continuous, and may already be aware of the many statistical reasons to prefer continuous variables. However, researchers often recommend limiting both the number of items, as well as the response options in EMA research (Shiffman, Stone, & Hufford, 2008). As a result, it may require some thought to decide what kind of response scale could produce plausibly continuous measures. In our ongoing work, we have tended to use a 0–100 sliding scale, but we have observed that some data-collection programs do not implement such a slider in an easy-to-use way; in such programs we might use a 0–10 Likert scale. If clinicians are planning to sum up several items, then narrower response scales may be feasible (e.g., because five items with a five-point scale becomes a subscale with a twenty-one-point range of possible scores). However, even multiple items are sometimes only useful provided that the participant uses the full range of the scale. That is, a participant who only answers using two points on the scale for each of five items will leave the researcher with a six-point range of scores, which is not continuous. Such patterns may also create problems in estimation as differences in variances across items may make relationships across items incomparable (Bulteel, Tuerlinckx, Brose, & Ceulemans, 2016). In such cases, within-person standardization of the variables can correct for differences in variances.

Item selection will also be influenced by how clinicians think about the relationship between constructs and items. Recently, idiographic analyses have been popularized by researchers who reference a network approach (cf., Borsboom & Cramer, 2013). Researchers who work within a network approach posit that individual items can represent constructs that directly change and influence one another within a network. In our experience, some clinical scientists assume that idiographic models are, by nature, network models. Further, some assume that one of the main analytic techniques used for idiographic analyses—vector autoregression (VAR)—is a network approach. In fact, VAR originated outside of and is used well beyond network frameworks. Rather, network techniques are tools for assessing the structure and dynamics of matrices, which can be defined in a number of ways (e.g. social networks, adjacency matrices, zero-order correlation matrices, partial correlations, etc.). VAR models are simply one technique for examining such matrices. Moreover, there are additional time series approaches that allow for the use of latent variables. Network approaches have often been pitted against the more traditional latent variable approaches, which posit that an underlying latent variable (e.g., depression) instigates change across individual depressive symptoms. Recent papers have offered an integration of these approaches (cf., Epskamp, Rhemtulla, & Borsboom, 2016). We know of no reason that we should consider network and latent variable approaches to be inevitably at odds. However, as the network framework has been used more widely in concert with idiographic designs, we will discuss a limitation of the network approach as it relates to variable selection.

One limitation of the network framework as typically applied concerns the common use of a single item to represent a single construct. For example, researchers may use the item "How down are you feeling right now" to represent depressed mood. Although one-item-per-construct is obviously easier on participants who are completing multiple surveys per day, there are multiple psychometric issues with using only one item to measure a construct and contrasting viewpoints regarding under what circumstances (if any) this makes sense (see, e.g., Epskamp et al., 2018; Fried & Cramer, 2017; Mõttus, 2016; Mõttus, Kandler, Bleidorn, Riemann, & McCrae, 2017; Seeboth & Mõttus, 2018). Clinicians interested in including more items without increasing participant burden are directed to further reading regarding planned missing data designs and imputation methods for multivariate time series data (Che, Purushotham, Cho, Sontag, & Liu, 2018; Graham, 2009; Graham, Taylor, Olchowski, & Cumsille, 2006; Honaker, King, & Blackwell, 2011; Liu & Molenaar, 2014; Silvia, Kwapil, Walsh, & Myin-Germeys, 2014). Furthermore, there are almost no available directions on how to assess or improve psychometrics in EMA studies (cf. Zimmerman et al., 2018, for an exception). Overall, we recommend that clinicians interested in collecting idiographic data balance concerns regarding participant burden with the careful selection of either (a) a few items that are best theorized to measure that construct or, if single-item measures are highly desirable for theoretical reasons, (b) single items that will produce an approximately continuous scale. In either case, having available evidence regarding the psychometric properties of the items used is highly desirable.

Recently, researchers have noted increasing concern about the construct validity of ambulatory scales. For example, removing items ad-hoc from an existing scale for use with EMA may change a construct's validity. Further, constructs might operate differently on a daily or hourly level versus a retrospectively assessed trait-like level (a concern that will be addressed in more detail below). Just because items measure *depressed mood* when assessed over the time frame of, say, the past 2 weeks, does not guarantee that they measure the same construct when assessed daily or hourly. Although these concerns remain largely unaddressed to date, there have been efforts to create dynamic scales for daily diary studies and test the validity of these scales for ambulatory assessment. The first of such scales, the Personality Dynamic Diary, has been successfully used in various populations (Zimmerman et al., 2018).

Additionally, several researchers have noted the role of context in analyzing trends from EMA data (cf., Myin-Germeys et al., 2009). We acknowledge that context may play a particularly important role in interpreting idiographic relationships between clinical symptoms. For example, a patient's report of anxiety in a phobic situation may be qualitatively different than his or her report of anxiety during the rest of the day. Yet, at this time, idiographic methods are limited in how well they can account for context in the model. A clinician could ask the patient to record the context that he or she is in during each survey throughout the day and (dichotomously) code each situation for analysis. However, this method would require many observations for each situation to reflect meaningful relationships between context and symptoms, which becomes increasingly difficult for behaviors or contexts that are less frequent or are actively avoided, as this greatly reduces statistical power to detect true differences across contexts. Such use of context within a model may often not be feasible given the complexity of an individual's life and limited resources and time for assessment. A more tractable method might be to measure context through the use of continuous measurements, such as the number of people who are currently in the patient's surrounding or evaluative ratings of the present context (cf., Kroeze et al., 2017), the psychological features of a situation (cf., Rauthmann et al., 2014), or lexical taxonomies of situation characteristics (cf., Parrigon, Woo, Tay, & Wang, 2017). Of course, such methods would only work when the important features of a context can be encoded in the measure being used.

Notably, even the most reliable and statistically pliable items are only useful if they measure constructs that are relevant to a particular individual. We would hope that many constructs used in previous group-based studies would prove to be relevant to many individuals, but, as mentioned above, there is no guarantee that constructs used in nomothetic research will necessarily apply to a specific individual (Allport, 1937; Nesselroade & Molenaar, 2016). The trouble is that although it may be obvious to say that we should measure variables that are relevant to an individual, we have limited empirical guidance for making this determination.

We are aware that this section of our guide might appear pessimistic, but we believe that clinicians interested in pursuing idiographic work would be well served to use clinical judgment in combination with hallmark symptoms of a disorder (e.g., depressed mood or anhedonia for major depressive disorder) and patient self-report to determine which variables would be most useful to include in a given patient's model (Borckardt et al., 2008; Kroeze et al., 2017). Although this is our best suggestion, we note that reliance on clinical judgment and patient self-report carries key limitations (e.g., self-report biases, amount of patient insight into sources of distress).

TIMING OF ASSESSMENTS

In addition to carefully selecting items and variables for an idiographic study, clinicians should also consider the time scale over which they believe the symptoms and constructs will vary. Researchers conducting EMA studies have measured constructs with varying frequency, ranging from once daily (i.e., a daily diary study) to multiple times a day (e.g., up to every 30 minutes; Shiffman et al., 2008). Measuring constructs multiple times a day will make it easier to collect a large amount of data within a shorter period of time and may capture the changes within the variable across the course of a day that would be missed at the daily level. However, assessing patients multiple times a day will increase patient burden, which may reduce compliance. More important, in many cases, these seemingly arbitrarily chosen time intervals may not capture the time scale on which each symptom or construct operates. For example, mood may change more frequently and (potentially) with less variability than impulsivity, which may change less frequently but with more variability. That is, mood might be different at each occasion, but show relatively smooth and steady changes, whereas impulsivity ratings might show infrequent, but strong spikes. Furthermore, if the variable of interest changes over a longer time range than predictor variables in the model, this situation could result in a lack of directed effects within the model, making the model appear sparse (Lane, Gates, Pike, Beltz, & Wright, 2018). That is, a relatively slow-to-change process (e.g., depressed mood in a depressed individual) may not appear to directly change in the model, as there is little variability within this symptom across the day.

One way to address variables that change at different time scales could be to examine trends at different lags. A lag refers to relationship over time (see Figure 2). For example, lag_1 refers to the relationship between a value at t-1 and the value at t. Different relationships between variables may be best exemplified through different lagged models (e.g., lag_1 versus lag_2). However, not all idiographic methods can test different lags within the same model and those that do require a considerable number of observations (e.g., Bringmann, Ferrer, Hamaker, Borsboom, & Tuerlinckx, 2018, Bringmann, L.F.,

Example Lagged Data Set

ID	day	time	y_t	y_{t-1}	y_{t-2}
1	1	1	4	NA	NA
1	1	2	2	4	NA
1	1	3	1	2	4
1	1	4	2	1	2
1	1	5	NA	NA	NA
1	1	6	NA	NA	NA
1	2	1	1	NA	NA
1	2	2	2	1	NA
1	2	3	2	2	1
1	2	4	1	2	2
1	2	5	NA	NA	NA
1	2	6	NA	NA	NA

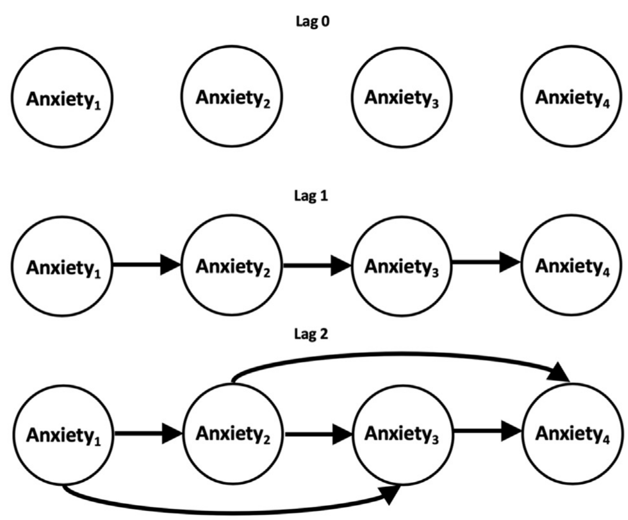

FIGURE 2 Example and illustration of lagged data. In the figure on the left, the gray highlighted rows show rows of missing values (represented as "NA") inserted in fixed interval sampling to prevent regression across day-night periods. The left columns on this figure show the originally measured time series, as well as lag_1 and lag_2 relationships (created by shifting each row 1 measurement point down). The black shaded values highlight how lagging a variable shifts the value. The figure on the right illustrates an example of the autoregressive effects of anxiety over time. Lag_0 includes the effect for anxiety across time points 1-4. Lag_1 demonstrates the relationship between t_{-1} and t (i.e., $Anxiety_1$ predicting $Anxiety_2$). Lag_2 demonstrates the relationship between t_{-2} and t (i.e., $Anxiety_1$ predicting $Anxiety_3$). Note that the Lag_2 example assumes that Lag_1 relationships are estimated as well, which is common when modeling Lag_2 relationships.

personal correspondence, July 6, 2018). Thus, researchers have tended to construct models at lag_1, most likely due to power constraints. Yet, methods that can accommodate multiple lags may be what are needed to accurately model the complexity of psychological processes. We believe that clinicians will need to strike a balance between choosing variables that are thought to operate on a similar time scale and choosing a time scale for assessment that will capture change in these variables over time. Of course, the trouble with this fact is that we have little knowledge so far as to how quickly psychological constructs should be expected to change and affect one another.

Furthermore, current idiographic methods are typically used with measurements taken at fixed intervals, because the most widely available methods require even spacing between assessments, which adds to the complexity of the timing issue discussed above. The use of fixed interval assessments may limit the ecological nature of the assessments, potentially increasing the risk of participant response bias due to the scheduled nature of assessments. However, some investigators have used methods that take irregularly spaced data and render it as if the data were measured at even intervals (cf., the use of cubic spline interpolation in Fisher, Reeves, Lawyer, Medaglia, & Rubel, 2017). Similarly, in Mplus, the *tinterval* command can be used to allow examination of data assessed at uneven intervals (Muthén & Muthén, 1998-2017). Such methods add a further level of complexity but can assist with the analysis of unevenly spaced data.

However, these solutions do not address issues with the frequency with which events occur, which may be a more pernicious problem. For example, a clinician may wish to assess the participant's psychological response to a binge eating episode or a self-injury event, which may occur with less frequency than other symptoms (e.g., changes in affect). Within a fixed interval schedule, the researcher would need to include a dichotomous item reflecting the presence or absence of events. However, when events are infrequent (or less frequent than the assessment schedule), there will be almost no variance in such a variable, which reduces the statistical power to model effects including this item.

Of course, it is feasible to ask participants to report after each event of this type. However, because most of the time series methods that are currently practical for clinicians work best with the collection of all variables at a given time point, the participant would need to report on all variables after the event of interest occurs. Thus, we recommend that clinicians interested in collecting data on less frequent events use both a fixed assessment schedule, as well as a user-initiated assessment schedule to collect data on potential predictors of this event, as well as contemporaneous information surrounding the event. Methods for dealing with unevenly spaced intervals can then be used to process and analyze the time series data.

Finally, it is important to note that clinicians who sample participants multiple times a day will need to address the issue of predicting across night periods. This problem refers to the prediction from the last assessment of one day to the first assessment of the following day (i.e., predicting relationships overnight). This raises both quantitative and qualitative issues. The quantitative issue only really arises when researchers ignore overnights. That is, because the amount of time between the last assessment of one day and the first of the next will almost always be longer than the interval between other adjacent assessments (e.g., the first and second assessments within a day), the resulting data will violate a fixed interval design unless this longer interval is accounted for. Researchers have addressed this issue by either treating the overnight period as an additional set of fixed intervals that are missing (e.g., Bringmann, Pe, et al., 2016) or by using spline interpolation methods (e.g., Fisher et al., 2017).

The need to deal with overnight periods is so ubiquitous that it is worth describing the methods available in more detail. First, to treat the overnight period as a set of fixed intervals, the researcher would insert lines of missing data in between the last assessment of the day and the first assessment of the following day. This is demonstrated in Figure 2. These rows of missing data should correspond to the number of fixed intervals that were (conceptually, at least) missed overnight, which introduces an additional issue. If, for example, a researcher uses a 5-hour interval and collects four assessments a day, the remaining 9 unaccounted hours during a night period would not be evenly divisible by five, making it impossible to insert the correct number of missing rows. Missing rows are vital because they prevent the model from incorporating overnight predictions into the model as if they represent a single lag. For example, in the figure, assessment points were 4 hours apart, which necessitates inserting two rows for the overnight period. Researchers have also used interpolation methods (e.g., cubic spline interpolation) to predict the missing values for the additional intervals, or produce estimates for a different set of intervals (e.g., four assessments as would have been obtained from 6-hour intervals). However, the insertion of missing values appears to have been used more frequently to address the overnight problem (Bringmann, Pe, et al., 2016), likely because it is easier to perform than current interpolation methods.

Methods for resolving the overnight issue assume that overnight periods simply represent an unobserved continuation of the same process that was occurring during the day. Because most participants are sleeping for at least part of the overnight period, it is plausible that the overnight interval may be qualitatively different from other intervals during the day. For example, there is evidence that sleep is critical for the consolidation of long-term memories (Drosopoulos, Schulze, Fischer, & Born, 2007; Smith & Lapp, 1991). It also seems implausible that many daytime psychological processes continue throughout the night, unimpeded by sleep (e.g., it seems strange to claim that participants with depression are equally depressed during sleep). In this case, our modeling strategies are treating sleep as something that can be ignored. Whether this is a problem (i.e., whether it causes inaccurate estimates) needs further attention.

ETHICAL CONSIDERATIONS AND PATIENT CONCERNS

Finally, ethical considerations and potential patient concerns with collecting idiographic data are worthy of consideration. Notably, many ethical considerations for collecting idiographic data are not significantly different than the considerations for other types of data collection and research for clinical problems. Ethical considerations range from secure storage of data, providing proper informed consent about the use of the data, the potential for a research agenda to influence treatment, and contraindications to this approach (Borckardt et al., 2008). Clinicians can utilize skills they routinely implement to handle secure and confidential data in accordance with HIPAA laws. However, clinicians should consider potential contraindications for the use of idiographic assessment in psychotherapy. For example, clinicians should assess patient comfort with technology and ability to provide informed consent before enrolling patients in an idiographic study. Likewise, patients who are in imminent distress or who exhibit severe emotion regulation deficits may not be suitable candidates for an intensive data collection schedule if plans for data collection involve a delay in treatment (Borckardt et al., 2008; Mumma & Fluck, 2016). Of course, serious and life-threatening behaviors often continue even after treatment commences, and clinicians might consider collecting data on such behaviors. However, if the clinician chooses to track behaviors that could clearly result in immediate harm (e.g., self-injurious behavior, suicidality), the patient and therapist should collaboratively discuss how the therapist will respond to items that assess risk of or intention for self-harm.

Additionally, patients may perceive an increase in symptoms as repeated assessments bring their attention to their symptomatology, potentially leading to distress. Previous research has demonstrated that repeated assessment does tend to increase the patient's distress at the beginning, perhaps because it highlights the depth of their impairment or suffering (Kahlon, Neal, & Patterson, 2014; Redhead, Johnstone, & Nightingale, 2015). However, it should be noted that this form of distress can be addressed and typically resolves over the course of treatment (Redhead et al., 2015). Indeed, research has shown that patients often appreciate and benefit from repeated assessment when there is a meaningful output, such as a model of their psychological distress (Beck, 2011; Kahlon et al., 2014; Kuyken, Padesky, & Dudley, 2011; Mumma & Fluck, 2016; Redhead et al., 2015). Additionally, technology is becoming increasingly idiographic in nature, giving power to the patient in terms of knowledge about their health (e.g., wearable fitness trackers, sleep apps, etc.). It is possible that idiographic assessments may be viewed similarly by the patient and may confer therapeutic benefit (Kroeze et al., 2017).

Data Cleaning and Analysis

Readers may note that the bulk of this article has concerned issues to think through before collecting any data. The meta-message here is completely intentional: Idiographic models require much planning up front to maximize the potential for gathering useful information, and clinicians and clinical scientists alike who want to utilize these methods should carefully consider the procedure of survey construction and data collection. Once the data have been collected, however, there remain key issues to consider.

DATA CLEANING

As with any other type of EMA data, the data collected for idiographic time series analyses require careful data cleaning. For example, as discussed previously, clinicians will need to consider how to handle missing data both on the assessment and item level. Depending on the method that was used for data collection, the clinician may also need to develop rules for handling duplicate responses within the same time frame (i.e., with separate time stamps), as well as responses made after the initial notification and potentially outside the appropriate assessment response window. Not all data collection applications can prevent data being collected outside of specified assessment intervals. As noted above, this issue is not a concern if the researcher has already decided to use a method that handles time more continuously. The use of automated data-collection software within smartphone applications may lull some clinicians into a false sense of security. Automated systems can eliminate human error, but should be checked carefully for their own potential errors. We urge clinicians to develop clear rules ahead of the data cleaning process for handling responses.

Stationarity

A special case regarding data cleaning involves searching for violations to stationarity. We have already described above why stationarity is important. One available test is the Phillips-Peron test, which is similar to the Kwiatkowski-Phillips-Shin-Shapiro test discussed earlier, and evaluates for linear trends (Phillips, 1987). This test is available as an R package and a significant p-value indicates that there is a linear trend (stats package; R Core Team, 2017). The clinician can then include an exogenous variable to model this trend or *detrend* the data (e.g., van der Krieke et al., 2015). The trouble with this test is its inability to deal with missing data. For example, when testing for stationarity, all specialty stationarity tests we are familiar with will either remove or ignore missing data, resulting in a lack of certainty as to whether stationarity was appropriately addressed. For example, the clinician could detrend the data by subtracting the previous timepoint, *t*-1, from the timepoint, *t*. Other methods, such as using the residuals from an ordinary least squares regression model, have also been used (e.g., Fisher et al., 2017). Time-varying methods, such as those discussed by Bringmann and colleagues (2016), can also be used to test for violations to stationarity, to model nonlinear trends, and can handle missing data appropriately. However, these methods are currently restricted to two variables, which make it difficult to analyze more complex models. However, it is plausible that additional multivariate time-varying methods will become available and ready for use in the near future as multiple researchers are actively examining and developing these methods. Our current preference is to use methods that allow the inclusion of time variables (number of days completing surveys; survey number of the day) in the model itself. Although this method does not account for all possible violations of stationarity, it allows a test of time trends while also accounting for them appropriately in the model.

VARIABLES

After collecting the data, it is important to first ensure that there is demonstrable separation between individual variables. For example, if a client responded to items assessing low mood in the same way as they responded to items assessing anxiety, including these items as separate predictors in the model may produce spurious findings (i.e., similar to the way that multicollinearity produces issues in a regression model). Thus, we suggest that clinicians run a correlation analysis before entering the separate, observed variables into the model. Clinicians should strongly consider compositing highly correlated variables before entering them into the model. Of course, if the intent was to use multiple items to assess latent variables or a subscale score, such correlations could simply support that such an approach is warranted.

MULTIPLE SOLUTIONS

Most of the current idiographic methods available have strategies for reducing overfitting the data. For example, within the network framework, a process referred to as *lasso* applies a penalty to regression weights in a model that shrinks the estimated coefficients and thresholds effects that are below a certain size to zero (Tibshirani, 1996). This lasso process reduces the likelihood of spurious findings. Similarly, another method called group iterative multiple model estimation (GIMME) method utilizes an automated search procedure to estimate the full set of relationships (Fisher et al., 2017; Gates, Molenaar, Hillary, Ram, & Rovine, 2010). As currently implemented in the *gimme* package in R, this procedure utilizes unified structural equation modeling, which uses model comparison tests to add and test additional model paths. Any unestimated paths in the model are set to zero. These methods reduce the chance of extraneous paths and may improve the construct validity of the model.

It is also important to recognize that some idiographic methods produce multiple outputs, which reflect different interpretations of the idiographic relationships. For example, using graphical VAR produces a contemporaneous network and a directed, lagged network. The contemporaneous network provides information about the partial correlations between variables within time (i.e., how two indicators relate to each other across the time series). The directed network provides information about the lagged regressive relationships between variables across time. It is also possible to examine partially directed networks that use the residuals from the contemporaneous network to map relationships over time.

Conclusion

Idiographic assessment and modeling promise an answer to the therapist's dilemma. By supplementing clinical insights with data obtained through individual-level models, clinicians may be able to increase the effectiveness of their case conceptualization for an individual patient. However, the complexity of the idiographic techniques available at present have created a new dilemma as the statistical foundations and nuances of individual-level modeling and VAR are complex and are not routinely taught in clinical science programs that often emphasize nomothetic approaches (Levine et al., 1992). We empathize with the reader who may desire to use idiographic methods but who feels overwhelmed with the idea of implementing such technology and statistical approaches in his or her clinical work. However, with the combination of EMA and smartphone technology, as well as the accessibility of open-access statistical software, we are increasingly optimistic that some of the technical aspects of idiographic modeling can be automated and simplified over time. Finally, with the increasing integration between research, clinical practice, and industry (e.g., Apple Toolkit, Google Verily), there is ample opportunity for third-party companies to develop and provide software for data collection and individual-level data analytic techniques.

Notably, most of the techniques referenced in Table 1 are undergoing improvements to promote ease of analysis and interpretation. We look forward to statistical models that are easier to use, but urge researchers and clinicians alike to continue to proceed with caution. The structure of this article focused heavily on issues that occur before one reaches the stage of using statistical models, and we believe that without consideration of the issues discussed here, statistical models that are easy to use may simply make it easier to be wrong. Thus, we urge interested clinicians to keep in mind that the importance of methodological rigor in both data collection and statistical methods.

We hope that the guidelines recommended here can support greater use of idiographic methods in clinical practice. The call for increased idiographic focus has continued for decades (Barlow & Nock, 2009) and in a field that was initially commissioned to identify idiographic truths or particularities of human nature (Lamiell, 1998), new idiographic methods represent a data-driven way forward in empirically studying the individual, even when that individual is a single patient in clinical practice. With the advent of EMA, smart-phone technology, and open-access software for analyzing time series data, a true integration of individual-level designs into clinical practice may now be achievable, with the potential for improvement of existing psychological models and psychological care.

Conflict of Interest Statement

Marilyn L. Piccirillo was supported by National Research Service Award from the National Institute of Mental Health [F31 MH 115641]. Thomas L. Rodebaugh was supported by the NARSAD Independent Investigator Grant from the Brain Behavior Research Foundation [Grant number is: 25785]. There are no other declarations of interest.

References

Allport, G. W. (1937). *Personality: A psychological interpretation.* Oxford, England: Holt. Retrieved from http://psycnet.apa.org/record/1938-01964-000.

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

Beck, J. (2011). *Cognitive Behavioral Therapy: Basics and Beyond* (2nd ed.). New York, NY: Guilford Press.

Beltz, A. M., & Gates, K. M. (2017). Network mapping with GIMME. *Multivariate Behavioral Research*, *52*, 789–804. https://doi.org/10.1080/00273171.2017.1373014

Borckardt, J. J., Nash, M. R., Murphy, M. D., Moore, M., Shaw, D., & O'Neil, P. (2008). Clinical practice as natural laboratory for psychotherapy research: A guide to case-based time-series analysis. *The American Psychologist*, *63*, 77–95. https://doi.org/10.1037/0003-066X.63.2.77

Borkenau, P., Ostendorf, F., Goldberg, L. R., Hemnel, S., Hofstee, W. K. B., & McCrae, R. R. (1998). The Big Five as states: How useful is the Five-Factor model to describe intraindividual variations over time? *Journal of Research in Personality*, *32*, 202–221. https://doi.org/10.1006/jrpe.1997.2206

Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*, 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203–219. https://doi.org/10.1037/0033-295X.110.2.203

Bringmann, L. F., Ferrer, E., Hamaker, E. L., Borsboom, D., & Tuerlinckx, F. (2018). Modeling nonstationary emotion dynamics in dyads using a Time-Varying Vector-Autoregressive Model. *Multivariate Behavioral Research*, *3171*, 1–22. https://doi.org/10.1080/00273171.2018.1439722

Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2016). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods*, *22*, 409–425. https://doi.org/10.1037/met0000085

Bringmann, L. F., Pe, M. L., Vissers, N., Ceulemans, E., Borsboom, D., Vanpaemel, W., & Kuppens, P. (2016). Assessing temporal emotion dynamics using networks. *Assessment*, *23*, 425–435. https://doi.org/10.1177/1073191116645909

Bulteel, K., Tuerlinckx, F., Brose, A., & Ceulemans, E. (2016). Using raw VAR regression coefficients to build networks can be misleading. *Multivariate Behavioral Research*, *51*, 330–344. https://doi.org/10.1080/00273171.2016.1150151

Che, Z., Purushotham, S., Cho, K., Sontag, D., & Liu, Y. (2018). Recurrent neural networks for multivariate time series with missing values. *Scientific Reports*, *8*, 60–85. https://doi.org/10.1038/s41598-018-24271-9

Drosopoulos, S., Schulze, C., Fischer, S., & Born, J. (2007). Sleep's function in the spontaneous recovery and consolidation of memories. *Journal of Experimental Psychopathology: General*, *136*, 169–183. https://doi.org/10.1037/0096-3445.136.2.169

Epskamp, S. (2018). Graphical VAR for experience sampling data. [Computer software]. Retrieved from https://cran.r-project.org/web/packages/graphicalVAR/graphicalVAR.pdf.

Epskamp, S., Kruis, J., & Marsman, M. (2017). Estimating psychopathological networks: Be careful what you wish for. *PLoS ONE*, *12*e0179891. https://doi.org/10.1371/journal.pone.0179891

Epskamp, S., Rhemtulla, M. T., & Borsboom, D. (2016). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika*, *82*, 904–927. https://doi.org/10.1007/s11336-017-9557-x

Epskamp, S., van Borkulo, C. D., van der Veen, D. C., Servaas, M. N., Isvoranu, A. -M., Riese, H., & Cramer, A. O. J. (2018). Personalized network modeling in psychopathology: The importance of contemporaneous and temporal connections. *Clinical Psychological Science*, *6*, 416–427. https://doi.org/10.1177/2167702617744325

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsmoom, D. (2018). The Gaussian Graphical Model in cross-sectional and time-series data. *Multivariate Behavioral Research*, *53*, 453–480. https://doi.org/10.1080/00273171.2018.1454823

Fisher, A. J. (2015). Toward a dynamic model of psychological assessment: Implications for personalized care. *Journal of Consulting and Clinical Psychology*, *83*, 825–836. https://doi.org/10.1037/ccp0000026

Fisher, A. J., & Boswell, J. F. (2016). Enhancing the personalization of psychotherapy with dyanmic assessment and modeling. *Assessment*, *23*, 496–506. https://doi.org/10.1177/1073191116638735

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*. https://doi.org/10.1073/pnas.1711978115

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*, 1044–1056. https://doi.org/10.1037/abn0000311

Fried, E. I., & Cramer, A. O. J. (2017). Moving forward: Challenges and directions for psychopathological network theory and methodology. *Perspectives on Psychological Science*, *12*, 999–1020. https://doi.org/10.1177/1745691617705892

Gates, K. M., & Molenaar, P. C. M. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage*, *63*, 310–319. https://doi.org/10.1016/j.neuroimage.2012.06.026

Gates, K. M., Molenaar, P. C. M., Hillary, F. G., Ram, N., & Rovine, M. J. (2010). Automatic search for fMRI connectivity mapping: An alternative to Granger causality testing using formal equivalences among SEM path modeling, VAR, and unified SEM. *NeuroImage*, *50*, 1118–1125. https://doi.org/10.1016/j.neuroimage.2009.12.117

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, *11*, 323–343. https://doi.org/10.1037/1082-989X.11.4.323

Hamaker, E., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science*, *26*, 10–15. https://doi.org/10.1177/0963721416666518

Hoenders, H. J. R., Bos, E. H., De Jong, J. T. V. M., & De Jonge, P. (2012). Temporal dynamics of symptom and treatment variables in a lifestyle-oriented approach to anxiety disorder: A single-subject time-series analysis. *Psychotherapy and Psychosomatics*, *81*, 253–255. https://doi.org/10.1159/000335928

Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, *45*, 1–47. https://doi.org/10.18637/jss.v045.i07

Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose-effect relationship in psychotherapy. *American Psychologist*, *41*, 159–164. https://doi.org/10.1037/0003-066X.41.2.159

Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy efficacy, effectiveness, and patient progress. *American Psychologist*, *51*, 1059–1064. https://doi.org/10.1037/0003-066X.51.10.1059

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12–19. https://doi.org/10.1037/0022-006X.59.1.12

Jones, C. J., & Nesselroade, J. R. (1990). Multivariate, replicated, single-subject, repeated measure designs and P-technique factor analysis: A review of intraindividual change studies. *Experimental Aging Research*, *16*, 171–183. https://doi.org/10.1016/0022-4596(78)90152-4

Jones, E. E., Ghannam, J., Nigg, J. T., & Dyer, J. F. (1993). A paradigm for single-case research: The time series study of a long-term psychotherapy for depression. *Journal of Consulting and Clinical Psychology*, *61*, 381–394. https://doi.org/10.1037//0022-006X.61.3.381

Kahlon, S., Neal, A., & Patterson, T. G. (2014). Experiences of cognitive behavioural therapy formulation in clients with depression. *The Cognitive Behaviour Therapist*, *7*, 1–15. https://doi.org/10.1017/S1754470X14000075

Kroeze, R., van der Veen, D. C., Servaas, M. N., Bastiaansen, J. A., Oude Voshaar, R. C., Borsboom, D., & Riese, H. (2017). Personalized feedback on symptom dynamics of psychopathology: A proof-of-principle study. *Journal for Person-Oriented Research*, *3*, 1–10. https://doi.org/10.17505/jpor.2017.01

Krone, T., Albers, C. J., & Timmerman, M. E. (2016). Bayesian dynamic modeling to assess differential treatment effects on panic attack frequencies. *Statistical Modeling*, *16*, 343–359. https://doi.org/10.1177/1471082X16650777

Kuyken, W., Padesky, C. A., & Dudley, R. E. J. (2011). *Collaborative case conceptualization: Working effectively with clients in cognitive-behavioral therapy*. New York, NY: Guilford Press.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, *54*, 159–178. https://doi.org/10.1016/0304-4076(92)90104-Y

Lamiell, J. T. (1998). "Nomothetic" and "idiograhic": Contrasting Windelband's understanding with contemporary usage. *Theory and Psychology*, *8*, 23–38. https://doi.org/10.1177/0959354398081002

Lane, S. T., Gates, K. M., Pike, H. K., Beltz, A. M., & Wright, A. G. C. (2018). Uncovering general, shared, and unique temporal patterns in ambulatory assessment data. *Psychological Methods*, *24*, 54–69. https://doi.org/10.17605/OSF.IO/D3NK9

Levine, F. M., Sandeen, E., & Murphy, C. M. (1992). The therapist's dilemma: Using nomothetic information to answer idiographic questions. *Psychotherapy*, *29*, 410–415. https://doi.org/10.1037/h0088544

Liu, S., & Molenaar, P. C. M. (2014). iVAR: A program for imputing missing data in multivariate time series using vector autoregressive models. *Behavior Research Methods*, *46*, 1138–1148. https://doi.org/10.3758/s13428-014-0444-4

Luborsky, L. (1953). Intraindividual repetitive measurements (P-technique) in understanding psychotherapeutic change. In O. Mowrer (Ed.), *Psychotherapy, theory, and research* (pp. 389–413). New York, NY: The Ronald Press. https://doi.org/10.1037/10572-015

Luborsky, L., & Mintz, J. (1972). The contribution of P-technique to personality, psychotherapy, and psychosomatic research. In *Multivariate Personality Research: Contributions to the Understanding of Personality in Honor of Raymond B. Cattell* Baton Rouge, LA: Claitor's Publishing Division.

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, *2*, 201–218. https://doi.org/10.1207/s15366359mea0204_1

Molenaar, P. C. M., De Goojier, J. G., & Schmitz, B. (1992). Dynamic factor anlaysis of nonstationary multivariate time series. *Pyschometrika*, *57*, 333–349. https://doi.org/10.1007/BF02295422

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, *50*, 181–202.https://doi.org/10.1007/BF02294246.

Mõttus, R. (2016). Towards more rigorous personality trait-outcome research. *European Journal of Personality*, *30*, 292–303. https://doi.org/10.1002/per.2041

Mõttus, R., Kandler, C., Bleidorn, W., Riemann, R., & McCrae, R. R. (2017). Personality traits below facets: The consensual validity, longitudinal stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology*, *112*, 474–490. https://doi.org/10.1037/pspp0000100

Mumma, G. H., & Fluck, J. (2016). How valid is your case formulation? Empirically testing your cognitive behavioural case formulation for tailored treatment. *The Cognitive Behaviour Therapist*, *9*, 1–25. https://doi.org/10.1017/S1754470X16000088

Muthén, B., & Muthén, L. (1998-2017). Mplus, version 8.0.

Myin-Germeys, I., Oorschot, M., Collip, D., Lataster, J., Delespaul, P., & van Os, J. (2009). Experience sampling research in psychopathology: Opening the black box of daily life. *Psychological Medicine*, *39*, 1533–1547. https://doi.org/10.1017/S0033291708004947

Nesselroade, J. R., & Molenaar, P. C. M. (2016). Some behaviorial science measurement concerns and proposals. *Multivariate Behavioral Research*, *51*, 396–412. https://doi.org/10.1080/00273171.2015.1050481

Parrigon, S., Woo, S. E., Tay, L., & Wang, T. (2017). CAPTION-ing the situation: A lexically-derived taxonomy of psychological situation characteristics. *Journal of Personality and Social Psychology*, *112*, 642–681. https://doi.org/10.1037/pspp0000111.supp

Piccirillo, M. L. & Rodebaugh, T. L. (in press). Foundations of idiographic methods in psychology and applications for psychotherapy. *Clinical Psychology Review* https://doi.org/10.1016/j.cpr.2019.01.002

Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica*, *55*, 277–301. https://doi.org/10.2307/1913237

R Core Team (2017). R: A language and environment for statistical computing. Vienna, Austria: Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., & Funder, D. C. (2014). The Situational Eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology*, *107*, 677–718. https://doi.org/10.1037/a0037250

Redhead, S., Johnstone, L., & Nightingale, J. (2015). Clients' experiences of formulation in cognitive behaviour therapy. *Psychology and Psychotherapy: Theory, Research and Practice*, *88*, 453–467. https://doi.org/10.1111/papt.12054

Revelle, W. R. (2018). psych: Procedures for personality and psychological research. [Computer software]. Evanston, IL: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Russell, R. L., Jones, M. E., & Miller, S. A. (2007). Core process components in psychotherapy: A synthetic review of P-technique studies. *Psychotherapy Research*, *17*, 271–288. https://doi.org/10.1080/10503300500529388

Seeboth, A., & Mõttus, R. (2018). Successful explanations start with accurate descriptions: Questionnaire items as personality markers for more accurate predictions. *European Journal of Personality*, *32*, 186–201. https://doi.org/10.1002/per.2147

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, *4*, 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Silvia, P. J., Kwapil, T. R., Walsh, M. A., & Myin-Germeys, I. (2014). Planned missing-data designs in experience-sampling research: Monte Carlo simulations of efficient designs for assessing within-person constructs. *Behavior Research Methods*, *46*, 41–54. https://doi.org/10.3758/s13428-013-0353-y

Smith, C., & Lapp, L. (1991). Increases in number of REMS and REM density in humans following an intensive learning period. *Sleep*, *14*, 325–330. https://doi.org/10.1093/sleep/14.4.325

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288. Retrieved from http://www.jstor.org/stable/2346178

Tschacher, W., & Ramseyer, F. (2009). Modeling psychotherapy process by time-series panel analysis (TSPA). *Psychotherapy Research*, *19*, 469–481. https://doi.org/10.1080/10503300802654496

Tschacher, W., Zorn, P., & Ramseyer, F. (2012). Change mechanisms of schema-centered group psychotherapy with personality disorder patients. *PLoS ONE*, *7*, 1–10. https://doi.org/10.1371/journal.pone.0039687

van de Leemput, I. A., Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., & Scheffer, M. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Science of the United States of America*, *111*, 87–92. https://doi.org/10.1073/pnas.1312114110

van der Krieke, L., Emerencia, A. C., Bos, E. H., Rosmalen, J. G. M., Riese, H., Aiello, M., & de Jonge, P. (2015). Ecological momentary assessments and automated time series analysis to promote tailored health care: A proof-of-principle study. *JMIR Research Protocols*, *4*, 1–12. https://doi.org/10.2196/resprot.4000

Zimmerman, J., Woods, W. C., Ritter, S., Happel, M., Masuhr, O., Jaeger, U., & Spitzer, C. (2018). Integrating structure and dynamics in personality assessment: First steps toward the development and validation of a Personality Dynamics Diary. *Psychological Assessment*. https://doi.org/10.31234/osf.io/5zcth

Received: September 14, 2018
Accepted: February 13, 2019
Available online: 1 March 2019