Journal of Contextual Behavioral Science 40 (2026) 100983

Contents lists available at ScienceDirect

Journal of Contextual Behavioral Science

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

Time series machine learning for idionomic process-based treatment planning: A tutorial on tsBoruta

Baljinder K. Sahdra [a,*], Mercedes G. Woolley [b], Cristóbal Hernández [c], William Li [a], Steven C. Hayes [d], Joseph Ciarrochi [a], Michael P. Twohig [b], Micheal E. Levin [b]

[a] Institue for Positive Psychology and Education, Australian Catholic University, North Sydney, Australia
[b] Department of Psychology, Utah State University, USA
[c] Universidad de los Andes, Chile, Facultad de Ciencias Sociales, Escuela de Psicología
[d] Department of Psychology, University of Nevada, Reno, USA

A R T I C L E I N F O

Keywords:
Time series machine learning
Trichotillomania
Idionomic analysis
Process based therapy
Treatment personalization

A B S T R A C T

Background: This study showcases how three advanced algorithms—iARIMAX, iBoruta, and tsBoruta—identify personalized treatment processes in clinical psychology, using hair-pulling (trichotillomania) as an empirical case.
Method: We compared these methods with previous findings and assessed their ability to detect linear and nonlinear associations. We predicted the methods to converge on cognitive fixation as a core predictor of hair-pulling and expected substantial heterogeneity in process-outcome associations—heterogeneity that, if systematic rather than random, could inform the design of personalized interventions. We also predicted tsBoruta would outperform iBoruta due to its consideration of time series elements.
Results: All three methods confirmed cognitive fixation as a key aggregate level predictor of hair-pulling. While iARIMAX initially showed stronger connections, these became more modest—but still meaningful—when accounting for multiple processes. The Boruta methods showed notable differences in idiographic conclusions, with tsBoruta proving more conservative in confirming significant effects. Notably, 61.11% of participants showed unique combinations of relevant process-outcome links. Targeting three key processes of cognitive fixation, valued action, and anxiety could potentially benefit 52 out of 54 individuals in the sample.
Conclusions: The findings support combining standardized protocols with personalized interventions that may be valuable for trichotillomania treatment. More broadly, this study provides a methodological tutorial and illustrates how tsBoruta offers a powerful, balanced approach to modeling complexity in clinical data for treatment planning.

1. Introduction

Despite decades of progress in developing disorder-specific protocols, the aggregate efficacy of psychological interventions appears to have reached a ceiling. While current protocols are effective for many, effect sizes have not substantially increased over the past three decades (Hofmann et al., 2025). Furthermore, a significant minority of patients, often estimated between 30% and 50%, do not achieve "meaningful improvement," defined here as sustained clinical remission or functional recovery rather than mere statistical change (Loerinc et al., 2015). These challenges reflect a long-standing observation, that people respond differently to psychotherapy, even when they present with similar problems and are treated by the same therapist using the same approach (Kiesler, 1966), highlighting the issue of heterogeneity that standardized protocols fail to account for (Varadhan et al., 2013).

Individual level differences in treatment response have renewed interest in personalized interventions. There is a growing movement to return to principles of functional analysis (cf., Hayes, 1981; Hayes et al., 2021), this time with an emphasis on creating replicable, research-supported, and scalable methods. Indeed, evidence is accumulating that tailoring treatment to the individual can lead to improved outcomes (Li et al., 2024; Nye et al., 2023).

This article is part of the following JCBS special issue: Beyond One-Size-Fits-All: Tutorials in Methods for Personalized Treatment and Evaluation.

* Corresponding author. Institute for Positive Psychology and Education, Australian Catholic University, PO Box 968, North Sydney, NSW 2059, Australia.
E-mail address: baljinder.sahdra@acu.edu.au (B.K. Sahdra).

https://doi.org/10.1016/j.jcbs.2026.100983
Received 4 September 2025; Received in revised form 16 January 2026; Accepted 17 February 2026
Available online 1 March 2026

An intervention's efficacy hinges on whether it alters the psychological or behavioral processes that are directly linked to the outcome (Kok et al., 2016). To ensure conceptual clarity, we distinguish between three related terms used in Process-Based Therapy (Hayes et al., 2022). A *process* refers to a dynamic, theoretically derived biopsychosocial variable (e.g., cognitive fixation, anxiety) that is hypothesized to influence the outcome. A *process of change* is a process that has been empirically demonstrated to mediate or predict an outcome of interest (i.e., a statistically significant predictor in our time-series models). *Principles* refer to the theoretical guidelines or heuristic rules (e.g., behavioral principles or functional analyses) used to select and influence these processes. The current study focuses on measuring *processes* to identify which ones function as *processes of change* for specific individuals. This means identifying the biopsychosocial processes most relevant to a given person, couple, or family that are functionally linked to the therapeutic outcome of interest, in the context of their goals and environment, and then determining how to shift those processes efficiently (Hofmann & Hayes, 2019). However, the field continues to rely primarily on aggregate-level statistical models that obscure meaningful within-person variability that would guide treatment personalization, making it difficult to empirically isolate the most relevant treatment targets for particular people.

This limitation is not just practical, but conceptual. It reflects the ergodicity problem: the mistaken assumption that average effects across individuals can be generalized to individuals themselves (Molenaar & Campbell, 2009). For this assumption to hold, two conditions must be met: psychological processes must be (1) stable within individuals over time, and (2) structurally identical across individuals (Molenaar & Campbell, 2009). In clinical contexts, neither assumption is tenable. Psychological processes are dynamic, context-dependent, and shaped by personal histories. As a result, what is often treated as noise in aggregated models may, in fact, be the very signal that matters most for particular people (e.g., Catts et al., 2025; Sahdra et al., 2023; Sahdra et al., 2024; Sahdra et al., 2025).

To advance towards empirically based and replicable idiographic methods to identify personalized processes as treatment targets, we are in urgent need of innovative data-analytic methods, particularly ones that are user-friendly for clinicians and ones that prioritize specificity over sensitivity (Hayes et al., 2019; Hofmann & Hayes, 2019). In this context, sensitivity refers to the ability to correctly identify relevant processes (true positives), whereas specificity refers to the ability to correctly reject irrelevant processes (true negatives). Prioritizing specificity ensures that interventions focus on processes truly relevant to the client's presenting problems, minimizing the risk of false positives, where irrelevant processes are mistakenly targeted, potentially leading to wasted therapeutic effort and lack of intervention focus. Conceptually, a “truly relevant process” is defined as a biopsychosocial variable that is functionally linked to the therapeutic outcome for a specific individual (Kok et al., 2016). Not all theoretically relevant processes are truly relevant in this sense. This is exemplified by recent findings in trichotillomania where theoretically sound processes, such as worrying about pulling, were found to be statistically unrelated to actual pulling behavior for the majority of participants (Woolley et al., 2025). Consequently, an “irrelevant process” is one that accounts for no unique variance in the outcome for a given person, regardless of its theoretical plausibility.

In clinical contexts, high specificity reduces unnecessary or ineffective interventions, preventing increased costs and potential client disengagement (Haynes & Williams, 2003). This need for specificity also helps explain why “one-size-fits-all” protocols, defined here as standardized interventions applied uniformly to all persons sharing a diagnostic label, often fall short. Such approaches are unable to account for the “unique constellation of processes,” or the distinct, person-specific combination of interacting biopsychosocial variables, driving a client's difficulties. Failure to address this idiographic complexity results in treatments that may be too broad or misaligned to be effective for many clients. While sensitivity is important to avoid missing relevant processes, the clinical costs of low specificity, undermine the model of process-based therapy (PBT), in which parsimony and precision are central to individualized care (Hayes et al., 2019).

Therefore, to advance personalized therapeutic frameworks like PBT, researchers have taken to the development of idionomic methods: approaches that model biopsychosocial processes at the level of particular people (individuals, couples, families) while incorporating aggregate-level information only when it enhances that understanding (Hayes et al., 2022). One such approach is Group Iterative Multiple Model Estimation (GIMME; Beltz & Gates, 2017). GIMME is based on unified Structural Equation Modeling (uSEM), which integrates features of SEM and time-series analysis. It begins by identifying relationships between variables within each individual's time series. Only after establishing person-level models does it identify shared (group or subgroup) connections, retaining only those that improve model fit for most people.

Researchers are now using GIMME to identify individual-level treatment targets and plan interventions within a PBT framework (e.g., Sanford et al., 2022; Woolley et al., 2025). For example, Woolley et al. (2025) applied GIMME to trichotillomania, a psychiatric disorder involving hair pulling at a rate that causes damage and negatively affects functioning. Their analysis revealed a common pattern across participants, in which sensory seeking and cognitive fixation on the urge to pull were functionally linked to hair-pulling. Sensory seeking referred to pulling in order to create or relieve particular tactile sensations in the hair or scalp, whereas cognitive fixation captured getting mentally ‘stuck’ on repetitive thoughts about urges to pull. Together, these processes appeared to help maintain the hair-pulling behavior for a majority of participants. However, aside from these processes, their study identified many person-specific differences in dynamic process relationships, revealing how a therapist may choose different processes to target based on networks for each individual. These results underscore the value of modeling change processes at the individual level to inform treatment.

Still, GIMME has several limitations (Gates & Molenaar, 2012). It does not include automated feature selection, requiring researchers to manually decide which variables to include, which is done often based on theory, group-level findings, or manual reduction procedures. In Woolley et al. (2025), this resulted in both excluding processes that might have been relevant to some individuals (due to limitations in number of variables to include in a model) as well as the inclusion of processes into individual networks that may not have been relevant for all individuals. For instance, the process of worrying about one's pulling had a direct (contemporaneous or lagged) connection to hair-pulling in only 12 of 54 participants. Moreover, based on centrality metrics, it was a highly influential process for only three participants. This suggests that manual item selection can lead to the inclusion of processes that add noise rather than clarity.

Another limitation is that GIMME assumes linear relationships between variables and does not account for interactions, nonlinear effects, or time-varying predictors (Beltz & Gates, 2017). Such features are common in psychological data. It also assumes that data are collected at equally spaced intervals, which is rarely the case with EMA, and that there are no systematic trends in the data, implying that variables do not change over time (Lane et al., 2019). Yet in clinical contexts, change is the point. As such, researchers must detrend the data prior to analysis, a process that can be challenging for clinicians without statistical training. Finally, although GIMME depicts directional arrows between nodes, the contemporaneous relationships it estimates are not strictly causal. This can make it difficult to interpret whether a process leads to an outcome or vice versa, further limiting the model's utility for clinical decision-making.

An alternative method that addresses some of these challenges is iARIMAX: idiographic Autoregressive Integrated Moving Average with Exogenous Variables (Sahdra et al., 2024; Ciarrochi et al., 2024).

iARIMAX models time-series data for each individual separately, identifying how a given process predicts a specific outcome over time. It extends traditional ARIMA models by incorporating predictors (the "X"), differencing the data (the "i" component) to handle non-stationarity (d), and modeling both autoregressive effects (p) and moving average components (q) to account for complex autocorrelation structures. This makes it more flexible than methods like GIMME, which typically control for only first-order autoregression (AR(1)) and do not address differencing or moving averages.

iARIMAX aligns with the aims of PBT by detecting time-sensitive links between change processes and outcomes applicable to particular people (Sahdra et al., 2024). After modeling each person separately, results can be aggregated using random-effects meta-analysis to estimate overall group-level effects while preserving the ability to also detect individual differences (Ciarrochi et al., 2024). However, iARIMAX is bivariate, modeling only one predictor and one outcome at a time, which restricts its ability to capture complex, interacting psychological systems. It also assumes linearity (Box et al., 2015) and may overlook non-linear or context-dependent dynamics.

Decision tree-based approaches may offer solutions to many of the limitations described previously. One such approach is Boruta, a feature selection algorithm based on random forests, a machine learning method that builds many decision trees and combines their results to improve prediction accuracy (Kursa & Rudnicki, 2010). A decision tree works like a flowchart: it asks a series of questions about different variables and makes predictions based on how these variables branch. From a process-based clinical angle, this means a tree might model how processes like avoidance or fatigue relate to outcomes such as sadness, stress, or hair-pulling. Each tree captures one possible way those relationships unfold, and random forests summarize the results of many such trees to yield more robust insights. Boruta uses this framework to test whether each variable adds meaningful information by comparing its importance to the best of the randomized shadow features (shuffled copies that destroy any genuine link to the outcome), thus reducing false positive errors and problems of over-fitting (where a model learns noise instead of the underlying pattern). Importantly, Boruta is not a bivariate method. It considers all potential predictors simultaneously in a multivariate framework, allowing it to identify the most important features in the context of all others.

Originally developed for fields like environmental science and economics (e.g., Gholami et al., 2021; Szul et al., 2021), Boruta has more recently gained traction in psychological research. It has been used to validate the Process-Based Assessment Tool (PBAT) by identifying which psychological processes most strongly predict outcomes such as stress and sadness (Ciarrochi et al., 2022; Larsson & Sundström, 2024). Boruta has also helped identify health-related factors (e.g., fatigue, BMI) that predict risk for depression and anxiety (Yang et al., 2025). Notably, Boruta has shown promise as a robust feature selection tool for identifying important predictors in complex, high-dimensional datasets (Kursa & Rudnicki, 2010). Although it has not been widely applied to individual-level intensive longitudinal data, its methodological strengths make it a natural candidate for adaptation into individualized feature selection approaches.

When applied separately to individuals in a sample, this approach is referred to as iBoruta (Li et al., 2026). It enables precise, person-specific identification of treatment-relevant processes without assuming linearity or requiring manual item selection. Compared to methods like GIMME or iARIMAX, iBoruta offers a scalable (easily adaptable to large numbers of participants), flexible framework that automatically selects key variables, captures complex interactions, and handles many predictors. It does so even in data that is noisy (containing random error) or high-dimensional (containing many potential predictors), making it well-suited to uncover individualized psychological process–outcome relationships. Despite its strengths in identifying nonlinear effects and operating without normal distribution assumptions (Kursa & Rudnicki, 2010), iBoruta has a critical limitation: it does not account for temporal dependencies inherent in time-series data. The algorithm assumes that observations are independent and identically distributed, a condition violated in EMA designs, where data are often autocorrelated and shaped by underlying trends. As a result, iBoruta may incorrectly identify some predictors as important, not because they truly relate to the outcome, but because the algorithm does not account for patterns in the data that unfold over time, like trends or autocorrelations (Shumway & Stoffer, 2017). These unmodeled time-based data structures can create misleading associations, which lowers the model's specificity and makes it harder to identify truly meaningful treatment targets.

To address this limitation, Li et al. (2026) developed tsBoruta, a hybrid model combining ARIMA-based time-series preprocessing with iBoruta's feature selection algorithm. The approach first applies ARIMA modeling (autoregressive, differencing, and moving average components) to each predictor to remove temporal dependencies like autocorrelation and trends. It then applies iBoruta to the residuals, which isolates non-temporal variance (the fluctuation in the data that remains after accounting for time-based trends) and enhances detection of meaningful process–outcome relationships.

In a simulation study with synthetic datasets, tsBoruta showed substantially higher specificity (fewer false positives) and precision (greater accuracy among selected predictors) than standard iBoruta across various conditions (Li et al., 2026). These results support tsBoruta as a conservative tool for isolating person-specific predictors in time-series data, particularly when minimizing false positives is clinically important. Our study aims to be among the first to test tsBoruta in a clinically relevant sample, extending beyond previous work that relied primarily on simulated data and nonclinical EMA datasets. This represents a methodological advancement in process-based therapy (PBT) idionomics, using an EMA dataset of biopsychosocial processes from adults with trichotillomania as a test case. From an idionomic perspective, it's essential to assess whether tsBoruta maintains its analytic integrity with real-world clinical data.

The present study examined these methods using trichotillomania as a clinically relevant example. Trichotillomania is a heterogeneous condition, that is, its features vary significantly across individuals. It is characterized by diverse pulling behaviors and a wide range of triggers, spanning from affective states to sensory regulation needs (Hicks et al., 2023; Grant et al., 2021; Woolley et al., 2025). To move toward process-based and personalized interventions that better address the heterogeneous nature of the disorder, Woolley et al. (2025) applied GIMME to identify both individual- and group-level network processes associated with hair pulling as a clinical outcome. In the present work, we build on these efforts by exploring additional idionomic methods to further clarify how person-specific treatment targets might be identified. This work not only contributes to the growing personalized treatment literature for trichotillomania but also offers broader insights into methodological approaches that may be applied across a range of psychological challenges. It also provides an accessible step-by-step tutorial for implementing tsBoruta. To further facilitate accessibility for readers new to these methods, we have provided a glossary of key methodological terms in Appendix A, and the tutorial code is provided in Appendix B.

1.1. The current study

The impact of idionomic research in clinical psychology depends on having multiple well validated methods (Creswell & Creswell, 2017; Hayes et al., 2022). In this study, we specifically compare three distinct algorithmic approaches: (1) iARIMAX, representing the standard linear time-series approach; (2) iBoruta, representing a machine learning approach capable of detecting nonlinearities but ignoring temporal dependency; and (3) tsBoruta, a novel hybrid algorithm designed to address the limitations of the former two. To establish the validity of iARIMAX, iBoruta, and tsBoruta, we will benchmark their nomothetic findings against the results previously obtained using GIMME on this same dataset (Woolley et al., 2025). We will examine how well iARIMAX, iBoruta, and tsBoruta identify both general predictors for trichotillomania and individual-specific processes, with the goal of informing a future treatment framework that combines aggregated and personalized processes. Through the novel machine learning approach of tsBoruta for analyzing time series data, we aimed to inform both general therapeutic principles and individualized interventions with adequate sensitivity and specificity.

In the context of this study, we operationally defined “utility” of a Boruta method based on three criteria: (1) *Nomothetic Convergence*, or the ability to replicate previously established group-level findings (validity check); (2) *Idiographic Performance*, defined by the balance of sensitivity and specificity relative to our best guess based on ARIMAX estimates; and (3) *Clinical Applicability*, defined as the capacity to characterize heterogeneity by identifying unique combinations of treatment targets. We further defined “adequate” performance based on the clinical necessity of parsimony. Given the cost of targeting irrelevant processes, we established that a method works “well” in this context if it prioritizes specificity (minimizing false positives) over sensitivity, thereby ensuring that identified targets are robustly linked to the outcome.

Our goal was to see whether general findings could inform standardized treatment protocols while individual findings could guide protocol customization to match each client's needs. We expected tsBoruta to be a promising method that balances sensitivity and specificity, suggesting its suitability for idionomics-based treatment planning. Our pre-registered research questions were as follows:

1. Do the algorithms of iARIMAX, iBoruta, and tsBoruta converge in identifying similar nomothetic conclusions about processes linked to the hair-pulling outcome?
2. Do their nomothetic conclusions align with previously known effects from GIMME using the same data?
3. To what extent do these methods differ in identifying idiographic conclusions about processes linked to the outcome?
4. How do these methods compare in assessing the importance of processes for outcomes with linear effects? Specifically, which methods are more reliable at detecting linear associations?
5. How do these methods compare in identifying nonlinear effects? Specifically, which methods are more effective at detecting nonlinear links between processes and outcomes?
6. To what extent are the bivariate associations between processes and outcomes heterogeneous?
7. To what extent are the multivariate associations between processes and outcomes heterogeneous? Specifically, as different numbers of processes may be relevant for different individuals, can the new tsBoruta algorithm help characterize heterogeneity in terms of which processes are clinically relevant across people?

We defined “differences” between methods as the divergence in classification decisions (confirmed vs. rejected), quantified using confusion matrices and McNemar's tests. Regarding linear effects (RQ4), we defined “reliable” and “effective” based on classification performance metrics of Sensitivity (Recall), Specificity, and Precision, treating statistically significant iARIMAX estimates as the baseline for linear detection. Regarding nonlinear effects (RQ5), “effectiveness” was operationalized as the probability of a method confirming a relationship that demonstrated a significant quadratic trend, analyzed via multilevel logistic regression. Finally, “heterogeneity” was operationalized using the I^2 statistic for bivariate associations (RQ6) and the frequency of unique process combinations for multivariate associations (RQ7).

Our pre-registered hypotheses were as follows: Based on the previously identified GIMME-based nomothetic link between cognitive fixation and the urge to pull (Woolley et al., 2025), we expected all methods to converge on the same nomothetic conclusions. Based on recent simulations (Li et al., 2026) which found that iARIMAX outperformed other methods including GIMME, indSEM, and iBoruta at identifying important processes for outcomes for linear effects, we expected iARIMAX to outperform other methods in characterizing the importance of processes for linear effects in the current study. In contrast, the simulations showed that when identifying nonlinear effects, iBoruta and tsBoruta were able to identify these effects whilst other methods did not. Consequently, we expected tsBoruta to outperform other methods in identifying non-linear effects. We expected to find high heterogeneity in the bivariate associations between processes and outcomes. Given the high degree of non-random heterogeneity in the network structures identified by Woolley et al. (2025), we expected tsBoruta to identify varying numbers of processes as important for predicting the outcome across different individuals. This would provide a novel index for characterizing heterogeneity as variability in the number of important variables across people.

2. Method

The study was pre-registered on Open Science Framework (OSF): link blinded for peer review: https://osf.io/2yuw9/overview.

2.1. Participants

This study used the sample reported in Woolley et al., 2025. Eligibility criteria included: (a) meeting DSM-5 diagnostic criteria for trichotillomania, (b) 18 years of age or older, (c) fluency in English, (d) residence in the United States, and (e) access to a smartphone. Recruitment was conducted through online platforms such as Reddit and Facebook. Recruitment advertisements informed participants that the study involved completing a 30-day intensive series of surveys (three per day for 30 days) designed to help researchers understand moment-to-moment patterns and trajectories of hair-pulling behavior. Participants were also informed they would partake a brief virtual interview to assess the present of trichotillomania according to DSM-5 criteria. In total, 224 individuals completed a preliminary screening survey to assess eligibility and hair-pulling behaviour.

2.2. Procedures

All procedures were approved by a university Institutional Review Board and followed the protocol established in Woolley et al. (2025). Of the individuals who completed the online screener, 134 met eligibility criteria and were contacted via email to schedule a virtual intake session. Fifty-eight individuals consented and completed the intake, which included the trichotillomania module of the Diagnostic Interview for Anxiety, Mood, OCD, and Neuropsychiatric Disorders (DIAMOND; Tolin et al., 2018). One participant did not meet diagnostic criteria, resulting in a final sample of 57. During intake, participants completed a sociodemographic questionnaire and received instructions on downloading and using *Avicenna*, a smartphone application used to deliver ecological momentary assessment (EMA) prompts.

Survey prompts began the following day and were delivered three times daily over a 30-day period (90 prompts in total). Prompts were randomly administered during the following windows: 9:00 a.m.–1:00 p.m., 1:00 p.m.–5:00 p.m., and 5:00 p.m.–9:00 p.m. Participants were given 2.5 h to respond to each prompt, with a reminder issued 30 min before expiration. Personalized reminders were sent weekly to those completing fewer than 67% of surveys. Participants received $50 for completing 60 or more surveys and $15 if they completed fewer than 60 across the study period.

2.3. Measures of psychological processes assessed via EMA

The EMA items used in the current study were originally developed for Woolley et al. (2025). Items were guided by the Extended Evolutionary Meta Model (EEMM; Hayes et al., 2020) and informed by established treatments for trichotillomania, such as, habit reversal training (HRT), Dialectal Behavior Therapy (DBT), and Acceptance and Commitment Therapy (ACT), specifically, processes these therapies aim to modify. The final EMA measure included 16 items: one outcome item assessing the number of minutes spent hair pulling since the previous prompt, and 14 process items spanning six domains: affect, cognition, sensory reinforcement, attention, values, and behaviour. The final item assessed perceived busyness between prompts.

Items were rated on a 5-point Likert scale (1 = Not at all to 5 = Extremely). In response to each item, participants were asked to consider how they felt since the last prompt. Affect items assessed emotional states that may trigger or maintain pulling, such as feeling anxious or bored. Cognition items captured maladaptive thoughts and cognitive fixation on urges, including worry that urges would not go away or simply becoming focused on an urge to pull. Sensory reinforcement items reflected the urge to engage in tactile stimulation or sensory exploration of hair, such as feeling compelled to touch or manipulate hair. Attention items measured awareness of internal or external cues linked to pulling, including noticing mirrors or thoughts like "I can't cope without pulling." Values items assessed motivation to act in line with personally meaningful goals rather than pulling, such as focusing on long-term benefits or attending to important life priorities. Finally, behavior items captured overt strategies to reduce pulling or manage urges, such as doing something else with one's hands or avoiding situations that increase urges to pull. Representative item examples include: affect ("I felt anxious"), cognition ("I got fixated on an urge to pull"), sensory reinforcement ("I felt the need to explore sensations related to my hair"), attention ("I was aware of things in my environment that usually lead me to pull"), values ("I attended to things more important than pulling"), and behavior ("I tried to ride out the urge by doing something else"). The within-person reliability of items across the study period was excellent (ICC(2)s = 0.93–0.99).

2.4. Analysis plan

2.4.1. Overview of the analytical strategy

Our analysis strategy employed iARIMAX, iBoruta, and tsBoruta to examine both group-level and individual-level patterns. We systematically assessed methodological convergence across these three approaches and with previous GIMME results, while evaluating idiographic divergence in identifying individual-specific processes with detailed performance metrics for clinical applicability. The analysis compared how the methods detected linear versus nonlinear relationships between processes and outcomes, characterized heterogeneity in both bivariate and multivariate process-outcome associations, and identified minimal combinations of processes that could potentially benefit the majority of individuals in the sample. This multi-faceted strategy allowed for systematic examination of both commonalities across individuals and unique process patterns within individuals, providing valuable insights for personalized treatment approaches in trichotillomania.

2.4.2. iARIMAX

iARIMAX is an idiographic adaptation of the Autoregressive Integrated Moving Average (ARIMA) model with an eXogenous covariate (Ciarrochi et al., 2024; Hyndman, 2025) designed to examine bivariate or multivariate time-series relationships for one individual at a time. An exogenous covariate in this case refers to an external predictor variable, such as anxiety predicting hair pulling. iARIMAX builds on ARIMA, which filters a person's timeline into three components: stable patterns from past history predictive of current values, long-term trends, and random noise. By separating these elements, the model isolates meaningful signals from random fluctuations. The method adds two key features to this framework. The "X" represents external predictors, allowing researchers to see how outside factors influence outcomes while controlling for a person's natural trends and patterns. The "i" stands for "idiographic," meaning the analysis is performed separately for each person to capture their specific dynamics, whether examining one influence or multiple predictors. Once individual models are calculated, results are combined through meta-analysis to create a group estimate that respects the unique differences between individuals. For technical details and equations, see Li et al. (2026).

i-ARIMAX (Ciarrochi et al., 2024) is an extension of the ARIMAX algorithm that estimates models one individual at a time. The individual-level estimates of the parameters of the exogenous regressor (process-outcome relationship) are then aggregated using random-effects meta-analysis to generate pooled estimates of effect size for each process-outcome link, their 95% confidence intervals and prediction intervals, and estimates of heterogeneity, I^2. The I^2 statistic indicates the proportion of observed variance in effect sizes that reflects true differences between particular persons rather than true sampling error; higher I^2 values (e.g., at or above 75%) suggest considerable non-random heterogeneity in how the process-outcome relationship manifests across different people. However, I^2 should be interpreted contextually alongside other factors such as effect size magnitude, confidence intervals, and prediction intervals, rather than relying on fixed cutoffs (Borenstein et al., 2017). This approach allowed us to account for heterogeneity in effect sizes across individuals and produce a group-level estimate that retained sensitivity to individual variation. In the current study, the bivariate iARIMAX was implemented using the *idionomics* package in R (Hernández et al., 2025) that creates a wrapper around the auto.arima function of the *forecast* package (Hyndman, 2025). The multivariate version of i-ARIMAX was implemented by iteratively calling the auto.arima function of the *forecast* package.

2.4.3. iBoruta

iBoruta is a wrapper around the Boruta feature selection algorithm (Kursa & Rudnicki, 2010), adapted for idiographic analysis by applying Boruta separately to each participant's data. Unlike traditional nomothetic approaches that pool data, iBoruta identifies person-specific predictors by evaluating the importance of each feature in predicting a given outcome. It differs from iARIMAX, which models linear bivariate relationships independently, by considering all features simultaneously and detecting nonlinear or interaction effects that might otherwise be missed.

Boruta is a feature selection algorithm, and it is better considered as a wrapper around the Random Forest algorithm. Random Forest calculates a defined number of decision trees, each of them differing in the input variables, and the instances (or rows) used for training to decrease overfitting (i.e., helping the model generalize to unseen data). In the context of classification, the ensemble of different trees are estimated and then the final classification is defined by the majority. In Random Forest, one way to estimate feature importance is by calculating the model prediction accuracy in an unseen dataset (out of bag sample or OOB) with the original model and then shuffling a particular feature (e.g., valued action) in the OOB accuracy calculation. The difference of accuracy (subtraction) between the trained tree with the original feature, and the same tree with a shuffled feature is calculated, as it serves as an indication of feature importance. It answers the question: How accurate is the model if that particular feature was not related to the outcome? Consequently, this measure is called "loss of accuracy", and it is usually transformed into a Z-score, by dividing the average loss across trees by its standard deviation.

Boruta takes advantage of this mechanism and adds another layer to identify relevant features. Before initializing the Random Forest algorithm, it creates shuffled copies of every feature. As the features are shuffled at random, they also lose their relationship with the dependent variable (e.g., hair-pulling). Then, all original and shadow features are used to estimate a random forest algorithm and the Z score of the accuracy loss is calculated for the original and the shadow features. As shadow features are already shuffled versions, they are reshuffled in the calculation of the accuracy loss too, just like their original counterparts.

Then, the key aspect of Boruta feature selection happens: Every original feature's importance Z score is compared to the Z score of the most important shadow feature (or MZSA for maximum Z score among shadow attributes) answering the question: Is a particular feature (e.g., valued action) more important than the best that one can expect at random? If the answer is yes, because it is more important, then the algorithm registers a hit. This procedure is done for each feature in the model. Following a statistical threshold (0.01 by default) and under a binomial distribution, it then calculates whether a feature's hit count (or number of times it was more important than its shadow) is significantly higher or lower than expected by chance. Those features that are statistically more frequently higher than the MZSA are considered as "confirmed", those that are more frequently lower than the MZSA are considered as "rejected" and deleted from all subsequent runs, and those that are not rejected or confirmed are considered as "tentative". The algorithm iterates using all non-rejected variables, until importance has been assigned to all features, or when a set number of iterations is reached (Kursa & Rudnicki, 2010). In the current study, 500 iterations were used to reduce tentative classifications.

By comparing predictors against shadow features across many iterations, iBoruta offers a permutation-based method for identifying statistically significant predictors, even in high-dimensional or noisy psychological data. Its flexibility, lack of distributional assumptions, and ability to detect complex patterns make it particularly suitable for idiographic psychological research. The current study implemented iBoruta using the *Boruta* package in R (Kursa & Rudnicki, 2023).

2.4.4. tsBoruta

Time Series Boruta (tsBoruta) is a hybrid approach that combines the strengths of iARIMAX and iBoruta. While iBoruta excels at identifying important predictors through robust feature selection, it does not account for temporal dependencies in time series data. Conversely, iARIMAX effectively models these dependencies using autoregressive (AR), integrated (I), and moving average (MA) components—making it ideal for handling autocorrelations and trends. tsBoruta involves first fitting an ARIMA model to each predictor for each individual using the auto-arima function (Hyndman & Khandakar, 2008). This process removes temporal structures through differencing (I) to address trends and models AR and MA components to account for autocorrelation and moving averages (Box et al., 2015). The residuals—variance not explained by temporal dependencies—are then extracted from the ARIMA model and fed into iBoruta. iBoruta then assesses each predictor's importance, one individual at a time, by comparing its performance against randomly shuffled "shadow" predictors, retaining only those with statistically robust predictive value (Kursa & Rudnicki, 2023).

By applying iBoruta to ARIMA-derived residuals, tsBoruta effectively filters out temporal noise and isolates meaningful predictor variance, enhancing both interpretability and accuracy of feature selection in time series data (Li et al., 2026). The implementation of tsBoruta in our study used the *forecast* package in R for ARIMA models (Hyndman, 2025) and the *Boruta* package for iBoruta (Kursa & Rudnicki, 2023). Appendix B contains a brief tutorial on tsBoruta, including step-by-step instructions for implementing the method in R.

2.4.5. Statistical power and sample size

Statistical power in idionomic analyses differs fundamentally from traditional cross-sectional methods such as Structural Equation Modeling (SEM). While SEM relies on between-person variance and typically requires large participant samples (e.g., $N > 200$) to achieve adequate power to estimate a large number of parameters, the iARIMAX and tsBoruta methods we employed derive power primarily from within-person variance and a focus on predicting a single outcome variable (Ciarrochi et al., 2024; Li et al., 2026). Consequently, the effective sample size for the primary model estimation is defined by the density of longitudinal observations (T) rather than the number of participants (N). Recent simulation studies indicate that for these algorithms, a within-person sample size of 50 to 70 time points offers a reasonable balance between sensitivity and specificity (Li et al., 2026). These simulations further suggest that while estimation is possible with as few as 30 points, samples larger than 150 time points may paradoxically reduce specificity and increase false positives. In the current study, the protocol included up to 90 ecological momentary assessment (EMA) prompts per participant. This density falls squarely within the relatively optimal range identified by Li et al. (2026), providing sufficient degrees of freedom to reliably estimate person-specific parameters while minimizing the risk of overfitting associated with excessively long time series. The participant count of 54 serves as the sample size for the second-stage random-effects meta-analysis, a number that is sufficient for aggregating individual effect sizes and estimating heterogeneity in intensive longitudinal designs (Ciarrochi et al., 2024; Sahdra et al., 2024).

3. Results

Analyses were conducted in R version 4.5.0 (R Core Team, 2025). The R code scripts and simulated data are available on OSF project: https://osf.io/qe7vs/overview. Due to IRB restrictions, we cannot provide open access to the raw data. To help other researchers use our R code, we created simulated data based on the original dataset. Using the MTS package (Tsay et al., 2022), we fit a first-order Vector Autoregressive (VAR) model to each participant's real data to capture their unique temporal and contemporaneous patterns. Since the simulated data differs from the real clinical data, results obtained from it will vary somewhat from those reported in this paper.

3.1. Preliminary analyses

Woolley et al. reported 22% missingness of all their data before they winnowed the processes to the final nine processes they used for GIMME. This study focused on the nine processes and the hair-pulling outcome variable. For these 10 variables, 18% of the data were missing (see Fig. S1 in Supplemental File), which we imputed using the copy-mean longitudinal imputation (Genolini et al., 2013).

In preliminary analyses, we compared linear and nonlinear (quadratic) regression models for each process-outcome association per person, resulting in 486 comparisons (54 participants $\times$ 9 process-outcome relationships). A quadratic trend suggests that a variable, such as hair pulling, may change at varying rates—either accelerating at a specific timepoint or following a pattern of increasing then decreasing, creating a curve-like pattern. These models provided a straightforward way of examining linearity idionomically through idiographic model comparisons followed by aggregated proportions. We calculated both the overall proportions across all 486 models and separate proportions for the 9 models per participant to assess the extent of linearity in the process-outcome associations. The details can be found in the "Linearity Exploration" section in the Supplemental File. As shown in Table S1, the process variable of cognitive fixation on the urge to pull showed the highest proportion of individuals with nonlinear (quadratic) relationships (29.63% of participants), and anxiety showed the lowest proportion of individuals with nonlinear relationships (14.81%). Fig. S2 shows the distribution of nonlinear relationships across individuals in the sample, depicting that while some individuals had nonlinear links for several variables whereas others had for fewer or none (range: 0% to 66.7% associations being nonlinear).

3.2. Main models

To test our pre-registered research questions, we conducted three kinds of idiographic models: iARIMAX, iBoruta, and tsBoruta. We ran 486 iARIMAX models in total, nine models for each of the processes predicting the hair-pulling outcome variable in each of the 54 participants. We next conducted 54 iBoruta and tsBoruta models each using the algorithms described in detail in the Method section. We merged the results of all the models to test our pre-registered research questions and hypotheses.

3.2.1. Methodological consistency in nomothetic findings

For Research Questions 1 and 2, we investigated whether iARIMAX, iBoruta, and tsBoruta algorithms identified consistent aggregate-level patterns related to hair-pulling outcomes, and how the findings from these three algorithms compared with previously established GIMME results using the same dataset. We hypothesized convergence across all approaches in identifying the nomothetic relationship between cognitive fixation on pulling urges and actual hair-pulling behavior, as previously demonstrated by GIMME analysis (Woolley et al., 2025).

Our hypothesis was largely confirmed. In their GIMME model, Woolley et al. (2025) identified only one process, cognitive fixation, that had a nomothetic link to the hair-pulling outcome. They reported its average beta weight of 0.45 ($SE = 0.04$). In our study, the pooled effect of iARIMAX estimates was 0.59 ($SE = 0.02$). Table 1 reports the results of the RE-MA of iARIMAX estimates for all processes predicting the outcome. The previously reported lower effect size of the GIMME estimate for cognitive fixation is understandable given that GIMME took into account other processes in the model whereas iARIMAX estimates reflect bivariate associations. When we ran a multivariate iARIMAX model (separately for each individual) where all processes were entered together as exogenous variables, thus controlling for each other, the pooled effect estimate for cognitive fixation was attenuated as expected: 0.32 ($SE = 0.03$) (See Fig. S3 for a comparison of the distributions of iARIMAX and multivariate iARIMAX estimates.). In examining the number of cases where the iARIMAX estimate showed the opposite sign of the pooled effect, only the cognitive fixation model yielded zero such cases, as shown in Table S2 in Supplemental File. That is, the iARIMAX effect of cognitive fixation on hair pulling was universally positive in this sample.

We next compared the extent to which the iBoruta and tsBoruta algorithms converged in identifying cognitive fixation's nomothetic link with hair-pulling. As shown in Table S3, there was a 100% agreement between the decisions of iBoruta and tsBoruta regarding cognitive fixation, whereas the rate of agreement for other processes varied from 61% to 91% (See Fig. S4 for a comparison of the distributions of iBoruta and tsBoruta importance scores.).

In examining the degree of agreement between iARIMAX and Boruta-based methods, the results revealed that for the 54 individuals in the data, only one person had statistically non-significant iARIMAX estimate for cognitive fixation on the urge to pull. The direction of the effect for all 53 statistically significant effects was positive, as shown in the histogram in Fig. S5. For the one person who had a statistically non-significant iARIMAX estimate for cognitive fixation, both Boruta-based methods rejected cognitive fixation. Of the 53 individuals for whom iARIMAX estimate for cognitive fixation was statistically significant, both iBoruta and tsBoruta confirmed that process as an important variable for 49 persons (92.45%), rejected in 3 cases (5.67%), and yielded a tentative decision for 1 case (1.89%). As shown in Table S4, a statistically significant chi-square test suggested a significant non-independence (convergence) of the methods' decisions. In sum, the three methods of iARIMAX, iBoruta, and tsBoruta converged in identifying cognitive fixation on the urge to pull as an important nomothetic predictor of hair pulling, which is in line with the GIMME-based findings reported by Woolley et al. (2025).

3.2.2. Methodological divergence in idiographic findings

For our Research Question 3, we asked the extent to which iARIMAX, iBoruta, and tsBoruta differed in identifying idiographic conclusions about processes linked to the outcome of hair pulling. For 87.03% of the individuals, the two algorithms of iBoruta and tsBoruta diverged in the decision of at least one of the nine processes. Fig. S6 in Supplemental File contains an example case where the two forms of Boruta converged. Fig. 1 contains an example case where the two forms of Boruta yielded different results. In this example case, the two algorithms diverged for 5 of the 9 processes such that the tsBoruta algorithm rejected 5 processes that were confirmed by iBoruta.

We next compared iBoruta and tsBoruta in cases where the iARIMAX estimate was statistically significant (Table S5) or non-significant (Table S6). For all the 344 effects where the iARIMAX estimate was statistically significant, the pattern of agreement between iBoruta and tsBoruta was as follows: 229 (66.57% of 344 effects) were confirmed by both iBoruta and tsBoruta, 265 (77.03%) confirmed by iBoruta, and 224 (65.11%) by tsBoruta (See Table S5 for the full confusion matrix and chi-square test.). For the 142 effects where iARIMAX was statistically non-significant (see Table S6), only 3 effects (2.11%) were confirmed by both iBoruta and tsBoruta, 28 effects (19.72%) were confirmed by iBoruta, and 9 effects (6.34%) were confirmed by tsBoruta. Both methods rejected 94 cases (66.20% of 142 cases total). iBoruta rejected 105 cases (73.94%) and tsBoruta rejected 124 (87.32%). In sum, tsBoruta was more conservative than iBoruta. tsBoruta confirmed 65.11%, compared to iBoruta confirming 77.03%, of the effects that had significant iARIMAX estimates. But it rejected 87.32%, compared to iBoruta rejecting 73.94%, of the effects that had non-significant iARIMAX estimates.

Next, we calculated model performance metrics where the assumed ground truth was whether the iARIMAX coefficient was statistically significant. To evaluate the general utility of the algorithms, we aggregated the classification results across all individuals. We calculated a confusion matrix for iBoruta and tsBoruta using only the confirmed and rejected cases. It is important to note that while the underlying models were estimated idiographically, these metrics represent a nomothetic summary of the methods' performance across the sample. All relevant performance metrics can be seen in Table 2. Overall, iBoruta showed a higher recall (sensitivity) and negative predictive value. It was better at finding actual positive cases (82% of the time) as it missed fewer of them. When it rejected a predictor, it was more likely to be a true negative (64% of the time). tsBoruta, showed higher precision and

Table 1
Summary of the results of random-effects meta-analyses of iARIMAX estimates of nine processes predicting the outcome of hair pulling.

Process predicting hair pulling in iARIMAX models	Pooled Effect	*SE*	CI.LB	CI.UB	PI.LB	PI.UB	I^2
Anxiety	0.26	0.03	0.21	0.32	−0.08	0.61	76.91
Boredom	0.18	0.02	0.13	0.23	−0.13	0.48	71.95
Thought Awareness	0.12	0.03	0.06	0.17	−0.26	0.50	80.83
Stimulus Control	−0.17	0.03	−0.23	−0.12	−0.53	0.19	79.40
Urge to Pull	0.59	0.02	0.54	0.64	0.28	0.90	79.69
Pulling Worries	0.38	0.03	0.32	0.44	−0.05	0.81	85.93
Sensory Seeking (Hair)	0.46	0.03	0.41	0.52	0.13	0.80	80.07
Sensory Seeking (Hands)	0.51	0.02	0.46	0.55	0.19	0.82	78.90
Valued Action	−0.34	0.03	−0.04	−0.29	−0.72	0.03	81.03

Note. SE = Standard error; CI.LB = Confidence interval, lower bound; CI.UB = Confidence interval, upper bound; PI.LB = Prediction interval, lower bound; PI.UB = Prediction interval, upper bound; I^2 = indicator of non-random heterogeneity.

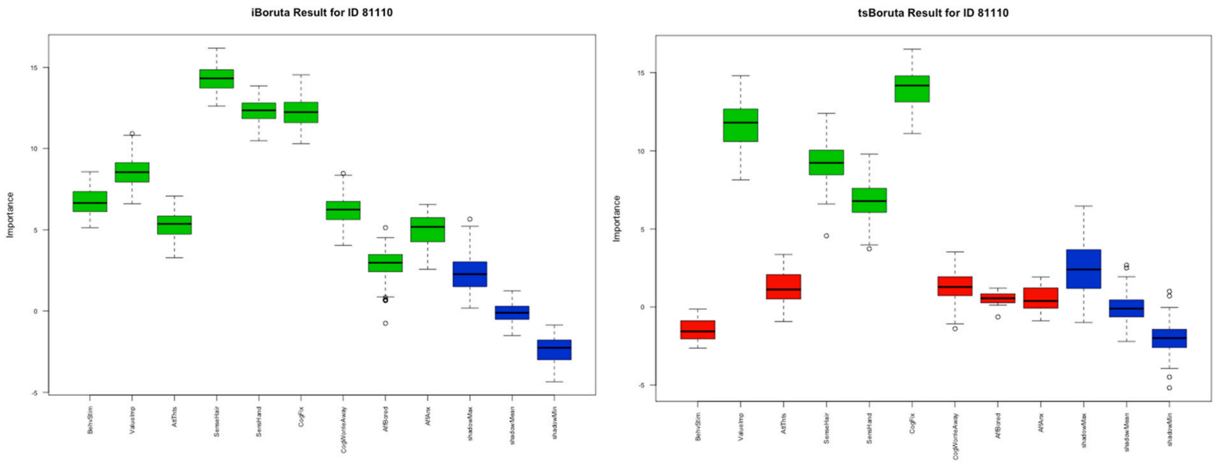

Fig. 1. The results of iBoruta and tsBoruta for an example case where the two algorithms yielded different decisions of important processes
Note. AffAnx: Anxiety; AffBored: Boredom; AttThts: Thought Awareness; BehvStim: Stimulus Control; CogFix: Urge to Pull; CogWorrieAway: Pulling Worries; SenseHair: Sensory Seeking (Hair); SensHand: Sensory Seeking (Hands); ValueImp: Valued Action. Green = confirmed; yellow = tentative; red = rejected; blue = shadow features (ignore).

Table 2
Performance metrics of iBoruta and tsBoruta.

Metric	iBoruta	tsBoruta	Interpretation
Accuracy	**0.81**	0.79	Overall proportion of correct predictions
Precision (Positive predictive value)	0.90	**0.96**	Of all predicted positives, how many are truly positives
Recall (Sensitivity)	**0.82**	0.74	Of all actual positives, how many were correctly identified
Negative predictive value	**0.64**	0.59	Of all predicted negatives, how many are truly negatives
Specificity	0.79	**0.93**	Of all actual negatives, how many were correctly identified
F1 Score	**0.86**	0.84	Harmonic mean of precision and recall

specificity, indicating that when it predicted a feature as confirmed, 96% of the time that case was a statistically significant iARIMAX estimate. tsBoruta also correctly identified a higher number of non-significant features (93% of them). iBoruta achieved a higher balance between precision and recall (F1 Score), and made a correct prediction 81% of the time, compared to 79% of tsBoruta. However, it is important to note that the dataset was highly imbalanced toward statistically significant effects, making accuracy a less informative metric.

From a clinical standpoint, precision, recall, negative predictive value, and specificity are important. Clinicians need to correctly identify meaningful processes that will likely affect clients' outcome, but they also need to correctly reject processes that are unlikely to be linked to the outcome of interest. To compare whether the errors of iBoruta and tsBoruta were similar at the aggregate level, we calculated a 2x2 contingency table of correct and incorrect predictions of both models pooled across all participants (excluding tentative decisions) and conducted a McNemar test (Agresti, 1990). This statistic serves as a population-average comparison of the algorithms' classification behavior. McNemar's statistic compares the number of instances that were correctly identified by one model and incorrectly identified by the second model. The null hypothesis is that these two counts of discordant predictions are equal, even though errors can be made on different instances or features in our case. The McNemar's test was not statistically significant (McNemar's $\chi 2(1) = 0.043$, $p = 0.716$), so no substantive differences were found in the number of unique errors made for each model. However, as established by the performance metrics (Table 2), the nature of these errors is different, with tsBoruta being more conservative. In summary, iBoruta appears better suited for capturing as many actual positive cases as possible, though at the cost of more false positives. Conversely, tsBoruta is more appropriate when the goal is to make positive case predictions more reliable or reduce false positives, though this comes at the expense of missing some genuine statistically significant effects.

3.2.3. Detecting linear vs. nonlinear effects

Our pre-registered Research Questions 4 and 5 focused on how iARIMAX, iBoruta, and tsBoruta compared when assessing processes with linear effects and when identifying nonlinear effects. Based on simulations studies (Li et al., 2026), we expected iARIMAX to outperform Boruta-based methods at identifying important processes with linear effects. Conversely, we expected Boruta-based methods to outperform iARIMAX in identifying nonlinear effects.

Preliminary analyses of 486 process-outcome pairs showed that linear models fit better than nonlinear (quadratic) models in 387 pairs (79.63%), indicating that linear effects were predominant. (See “Linearity Exploration” section in Supplemental File.) Still, of the 54 participants, 44 (81.48%) demonstrated at least one nonlinear relationship between process and outcome. We calculated odds ratios using the three methods of iARIMAX, iBoruta, and tsBoruta to evaluate their detection of nonlinear effects through statistical significance for iARIMAX or confirmation decision for the Boruta methods. To account for the fact that the effects were nested within participants, we employed multilevel logistic regression models, where each method's result (confirmed/significant or not) predicted the presence of nonlinear effects. Fig. S7 displays the odds ratio point estimates and their 95% confidence intervals (CIs). The substantial overlap in CIs across all three methods indicated no statistically significant differences in their detection of nonlinear versus linear effects as significant or confirmed in this sample.

3.2.4. Heterogeneity of bivariate process-outcome associations

Research Question 6 focused on heterogeneity in bivariate associations between processes and outcomes. We expected to find high variability. For each of the process-outcome pairs of bivariate models of iARIMAX, the random-effects meta-analyses results are summarized in Table 1, as reported earlier. The I^2 and 95% prediction intervals convey the degree of heterogeneity. The mean I^2 was 79.41 (range: 71.95 to 85.93).

Of the total 486 estimates of iARIMAX, 71 effects were negative and statistically significant, and 273 effects were positive and statistically significant. Of these two categories of iARIMAX effects, we examined the proportion of effects that were from each of the different processes. The results are summarized in Table S7. For example, of all the negative significant iARIMAX effects, 1.41% were that of the process of anxiety. Of all the positive ones, 12.82% were the effects of anxiety. In this case, a negative effect indicates reduced hair pulling. Of all the negative effects, valued action had the highest proportion (59.15%) and did not have any significant positive effect. Stimulus control had the next highest percentage of negative effects (30.99%) and very low percentage of positive effects (0.37%). We might be tempted to conclude that valued action, and possibly stimulus control, might be universally beneficial in reducing hair pulling. However, caution is needed because the 95% PIs, an important indicator of heterogeneity, for both processes included both negative and positive values, as shown in Table 1 of the summary of RE-MA of iARIMAX estimates. All processes showed high heterogeneity in RE-MAs.

3.2.5. *Heterogeneity of multivariate processes-outcome associations*

Our final Research Question 7 focused on the heterogeneity in multivariate associations between processes and outcomes. Based on Woolley et al.'s (2025) findings of heterogeneous network structures, we expected tsBoruta to identify different numbers of important predictive processes across individuals, potentially creating a new index for characterizing heterogeneity as variability in important variables across people. As expected, tsBoruta identified different numbers of processes as important for different persons in the sample: $M = 4.69$ ($SD = 1.84$; range $= 0$ to 9). Fig. S8 depicts the heterogeneity of important processes as a histogram of number of processes confirmed as important by tsBoruta for different people.

The total number of possible combinations of intersections among nine processes is 512 (calculated as 2^9 as per set theory). These numbers represent every possible combination of intersections among nine processes that can intersect, from having no processes being important (empty set) to all nine processes being important, and every combination in between. The complete space of all possible combinations of nine processes for 54 individuals is visually depicted in Fig. S9 in Supplemental File. The empirically observed combinations in our sample are shown in Fig. 2. A striking feature of this figure is the number of individuals showing unique combinations of the nine processes, with an intersection size of one. Of the 54 participants, 33 (61.11%) exhibited unique process combinations as important for trichotillomania. This suggests that a one-size-fits-all treatment approach may not be optimal for trichotillomania. Two individuals showed none of the nine processes when confirmed by tsBoruta, which suggests that other processes may be important for these two individuals.

Importantly, we identified a minimal combination of processes that, when targeted in an intervention, would likely benefit all but two individuals in this sample. The Venn diagram of cognitive fixation on the urge to pull, valued action, and anxiety in Fig. 3 demonstrates that 52 out of 54 individuals had at least one of these three potentially targetable processes confirmed by tsBoruta. The result suggests that clinicians could develop more effective, streamlined interventions by focusing on these three key processes while still acknowledging individual differences in how these processes manifest in the context of other processes.

Fig. 4 provides a final visual synthesis of the comparative performance metrics between the two machine learning approaches. This visualization highlights a distinct methodological trade-off: while iBoruta demonstrates higher Sensitivity (Recall), making it advantageous for detecting a broader range of potential predictors for hypothesis generation, tsBoruta exhibits superior Specificity and Precision. This distinction suggests that tsBoruta is the more advantageous method for identifying a conservative, high-confidence set of treatment targets, particularly in clinical contexts where minimizing false positives is a priority.

4. Discussion

A central challenge in process-based therapy research is identifying which individual-level predictors meaningfully influence outcomes over time, especially in noisy ecological momentary assessment (EMA) data. In the present study, we sought to investigate how three algorithms—iARIMAX, iBoruta, and tsBoruta—compared in identifying processes linked to the hair-pulling outcome, examining both nomothetic and idiographic conclusions. Speaking generally, iARIMAX offers

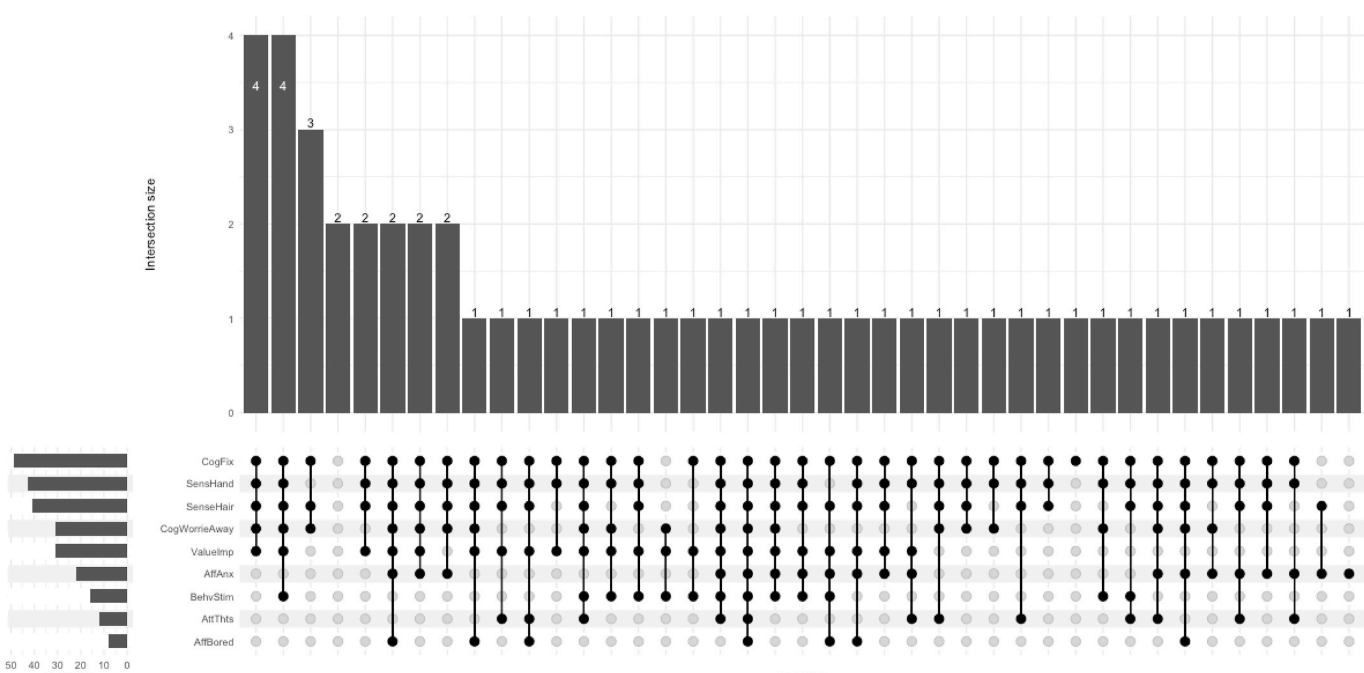

Fig. 2. Empirically observed patterns of overlap among processes identified as important by tsBoruta in 54 individuals in the sample

Note. AffAnx: Anxiety; AffBored: Boredom; AttThts: Thought Awareness; BehvStim: Stimulus Control; CogFix: Urge to Pull; CogWorrieAway: Pulling Worries; SenseHair: Sensory Seeking (Hair); SensHand: Sensory Seeking (Hands); ValueImp: Valued Action.

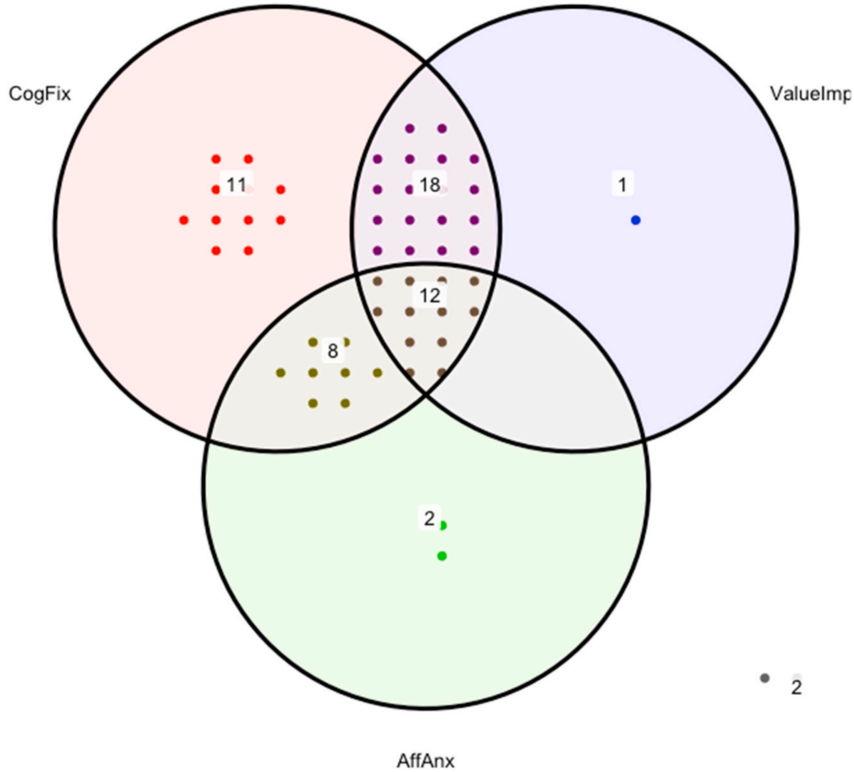

Fig. 3. Venn diagram showing the overlap of individuals for whom tsBoruta identified three key processes affecting hair pulling behavior: cognitive fixation on the urge to pull, anxiety (anxiety), and valued action (valued action)
Note. CogFix: Urge to Pull; AffAnx: Anxiety; ValueImp: Valued action.

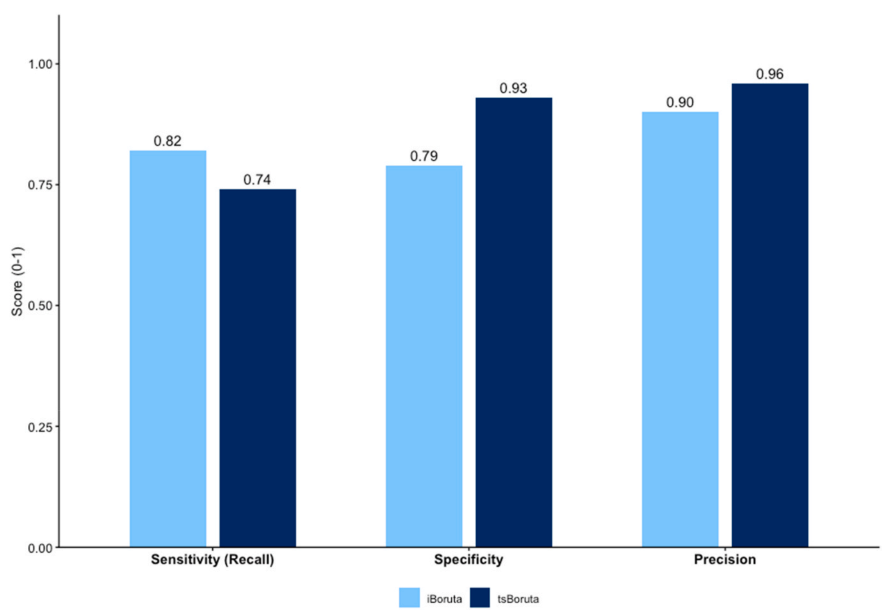

Fig. 4. Comparative performance metrics of iBoruta and tsBoruta
Note. Sensitivity (Recall) refers to the ability of the algorithm to correctly identify a process as important when it is truly statistically significant (True Positive Rate). Specificity refers to the ability of the algorithm to correctly reject a process when it is truly non-significant (True Negative Rate). Precision refers to the probability that a process identified as "confirmed" by the algorithm is truly statistically significant (Positive Predictive Value).

the advantage of modeling temporal dependencies directly and estimating lagged effects on a single outcome, making it well-suited for hypothesis-driven investigations of linear effects. In contrast, tsBoruta appears well suited for exploratory, data-driven analysis by detecting complex, potentially nonlinear and interaction-based relationships among many predictors. Importantly, tsBoruta is agnostic to the type of dependent variable—it can be applied to continuous, categorical, or count outcomes—and does not assume a fixed functional form. This flexibility makes it especially useful when working with rich, multivariate EMA data where theoretical models may not fully specify all relevant interactions or time lags. Together, these complementary approaches appeared to provide a more comprehensive toolkit for advancing personalized intervention science.

Based on prior research (Woolley et al., 2025), we hypothesized that all methods would converge on the same nomothetic conclusion regarding the link between cognitive fixation on the urge to pull and hair-pulling. We also expected to find substantial heterogeneity in both bivariate and multivariate associations between processes and outcomes. Based on recent simulations (Li et al., 2026), we expected tsBoruta to outperform iBoruta because tsBoruta accounts for the time series elements such as autocorrelation and trends. Given the heterogeneous network structures previously identified by Woolley et al. (2025), we expected tsBoruta to reveal individualized patterns of clinically relevant processes—highlighting not only variation across individuals but also offering a pathway toward personalized models of process–outcome linkage in psychological treatment.

In their GIMME model, Woolley et al. (2025) found that cognitive fixation on the urge to pull hair was the only process consistently linked to hair-pulling outcomes. While our study initially found a stronger connection using iARIMAX estimates, this difference was expected, as iARIMAX examines pairs of variables in isolation, whereas GIMME identifies connections while controlling for the influence of other variables in the network. When we accounted for other processes, the relationship was more modest but still meaningful. Importantly, cognitive fixation showed reliable positive effects across all our analyses: it was the only process that maintained consistent positive relationships in iARIMAX estimates and showed perfect agreement between the results of iBoruta and tsBoruta, while other processes showed much lower agreement. Almost all participants showed significant positive effects for cognitive fixation, with both statistical methods confirming its importance in the vast majority of cases. In essence, all three methods we used supported cognitive fixation on the urge to pull as a key predictor of hair pulling, aligning with Woolley et al.'s GIMME-based findings.

While the methods revealed identical nomothetic conclusions, they differed in idiographic conclusions. Our analysis of the two machine learning methods, iBoruta and tsBoruta, revealed notable differences in how they identified processes linked to outcomes. The algorithms disagreed in the pattern of important processes in the vast majority of individuals. For all the iARIMAX effects that were statistically significant, both Boruta methods agreed on confirming about two-thirds of them, though iBoruta more frequently identified additional effects that iARIMAX did not detect. For non-significant effects, both methods were much more conservative, almost never confirming these cases. While both Boruta-based methods made a similar number of unique errors, they differed in the *type* of error: iBoruta was more likely to include additional potential predictors, whereas tsBoruta was more selective.

From a clinical perspective, this trade-off has practical implications. When the priority is to detect all potentially relevant processes—for example, identifying every possible factor involved in hair-pulling—iBoruta may be preferred due to its greater sensitivity. In contrast, if the goal is to confidently rule out irrelevant processes and avoid unnecessary interventions, tsBoruta offers a stricter selection, making it well-suited for time-limited or high-stakes clinical settings. A key takeaway is that iBoruta expands the pool of potentially important processes, supporting broad hypothesis generation, while tsBoruta narrows it down to those most likely to matter, supporting more targeted intervention planning.

The findings also emphasize the importance of balancing sensitivity and specificity when identifying processes linked to psychopathological outcomes. This balance has direct implications for treatment planning. The convergence of iARIMAX, iBoruta, and tsBoruta in identifying cognitive fixation as a nomothetic predictor of hair-pulling supports the consistency of this finding. In a situation where no more information than the hair-pulling behavior is available, clinicians can confidently target cognitive fixation in interventions, knowing it is a reliable factor across a vast majority of individuals. However, when more information is available about an individual, the differences between iBoruta and tsBoruta in identifying idiographic (individual-specific) processes highlight the need for personalized treatment approaches. While iBoruta is more sensitive and may identify more potential factors, tsBoruta is more conservative and may help avoid over-targeting irrelevant processes. Such a balanced approach that is more conservative with tsBoruta may be beneficial in focusing empirically derived idiographic case conceptualizations on the most important, meaningful processes to target in treatment, while avoiding unnecessary complexity.

Evaluating our findings against the operational definitions of utility of a method established at the outset, tsBoruta emerges as the most clinically viable method for idionomic analysis in this context. Regarding *nomothetic convergence*, all three algorithms successfully replicated the established group-level link between cognitive fixation and hair pulling, satisfying our validity check against prior GIMME-based findings. However, the methods diverged meaningfully on *idiographic performance*, where our definition of "adequacy" prioritized specificity to minimize the clinical cost of targeting irrelevant processes. While iBoruta demonstrated higher sensitivity (0.82), its lower specificity (0.79) implies a greater risk of false positives, potentially leading clinicians to intervene on processes that do not functionally influence the outcome. In contrast, tsBoruta achieved a specificity of 0.93 and a precision of 0.96, aligning closely with our requirement for parsimony and reducing the likelihood of wasted therapeutic effort. Finally, regarding *clinical applicability*, tsBoruta successfully characterized the sample's heterogeneity, revealing that 61.11% of participants exhibited unique process combinations while simultaneously identifying a pragmatic three-process target set (cognitive fixation, valued action, anxiety) relevant for 96% of the sample. Thus, by balancing rigorous feature selection with the capacity to map unique clinical presentations, tsBoruta demonstrates the highest utility for informing personalized process-based interventions.

4.1. Process-based treatment planning for trichotillomania

The three methods of iARIMAX, iBorutra, and tsBoruta were effective at identifying true nomothetic effects—cases where a process variable consistently affects nearly all individuals in a sample. The effect for cognitive fixation on the urge to pull serves as an excellent example in PBT for how idionomic methods can discover genuine nomothetic effects while still maintaining a rigorous appreciation for idiographic patterns. This is particularly noteworthy because it demonstrates that some psychological processes can reliably predict outcomes across almost all individuals, strengthening our confidence in treating them as universal mechanisms in conceptualizing and treating trichotillomania.

Our analysis also revealed substantial variation in how different processes related to outcomes. We examined the relationships between processes and outcomes using bivariate iARIMAX models, finding considerable heterogeneity across all processes. Most effects were positive and statistically significant, with a smaller number showing negative significance. Among the different processes, anxiety showed mixed results—rarely appearing among negative effects but more commonly among positive ones. Notably, valued action stood out by showing only negative effects (indicating reduced hair pulling) with no significant positive effects. The process of stimulus control similarly showed mostly negative effects with very few positive ones. While these findings might suggest that valued action and stimulus control were consistently beneficial in reducing hair pulling, we must be cautious with this interpretation. The statistical measures of variability (I^2 and 95% prediction intervals) showed that all processes, including these two and fixation on the urge to pull, could have either positive or negative effects depending on the situation. The results of tsBoruta also confirmed a high degree of heterogeneity—the method identified different numbers of processes as important for different persons, ranging from zero for some to all nine processes for others.

A key challenge in implementing idionomic methods in tailored digital or in-person interventions lies in distinguishing meaningful signals from noise, given the multiple variables tested per person and various methodological choices that could lead to false positives. For treatment targeting, prioritizing specificity over sensitivity makes practical sense—we cannot target everything, so it is better to be more selective. The results indicate that while different methods yield varying idiographic effects, tsBoruta offers a good balance between sensitivity and specificity, making it particularly suitable for personalizing PBT interventions.

Our analysis using tsBoruta revealed significant individual variation in relevant processes across participants with trichotillomania, including cases where none of the nine processes appeared significant. This suggests there may be additional crucial processes not captured in our current framework—an important consideration for future research and clinical practice. The high percentage (61.11%) of individuals showing unique combinations of processes further indicates that a one-size-fits-all treatment approach may not be optimal.

Building on these insights, we propose examining process-specific effects rather than clustering individuals based on their entire network of processes. This targeted approach would help identify which processes are most relevant for specific individuals. For example, we could determine which individuals show strong relationships between valued action and hair pulling, then tailor values-focused interventions accordingly. This could be implemented either by selectively including values-based components for responsive individuals or by testing process variables as moderators of intervention effectiveness.

Despite the individual variations, our research revealed an efficient intervention strategy: targeting just three processes—fixation on the urge to pull, valued action, and anxiety—could potentially benefit 52 out of 54 individuals in the sample. While this approach may not represent the most efficient individualized plan (as not all processes are relevant to every person), it offers a practical advantage when designing universal interventions. By using tsBoruta to identify important processes across large samples and focusing on areas of overlap, this method ensures that at least some intervention content is likely to be relevant for nearly everyone—an important consideration in settings where fully personalized treatment is not feasible.

4.2. Implications for study design and clinical implementation

To successfully employ these idionomic methods in clinical practice, the design of the assessment is just as critical as the analysis. First, regarding sample density, these time-series methods rely on within-person variance. Based on a recent simulation study (Li et al., 2026), a reasonable recommendation is a minimum of 50 to 60 time points per individual to ensure stable parameter estimates for iARIMAX and sufficient training data for tsBoruta. In a clinical context, this might look like 3–4 prompts per day over a 2–3 week baseline period, or a more substantial baseline period if that is a viable option. Second, regarding sampling frequency, the interval between prompts should ideally match the temporal dynamics of the behavior that is known. For rapidly fluctuating behaviors like hair pulling, momentary assessments (e.g., every 3–4 h) may be superior to daily diaries, which may smooth over critical antecedents. Conversely, behaviors that have meaning only over longer time periods (e.g., arguments within a couple) will need larger temporal windows.

Finally, these methods can be seamlessly adapted to evaluate intervention effects (e.g., comparing a baseline phase to a treatment phase). In iARIMAX, this can be achieved by adding a binary "Phase" variable (coded 0 for baseline, 1 for intervention) as an exogenous predictor. A significant coefficient for this variable indicates a shift in the mean level of the feature of interest. Additionally, including an interaction term (Process x Phase) allows practitioners to test if the *process* of relevance to the outcome has changed, for example, determining if the link between anxiety and hair pulling has weakened during treatment. Similarly, in tsBoruta, the "Phase" variable can be included as a feature; if the algorithm confirms it as important, it indicates that the intervention status itself is a robust predictor of the outcome, distinct from the natural fluctuations of other psychological processes.

4.3. Limitations and future directions

A key limitation of our study was insufficient statistical power to thoroughly evaluate how iARIMAX, iBoruta, and tsBoruta methods compare in detecting linear versus nonlinear effects. Recent simulations by Li et al. (2026) indicated that iARIMAX excels at detecting bivariate linear effects while tsBoruta may be better in identifying nonlinear relationships in a multivariate context. However, our limited sample size prevented us from validating these findings. Future studies should use larger samples to test more complex patterns in multivariate time series. Longer time series and larger sample sizes will allow more sophisticated tests beyond simple linear and quadratic comparisons using methods that can capture different forms of nonlinear relationships while accounting for temporal data patterns.

We also acknowledge potential methodological challenges in the application of tsBoruta that warrant careful consideration. While the algorithm demonstrated success in identifying varying numbers of relevant processes across individuals, there were notable instances where no measured processes reached statistical significance for an individual. This finding has important implications. First, it suggests the potential existence of unmeasured yet crucial psychological processes that may significantly influence hair-pulling behavior. Second, it highlights the importance of considering contextual factors unique to each individual's lived experience.

To address these limitations, we recommend implementing a more comprehensive assessment protocol that captures a broader range of psychological processes and developing a structured method for clients to identify and report idiosyncratic processes unique to their socio-economic-cultural context. This person-centered approach could include regular feedback sessions where clients can articulate meaningful processes not captured by standardized measures, potentially revealing important cultural, social, or environmental factors that influence their hair-pulling behavior. Such an approach would not only enhance the ecological validity of our measurements but also align with principles of culturally responsive clinical practice.

A critical limitation of the present study is its reliance on observational data to model the relationships between psychological processes and hair pulling. While our application of iARIMAX and tsBoruta successfully identified strong statistical associations, such as the nomothetic link between cognitive fixation and hair pulling, or the idiographic importance of valued action for specific individuals, these findings reflect how these targets naturally vary over time rather than how they may respond to deliberate manipulation. It should not be assumed that the predictive validity observed in a naturalistic context will translate directly into therapeutic utility because the dynamics of a system under intervention may differ significantly from its natural state. For example, interventions designed to actively target processes like anxiety or sensory seeking might alter the underlying structural relationship between the predictor and the outcome, a complexity that an observational analysis, such as the one conducted in this study, cannot capture. Consequently, while these idionomic methods provide a powerful heuristic for selecting potential treatment targets, future research must bridge the gap between prediction and influence by experimentally manipulating these identified processes, possibly through micro randomized trials or single case experimental designs, to verify that changing these targets produces the expected reduction in symptom severity.

Finally, while the methods we have examined show promise for personalizing interventions, several practical challenges must be addressed before widespread clinical implementation. The immediate priorities include developing standardized clinical guidelines and user-friendly tools for practitioners, along with rigorous evaluation of real-world effectiveness and cost-feasibility across different clinical settings. Looking ahead, research should focus on creating robust decision-support systems that can help clinicians apply these analytical approaches effectively. This includes establishing standardized protocols for clinical decision-making, developing integrated analysis tools, and investigating how these methods can be used to dynamically adjust interventions based on patient progress. Crucially, long-term outcome studies will be essential to validate the sustained effectiveness of interventions guided by these analytical approaches.

5. Conclusion

Our study represents a significant methodological advancement by systematically comparing three algorithms of iARIMAX, iBoruta, and tsBoruta in detecting both nomothetic and idiographic effects in psychological EMA data. There is a dearth of machine learning applications in clinical psychology that explicitly account for time series data. tsBoruta represents a significant step forward in this area. Our comparative approach reveals the relative strengths of each algorithm. Notably, tsBoruta was more conservative, making it useful in clinical contexts where avoiding false positives (e.g., under time constraints) is particularly important.

More broadly, our research bridges the gap between nomothetic and idiographic approaches in psychological intervention research, using trichotillomania as a clinically relevant example. The results of our study clearly demonstrate how to identify both universal predictors like cognitive fixation and individual-specific processes like valued-action and anxiety, enabling a framework that combines standardized treatment protocols with personalized interventions. This integrated approach can allow clinicians to leverage general therapeutic principles while customizing treatments to match each client's unique needs. tsBoruta offers a promising method for such idionomics-based treatment planning.

CRediT authorship contribution statement

Baljinder K. Sahdra: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization. **Mercedes G. Woolley:** Writing – review & editing, Writing – original draft, Resources, Project administration, Investigation, Funding acquisition, Data curation, Conceptualization. **Cristóbal Hernández:** Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis. **William Li:** Writing – original draft, Formal analysis. **Steven C. Hayes:** Writing – review & editing, Writing – original draft, Conceptualization. **Joseph Ciarrochi:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Michael P. Twohig:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization. **Micheal E. Levin:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Investigation, Funding acquisition, Data curation, Conceptualization.

Author note

This study conducted pre-registered secondary data analysis (https://osf.io/2yuw9/overview) using a sample from a previously published study (Woolley et al., 2025). The university's Institutional Review Board approved the study, and all participants provided informed consent. However, since the consent form lacked statements about sharing de-identified data on open science platforms, we cannot make the raw data openly accessible. We have included a Supplemental File, our R code analyses, and simulated data based on first-order VAR estimates of the real data. Note that results from this simulated data will differ somewhat from those reported in this paper. We have also included an Appendix with a tsBoruta tutorial. While we did not use generative artificial intelligence tools to write the first draft, we employed them during editing to enhance this manuscript's readability.

Declaration of competing interest

Given their role as Editor, Sahdra B.K., Editorial Board member, Levin M., and Student Editorial Board member Li W., had no involvement in the peer-review of this article and had no access to information regarding its peer-review. All other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix C. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jcbs.2026.100983.

Appendix A

Glossary of key methodological terms.

Term	Definition
Autoregression (AR)	A process where a variable's current value is predicted by its own past values. A "First-order" (AR1) process means the current value is predicted by the last observed value.
Differencing (d)	A transformation used to stabilize a time series by subtracting the current observation from the previous one. This helps in removing trends (e.g., a gradual increase in anxiety over a month) to achieve stationarity.
Exogenous Covariate	An external variable used to predict the outcome. In this study, psychological processes (e.g., "Cognitive Fixation") serve as exogenous covariates predicting the outcome ("Hair Pulling").
Feature Selection	The process of automatically selecting which variables (features) are most relevant for predicting an outcome, while discarding irrelevant ones (noise or unrelated).
Moving Average (MA)	A component of time-series modeling that accounts for past prediction errors (shocks) to improve future forecasts. It helps smooth out random fluctuations.
Non-Stationarity	A condition where the statistical properties of a time series (like the mean or variance) change over time (e.g., a client getting progressively better). Most analyses require data to be made "stationary" (stable) first.
Overfitting	The model learns patterns that fit the training data too closely (including noise), so it performs worse on new data.
Shadow Feature	A shadow feature is a copy of a real variable with its values randomly shuffled so it has no real link to the outcome. Boruta then keeps a real feature only if it beats the best shadow features.
Stationarity	A state where the statistical properties of a time series (mean, variance, autocorrelation) remain constant over time. This is a standard assumption for many time-series analyses.

Appendix B

tsBoruta Tutorial: Person-Specific Feature Selection for Time Series Data in R

tsBoruta is an advanced, yet accessible, method for identifying which psychological processes are truly important for predicting an outcome in time series data—especially when data are collected intensively over time from individuals, e.g., in clinical psychology, behavioral medicine, or digital health. tsBoruta combines two powerful tools: (1) ARIMA modeling, which removes trends and autocorrelation from each predictor, so that only the "new" information at each time point is used; and (2) Boruta feature selection, which identifies which predictors are genuinely important for the outcome, using a robust approach. This tutorial will guide you through the rationale, workflow, and practical steps for using tsBoruta on your own data, even if you have limited programming experience.

What is tsBoruta and why use it?

In time series data, variables often show patterns over time, e.g., mood today can be similar to mood yesterday. If you ignore these patterns, you might mistakenly think a variable is important just because it follows a trend, not because it truly predicts your outcome. tsBoruta offers a solution. It first "cleans" each predictor by removing these time-based patterns using ARIMA, then uses Boruta to find all predictors that are truly relevant for your outcome. This is especially useful for personalized analysis: tsBoruta is run separately for each individual, revealing person-specific predictors. In clinical and behavioral research, tsBoruta helps identify which processes (e.g., anxiety, experiential avoidance) are most relevant for each person to predict a specific outcome (e.g., depressed mood).

How does tsBoruta work?

1. ARIMA Preprocessing: For each predictor/process variable (e.g., anxiety, stress), tsBoruta fits an ARIMA model to remove trends, autocorrelations, and specific unpredictable events. The leftover "residuals" represent the new, unpredictable part of the variable at each time point that wasn't explained by its own past values.
2. Feature Selection with Boruta: Boruta compares the importance of each predictor using the ARIMA residuals to that of random "shadow" predictors. If a real predictor is consistently more important than the shadows, it is marked as "Confirmed."
3. Interpretation: The output tells you which predictors are Confirmed (important), Tentative (maybe important), or Rejected (not important) for your outcome.

Step-by-Step tsBoruta workflow in R

Prerequisites

- Install R and RStudio (if not already installed).
- Install required packages (run this once):

```
install.packages("forecast")  # For ARIMA modeling
install.packages("Boruta")    # For feature selection
```

Simulate example data (or load your own "long" format data)

For this tutorial demonstration, the code below simulates simple time series data for 3 individuals.

```
set.seed(123)
n <- 100  # Number of time points per individual
n_ind <- 3  # Number of individuals

sim_data <- do.call(rbind, lapply(1:n_ind, function(id) {
  pred1 <- as.numeric(arima.sim(model = list(ar = 0.6), n = n))
  pred2 <- rnorm(n)
  pred3 <- as.numeric(arima.sim(model = list(ma = 0.5), n = n))
  outcome <- 0.7 * pred1 - 0.3 * pred3 + rnorm(n)
  data.frame(ID = id, Time = 1:n, outcome, pred1, pred2, pred3)
}))
head(sim_data)
```

If using your own data, your data should be in "long" format, with columns for ID (individual), Time, outcome variable, and predictors.

tsBoruta Template Code

Step 1: Calculate ARIMA residuals for each predictor and individual

```
# Function to calculate ARIMA residuals for each individual
calculate_residuals <- function(data) {
  residuals_list <- list()
  for (id in unique(data$ID)) {
    individual_data <- subset(data, ID == id)
    individual_residuals <- data.frame(ID = rep(id, nrow(individual_data)), Time =
individual_data$Time)
    vars_to_model <- setdiff(colnames(individual_data), c("ID", "Time"))
    for (var in vars_to_model) {
      arima_model <- auto.arima(individual_data[[var]])
      individual_residuals[[var]] <- residuals(arima_model)
    }
    residuals_list[[as.character(id)]] <- individual_residuals
  }
  return(residuals_list)
}

# Apply the function to your data
residuals_results <- calculate_residuals(sim_data)
```

Step 2: Convert residuals to numeric

```
# Ensure all columns are numeric (Boruta requires this)
residuals_num <- lapply(residuals_results, function(x) {
  for (col in colnames(x)) {
    if (is.ts(x[[col]])) {
      x[[col]] <- as.numeric(x[[col]])
    }
  }
  return(x)
})
```

Step 3: Run Boruta on ARIMA residuals (tsBoruta)

```
# Replace 'outcome' and predictor names as appropriate for your data
TS_boruta_results <- lapply(residuals_num, function(x) {
  set.seed(123)
  Boruta(
    outcome ~ pred1 + pred2 + pred3,
    data = x,
    maxRuns = 500,
    doTrace = 3
  )
})
```

Step 4: Review and visualize results

```r
# Print results for the first individual
print(TS_boruta_results[[1]])

# Plot feature importance for each individual
for (i in 1:length(TS_boruta_results)) {
  plot(
    TS_boruta_results[[i]],
    las = 2,
    cex.axis = 0.7,
    sort = FALSE,
    xlab = "",
    main = paste("tsBoruta Result for ID", names(TS_boruta_results)[i])
  )
}
```

How to interpret tsBoruta output

- Confirmed (green): Predictor is robustly important for the outcome.
- Tentative (yellow): Predictor may be important; more data or investigation needed.
- Rejected (red): Predictor is not important for the outcome.

For example, if pred1 is Confirmed for ID 1, it means that, after removing time-based patterns, pred1 is a strong predictor of the outcome for that individual. Use your clinical or research knowledge to interpret which predictors are meaningful for each individual.

Tips and troubleshooting

- Data format: Ensure your data is in long format with columns for ID, Time, outcome, and predictors.
- Missing values: Handle missing data before running tsBoruta (e.g., impute or remove, as appropriate).
- Variable names: Update the formula in the Boruta call to match your variable names.
- Computation time: tsBoruta can be slow for large datasets or many predictors; start with a small example.

References

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.

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

Borenstein, M., Higgins, J. P. T., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Research Synthesis Methods, 8* (1), 5–18. https://doi.org/10.1002/jrsm.1230

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). John Wiley & Sons.

Catts, N. V., Sahdra, B. K., Ciarrochi, J., Fraser, M. I., Hernández, C., Hayes, S. C., & Gloster, A. T. (2025). Distinct pathways to well-being: Exploring valued action and mood among stoics and non-stoics. *Journal of Contextual Behavioral Science*. , Article 100898. https://doi.org/10.1016/j.jcbs.2025.100898

Ciarrochi, J., Sahdra, B., Hayes, S. C., Hofmann, S. G., Sanford, B., Stanton, C., Yap, K., Fraser, M. I., Gates, K., & Gloster, A. T. (2024). A personalised approach to identifying important determinants of well-being. *Cognitive Therapy and Research, 48* (4), 1–22. https://doi.org/10.1007/s10608-024-10486-w

Ciarrochi, J., Sahdra, B., Hofmann, S. G., & Hayes, S. C. (2022). Developing an item pool to assess processes of change in psychological interventions: The process-based assessment tool (PBAT). *Journal of Contextual Behavioral Science, 23*, 200–213. https://doi.org/10.1016/j.jcbs.2022.02.001

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Gates, K. M., & Molenaar, P. C. M. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage, 63*(1), Article 310319. https://doi.org/10.1016/j.neuroimage.2012.06.026

Genolini, C., Écochard, R., & Jacqmin-Gadda, H. (2013). Copy mean: A new method to impute intermittent missing values in longitudinal studies. *Open Journal of Statistics, 3*(4), 26–40. https://doi.org/10.4236/ojs.2013.34A004

Gholami, H., Mohammadifar, A., Golzari, S., Kaskaoutis, D. G., & Collins, A. L. (2021). Using the Boruta algorithm and deep learning models for mapping land susceptibility to atmospheric dust emissions in Iran. *Aeolian Research, 50*, Article 100682. https://doi.org/10.1016/j.aeolia.2021.100682

Grant, J. E., Peris, T. S., Ricketts, E. J., Lochner, C., Stein, D. J., Stochl, J., Chamberlain, S. R., Scharf, J. M., Dougherty, D. D., Woods, D. W., Piacentini, J., & Keuthen, N. J. (2021). Identifying subtypes of trichotillomania (hair pulling disorder) and excoriation (skin picking) disorder using mixture modeling in a multicenter sample. *Journal of Psychiatric Research, 137*, 603–612. https://doi.org/10.1016/j.jpsychires.2020.11.001

Hayes, S. C. (1981). Single case experimental design and empirical clinical practice. *Journal of Consulting and Clinical Psychology, 49*(2), 193–211. https://doi.org/10.1037/0022-006X.49.2.193

Hayes, S. C., Ciarrochi, J., Hofmann, S. G., Chin, F., & Sahdra, B. (2022). Evolving an idionomic approach to processes of change: Towards a unified personalized science of human improvement. *Behaviour Research and Therapy, 156*, Article 104155. https://doi.org/10.1016/j.brat.2022.104155

Hayes, S. C., Hofmann, S. G., & Ciarrochi, J. (2020). A process-based approach to psychological diagnosis and treatment: The conceptual and treatment utility of an extended evolutionary meta model. *Clinical Psychology Review, 82*, Article 101908. https://doi.org/10.1016/j.cpr.2020.101908

Hayes, S. C., Hofmann, S. G., Stanton, C. E., Carpenter, J. K., Sanford, B. T., Curtiss, J. E., & Ciarrochi, J. (2019). The role of the individual in the coming era of process-based therapy. *Behaviour Research and Therapy, 117*, 40–53. https://doi.org/10.1016/j.brat.2018.10.005

Hayes, S. C., Merwin, R. M., McHugh, L., Sandoz, E. K., A-Tjak, J. G., Ruiz, F. J., … McCracken, L. M. (2021). Report of the ACBS task force on the strategies and tactics of contextual behavioral science research. *Journal of Contextual Behavioral Science, 20*, 172–183. https://doi.org/10.1016/j.jcbs.2021.03.007

Haynes, S. N., & Williams, A. E. (2003). Case formulation and design of behavioral treatment programs: Matching treatment mechanisms to causal variables for behavior problems. *European Journal of Psychological Assessment, 19*(3), 164–174. https://doi.org/10.1027/1015-5759.19.3.164

Hernández, C., Ciarrochi, J., Hayes, S., & Sahdra, B. K. (2025). *Idionomics* (R package). *GitHub*. https://github.com/cristobalehc/idionomics.

Hicks, M. M., Mouton-Odum, S., Patyk, K. C., Zamora, R. J., & De Nadai, A. S. (2023). Latent class analysis of emotions experienced during compulsive hair-pulling episodes. *Journal of Behavior Therapy and Experimental Psychiatry, 81*, Article 101882. https://doi.org/10.1016/j.jbtep.2023.101882

Hofmann, S. G., & Hayes, S. C. (2019). The future of intervention science: Process-based therapy. *Clinical Psychological Science: A Journal of the Association for Psychological Science, 7*(1), 37–50. https://doi.org/10.1177/2167702618772296

Hofmann, S. G., Kasch, C., & Reis, A. (2025). Effect sizes of randomized-controlled studies of cognitive behavioral therapy for anxiety disorders over the past 30 years.

Clinical Psychology Review, 117, Article 102553. https://doi.org/10.1016/j.cpr.2025.102553

Hyndman, R. J. (2025). Forecast: Forecasting functions for time series and linear models (R package version 8.22) https://pkg.robjhyndman.com/forecast/.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software, 27*(3), 1–22. https://doi.org/10.18637/jss.v027.i03

Kiesler, D. J. (1966). Some myths of psychotherapy research and the search for a paradigm. *Psychological Bulletin, 65*(2), 110–136. https://doi.org/10.1037/h0022911

Kok, G., Gottlieb, N. H., Peters, G. J., Mullen, P. D., Parcel, G. S., Ruiter, R. A., Fernández, M. E., Markham, C., & Bartholomew, L. K. (2016). A taxonomy of behaviour change methods: An intervention mapping approach. *Health Psychology Review, 10*(3), 297–312. https://doi.org/10.1080/17437199.2015.1077155

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the boruta package. *Journal of Statistical Software, 36*(11). https://doi.org/10.18637/jss.v036.i11

Kursa, M. B., & Rudnicki, W. R. (2023). Boruta: A feature selection algorithm based on random forest. *R package version 7.0.0*. https://CRAN.R-project.org/package=Boruta .

Lane, S. T., Gates, K. M., Pike, H. K., Beltz, A. M., & Wright, A. G. C. (2019). Uncovering general, shared, and unique temporal patterns in ambulatory assessment data. *Psychological Methods, 24*(1), 54–69. https://doi.org/10.1037/met0000192

Larsson, A. B., & Sundström, F. T. A. (2024). Optimizing process-based therapy: A boruta-driven approach to identifying key clinical outcomes in a Swedish validation of the process-based assessment tool (PBAT). *Journal of Contextual Behavioral Science, 33*, Article 100794. https://doi.org/10.1016/j.jcbs.2024.100794

Li, W., Ciarrochi, J., Fraser, M. I., Hayes, S. C., Hernández, C., Ong, C. W., & Sahdra, B. (2026). Advancing idiographic methodology for data-driven treatment personalization: Comparing ARIMAX, network, and tree-based approaches. *PsyArXiv Preprint*. https://doi.org/10.31234/osf.io/7mxn3_v1

Li, W., Gleeson, J., Fraser, M. I., Ciarrochi, J., Hofmann, S. G., Hayes, S. C., & Sahdra, B. (2024). The efficacy of personalized psychological interventions in adolescents: A scoping review and meta-analysis. *Frontiers in Psychology, 15*. https://doi.org/10.3389/fpsyg.2024.1470817

Loerinc, A. G., Meuret, A. E., Twohig, M. P., Rosenfield, D., Bluett, E. J., & Craske, M. G. (2015). Response rates for CBT for anxiety disorders: Need for standardized criteria. *Clinical Psychology Review, 42*, 72–82. https://doi.org/10.1016/j.cpr.2015.08.004

Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science, 18*(2), 112–117. https://doi.org/10.1111/j.1467-8721.2009.01619.x

Nye, A., Delgadillo, J., & Barkham, M. (2023). Efficacy of personalized psychological interventions: A systematic review and meta-analysis. *Journal of Consulting and Clinical Psychology, 91*(7), 389–397. https://doi.org/10.1037/ccp0000820

R Core Team. (2025). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing.

Sahdra, B. K., Ciarrochi, J., Fraser, M. I., Yap, K., Haller, E., Hayes, S. C., Hofmann, S. G., & Gloster, A. T. (2023). The compassion balance: Understanding the interrelation of Self- and other-compassion for optimal well-being. *Mindfulness, 14*(8), 1997–2013. https://doi.org/10.1007/s12671-023-02187-4

Sahdra, B. K., Ciarrochi, J., Klimczak, K., Krafft, J., Hayes, S. C., & Levin, M. (2024). Testing the applicability of idionomic statistics in longitudinal studies: The example of 'Doing What Matters'. *Journal of Contextual Behavioral Science, 32*. https://doi.org/10.1016/j.jcbs.2024.100728

Sahdra, B. K., Shin, A., Fraser, M., Levin, M. E., Klimczak, K. S., Krafft, J., Hayes, S. C., Hernández, C., & Ciarrochi, J. (2025). One size does not make all happy: Idionomic links between striving for positive states and happiness in experience sampling. *Journal of Happiness Studies, 26*, 101. https://doi.org/10.1007/s10902-025-00933-0

Sanford, B. T., Ciarrochi, J., Hofmann, S. G., Chin, F., Gates, K. M., & Hayes, S. C. (2022). Toward empirical process-based case conceptualization: An idionomic network examination of the process-based assessment tool. *Journal of Contextual Behavioral Science, 25*, 10–25. https://doi.org/10.1016/j.jcbs.2022.05.006

Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications: With R examples* (4th ed.). Springer.

Szul, T., Tabor, S., & Pancerz, K. (2021). Application of the BORUTA algorithm to input data selection for a model based on rough set theory (RST) to prediction energy consumption for building heating. *Energies, 14*(10), 2779. https://doi.org/10.3390/en14102779

Tolin, D. F., Gilliam, C., Wootton, B. M., Bowe, W., Bragdon, L. B., Davis, E., Hannan, S. E., Steinman, S. A., Worden, B., & Hallion, L. S. (2018). Psychometric properties of a structured diagnostic interview for DSM-5 anxiety, mood, and obsessive-compulsive and related disorders. *Assessment, 25*(1), 3–13. https://doi.org/10.1177/1073191116638410

Tsay, R. S., Wood, D., & Lachmann, J. (2022). *MTS: All-purpose toolkit for analyzing multivariate time series (MTS) and estimating multivariate volatility models*. CRAN [R package] Version 1.2.1. https://CRAN.R-project.org/package=MTS.

Varadhan, R., Segal, J. B., Boyd, C. M., Wu, A. W., & Weiss, C. O. (2013). A framework for the analysis of heterogeneity of treatment effect in patient-centered outcomes research. *Journal of Clinical Epidemiology, 66*(8), 818–825. https://doi.org/10.1016/j.jclinepi.2013.02.009

Woolley, M. G., Sahdra, B. K., Bowers, E. M., Capel, L. K., Twohig, M. P., & Levin, M. E. (2025). An idionomic network analysis of trichotillomania treatment processes: Gathering group-level and individual-level insights to inform a personalized therapeutic framework. *Behavior Therapy*. https://doi.org/10.1016/j.beth.2025.01.006

Yang, K., Gong, Y., Xu, X., Sun, T., Qu, X., He, X., & Liu, H. (2025). Prediction model for psychological disorders in ankylosing spondylitis patients based on multi-label classification. *Frontiers in Public Health, 13*, 1497955. https://doi.org/10.3389/fpubh.2025.1497955