



**Universiteit
Leiden**
The Netherlands

Dynamic Time Warp (DTW) as a scalable, data-efficient, and clinically relevant analysis of dynamic processes in patients with psychiatric disorders: a tutorial

Kopland, M.C.G.; Giltay, E.J.

Citation

Kopland, M. C. G., & Giltay, E. J. (2025). Dynamic Time Warp (DTW) as a scalable, data-efficient, and clinically relevant analysis of dynamic processes in patients with psychiatric disorders: a tutorial. *Journal Of Eating Disorders*, 13(1). doi:10.1186/s40337-025-01414-8

Version: Publisher's Version

License: [Creative Commons CC BY-NC-ND 4.0 license](#)

Downloaded from: <https://hdl.handle.net/1887/4298253>

Note: To cite this publication please use the final published version (if applicable).

METHODOLOGY

Open Access



Dynamic Time Warp (DTW) as a scalable, data-efficient, and clinically relevant analysis of dynamic processes in patients with psychiatric disorders: a tutorial

Maren C. G. Kopland^{1,2*} and Erik J. Giltay^{3,4}

Abstract

Dynamic Time Warping (DTW) is an emerging analytic technique that offers a flexible approach to modeling symptom dynamics in psychological and psychiatric research. Unlike traditional network models, which often rely on linear associations, DTW aligns symptom trajectories even when changes unfold at slightly different speeds or time intervals. This tutorial offers a brief introduction into DTW and demonstrates how to apply DTW to panel or time series data. We illustrate the workflow using clinical case data from patients with eating disorders, to capture temporal patterns that cannot be detected with conventional network analysis techniques, as these require more intensive time-series data. Key advantages include its applicability to non-stationary data, flexibility in handling irregular time intervals, and reduced reliance on frequent assessments, which patients often cannot maintain due to the burden. We also discuss some of the limitations such as noise, scaling decisions and lack of Granger causality associations. Finally, we outline directions for future research. By expanding the methodological toolkit available for studying therapy processes, DTW holds promise for advancing both research and clinical practice in personalized mental health care.

Introduction

In psychological and psychiatric research as well as in clinical practice, the dominant paradigm has long centered around group-level analyses to guide therapeutic decision-making [27, 28]. Most research on eating disorder treatment has relied on cross-sectional studies,

cohort analyses, and randomized controlled trials on a group level. Yet, patients with eating disorders differ not only in the symptoms they present (e.g., [25]) but also in the way these symptoms evolve over time during therapy (e.g., [23]) and in their potential responses to interventions. Thus, group-level change can obscure patterns of individual change crucial to the recovery process [14]. As a result, there is now a shift in treatment research due to a growing consensus that research on treatment effectiveness needs to focus on individual change rather than group-level change [8, 24, 25].

Central to this shift is the need to better understand intraindividual dynamics—how symptoms, emotions, cognitions, and behaviors fluctuate over time within a person, and how these temporal patterns shape

*Correspondence:

Maren C. G. Kopland
marencgk@gmail.com

¹Modum Bad Psychiatric Hospital, Badeveien 287, 3370 Vikersund, Norway

²Department of Psychology, University of Oslo, Oslo, Norway

³Department of Psychiatry, Leiden University Medical Center (LUMC), Leiden, the Netherlands

⁴Health Campus The Hague, Leiden University, The Hague, The Netherlands



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

treatment progress and outcomes. Over the past decades, psychotherapy research has increasingly focused on disentangling between-person (group-level) effects from within-person processes in an effort to identify the mechanisms that drive therapeutic change [20, 21]. More recently, this pursuit has evolved toward the goal of personalization, recognizing that therapeutic change is often idiosyncratic and shaped by the unique interplay of psychological processes within an individual. Within this emerging paradigm, the network theory of mental disorders has gained prominence as a promising framework for modeling these intra-individual dynamics (e.g., [6, 10]), also in eating disorder therapy [25, 26].

By conceptualizing symptoms and psychological constructs as interconnected *nodes* in a dynamic system, network models offer a novel approach to understanding and personalizing the process of change in psychotherapy [5, 9, 31]. Olthof and colleagues [31] propose that psychopathology can be conceptualized as a complex dynamic pattern of 'self-organized interactions between interdependent biopsychosocial processes in a complex adaptive system comprising a person in its environment' (p. 314). In other words, they suggest that a patient's symptoms may form an interconnected, potentially self-reinforcing system of experiences and behaviors. Unpacking each patient's complex set of symptoms may help identify more targeted intervention points in therapy- and hopefully to better treatment outcomes for patients.

However, studying these dynamics is methodologically demanding. Many of the most powerful analytic tools available, such as multilevel vector autoregressive [13] models and or dynamic structural equation modeling (DSEM; [1], Hamaker & Muth en, 2017), require extensive longitudinal data—often dozens of or a hundred time points per person—to produce reliable estimates (e.g., [34]). Such dense data are rarely available in real-world therapy settings, where measurement frequency is limited by both practical and ethical constraints. Experience Sampling Method (ESM) data collection can be demanding for patients with severe psychopathology, where initiative and energy are often limited by the condition. This creates a significant barrier to implementing personalized care at scale. Without access to detailed temporal models, clinicians may miss critical insights into how their clients are changing over time. At the same time, researchers struggle to generalize individual-level findings (i.e., idiographic results) to larger populations, limiting their ability to inform theory and practice.

Dynamic Time Warping (DTW) can be a promising alternative for investigating individual-level as well as group-level temporal network dynamics (i.e., networks based on data collected across time) in eating disorder therapy. DTW is a technique originally developed for time series alignment in speech and signal processing

and has more recently found applications in psychology and behavioral science (e.g., [11, 18, 22, 23, 33, 34]). Unlike model-based approaches, DTW does not require parametric assumptions or extensive data. Instead, it provides a flexible, non-linear way to measure similarity between time series, even when they differ in length or timing [18]. We propose that DTW can serve as a practical method for analyzing idiographic data capturing the ebb and flow of symptoms or behaviors within a person over time and subsequently aggregating these patterns to the group level. This allows us to retain the granularity of individual trajectories while still identifying commonalities across cases, allowing an overview of *pattern of change* in psychotherapy. In doing so, DTW offers a viable path toward scalable, data-efficient, and clinically relevant analysis of dynamic processes in therapy.

This tutorial demonstrates the 'how' and 'why' of DTW, in an eating disorder context. Although our focus here is on eating disorders, the analytical approach we present is broadly applicable to other stress-related disorders, such as anxiety disorders, depression, and post-traumatic stress disorder, conditions for which a complex systems framework is appropriate, as a wide range of biopsychosocial factors influence both their onset and course [6, 10, 32]. We outline the theoretical rationale for using DTW with temporal data, describe the methodological workflow, and provide empirical illustrations using clinical eating disorder case data. The tutorial is implemented in R statistical software (R Foundation for Statistical Computing, Vienna, Austria, 2016. URL: <https://www.R-project.org/>), a widely used language for scientific analysis today. We provide R scripts (in Supplementary files) that guide the execution of individual- and group-level DTW analyses, which can be applied to users' own datasets. We provide a tutorial on *undirected* networks, which capture patterns of similarity in symptom change over time, and *directed* networks, which model how change in one symptom may precipitate change in others, thereby representing potential mechanisms of change.

The clinical case data used in this tutorial are drawn from a randomized controlled trial (RCT) conducted in Norway. The overarching RCT is registered at ClinicalTrials.gov (NCT02649114). For more information on the study protocol (with ethical approval etc.), please consult Vrabel and colleagues [36]. The group-level DTW analysis was applied to real-world data with patients receiving treatment at a residential ward for severe eating disorders. For a full description of the study and its findings resulting from DTW analyses, please consult Kopland and colleagues [22, 23].

Methods

The general mechanism of DTW

Most longitudinal, idiographic network research to date has used vector autoregressive (VAR) models [13]. Despite various strengths, these traditional models need many assessments (intensive time-series data) and rely on several assumptions including linearity of associations and fixed time intervals between successive assessments (Haslbeck et al., 2021). This means the analysis assumes that symptom scores at one time point remain stable over time, so past scores from a patient in therapy are equally informative for future scores. Furthermore, if changes in symptoms occur over longer periods or abruptly, they cannot always be recognized as being associated within these traditional models [34]. This could lead to difficulties when working with data from clinical practice, which are rarely intensive times series data, and many of the assumptions are difficult to be met. For example, we expect change to happen in therapy and for some patients such changes may happen suddenly. The method for analysis needs to be able to handle these sudden changes.

DTW is a good alternative as it uses a non-linear, shape-based approach. This means that the shape-based algorithm compares the overall form or trajectory of two time series (of symptoms) rather than matching them strictly point-by-point in time, which might obscure meaningful symptom change across therapy. The shape-based algorithm uses ‘Elastic’ distance measures [4, 18], which does not assume linearity. An elastic distance measure allows *flexibility* in matching: one point in one series can align with several points in the other, or the matching can “stretch” and “compress” in time. In psychotherapy research, this means that if two patients with eating disorders work on their fear of eating—one showing much less restraint in week 2 and the other in week 6—DTW can recognize that their underlying processes are similar, even though the changes occur at different times. The key point is that “less fear of eating,” is associated with less “restraint”. DTW captures this by using a shape-based approach and an elastic, flexible algorithm. In the following, we will outline how DTW works by presenting the steps in DTW, from investigating a single item pair, aggregating individual data, making undirected and directed networks both at an individual level and group level. In this tutorial we will use data from a real-world setting with items from the Eating Disorder Questionnaire (EDE-Q). Each node in the network represents one item. Because many EDE-Q items in the network are symptoms (but not always), we use the terms ‘items’ and ‘symptoms’ interchangeably when describing the nodes.

How undirected DTW works—for a single item pair over time

Symptoms can be repeatedly measured within a person over time. Each point represents an assessment of where the intensity of the symptoms is recorded. By tracking these symptoms across multiple time points, we can examine whether they tend to rise and fall together, suggesting a potential relationship, or whether they fluctuate independently. This type of within-person temporal data allows for detailed analysis of how symptoms co-occur or diverge over time, offering valuable insights into individual symptom dynamics. DTW can be helpful in this context, as it calculates the distance for each pair of symptoms (between each symptom and all the other symptoms) in the model across several time points (with panel data or within a more extensive time series) [6]. In Fig. 1 examples of such time series are presented, showing the scores of two symptoms (Symptom A: ‘restraint’, and Symptom B: ‘fear of eating’) across 20 time points.

How undirected DTW works—for a set of items over time

An undirected network is a network of how several symptoms change together across time. Often such comparisons are made point by point. In other words, symptom scores at time point 1 in Symptom A are compared only with time point 1 in Symptom B, assuming perfect time alignment. This is called Euclidean distance measure. While calculating the Euclidean distance is straightforward, it can be misleading if similar symptom patterns occur at slightly different time points.

However, as two symptoms will not always change at the same pace, DTW looks for the best alignment between the two time series. It allows for more flexibility in how the time series are aligned. Instead of enforcing a strict point-by-point comparison, DTW enables non-linear stretching and compression along the time axis. As previously mentioned, this means that changes in one symptom can be aligned with similar changes in the other symptom, even if they occur a few timepoints earlier or later. Figure 2 illustrates the difference between Euclidean distance and DTW distance when comparing the temporal patterns of two symptoms across time. In the top panel, the Euclidean distance is shown. The vertical dashed lines highlight how even small shifts in timing can result in large perceived distances, despite the overall shape of the patterns being quite similar. The dashed lines in the bottom panel demonstrate how DTW finds the most meaningful alignment between the two patterns, despite differences in timing.

Other methods for network analysis typically rely on mean scores across time points. In contrast, DTW operates directly on raw scores, preserving the original variability in symptom dynamics. However, using raw scores introduces the challenge of comparing data across

Assessment	Symptom A	Symptom B
1	2	1
2	3	4
3	5	4
4	4	2
5	4	3
6	3	1
7	2	2
8	1	2
9	1	2
10	1	2
11	2	4
12	2	3
13	5	6
14	6	5
15	5	3
16	1	2
17	3	3
18	2	2
19	5	3
20	4	3

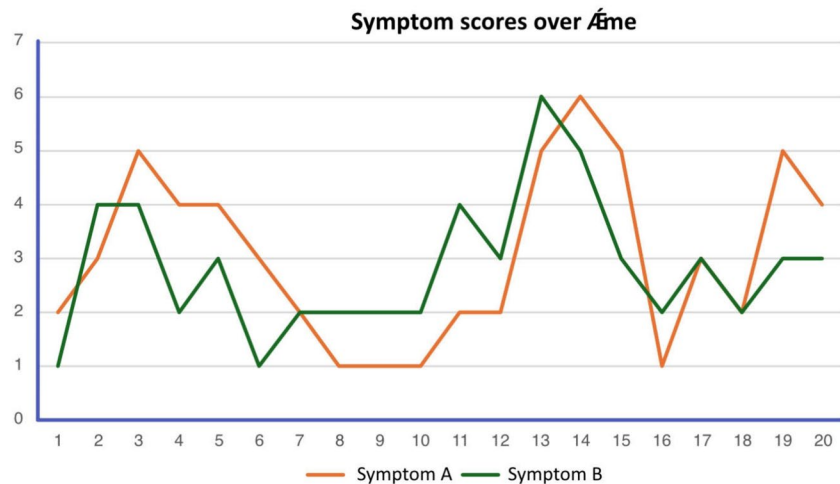


Fig. 1 Symptom scores over time. The chart and graph show how Symptom A (e.g., restraint) and Symptom B (e.g., fear of eating) change across assessment time points

different questionnaires with varying scoring systems. Sometimes, nodes in the network represent sum scores with a completely different range than nodes based on mean values or single-item scores. To address this, DTW applies scaled data during analysis, standardizing scores to enable meaningful comparisons across measures (see Fig. 3; [29]). When scores represent very rare occurrences in the dataset, such as symptoms like depersonalization or suicidal ideation, scaled scores can result in extreme values. In these cases, it may be advisable to cap scores at plus or minus 3 standard deviations.

Data can be scaled by using the ‘scale()’ function in R, which centers each variable around its mean and dividing by its standard deviation to ensure comparability across measures.

Aggregating individual data to generate an undirected group-level network

Undirected DTW can be used to compare each person’s symptom patterns with those of others, identifying similarities in how symptoms change over time. Importantly, DTW does not assume that one person is “ahead” or “behind” another; instead, it flexibly aligns symptom trajectories, regardless of differences in timing. This makes it possible to detect subgroups of patients with similar patterns of change, even when those patterns unfold at different speeds or times.

DTW relies on a dynamic programming approach that stretches and compresses time series to minimize a predefined distance measure [2, 15, 18]. For each pair of time series, DTW constructs a *cost matrix*, where each cell represents the local cost of aligning one observation from

the first series with one observation from the second. The algorithm then searches this matrix for the *optimal warping path*—the sequence of alignments that minimizes the cumulative distance while allowing for temporal stretching or compression. The resulting DTW distance summarizes the overall similarity between the two sequences after optimal alignment, and these distances can then be averaged across patients.

In this tutorial we report the results from a previous article; Kopland and colleagues (2024). In this study, the distance matrix contains $n*(n-1)/2 = 378$ distinct distances per patient (i.e., $(28*27)/2 = 378$ in our example, due to 28 questions in the EDE-Q), which, in the total group led to $122 * 378 = 46,872$ calculated DTW distances (as $N = 122$ in our study). In simple terms, the *cost matrix* shows every point-to-point comparison between two time series (e.g., restraint vs. fear of eating), while the *distance matrix* summarizes the overall similarity between entire symptoms and their relations to all other symptoms in the network (see Fig. 4).

The distance matrix was subsequently analyzed using hierarchical cluster analysis to identify groups of symptoms with similar change profiles. The optimal number of clusters was determined using the elbow method, based on explained variance as a function of the number of clusters (hierarchical Ward.D2 clustering). This analysis identified three symptom clusters, which were color-coded consistently throughout all figures.

Before DTW analysis, all EDE-Q items were standardized at the group level to ensure that distances reflected change dynamics over time rather than differences in symptom scale. Only significant edges are displayed in

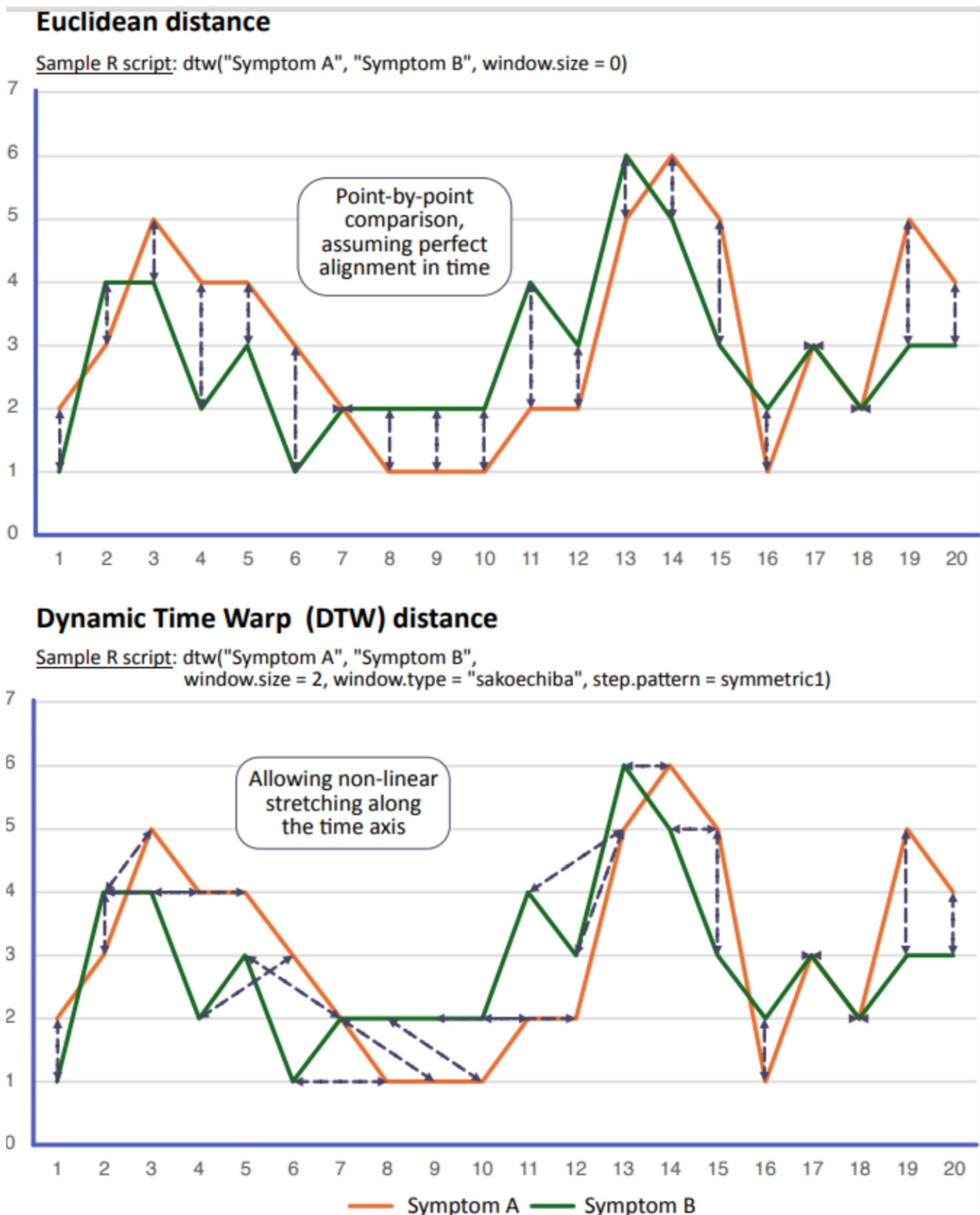
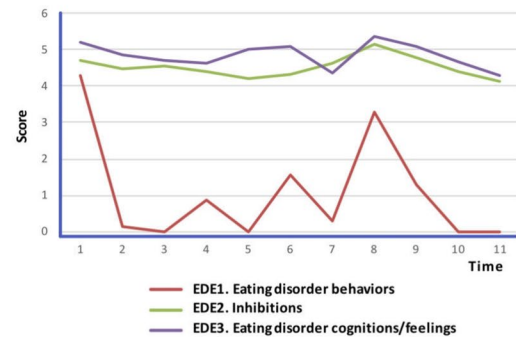


Fig. 2 Euclidean distance and DTW distance. Comparison of symptom time series using two methods. Top panel: direct, point-by-point comparison using Euclidean distance. Bottom panel: non-linear comparison using Dynamic Time Warping (DTW), which allows local “stretching” of the time axis. The R code for both methods is provided below the heading for tutorial purposes

Original data

time	EDE1. Eating disorder behaviors	EDE2. Inhibitions	EDE3. Eating disorder cognitions/feelings
1	4,29	4,71	5,21
2	0,14	4,47	4,86
3	0,00	4,54	4,71
4	0,86	4,38	4,62
5	0,00	4,19	5,00
6	1,57	4,33	5,07
7	0,29	4,63	4,36
8	3,29	5,14	5,36
9	1,29	4,79	5,07
10	0,00	4,38	4,64
11	0,00	4,13	4,29



Scaled data

time	EDE1. Eating disorder behaviors	EDE2. Inhibitions	EDE3. Eating disorder cognitions/feelings
1	2,19	0,68	1,10
2	-0,63	-0,16	0,06
3	-0,72	0,09	-0,35
4	-0,14	-0,48	-0,64
5	-0,72	-1,12	0,48
6	0,34	-0,63	0,69
7	-0,53	0,37	-1,39
8	1,51	2,14	1,52
9	0,15	0,94	0,69
10	-0,72	-0,48	-0,56
11	-0,72	-1,34	-1,60

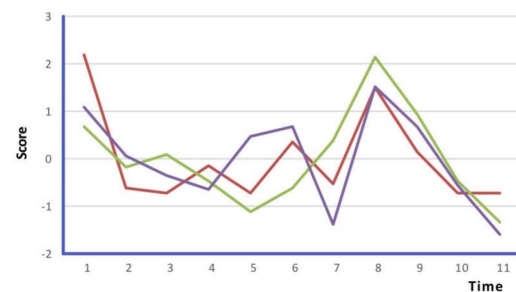


Fig. 3 Original and scaled data in DTW. Table and graph showing the difference between raw scores and scaled scores in three different eating disorder items (symptom dimensions)

the resulting undirected DTW network. In this network, smaller distances between two symptoms indicate that they tend to co-occur more similarly over time. The network was generated using the *qgraph* R package [12]. Figure 5 shows the undirected symptom network for 122 patients across therapy. To aid replication, we also provide an example R script with simulated data from nine patients across ten measurements (see Supplement 1 and <https://osf.io/fx8y5>).

In the network, *edges* (i.e., links between nodes) represent significant relationships ($p < 0.05$) based on the shortest distances between symptoms, adjusted for each patient’s average item scores over time. This adjustment helps to avoid spurious edges that might otherwise arise from symptoms that are generally rated at similar levels across assessments. The thickness of edges indicates the strength of similarity in change patterns, while the size of each node reflects its connectivity to other nodes.

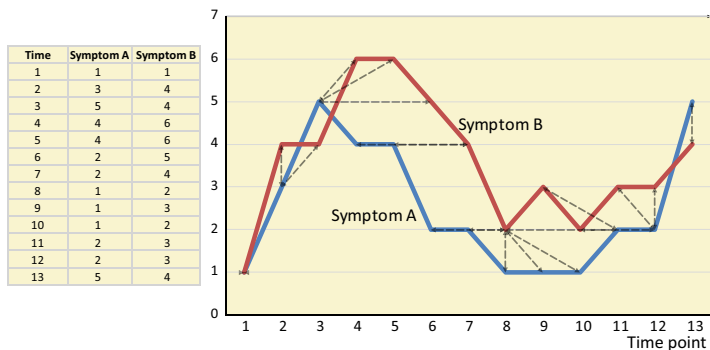
Finally, the standardized centrality of each of the 28 EDE-Q symptoms was calculated and presented in a bar graph. Symptoms with high centrality are more influential in the network, as they are strongly connected to many other symptoms. Such central symptoms are often considered valuable therapeutic targets, since changing them may influence other symptoms and thereby support recovery. In our network [22], three dimensions of change similarities were identified: eating disorder inhibition (yellow), eating disorder cognitions and feelings (red) and lastly, eating disorder behavior (blue).

Overvaluation of shape and weight showed the highest centrality and out-strength. This suggests that addressing thoughts and feelings about shape and weight could affect multiple other symptom clusters, potentially “dissolving” the broader symptom network of eating disorders. For more thorough discussion on the clinical implications of this study, please consult [22].

How directed DTW works—for a single item pair over time

Directed DTW estimates temporal lag patterns among nodes. In the context of therapy data, this approach can indicate which symptom changes are more likely to precede changes in other symptoms, thereby highlighting potential targets for intervention. One of the key conditions for causality is the presence of a temporal relationship, where the causal factor must occur before the consequence. Although directed DTW is based on observational data, it can bring us closer to understanding how change in one symptom may influence or lead to change in another. Figure 6 shows time-series data for three items (A, B, and C) across 40 measurements, along with the resulting undirected and directed DTW illustrations. The green line (Item C) follows a different pattern compared to the blue (Item A) and red (Item B) lines, fluctuating independently. In contrast, Items A and B show a more similar dynamic over time, moving roughly together in their ups and downs. As a result, in the undirected DTW network, an *edge* is drawn between Item A and Item B, indicating their similarity, while Item C

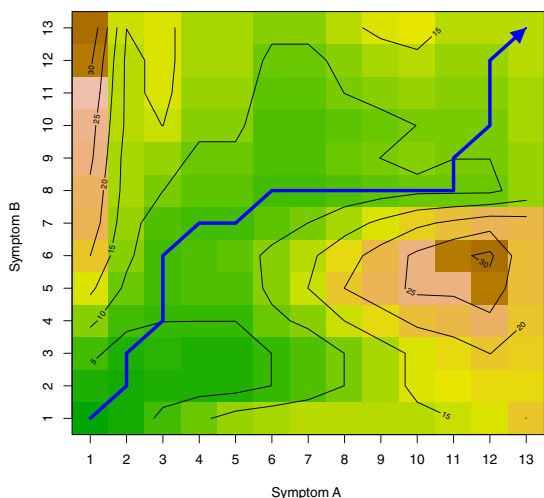
A. Symptom scores over time



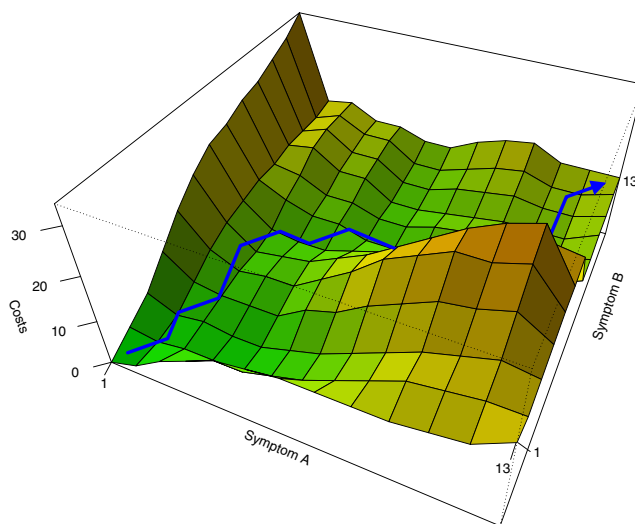
B. Sample R script

```
library("dtw")
SymptomA <- c(1,3,5,4,4,2,2,1,1,1,2,2,5)
SymptomB <- c(1,4,4,6,6,5,4,2,3,2,3,3,4)
route <- dtw(SymptomA, SymptomB, keep = TRUE)
route$stepPattern
route$distance
plot(route, type = "density")
```

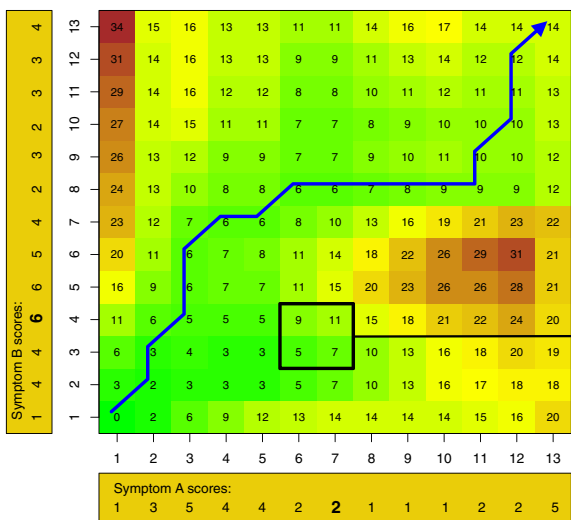
C. Density heatmap of cost matrix



D. 3D landscape plot of cost matrix



E. Cost matrix



F. Calculate one cumulative alignment cost

"Symmetric2" Step pattern recursion:

$$g[i,j] = \min($$

- $g[i-1,j-1] + 2 * d[i,j]$,
- $g[i,j-1] + d[i,j]$,
- $g[i-1,j] + d[i,j]$)

Step pattern: "symmetric2"

$g[i-1,j-1] = 5$
 $2 * d[i,j] = 2 * (6 - 2) = 8$
 $g[i-1,j-1] + 2 * d[i,j] = 5 + 8 = 13$

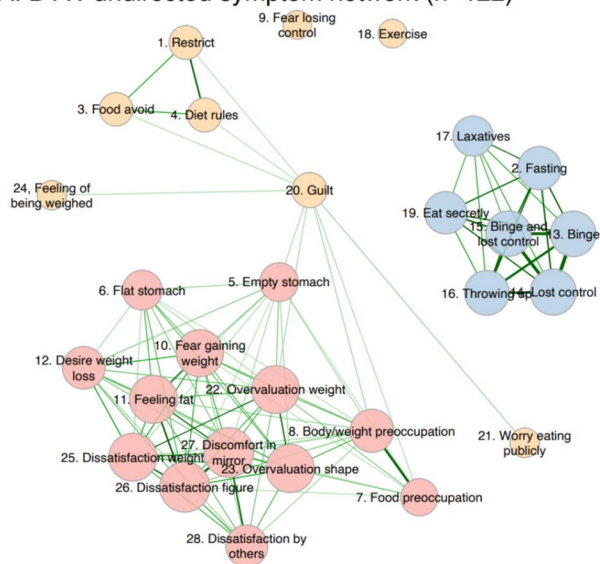
$g[i,j-1] = 9$
 $d[i,j-1] = (6 - 2) = 4$
 $g[i,j-1] + d[i,j] = 9 + 4 = 13$

$g[i-1,j] = 7$
 $d[i-1,j] = (6 - 2) = 4$
 $g[i-1,j] + d[i,j] = 7 + 4 = 11$

$g[i,j] = \min(13, 13, 11) = 11$

Fig. 4 Cost matrix in DTW. illustrates this process. Panel **A** shows two symptom trajectories over time and their DTW distance. Panel **B** provides the R script used to calculate the DTW distance. Panels **C** and **D** display the optimal alignment between two symptoms using a heatmap and a 3D plot. Panel **E** presents the cost matrix, representing the cumulative distance between two symptoms across all patients. Finally, Panel **F** outlines the mathematical definition of a single cell in the cost matrix

A. DTW undirected symptom network (n=122)



B. Standardized centrality of EDE symptoms

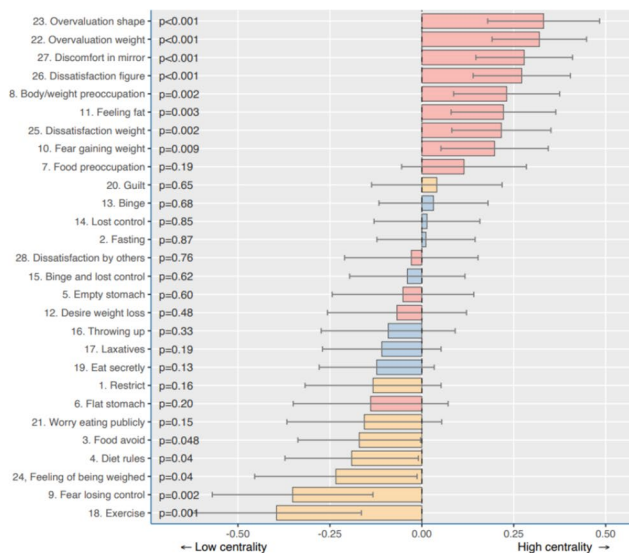


Fig. 5 Undirected DTW network of 122 patients with eating disorders. **A.** Showing an undirected network of how symptom clusters change similarly over time. The three colors represent clusters of symptoms with similar change profiles: blue indicates eating disorder behaviors, yellow reflects eating disorder-related inhibition, and red represents eating disorder-related cognitions and feelings. **B.** Showing the centrality of each item of the EDE-Q. Overvaluation of shape and weight being the most central items. The figure is retrieved from [22] (doi: <https://doi.org/10.1002/eat.24097>) with the general permission of Wiley

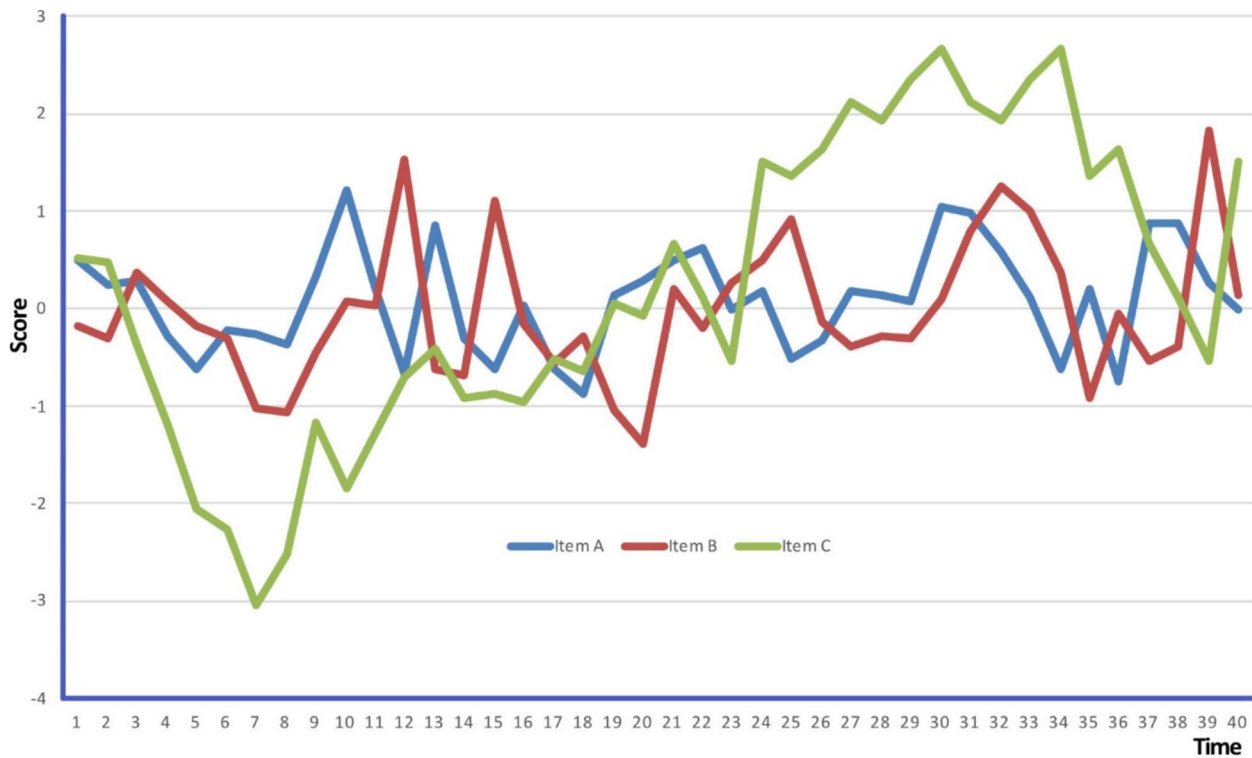
remains disconnected. In the directed DTW network, the analysis also considers the temporal ordering of changes: it is observed that changes in Item A tend to precede similar changes in Item B. Therefore, a directed edge (arrow) is drawn from Item A to Item B, suggesting that dynamics in A precedes and those of the dynamics in B, while Item C again remains isolated. In the directed DTW network, *temporal precedence* yields a directed edge from A to B. Temporal precedence means that one event or change occurs before another, establishing the correct time order between variables (e.g., in therapy data, a reduction in fear of eating must occur before a reduction in restraint for the former to be considered a possible driver of change).

Figure 7 builds on Fig. 6 by focusing only on the blue and red item scores (Item A and Item B), showing their point-by-point DTW alignments between the two time series. The analysis here uses a directed warping path, which only looks forward in time, comparing the relationship between Item A and Item B at lag 0 (the same time point) and lag 1 (one time point ahead). Many of the arrows point slightly forward in time, meaning that the best match between a value of Item A and a value of Item B often occurs shortly after, rather than exactly at the same time. This suggests that the distance to the "next" time point is mathematically shorter (i.e., a better match) than the distance at the current time point, which supports the idea that changes in Item A slightly precede changes in Item B. This directional alignment therefore

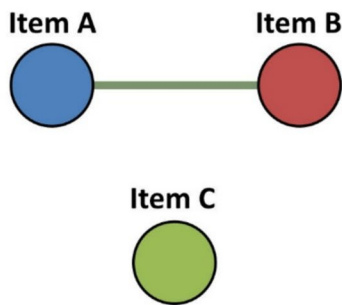
provides information about the degree of potential lead-lag relationships between the two items. A clinical example is that directed DTW can detect patterns of e.g. a reduction of 'fear of eating' (Item A) precedes a decrease in 'restraint' (Item B), thereby indicating temporal precedence. Issues of temporal precedence and model specification (bivariate vs. multivariate) will be addressed in the Discussion section.

As noted for the undirected DTW analysis, a 'cost matrix' underlies the point-by-point alignment of two time series. It is used to compare the similarity between two symptom time series, which essentially map how one pattern of change can be aligned with another over time (see Fig. 4). You can think of the cost matrix as a grid where each cell represents the 'cost' or difference between symptom scores at two time points, one from each series. DTW then finds the optimal path through this grid that minimizes the total cost, effectively identifying how the patterns align, even if one person's symptoms change faster or slower than another's.

In contrast to the undirected DTW analysis, we now aim to explain the cost matrix that underlies the directed (temporal) DTW analysis. This was done using a revised Sakoe-Chiba window band [30], which is a part of the DTW algorithm that constrains the window "searching" for optimal alignment. The Sakoe-Chiba was specified as being asymmetric. This means that the dynamic alignment was constrained to a single direction (i.e., only toward later time points) to examine directionality (see



Undirected DTW network



Directed DTW network

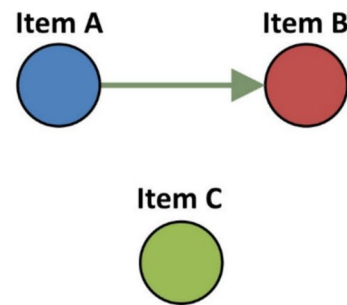


Fig. 6 Undirected and directed DTW of symptom change. Time-series data for three items **A**, **B**, **C** across 40 measurements and the resulting DTW networks. Items **A** (blue) and **B** (red) show similar dynamics, leading to an edge between them, while Item **C** (green) remains isolated. Because changes in Item **A** (blue) frequently precede similar changes in Item **B** (red), the directed DTW analysis identifies a directional relationship, represented by an arrow from **A** to **B** in the directed DTW network

Fig. 8). The Sakoe-Chiba also narrows the warping path to a band around the diagonal of the matrix, preventing excessive matching to time points very far away, as well as making the computation more efficient and accurate [30]. In Fig. 8, the optimal alignment path from Item A to Item B is shown as the red route, within this asymmetric window. The final directed distances between the two items A and B are calculated at the bottom. So, we use both distances, from Item A to Item B (as shown in Fig. 6 and 7) and the distance from Item B to Item A that was much longer (i.e., 25.1 versus 11.9). The relative difference in distance is calculated and highlights how temporal directionality between the symptom pair can be quantified.

How directed DTW works—for a group of individuals

For the tutorial on directed DTW, we again present the results from Kopland and colleagues [22]. Similarly, we conducted directed group-level DTW analyses. For each of the 122 patients, a directed distance matrix was estimated. Thus, for each person we first analyze every possible symptom pair to see if one symptom tends to change before the other. This produces a distance matrix for that individual. We then combine these individual networks from all participants into a single aggregated group-level network by averaging the results and testing which connections are statistically significant. Only the connections that pass this test are shown as arrows in the final network, indicating the direction from the symptom that

Item scores over time, with DTW warping

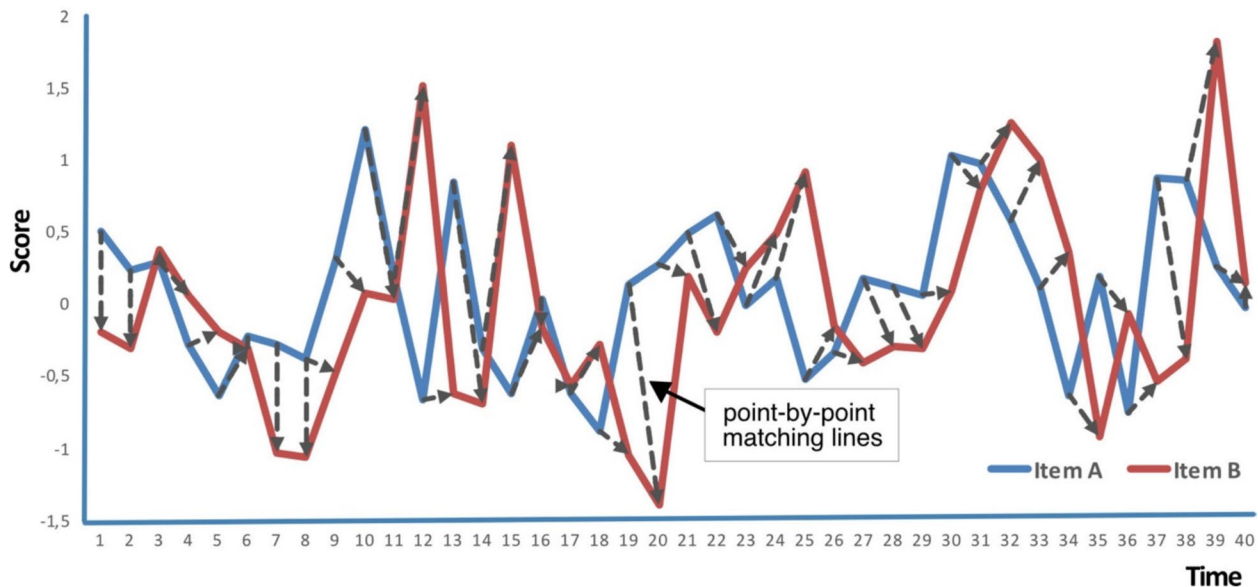


Fig. 7 DTW warping of item scores across therapy Item scores over time with DTW warping. The point-by-point DTW alignments are shown as arrows between Item A (blue, 'fear of eating') and Item B (red, 'restraint')

tends to change first to the symptom that tends to change later. Subsequently, all 122 distance matrices were combined to yield standardized *out-strength* (i.e., temporal lead, where changes in the symptom influence changes in other symptoms) and *in-strength* (i.e., temporal lag, where changes in other symptoms influence a symptom) centrality values, for which the confidence intervals were assessed through bootstrapping. See Fig. 9 for the network.

Large nodes that have thick edges with arrows pointing outwards indicate items (symptoms) that have high centrality and out-strength and thus have high importance and influence in the network. These symptoms could be potential mechanisms of change. Panel B shows the significance level of the out- and in-strength of the different EDE-Q items. In this network, change in overvaluation of shape seem to precede and affect change in other symptoms, especially overvaluation of weight, the wish for a flat/empty stomach and desire for weight loss. For a thorough discussion on the clinical implications of this network, please consult Kopland and colleagues [22]. For directed DTW, a sample R script is provided that simulates data from 6 patients across 10 measurements (see Supplement 2 and <https://osf.io/fx8y5>).

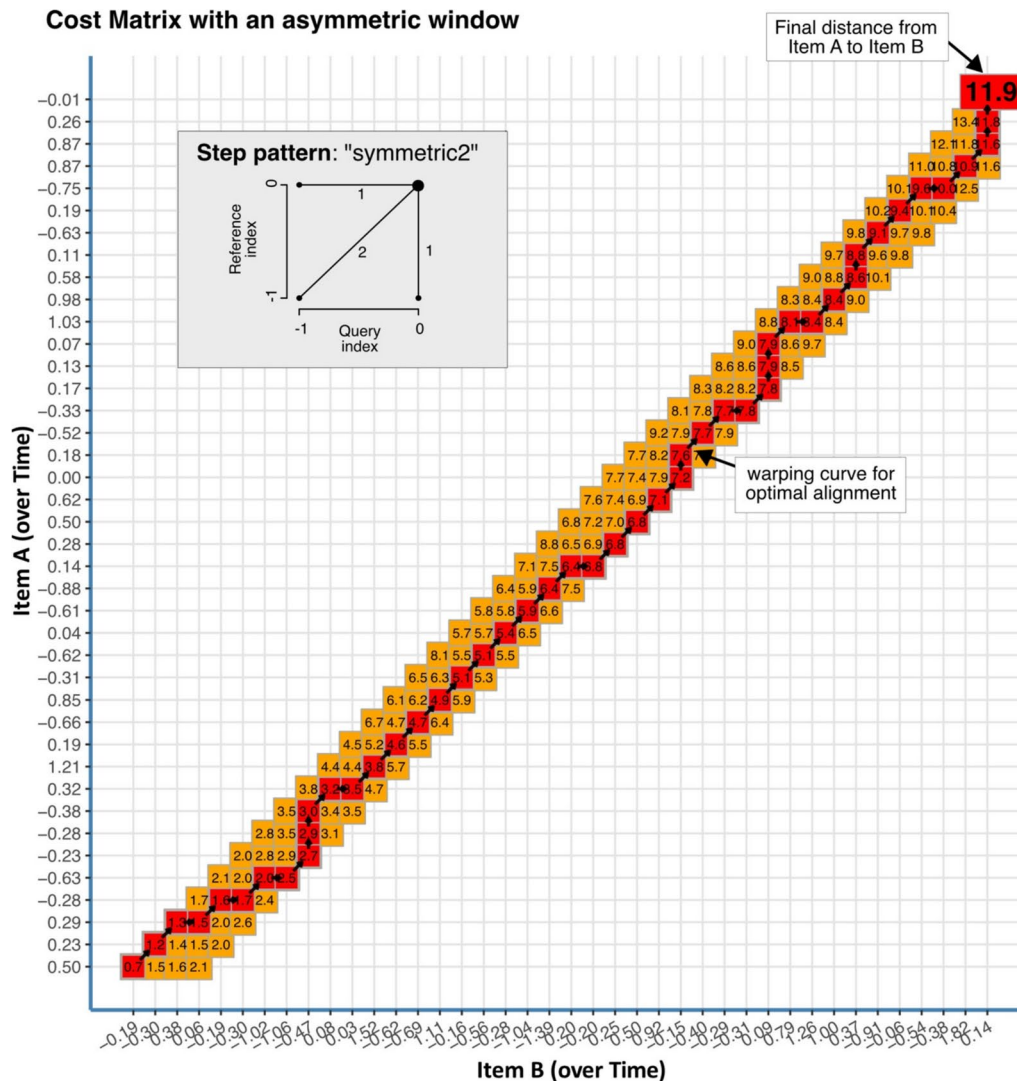
Discussion

This tutorial demonstrates the potential of DTW as a flexible and accessible method for investigating temporal dynamics in psychotherapy. By aligning and comparing individual time series data, DTW enables researchers and clinicians to examine processes within a person while

still allowing meaningful aggregation across individuals. This is particularly important given the challenge of reconciling individual and group-level findings in the move toward more personalized medicine [19].

Eating disorders can be conceptualized as disorders of a complex dynamic system, in which multiple symptoms, stressors, and biological characteristics interact over time in non-linear ways. DTW is well suited to this complexity because it captures the dynamic relationships between variables rather than focusing on them in isolation. By identifying lead-lag patterns among symptoms and external factors, DTW allows researchers to map the interplay of cognitive, emotional, behavioral, and physiological processes within a single analytic framework. In many contexts, this capacity makes DTW sufficient as a standalone method, particularly when the goal is to understand the temporal structure of symptom dynamics. At the same time, DTW can complement sophisticated modeling approaches, such as mIVAR, DSEM, and Latent Growth Curve Models [3], serving as an exploratory tool to detect patterns worthy of further modeling or as a fallback when data do not meet traditional assumptions.

A key advantage of DTW is its lower data requirements than mIVAR and other parametric network analysis techniques [34]. It can be applied to panel data, that has relatively few time points per participant. More work is needed on power calculations as there are no clear guidelines. However, both the number of measurements per individual and the number of participants is important for obtaining reliable DTW interpretations.



Calculation of the directed distances

$$\text{Distance}_{\text{item A to item B}} = \frac{\text{Distance}_{B \rightarrow A} - \text{Distance}_{A \rightarrow B}}{\text{Distance}_{B \rightarrow A} + \text{Distance}_{A \rightarrow B}} = \frac{25.1 - 11.9}{25.1 + 11.9} = \frac{13.2}{37.0} = 0.36$$

$$\text{Distance}_{\text{item B to item A}} = \frac{\text{Distance}_{A \rightarrow B} - \text{Distance}_{B \rightarrow A}}{\text{Distance}_{A \rightarrow B} + \text{Distance}_{B \rightarrow A}} = \frac{11.9 - 25.1}{11.9 + 25.1} = \frac{-13.2}{37.0} = -0.36$$

Fig. 8 Cost matrix of two items with warping curve alignment. Cost matrix between Item A and Item B. The red line shows the optimal warping curve computed using dynamic time warping (DTW). Each cell in the cost matrix corresponds to the local cost of aligning a point in Item A (y-axis) with a point in Item B (x-axis). The warping path is determined by selecting the sequence of cells (shown in red) that minimizes the cumulative alignment cost from the lower-left to the upper-right corner of the matrix. Because an asymmetric window is applied, the warping path is constrained along the diagonal but only allows forward progression in time. This means that each point in Item A can only be aligned with current or future points in Item B, ensuring a temporal alignment. In other words, the alignment avoids "backward" matches, which yields a directional relationship between the two time series. The final cumulative cost at the end of the warping path (i.e., 11.9) represents the DTW distance from Item A to Item B. This value, combined with the reverse distance (from Item B to Item A, being 25.1), allows for the calculation of the directed distance, which quantifies that time series A systematically preceded changes in time series B with a strength of 0.36

When the number of measurements within a single person is very high (e.g. >60), it is possible to conduct an 'n=1 study', allowing for identification of symptoms or stressors with a substantial temporal lead that may play

a potential causal role in worsening other symptoms. However, when the number of measurements per person is small, a larger sample size is required to obtain stable and reproducible group-level results. While a universal

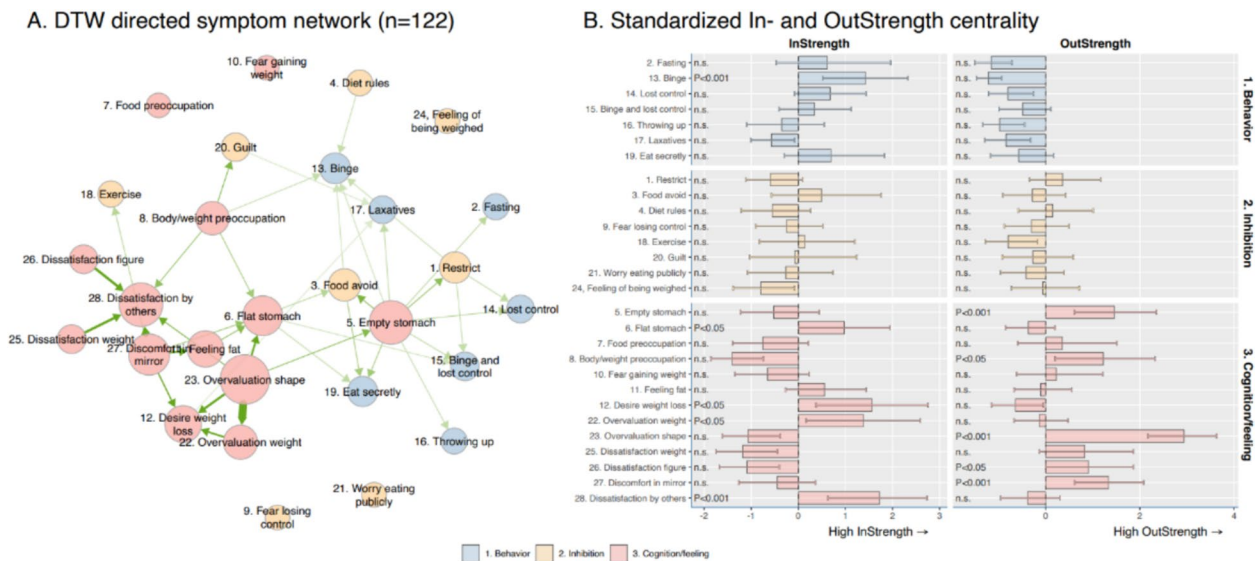


Fig. 9 Directed DTW network of 122 patients with eating disorders A directed DTW network. Panel A shows the network with arrows. Large nodes that have thick edges pointing outwards indicate items (symptoms) that have high out-strength and thus have high importance and influence in the network. Panel B shows the significance level of the out- and in-strength of the different EDE-Q items. The figure is retrieved from [22] (doi: <https://doi.org/10.1002/eat.24097>) with the general permission of Wiley

rule is not yet established, based on our experience, when there are around eight measurements per person, a sample size of approximately 100 participants is needed to produce reliable group-level directed DTW networks. The choice of measurement interval should be closely matched to the expected rate of change in the target variables. As Bringmann et al. [7] emphasize, variables that fluctuate quickly—such as eating disorder urges or mood states—require dense sampling to capture their dynamics, whereas slower-changing variables—such as BMI—necessitate longer intervals to allow meaningful change to occur. Aligning the measurement interval with the temporal nature of the variable is critical for obtaining interpretable DTW results.

In contrast to multivariate time series models such as mIVAR, DTW do not require long, stationary time series, or parametric assumptions. A recent simulation study showed that unlike mIVAR, DTW did not generate spurious edges in the presence of colliders [34]. A collider is a variable that is influenced by two (or more) other variables, such that conditioning on it can create a spurious association between those variables. These features make DTW particularly suitable for clinical applications, where frequent measurement may not be feasible and where nodes may have some content overlap. In this sense, DTW represents a methodological middle ground: more sensitive to individual dynamics than cross-sectional analyses, yet more scalable and practical than complex modeling techniques.

Beyond its feasibility, DTW also provides intuitive though complex visual and quantitative outputs that can be directly interpreted. For instance, the matrices

of similarity and alignment paths can reveal how closely individuals follow similar trajectories over time, despite possible differences in timing or phase. This can be especially valuable in clinical work, where understanding the temporal structure of change—whether symptoms improve steadily, fluctuate, or follow delayed patterns—can inform intervention planning and monitoring.

While DTW offers a flexible approach to analyzing therapy dynamics, several methodological limitations warrant consideration. First, unlike autoregressive models, DTW cannot infer true Granger causal relationships as it is a bivariate model. *Granger causality* [16] is an important concept in psychotherapy research where one wants to move closer to causality and answer questions like which symptoms changes happen before others. It is a statistical method used to examine whether the values in one time series (A) at a particular time point (t -1) can predict the future values in another time series (B) at a different time point (t) while accounting for the autocorrelation in series B between time points (t-1) and (t). By applying DTW analysis in both directions (A predicting B and B predicting A), we can determine whether there is predictive relationship between time series A and B. In simpler terms, it is like noticing that every time a patient's stress level rises, their sleep quality tends to drop a day later, hinting at a potential link. However, if stress (A) causes poor sleep (B), and poor sleep causes low mood (C), DTW may also make it appear as if stress also directly causes low mood — even though the relationship is actually indirect, while stress only precedes low mood. This illustrates DTW's limitation: it identifies temporal order, but does not establish causality. Unlike

multivariate Granger causality models (e.g., mIVAR [34], DTW does not adjust for confounders or account for autocorrelation structures, so its result should be interpreted as suggestive patterns rather than evidence of strict causal relationships. However, if we can establish temporal order (i.e., directionality), we can screen symptoms to experimentally manipulate in therapy to help inform personalized care.

Second, DTW is sensitive to noise and may overfit idiosyncratic (individual) fluctuations, particularly when the number of participants is small, or when time series are short, irregular, or affected by substantial measurement error. Third, DTW requires sufficient variation in symptom scores over time within individuals. If a node/symptom trajectory shows little to no fluctuation—for example, consistently recording zero scores—it tends to form very strong edges with other non-fluctuating nodes, even though no meaningful dynamic information is actually present for that symptom. Fourth, aggregating individual DTW-based results to the group level remains an open methodological challenge. Because DTW captures highly individualized patterns of symptom alignment, aggregating results to the group level remains a methodological challenge, as it risks obscuring the very person-specific dynamics that DTW is designed to detect. Furthermore, because DTW emphasizes overall pattern similarity rather than strict time-point correspondence, it may obscure clinically meaningful short-term changes. Fifth, when time points are spaced far apart, DTW may fail to detect temporal relationships that occur within these intervals, such as rapid symptom shifts following therapeutic interventions or acute stressors.

Future directions

Looking forward, continuously collecting and analyzing data streams with DTW could support adaptive, real-time monitoring systems that provide rapid, interpretable feedback on individual trajectories. This makes DTW suitable for integration into wearable devices or smartphone algorithms, identifying personalized temporal lead–lag relationships—such as patterns of activity, mood, or physiological changes that precede symptom escalation—and delivering timely, tailored interventions in the form of prompts or “nudges” to encourage adaptive behaviors or prevent deterioration. For example, if DTW detects that an increase in daily activity consistently precedes a spike in eating disorder thoughts, the app could prompt the patient to increase food intake, reduce activity or contact a health professional as a preventive strategy. Such an approach could make DTW a practical tool for personalized, real-time clinical support outside of traditional therapy sessions. Just-in-time adaptive interventions (JITAs) delivered via mobile health technologies can provide scalable, context-aware, and individualized

support [37], and DTW might further strengthen such adaptive systems by identifying personalized lead-lag dynamics.

In sum, DTW provides an interpretable and feasible approach that can both stand alone and integrate with other methods. Future research should continue to refine DTW-based techniques, explore their integration with complementary approaches and new approaches (such as wearable devices and smartphone algorithms), and test their utility across diverse clinical contexts and populations.

Abbreviations

DTW	Dynamic time warp
DSEM	Dynamic structural equation modeling
EDE-Q	Eating disorder examination questionnaire
LGCM	Latent growth curve model
mIVAR	Multilevel vector autoregression model
RCT	Randomized controlled trial

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40337-025-01414-8>.

Additional file 1.

Additional file 2.

Author contributions

M.K. took the lead in writing the manuscript in close collaboration with E.G. E.G. has developed the code shared in this manuscript. Both authors have read and approved the final manuscript.

Funding

Open access funding provided by University of Oslo (incl Oslo University Hospital). No funding.

Data availability

The tutorial codes for undirected (<https://osf.io/gjveb>) and directed (<https://osf.io/gjveb>) individual and group-level analyses, as well as the sample code used in our analyses of patient data is available here: [<https://osf.io/tx8y5>] (<https://osf.io/4h6be>) and is also added in this paper (supplement 1 and 2).

Declarations

Ethics approval and consent to participate

The study was preregistered in Clinical Trial Registration [ClinicalTrials.gov](https://www.Clinicaltrials.gov), <https://www.Clinicaltrials.gov/ct2/show/NCT02649114>. The study was approved by the Norwegian Regional Committee for Ethics in Medical and Health Research (REK). The project's identification number is 2014/836/REK. Patients were informed of the repeated weekly questionnaire and the purpose. We requested patients' permission to use their data for research. To secure patients' confidentiality, all patient-identifiable data were removed from the database.

Competing interests

The authors declare no competing interests.

Received: 29 April 2025 / Accepted: 18 September 2025

Published online: 21 October 2025

References

- Asparouhov T, Hamaker EL, Muthén B. Dynamic structural equation models. *Struct Equ Modeling*. 2017;25(3):359–88. <https://doi.org/10.1080/10705511.2017.1406803>.
- Berndt, D. J., & Clifford, J. Using dynamic time warping to find patterns in time Series. Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining, Seattle, WA. 1994.
- Bollen KA, Curran PJ. *Latent curve models: a structural equation perspective*. Hoboken, NJ: Wiley; 2006.
- Booij MM, van Noorden MS, van Vliet IM, Ottenheim NR, van der Wee NJA, Van Hemert AM, et al. Dynamic time warp analysis of individual symptom trajectories in depressed patients treated with electroconvulsive therapy. *J Affect Disord*. 2021;293:435–43. <https://doi.org/10.1016/j.jad.2021.06.068>.
- Borsboom D. A network theory of mental disorders. *World Psychiatry*. 2017;16(1):5–13. <https://doi.org/10.1002/wps.20375>.
- Borsboom D, Deserno MK, Rhemtulla M, Epskamp S, Fried EI, McNally RJ, et al. Network analysis of multivariate data in psychological science. *Nat Rev Methods Primers*. 2021;1(1):58.
- Bringmann LF, Albers C, Bockting C, Borsboom D, Ceulemans E, Cramer A, et al. Psychopathological networks: theory, methods and practice. *Behav Res Ther*. 2022;149:104011. <https://doi.org/10.1016/j.brat.2021.104011>.
- Bryant E, Marks P, Griffiths K, Boulet S, Pehlivan M, Touyz S, et al. Treating the individual: moving towards personalized eating disorder care. *J Eat Disord*. 2025;13:63. <https://doi.org/10.1186/s40337-025-01246-6>.
- Cramer AO, Waldorp LJ, Van Der Maas HL, Borsboom D. Comorbidity: a network perspective. *Behav Brain Sci*. 2010;33(2–3):137–50. <https://doi.org/10.1017/S0140525X09991567>.
- Cramer AO, van Borkulo CD, Giltay EJ, van der Maas HL, Kendler KS, Scheffer M, et al. Major depression as a complex dynamic system. *PLoS ONE*. 2016;11(12):e0167490. <https://doi.org/10.1371/journal.pone.0167490>.
- Dingemans AE, Giltay EJ, Rohrbach PJ, van Furth EF, Slof-Op 't Landt MCT. Understanding personalized dynamics in eating disorders: a dynamic time warp analysis. *Int J Eat Disord*. 2025. <https://doi.org/10.1002/eat.24407>.
- Epskamp S, Cramer AO, Waldorp LJ, Schmittmann VD, Borsboom D. Qgraph: network visualizations of relationships in psychometric data. *J Stat Softw*. 2012;48:1–18. <https://doi.org/10.18637/jss.v048.i04>.
- Epskamp S, Deserno M, K, & Bringmann, L. F. (2020). *mlVAR: Multi-level vector autoregression*. R package version 0.4.4. <https://cran.r-project.org/web/packages/mlVAR/mlVAR.pdf>
- Fisher AJ, Medaglia JD, Jeronimus BF. Lack of group-to-individual generalizability is a threat to human subjects research. *Proc Natl Acad Sci USA*. 2018;115(27):E6106. <https://doi.org/10.1073/pnas.1711978115>.
- Giorgino T. Computing and visualizing dynamic time warping alignments in R: the dtw package. *J Stat Softw*. 2009;31(7):1–24. <https://doi.org/10.18637/jss.v031.i07>.
- Granger CW. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*. 1969;37(3):424–38. <https://doi.org/10.2307/1912791>.
- Haslbeck JMB, Borsboom D, Waldorp LJ. Moderated network models. *Multivar Behav Res*. 2019;56(2):256–87. <https://doi.org/10.1080/00273171.2019.1677207>.
- Hebbrecht K, Stuivenga M, Birkenhager T, Morrens M, Fried EI, Sabbe B, et al. Understanding personalized dynamics to inform precision medicine: a dynamic time warp analysis of 255 depressed inpatients. *BMC Med*. 2020;18(1):400. <https://doi.org/10.1186/s12916-020-01867-5>.
- Hekler EB, Klasnja P, Chevance G, Golaszewski NM, Lewis D, Sim I. Why we need a small data paradigm. *BMC Med*. 2019;17(1):133. <https://doi.org/10.1186/s12916-019-1366-x>.
- Kazdin AE. Mechanisms of change in psychotherapy: advances, breakthroughs, and cutting-edge research (Do Not Yet Exist). In: Bootzin RR, McKnight PE, editors. *Strengthening Research Methodology: Psychological Measurement and Evaluation*. Berlin: American Psychological Association; 2006. p. 77–101.
- Kazdin AE. Understanding how and why psychotherapy leads to change. *Psychother Res*. 2009;19(4–5):418–28. <https://doi.org/10.1080/10503300802448899>.
- Kopland M, Vrabel K. Breaking the cycle: identifying key symptom pathways of eating disorders and the influence of childhood maltreatment. *Int J Eating Dis*. 2024;57(2):316–26. <https://doi.org/10.1002/eat.24097>.
- Kopland MCG, Vrabel K, Landt MS-O. Network dynamics of self-compassion, anxiety, and depression during eating disorder therapy. *Eur Eating Dis Rev*. 2025;33(1):35–52. <https://doi.org/10.1002/erv.3121>.
- Levinson H, Keshishian B, Vanzhula C. Using individual networks to identify treatment targets for eating disorder treatment: a proof-of-concept study and initial data. *J Eat Dis*. 2021;9(1):147. <https://doi.org/10.1186/s40337-021-00504-7>.
- Levinson CA, Williams BM, Christian C, Hunt RA, Keshishian AC, Brosorf LC, et al. Personalizing eating disorder treatment using idiographic models: an open series trial. *J Consult Clin Psychol*. 2023;91(1):14. <https://doi.org/10.1037/ccp0000785>.
- Levinson CA, Christian C, Becker CB. How idiographic methodologies can move the clinical-science field forward to integrate personalized treatment into everyday clinical care and improve treatment outcomes. *Clin Psychol Sci*. 2025;13(1):69–82.
- Molenaar (2004). A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever. *Measurement: Interdisciplinary Research and Perspectives*. 2(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1
- Molenaar PCM, Campbell CG. The new person-specific paradigm in psychology. *Curr Dir Psychol Sci*. 2009;18(2):112–7. <https://doi.org/10.1111/j.1467-8721.2009.01619.x>.
- Mueen, A., Keogh, E. J., (2016): Extracting Optimal Performance from Dynamic Time Warping. Conference on Knowledge Discovery and Data Mining (KDD) 2016: 2129–2130. <https://www.cs.unm.edu/~mueen/DTW.pdf>
- Murtagh F, Legendre P. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *J Classif*. 2014;31:274–95. <https://doi.org/10.1007/s00357-014-9161-z>.
- Olthof M, Hasselman F, Oude Maatman F, Bosman AM, Lichtwarck-Aschoff A. Complexity theory of psychopathology. *J Psychopathol Clin Sci*. 2023;132(3):314–23. <https://doi.org/10.1037/abn0000740>.
- Scheffer M, Bockting CL, Borsboom D, Cools R, Delecroix C, Hartmann JA, et al. A dynamical systems view of psychiatric disorders—practical implications: a review. *JAMA Psychiatr*. 2024;81(6):624–30. <https://doi.org/10.1001/jamapsychiatry.2024.0228>.
- Slof Landt MCT, Dingemans AE, Giltay EJ. Eating disorder psychopathology dimensions based on individual co-occurrence patterns of symptoms over time: a dynamic time warp analysis in a large naturalistic patient cohort. *J Eating Weight Dis*. 2022;27(8):3649–63. <https://doi.org/10.1007/s40519-022-01504-5>.
- van der Does F, van Eeden W, Bringmann LF, et al. Dynamic time warp versus vector autoregression models for network analyses of psychological processes. *Sci Rep*. 2025;15:11720. <https://doi.org/10.1038/s41598-025-94782-9>.
- Vrabel K, Hoffart A, Ro Ø, Martinsen EW, Rosenvinge JH. Co-occurrence of avoidant personality disorder and child sexual abuse predicts poor outcome in long-standing eating disorder. *J Abnorm Psychol*. 2010;119(3):623–9. <https://doi.org/10.1037/a0019857>.
- Vrabel KR, Wampold B, Quintana DS, Goss K, Waller G, Hoffart A. The modum-ED trial protocol: comparing compassion-focused therapy and cognitive-behavioral therapy in treatment of eating disorders with and without childhood trauma: protocol of a randomized trial. *Front Psychol*. 2019;10:1638. <https://doi.org/10.3389/fpsyg.2019.01638>.
- Zainal NH, Liu X, Leong U, Yan X, Chakraborty B. Bridging innovation and equity: advancing public health through just-in-time adaptive interventions. *Annu Rev Public Health*. 2025. <https://doi.org/10.1146/annurev-publhealth-071723-103909>.

Publisher's Note

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