Journal of Behavioral Data Science, 2024, *4* (1), 19–44.
DOI:https://doi.org/10.35566/jbds/daniel

Stability and Spread: Transition Metrics that are Robust to Time Interval Misspecification

Katharine E. Daniel[1] [0000−0002−4371−2740],
Robert G. Moulder[2] [0000−0001−7504−9560],
Matthew W. Southward[3] [0000−0002−5888−2769],
Jennifer S. Cheavens[4] [0000−0002−3778−9346], and
Steven M. Boker[1] [0000−0003−2117−8548]

[1] Department of Psychology, University of Virginia, Charlottesville, USA
ked4fd@virginia.edu, smb3u@virginia.edu
[2] Institute of Cognitive Science, University of Colorado Boulder, Boulder, USA
robert.moulder@colorado.edu
[3] Department of Psychology, University of Kentucky, Lexington, USA
southward@uky.edu
[4] Department of Psychology, The Ohio State University, Columbus, USA
cheavens@osu.edu

Abstract. Intensive longitudinal data collected via ecological momentary assessment (EMA) are often sampled with unequal time spacing between surveys. Given the popularity of EMA data, it is important to understand whether time series methods are robust to such time interval misspecification. The present study demonstrates via simulation that stability and spread—two metrics for quantifying different aspects of transitioning behavior within multivariate binary time series data—are unbiased when applied to data that are collected along an off/on burst sampling schedule, a between-person random sampling schedule, and a within-person random sampling schedule. These results held in randomly generated data with differing numbers of time series variables (k=10 and k=20) and in data simulated based on the proportions of observed data from a prior EMA study. Further, stability and spread demonstrated approximately 95% coverage for all between- and within-person random sampling schedules. However, coverage for stability and spread was poor in the off/on burst sampling schedules (around 67%). We also applied these transition metrics—which measure repetitiveness and diversity of transitions, respectively—to a foundational EMA dataset that was among the first to show that adults regularly use many different emotion regulation strategies throughout their daily life (Heiy & Cheavens, 2014). As hypothesized, we found a stronger positive relation between mood and higher stability/lower spread in emotion regulation among people with fewer depressive symptoms than those with more depressive symptoms. Taken together, stability and spread appear to be appropriate metrics to use with data collected using common unequal time spacing conditions and can be used to uncover theoretically consistent insights in real psychosocial data.

Keywords: Time series · Ecological momentary assessment · Transitions · Binary data · Emotion regulation · Switching

1 Introduction

Recent advances in technology have contributed to an explosion in intensive longitudinal data (ILD) collection. ILD studies involve repeated data collection throughout participants' days, often for multiple days at a time, and may involve passive sensing (i.e., automatic data reports from electronic devices participants have with them) or participant-reported data. Researchers may collect data according to a variety of schedules, depending on the research question. Three common methods include fixed interval sampling (e.g., every 15 minutes; Ebner-Priemer & Sawitzki, 2007), random interval sampling (e.g., Burke et al., 2017), or burst designs (i.e., sampling intensively for a relatively brief duration, pausing for a relatively longer time, before sampling intensively again; Stawski, MacDonald, & Sliwinski, 2015)

ILD may be categorical, ordinal, continuous, or qualitative, depending on the research question. Here, we focus on binary data indicating the presence or absence of a construct. For instance, researchers may be interested in whether or not participants were experiencing an emotion (e.g., Heiy & Cheavens, 2014), used an emotion regulation strategy (Southward & Cheavens, 2020), or were in an interpersonal context (Daros et al., 2019). Binary data is important because it allows for the study of transitions between two states—from present to absent and vice versa—which underlie many theories of human behavior (e.g., attentional focus under conditions of threat: Johnson, Zaback, Tokuno, Carpenter, & Adkin, 2019) and facilitate the development of machine learning models (e.g., Pereira, Mitchell, & Botvinick, 2009; Steinwart & Christmann, 2008). Furthermore, collecting many different binary variables allows for the study of transitions between one endorsed construct to another—a transition type called a 'switch.'

Daniel, Moulder, Teachman, and Boker (2023) presented a method for quantifying two aspects of switching behavior—stability and spread—within multivariate binary time series data to characterize how a system moves between many possible binary states over time. Specifically, this method functions by building a sliding series of transition matrices where each transition matrix reflects all pairwise transitions that occurred between back-to-back observations within that pre-defined window of observations.[5] Two matrix algebra equations are then applied to each transition matrix to arrive at stability and spread values, which can be used in analysis. Stability measures the extent to which a system transitions from one binary variable to the same binary variable at the next time point relative to all observed transitions. Spread measures how many

[5] Taking this windowing approach offers researchers the option to calculate stability and spread values multiple times at different parts of the full time series.

unique pairwise transitions are observed relative to all possible transitions afforded by the system.

Stability and spread are especially useful when researchers want to measure how rigid and diverse, respectively, transitions are within a high dimensional system. For example, panic disorder researchers could ask participants to report where they were located during the onset of each panic attack. One person with panic disorder might tend to have panic attacks in the same location (high stability) whereas a second person might variably have panic attacks across different locations (low stability). Two other people with equally low stability might tend to either alternate between having panic attacks at home and at work (low spread) or tend to have panic attacks across many different locations with minimal regularity (high spread).

Although stability and spread indices characterize the *sequence of transitions* between observations, they do not account for *elapsed time* between observations. This is unproblematic if every transition is captured in the data or if the process of interest is sampled at equal intervals. However, this may pose a limitation to the method if transitions are sampled at unequal time intervals. Time misspecification may be especially pertinent to psychosocial researchers because ILD are often collected with unequal intervals between observed data points (Myin-Germeys et al., 2018), but unequal sampling intervals are not well-suited to analytic tools that assume equal time intervals between observations and participants (Stone & Shiffman, 1994). Therefore, stability and spread may be biased or demonstrate poor coverage if they are applied to ILD sampled with unequal time spacing.

However, stability and spread are calculated with a sliding window approach (Daniel et al., 2023), which has previously been shown to be an effective way to handle time misspecification in other use cases (Boker, Tiberio, & Moulder, 2018). Thus, these measures may be inherently robust to time interval misspecification, a quality that would make them optimally useful in psychosocial research without needing to further expand the method. Therefore, in this paper we aim to establish the robustness of stability and spread to time interval misspecification through simulation. We then demonstrate the method's potential utility to advance theory in psychosocial research by applying stability and spread to unequally sampled emotion regulation (ER) strategy self-reports throughout the daily lives of adults with varying levels of depression.

2 Methods

2.1 Data Simulation

To explore how robust stability and spread are when applied to data with time interval misspecification, we simulated three datasets—two simulated randomly and one simulated based on proportions of endorsements observed in a prior EMA study. We describe each set of simulations below.

Randomly Generated Simulations. First, we simulated data for 100 people where each person reported on 10 binary time series variables 200 times. Responses to each binary variable (1 vs. 0) were randomly determined. We consider this our k = 10 parent dataset.[6]. From this parent dataset, we generated four datasets per person, each with 25 observations per person: one reference dataset with equally spaced observations and three further datasets, each with a unique time-interval misspecification. To create the equally spaced dataset, we selected every eighth observation from the 200-observation-long parent dataset. To impose three common types of time interval misspecification from which to compare, we selected from the original 200: (1) observations 36-40, 76-80, 116-120, 156-160, 196-200 (to mimic a burst sampling schedule when equally spaced sampling occurs during the "on burst" and no sampling happens during the "off burst"), (2) 25 random observations held constant between persons (to mimic a sampling schedule with unequal spacing between observations but identical spacing between person), and (3) 25 random observations for each person (to mimic unequal spacing within and between person). We conducted 1,000 replications in each condition of this simulation to arrive at the simulations used to calculate bias and coverage. Although we allowed the equally spaced dataset to vary across all simulations used to calculate bias (i.e., it could start its equal sampling on any of the first 8 observations), we fixed the equal sampling dataset to be the same across all simulations used to calculate coverage so that there would be one set of "true" stability or spread values against which to compare.

We then repeated these steps, making only one change—we simulated a parent dataset with 20 time series. We refer to this as our k = 20 parent dataset.

Data-Informed Simulation. Because stability and spread have been proposed as potentially meaningful characterizations of ER strategy use in daily life (Daniel, Moulder, Boker, & Teachman, 2024), we simulated data for 100 people where each person was measured 200 times on nine binary variables based on real data proportions from a two-week EMA study measuring use of nine different ER strategies (Daros et al., 2019). In this study, 140 undergraduates reported up to six times per day whether they used any of nine ER strategies. Responses to each binary variable at each time point were drawn from a binomial distribution with the following proportions of successes: .618, .048, .110, .107, .072, .052, .014, .055, .076. These values reflect the proportion of observations where each of the following responses were selected out of all submitted EMA surveys: not regulating one's emotions, cognitive reappraisal, problem solving, introspection, acceptance, advice-seeking, expressive suppression, emotional suppression, and distraction, respectively (Daros et al., 2019). Proportions do not sum to 1 because multiple ER strategies could be endorsed simultaneously. We refer to this

[6] We focused on k = 10 variables and k = 20 variables because stability and spread were designed to calculate transition information from higher dimensional data sources than existing approaches typically handle (Daniel et al., 2023) and yet are still likely to be measured by psychosocial researchers (e.g., Heiy & Cheavens, 2014)

as our data-informed parent dataset. All remaining steps were identical to those detailed above.

Calculating Stability and Spread. We calculated stability and spread for each sub-dataset using Daniel et al. (2023)'s methods. Specifically, we constructed transition matrices using the 'buildTransArray' function in the *TransitionMetrics* package (Daniel & Moulder, 2020). To demonstrate how the transition matrices are built from data, imagine a participant i rated whether or not they used each of four different ER strategies (ER1, ER2, ER3, ER4) at 6 time points (T1, T2, T3, T4, T5, T6,). With these data (given on the top of Table 1) and assuming a window of 5 and a lag of 1, we can construct two transition matrices (X_{i1}, X_{i2}).

To construct X_{i1}, we would start by creating a 4 x 4 matrix for which all elements are initialized to zero. The example data show that ER1 occurred at the first observation (T1) and ER3 occurred at the second observation (T2). This means that a pairwise transition from ER1 to ER3 occurred between the first two time points. Thus, we increment the (3,1) element of X_{i1} by one (to reflect the transition from ER1 to ER3). All other elements remain at 0. At the next observation (T3), ER3 was endorsed again, indicating a pairwise transition from ER3 to ER3 between T2 and T3. To reflect this pairwise transition, we increment the (3,3) element of X_{i1} by one. At the next observation (T4), ER3 was once again endorsed, indicating a pairwise transition from k3 to k3 between T3 and T4. Thus, we increment the (3,3) element of X_{i1} by one such that the (3,3) element now equals two. At the next observation (T5), ER2 was endorsed, indicating a transition from ER3 to ER2 between T4 and T5. Thus, we increment the (2,3) element of X_{i1} by one, such that the (2,3) element now equals one. At this point, all transitions between the four binary variables across the first five time points are reflected in X_{i1} (see Table 1).

To construct X_{i2} we would start with a second 4 x 4 matrix, also initialized to zero. The window of observations being read into X_{i2} would be shifted down the time series by one compared to what was read into X_{i1}, such that the transitions between T1 and T2 described above would not be captured by the new matrix. The transitions between T2 and T3, T3 and T4, and T4 and T5, however, would be incremented into the new matrix like in X_{i1}. Finally, because the window of observations was shifted down one, there would be one new transition to add to X_{i2} (i.e., the transition between T5 and T6). In these example data, ER4 was endorsed at T6, indicating that a transition from ER2 to ER4 occurred between T5 and T6. Thus, we increment the (4,2) element of X_{i2} by one. At this point, all transitions between the four binary variables across the next five time points are reflected in X_{i2} (see Table 1). The stability and spread equations, which are described further below, would then be applied to these two populated transition matrices.

In the present study, the dimension of each transition matrix was determined by the number of variables included in the data (e.g., when k = 9, the transition matrices were 9-by-9). Five observations were included in each matrix with a

Table 1. Visual Demonstration of Method

	Example Data			
	ER1	ER2	ER3	ER4
T1	1	0	0	0
T2	0	0	1	0
T3	0	0	1	0
T4	0	0	1	0
T5	0	1	0	0
T6	0	0	0	1
	X_{i1}			
	ER1	ER2	ER3	ER4
ER1	0	0	0	0
ER2	0	0	1	0
ER3	1	0	2	0
ER4	0	0	0	0
	Stability = 2/4 Spread = 3/16			
	X_{i2}			
	ER1	ER2	ER3	ER4
ER1	0	0	0	0
ER2	0	0	1	0
ER3	0	0	2	0
ER4	0	1	0	0
	Stability = 2/4 Spread = 3/16			

Note. Two transition matrices constructed from example data with 6 observations (T1 through T6), window size of 5 observations per transition matrix, four binary time series (ER1 through ER4), and a windowing lag of one. We chose not to reduce the stability and spread fractions, when appropriate, to avoid obscuring the relationship between the matrices and the resulting values.

windowing lag of one. Phrased differently, each transition matrix is composed of all pairwise transitions between "on" states that occurred between successive observations within a window of five observations at a time. We calculated stability and spread using the 'transStats' function in the *TransitionMetrics* package.

Stability, which measures the extent to which a system transitions from one binary variable to the same binary variable at the next time point relative to all observed transitions, makes use of one important characteristic of the transition matrix. Namely, if the same time series variable is "on" in two back-to-back surveys, then a cell along the on-diagonal of the matrix would be incremented to reflect this "repeat." Conversely, if a switch from one time series variable to another time series variable occurs in two-back-back surveys, then a cell along the off-diagonal of the matrix would be incremented to reflect this "switch." As such, stability is given by

$$St_{ij} = \frac{tr(\mathbf{X}_{ij})}{\sum\sum \mathbf{X}_{ij}}, \tag{1}$$

where $\text{tr}(\mathbf{X}_{ij})$ is the sum of the elements along the on-diagonal of a transition matrix X_{ij} for the jth transition matrix of the ith person and $\sum\sum \mathbf{X}_{ij}$ is the sum of all elements of X_{ij}. Stability is bounded between 0 and 1 where values closer to 1 indicate a tendency to repeat the same variable between successive surveys and values closer to 0 indicate a tendency to switch between at least two different variables between successive surveys. This process resulted in 4 stability vectors per dataset (12 vectors total), where each vector was a string of stability values calculated within each of the sub-datasets simulated 1000 times for each participant.

Spread, which measures how many unique pairwise transitions are observed relative to all possible transitions afforded by the multivariate system, makes use of a second important characteristic of the transition matrix. Namely, each cell in the transition matrix represents a unique pairwise transition between two successive observations and all possible pairwise transitions are represented in the full transition matrix. As such, spread is given by

$$Sp_{ij} = \frac{nz(X_{ij})}{k^2}, \tag{2}$$

where $\text{nz}(\cdot)$ is a count of the number of non-zero elements in $\cdot$ and k^2 is the number of elements in X_{ij}. Spread is bounded between 0 and 1 where values closer to 1 indicate that a higher proportion of all possible transition pairings were observed and values closer to 0 indicate that very few of the possible transition pairings were observed. This process resulted in 12 spread vectors.

Calculating Bias. To calculate the bias in average stability scores introduced by each time interval misspecification sampling schedule, we subtracted the stability vector calculated from the equal time spacing data set (i.e., the expected value) from each unequal time spacing stability vector (i.e., the observed values). We then repeated this process for the spread vectors. We calculated these difference vectors separately for each parent dataset, resulting in nine vectors of difference scores: three for stability and three for spread, for each parent dataset. Difference scores indicate how different the transition metric was when calculated on data with unequal time spacing compared to when it was calculated on data collected with equal time spacing. Difference scores closer to zero indicate less bias in the values obtained from equal and unequal time spacing sampling schedules.

We first plotted the distributions of these difference scores. We then calculated the mean (average bias) and standard deviation of the mean (standard error of bias) difference score observed in all 1000 simulations. We used the average bias and standard error of bias from each sampling schedule to calculate z-scores to determine if the observed bias was significantly larger than zero. We also calculated the mean and standard deviation of the standard deviation in difference scores observed in all 1000 simulation runs. Whereas average bias indicates how much systematic bias likely results from unequal sampling across many samples, average standard deviation indicates how much a single stability or spread value calculated in unequally spaced data may deviate from what would have been observed in an equal spacing design. Finally, we calculated the root mean square error (RMSE), which measures the degree to which the observed scores vary from the best fit line of the expected scores. Low variability of error and low bias both contribute to RMSE values that are closer to zero.

Calculating Coverage. To calculate 95% coverage in stability, we determined whether each equal time-spacing stability vector (i.e., the true values) fell within +/- 1.96 standard deviations of the average[7] stability value calculated from a given time interval misspecification stability vector. We then calculated the percentage of the time this was true. A measure with 95% coverage would expect this percentage to be 95%. We repeated this process for the spread vectors. We calculated these separately for each parent dataset, resulting in nine vectors of TRUE/FALSE decisions: three for stability and three for spread, for each parent dataset.

Open Data Statement. All simulations and code are available at https://osf.io/xf82h/.

3 Results

3.1 Bias

The average bias in both stability and spread was not significant for any time interval misspecification sampling schedule in either randomly generated parent

[7] Per Daniel et al. (2023), matrices that showed no transitions (i.e., where no time series were observed in the "on" state across five successive observations) received a stability score of "noUse." Since "noUse" cannot contribute to the mean stability value, these observations were dropped.

dataset (Tables 2 and 3). Whereas the average bias for all cases was near-zero, the average standard deviation between observed and expected stability was .058-.059 when k = 10, depending on the sampling schedule, and .023 for all sampling schedules when k = 20. The average standard deviation between observed and expected spread scores was .139-.140 when k = 10 and .098 for all sampling schedules when k = 20. RMSE scores for these comparisons were near zero.

The average bias in both stability and spread were also not significant for any time interval misspecification case when the data were simulated based on observed proportions from a prior EMA study (Table 4). The average bias for all cases from the data-informed simulations was near-zero and the average standard deviation between observed and expected spread scores (SD = .088) was within the range observed in the randomly generated data. However, the average standard deviation between observed and expected stability scores was higher than previously seen (i.e., approximately .263 on a measure ranging from 0 to 1). That said, RMSE scores for all stability and spread sampling schedules were low.

3.2 Coverage

Across all parent data sets, stability and spread demonstrated approximately 95% coverage of the random sampling schedules. However, coverage for stability and spread was poor in all off/on burst sampling schedules, ranging between 64.5% and 67.6% (Tables 2-4).

Table 2. Bias and Coverage Estimates for Stability and Spread Calculations with Time Interval Misspecification when k = 10

	Average Bias	SE of Bias	Z Score of Bias	Average SD (SE)	RMSE	Coverage
Stability						
Off/On Burst Sampling	-7.24 e–5	0.002	-0.016	.059 (.010)	0.002	67.77%
Between Random Sampling	3.91 e–5	0.002	0.009	.058 (.010)	0.002	93.48%
Within Random Sampling	-2.48 e–5	0.002	-0.006	.058 (.010)	0.002	93.35%
Spread						
Off/On Burst Sampling	6.24 e–5	0.006	0.005	.140 (.007)	0.006	67.90%
Between Random Sampling	-2.62 e–5	0.006	-0.002	.139 (.007)	0.006	95.67%
Within Random Sampling	7.25 e–5	0.006	0.006	.140 (.006)	0.006	96%

Note. k = the number of time series included in the simulation. SE = standard error; SD = standard deviation; RMSE = root mean standard error. All Z scores are all non-significant at $p > .05$, suggesting that the average amount of bias is non-significant for both stability and spread metrics across all three misspecification sampling schedules. Coverage reflects the percentage of time when the true value (given by the equal time spacing series) falls within +/-1.96 standard deviations of the estimate (given by the relevant time misspecification series).

Table 3. Bias and Coverage Estimates for Stability and Spread Calculations with Time Interval Misspecification when k = 20

	Average Bias	SE of Bias	Z Score of Bias	Average SD (SE)	RMSE	Coverage
Stability						
Off/On Burst Sampling	-2.40 e–5	0.001	-0.013	.023 (.006)	0.001	66.86%
Between Random Sampling	3.27 e–5	0.001	0.018	.023 (.006)	0.001	94.57%
Within Random Sampling	-3.47 e–5	0.001	-0.019	.023 (.006)	0.001	94.47%
Spread						
Off/On Burst Sampling	2.48 e–4	0.004	0.031	.098 (.005)	0.004	65.57%
Between Random Sampling	-5.84 e–5	0.004	-0.007	.098 (.005)	0.004	94.67%
Within Random Sampling	9.30 e–5	0.004	0.011	.098 (.004)	0.004	94.86%

Note. k = the number of time series included in the simulation. SE = standard error; SD = standard deviation; RMSE = root mean standard error. All Z scores are all non-significant at $p > .05$, suggesting that the average amount of bias is non-significant for both stability and spread metrics across all three misspecification sampling schedules. Coverage reflects the percentage of time when the true value (given by the equal time spacing series) falls within +/-1.96 standard deviations of the estimate (given by the relevant time misspecification series).

Table 4. Bias and Coverage Estimates for Stability and Spread Calculations with Time Interval Misspecification for k = 9 when the Probability of Each Time Series' Endorsement is Given by Previously Collected Data

	Average Bias	SE of Bias	Z Score of Bias	Average SD (SE)	RMSE	Coverage
Stability						
Off/On Burst Sampling	-4.68 e–4	0.012	-0.021	.265 (.011)	0.012	64.62%
Between Random Sampling	6.27 e–4	0.011	0.028	.263 (.010)	0.012	96.29%
Within Random Sampling	-1.14 e–4	0.011	-0.005	.263 (.008)	0.011	96.52%
Spread						
Off/On Burst Sampling	-6.82 e–5	0.004	-0.009	.088 (.006)	0.004	64.48%
Between Random Sampling	-2.24 e–4	0.004	-0.028	.087 (.006)	0.004	94.33%
Within Random Sampling	-8.22 e–5	0.004	-0.011	.087 (.005)	0.004	94.33%

Note. k = the number of time series included in the simulation. SE = standard error; SD = standard deviation; RMSE = root mean standard error. All Z scores are all non-significant at $p > .05$, suggesting that the average amount of bias is non-significant for both stability and spread metrics across all three misspecification sampling schedules. Coverage reflects the percentage of time when the true value (given by the equal time spacing series) falls within +/-1.96 standard deviations of the estimate (given by the relevant time misspecification series

4 Discussion

Through simulation, we explored the robustness of stability and spread indices when applied to data with common time interval misspecifications. In the aggregate, stability and spread values were unbiased when applied to data collected using a between-person random sampling schedule, a within-person random sampling schedule, and an off/on burst sampling schedule. These results held in randomly generated data with differing numbers of binary variables and in data simulated based on the proportions of observed data from a prior EMA study. Further, stability and spread demonstrated approximately 95% coverage for all between- and within-person random sampling schedules. However, coverage for stability and spread was poor in the off/on burst sampling schedules. Taken together, these findings suggest that stability and spread are appropriate metrics to use with ILD that are collected using unequal time spacing conditions, although researchers should be cautious when interpreting any given stability or spread estimate if sampling a continuously transitioning process with a burst design. This pattern of results supports the appropriateness of applying stability and spread indices across a range of common EMA sampling schedules (Myin-Germeys et al., 2018). With the proliferation of EMA across research disciplines in psychology in the past two decades (Wrzus & Neubauer, 2023), our findings provide more confidence in the general applicability of these transition metrics.

4.1 Stability and Spread are Unbiased on Aggregate, but Degree of Bias Varies

It is possible that stability and spread appear unbiased because the method uses a sliding window with a lag of one. This approach is consistent with time delay embedding procedures, which have shown robustness to time interval misspecification (Boker et al., 2018). Given this shared procedure between time delay embedding and our windowing approach, researchers who choose to only calculate stability and spread once per person, or repeatedly but with non-overlapping windows, may be more vulnerable to bias. Future simulations that test boundary conditions of the apparent robustness of stability and spread will be useful for researchers who wish to set a windowing lag that is not one. Further, our results support the idea that time delay embedding procedures might help address time interval misspecification across a broader range of time series methods than were originally tested by Boker et al. (2018).

Although unbiased in the aggregate, there was variation in how close specific stability or spread values were when calculated from data with and without time interval misspecification. Bias measures how well the stability and spread values that are calculated from time interval misspecification data (i.e., observed scores) approximate the stability and spread values that are calculated from data with equal time spacing (i.e., expected scores). Variance, on the other hand, measures how tightly clustered the observed scores are relative to the expected scores. Thus, the true amount of error in stability and spread in a system can be decomposed as the sum of the bias term and the variance term. Given that the total error is the sum of these two terms, there is a trade-off between bias and variance (i.e., if total error stays constant, less bias implies more variance and vice versa; Mehta et al., 2019). As such, it is not surprising that with no significant bias, some degree of variance remained, although this observed variance still tended to be low.

Across both k = 10 and k = 20 datasets, variance in spread was greater than variance in stability. However, a visual inspection of scatter plots revealed that stability was prone to more frequent edge cases, possibly because spread values are determined at the unique cell level (the denominator is the number of cells in the transition matrix) whereas stability values are determined at the transition level (the denominator is the number of observed transitions). We observed the opposite pattern in the data-informed datasets: variance was higher in stability relative to spread. Further, the data-informed variances were notably higher than those observed across all other cases. It could be that when every time series is equally probable to be observed at a given measurement occasion, the difference in stability calculated using equal time spacing relative to any of the unequal time spacing measures is relatively small (given that random endorsement patterns make high stability values less likely), but when a particular time series is significantly more likely to be endorsed, there is a wider range of possible stability values and thus greater variability between sampling schedules. Researchers interested in using the stability value of any one transition matrix might do so cautiously when using a sampling process expected to have pockets of high underlying stability; however, coverage for between- and within-person random sampling remained good.

4.2 Coverage Depends on Sampling Schedule

Whereas coverage was consistently around 95% for stability and spread when random sampling was imposed between- and within-people, coverage was poor for these metrics when an off/on burst sampling method was used. Specifically, our burst design alternated between sampling five observations in a row and not sampling for 35 observations in a row. It is likely that this design imposed too long of a gap between measurement bursts such that the likelihood for the true value to be contained within the confidence intervals from these samples was substantially reduced. Researchers using burst designs should be cautious when applying stability and spread to their data, though the amount of time between measurement bursts that may be detrimental likely depends on the frequency at which the system being studied changes. Moreover, this caution may only be warranted if transition matrices are allowed to span samples collected across two different bursts. In the current simulations, we allowed observations 38, 39, 40, 76, and 77, for example, to contribute to a given stability and spread value. Researchers using burst designs when the process is expected to continue transitioning between bursts might instead prevent transition matrices from spanning long gaps in samples. This constraint would likely mitigate these issues.

5 Applied Example: Emotion Regulation

How diverse a set of ER strategies that a person switches between (Wen, Quigley, Yoon, & Dobson, 2021) and how often a person transitions between different ER strategies—whether too often (Southward, Altenburger, Moss, Cregg, & Cheavens, 2018) or not often enough (Bonanno & Burton, 2013) —is expected to be associated with mental health (Gross, 2015). Being appropriately responsive to changing emotional intensity likely promotes effective ER switching decisions (Bonanno & Burton, 2013). However, people with more depressive symptoms may exhibit particular deficits in effectively switching ER strategies in response to variations in their mood (Cheng, 2001; Coifman & Bonanno, 2010; Southward & Cheavens, 2017; Southward, Eberle, & Neacsiu, 2022), given that depression is associated with "stereotyped and inflexible responses to a variety of emotional stimuli" (Rottenberg, Kasch, Gross, & Gotlib, 2002). To demonstrate the utility of spread and stability in real data, we conducted an applied example to test how depressive symptoms moderate the relation between momentary ER strategy switching patterns and mood throughout the daily lives of adults oversampled to be elevated in neuroticism.

We expected that participants with higher depressive symptoms would report more unique types of ER strategy switches (higher spread) and more frequent strategy switches (lower stability; e.g., rumination → expressive suppression → substance use → reappraisal → denial) because depression has been associated with greater diversity in ER strategy use (Wen et al., 2021). We also expected that periods characterized by a more positive average mood state would be associated with fewer types of unique ER strategy switches (lower spread) and less frequent switching (higher stability; e.g., positive refocusing → positive refocusing → positive refocusing → positive refocusing → positive refocusing) because experiencing positive mood reduces the need to search for strategies designed to reduce negative emotions (Gross, 2015). Further, relatively improved mood might indicate that an effective strategy has been found and switching is not warranted (Daniel et al., 2024). Finally, we expected that depressive symptoms would moderate the relation between average state mood and spread/stability in ER strategies, such that individuals with less depression would show a stronger relation between more positive mood and lower spread/higher stability in ER strategies.

5.1 Method

Ninety-two undergraduate students at a large Midwestern university participated for course credit in this ethics board-approved study (protocol 2008B0320). Participants ranged from 18-31 years old (M = 19.73, SD = 2.25) and were predominantly female (54%) and White (81%). After providing informed consent, participants completed a range of ER and psychopathology questionnaires at baseline and were then shown how to answer an EMA survey using a study-provided palm pilot. The EMA surveys were sent three times per day for 10 days. Surveys were sent randomly within four-hour intervals, typically occurring around 1pm, 5pm, and 9pm. We report how we determined our sample size, all data exclusions, and all relevant measures. See Heiy and Cheavens (2014) for more details.

Measures To assess the presence and severity of past-week depressive symptoms, participants completed the Center for Epidemiologic Studies Depression scale (CES-D; Radloff, 1977) at baseline. The CES-D is a 20-item self-report scale where higher scores indicate greater depressive symptom severity, and scores $\geq$ 16 indicate the likely presence of major depressive disorder (MDD). Cronbach's alpha for this sample was .92, and participants reported somewhat elevated depression scores, M = 18.30, SD = 11.61, with 49 participants (55.1%) likely meeting criteria for MDD.

At each EMA survey, participants rated their current mood from 0 to 100, with 100 reflecting the best possible mood and 0 reflecting the worst possible mood. Participants were then asked to identify the strongest negative emotion (if any) they had experienced in the prior four hours. If they reported a negative emotion, participants received a list of 20 different ER strategies presented in a random order and were asked if they had used any of the strategies to lessen or decrease the intensity of the negative emotion.[8] Response options for each strategy were: no; yes, but it did not change the intensity of the emotion; and yes, and it did change the intensity of the emotion. In the present study, we collapsed the two "yes" responses into one such that each strategy was either endorsed (yes = 1) or not endorsed (no = 0). If participants did not report experiencing a negative emotion, ER strategy use was not assessed.

Analytic Approach All analyses were conducted in R version 4.1.3 (R Core Team, 2022). Analysis scripts are openly provided (https://osf.io/xf82h). Data are available upon reasonable request. The analysis scripts detail all data preprocessing, stability and spread calculation, and model fitting steps in both structural equation and Bayesian modeling frameworks.

We calculated spread and stability in ER strategies according to the steps outlined in Daniel et al. (2023) using the *TransitionMetrics* package (Daniel & Moulder, 2020). We set window size (W) to nine and used a windowing lag of one. We set $W = 9$ so that each 20-by-20 transition matrix would be built with surveys from at least three different days, thereby increasing the likelihood that participants would be reporting ER attempts in response to multiple, distinct events. The number of transition matrices built per participant varied by how many total observations they submitted throughout the study. Participants who submitted fewer than nine surveys (n = 3) were excluded from analyses given insufficient observations upon which to build a minimum of one complete transition matrix. These analytic choices are consistent with previous pre-registered

[8] The strategies assessed were acceptance, behavioral activation, rumination, problem-solving, perspective, social support, benefit finding, consequences, self-blame, generalizing, other-blame, expressive suppression, positive refocusing, reappraisal, sleep, emotional suppression, substance use, denial, exercise, non-suicidal self-injury.

decisions in the same data (https://osf.io/d3tyn). We built two models, one predicting spread in ER strategies and another predicting stability in ER strategies.

To test whether depression moderated the relation between average state mood and spread in ER strategy choices, we fit the structural equation model presented in Figure 1. We used a time delay embedding matrix with nine dimensions and a windowing lag of one. Spread in ER strategies was entered as a time-varying manifest variable. We divided grand-mean-centered depression severity scores by 100 and entered them as a time-invariant manifest variable.[9] State mood was person-mean-centered and then entered into a nine-dimension time delay embedding matrix to conform to the number of surveys contributing to each spread calculation. Average state mood was specified along the nine dimensions of that time delay embedding matrix. Specifically, average state mood was entered as a latent variable with loadings to all nine state mood indicators fixed to one. We used an identity variable to include the moderation effect of depression severity on the relation between average state mood and spread in ER strategy choices. We conducted this analysis in *OpenMx* version 2.20.6 (Neale et al., 2016) and used likelihood-based confidence intervals (the "mxCI" argument in the "mxModel" statement) to test the significance of all hypothesized paths. Paths with 95% CIs that did not include zero were considered significant.

We used similar procedures to test whether depression moderated the relation between average state mood and stability in ER strategy choices. However, because stability returns a "noUse" result if no transition was observed—which occurs if ER strategies were not endorsed between two back-to-back surveys within a given transition matrix—we opted to use multilevel mixture modeling (e.g., Muthén & Asparouhov, 2009). Specifically, we predicted whether some numerical stability value was returned (vs. "noUse"; a binary outcome) and, if some stability value was returned, the level of that value (a continuous outcome). We therefore fit this mixture model with logistic regression and linear regression components, respectively. Predictors for both model components included grand-mean-centered depression symptoms, person-mean-centered average state mood (which was calculated outside of the model on nine EMA surveys at a time), and their interaction. We also included person-level random intercepts for both model components. We fit this model in the Bayesian framework using *rjags* version 4-13 (Plummer, Stukalov, & Denwood, 2023) with uninformative priors. All parameters converged[10] using two chains on 5,000 iterations after a burn in period of 15,000 iterations. Bayesian credible intervals that do not include 0 are considered significant.

[9] We elected to estimate depression outside of the model, rather than to estimate it in the model as a latent variable, because estimating latent-by-latent interactions within structural equation modeling is relatively understudied and often considered infeasible. We divided grand-mean-centered depression severity by 100 to aid model convergence.

[10] The variance parameter for the random intercept of continuous stability did not converge due to minimal average between-person variance on this outcome. However, removing this random intercept did not change other model outcomes so we decided to retain this parameter to maintain the multi-level structure present by design.

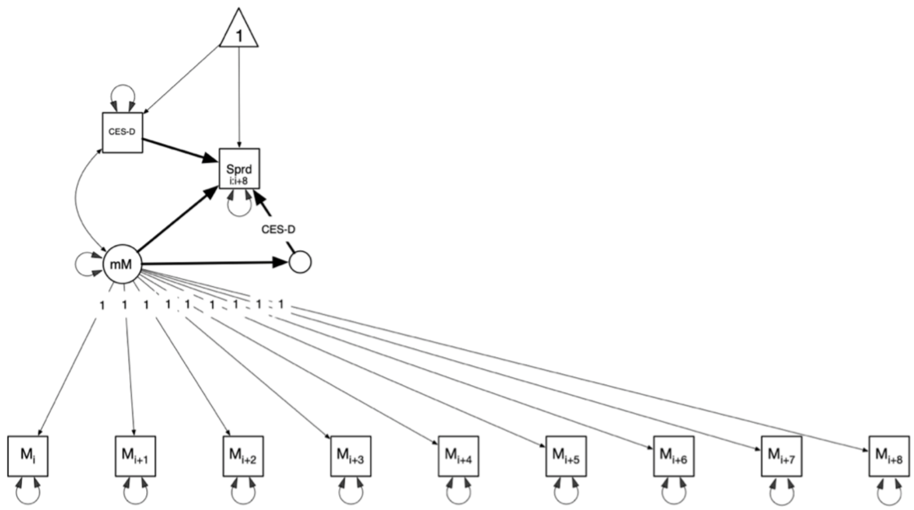

Figure 1. Spread in Emotion Regulation Strategy Choices Path Model

Note. Paths tested for path significance are bolded. CES-D = Center for Epidemiologic Studies Depression, measure of trait depression symptom severity; Sprd = spread in emotion regulation strategy choices solved for along a sliding series of nine successive surveys inclusive; mM = average state mood score given by the nine state mood scores which are labeled M_i through M_{i+8}.

5.2 Results

Predicting Spread in ER Strategy Transitions. We found a significant interaction between depressive symptom severity and mean state mood in predicting spread in ER strategies (Table 5). Specifically, participants made fewer unique switches between ER strategies (lower spread) when they were experiencing more positive mean state mood than usual, but this effect was especially pronounced in those who endorsed fewer depressive symptoms (Figure 2). This is in line with our hypothesis that people with fewer depression symptoms would be more likely to demonstrate an association between better state mood and less of a need to switch between many different ER strategies.

Predicting Stability in ER Strategy Transitions. We found a significant main effect of mean state mood in predicting whether some stability value was returned (Table 6). Specifically, when participants reported better mean state mood than usual, they were less likely to report back-to-back ER strategy use ("noUse" was returned more often). However, depression was neither a significant main effect nor a significant moderator in the logistic regression component of the model.

Table 5. Summary of Free Parameters in Spread in Emotion Regulation Strategies Model

	Estimate	**Standard Error**	**95% CI LB**	**95% CI UB**
Paths Assessed for Significance				
mMood → Spread	**-0.02**	**0.005**	**-0.04**	**-0.02**
CES-D → Spread	-0.05	0.05	-0.15	0.04
mMood × CES-D → Spread	**0.06**	**0.03**	**0.01**	**0.12**
Additional Paths Included in the Model				
covariance (mMood, CES-D)	-0.04	0.02		
(Error) Variances ρ				
mMood_v	7.97	1.53		
Spread_e	0.02	0.001		
CES-D_e	0.01	0.0004		
Means				
CES-D	6.3 e–6	0.002		
Spread	0.13	0.005		

Notes. LB = Lower bound and UB = Upper bound. Mood = mean state mood; CES-D = Center for Epidemiologic Studies Depression, measure of trait depression symptom severity; Spread = spread in emotion regulation strategy choices (higher spread denotes a higher proportion of unique types of switches between ER strategies and lower spread denotes fewer unique types of switches); appending a variable name with _v denotes a variance; appending a variable name with _e denotes an error variance. Paths are significant if its upper and lower 95% Confidence Interval bounds do not cross 0. Significant paths are bolded.

× is used to reflect an interaction effect between two variables on Spread.

ρ Estimates and standard errors for the nine state mood indicators along the time delay embedded matrix are not included to improve readability of model output.

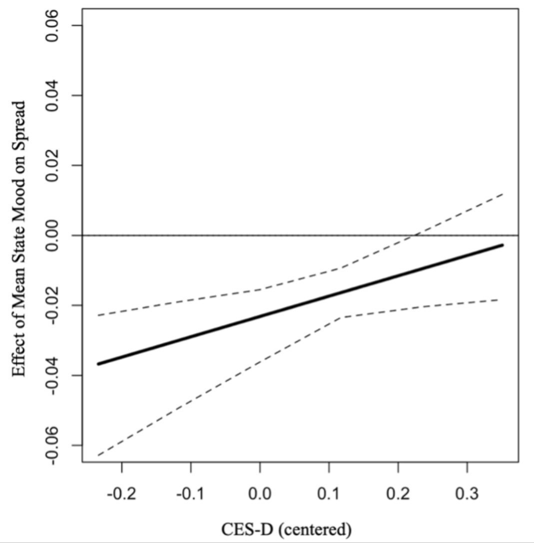

Figure 2. Johnson-Neyman Plot Depicting the Significant Interaction between Trait Depression Severity and Mean State Mood Intensity on Spread in Emotion Regulation Strategy Choices

Note. Increasingly negative values on the y-axis indicate that the association between more positive mood and less unique transitions between emotion regulation strategies is becoming stronger. CES-D = Center for Epidemiologic Studies Depression, measure of trait depression symptom severity.

We found a significant interaction between depression severity and mean state mood in predicting continuous stability in ER strategy choices (Table 6). Specifically, participants with fewer depression symptoms tended to switch between ER strategies less frequently (higher stability) when experiencing periods of more positive state mood than usual, whereas participants with more depression symptoms tended to switch between ER strategies more frequently (lower stability) when experiencing periods of more positive state mood than their usual (Figure 3). This is in line with our hypothesis that people with fewer depression symptoms would be more likely to demonstrate an association between better state mood and less frequent switching between ER strategies.

5.3 Discussion

Using empirical data that followed a within-person random sampling schedule, stability and spread metrics led to both sensible and novel insights into ER strategy use. In line with expectations, all participants evidenced more unique strategy switches (higher spread) and used any strategies more consecutively (fewer 'noUse' results for stability) when feeling more intense negative moods,

Table 6. Stability in Emotion Regulation Strategies Model

	2.5% Credible Interval	**50% Credible Interval**	**97.5% Credible Interval**
Logistic Regression Component			
mMood	**-0.2**	**-0.15**	**-0.12**
CES-D	-2.6 e–2	3.7 e–2	0.11
mMood × CES-D	-1.0 e–3	1.4 e–3	3.9 e–3
Linear Regression Component			
mMood	-3.3 e–4	8.7 e–6	3.4 e–4
CES-D	**-2.2 e–3**	**-1.6 e–3**	**-1.1 e–3**
mMood × CES-D	**-2.0 e–4**	**-1.7 e–4**	**-1.4 e–4**

Notes. The outcome for the logistic regression component of the model is whether a numerical stability value was returned (1) or "noUse" was returned (0). The outcome for the linear regression model is the numerical stability value that was returned (where a higher stability value denotes less switching, and a lower stability value denotes more switching). A random intercept per participant was included for both components of the model. mMood = mean state mood; CES-D = Center for Epidemiologic Studies Depression, measure of trait depression symptom severity. Effects are significant if its 95% Credible Interval does not cross 0. Significant effects are bolded.
× is used to reflect an interaction effect between two variables on the outcome.

potentially indicating a type of searching for the "right" strategy in response to greater distress (Aldao & Nolen-Hoeksema, 2013; Southward et al., 2018). Interestingly, depression moderated these effects in novel ways. People with greater depression were more likely to exhibit more unique strategy switches (higher spread) and more frequent switching (lower stability) in response to more positive moods than those with less depression. These results suggest that ER in depression may be characterized as consistently inconsistent – that is, more frequent and more diverse regardless of the emotional context. Of course, replication in larger and more diverse samples, including those that were not selected to be relatively high in trait neuroticism, is needed to strengthen these claims. Future researchers may use stability and spread indices to test if unique patterns of ER distinguish different disorders and identify specific clinical targets (i.e., switching strategies too quickly or too often).

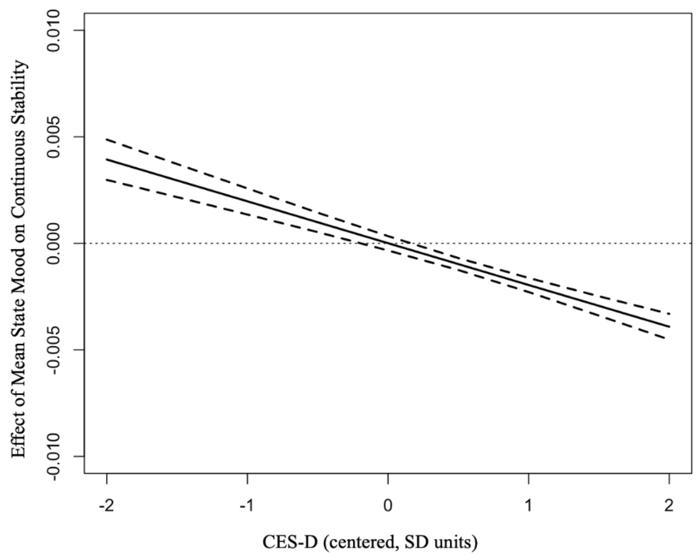

Figure 3. Johnson-Neyman Plot Depicting the Significant Interaction between Trait Depression Severity and Mean State Mood Intensity on Continuous Stability in Emotion Regulation Strategy Choices

Note. Increasingly positive values on the y-axis indicate that the association between more positive mood and less frequent switches in emotion regulation strategies is becoming stronger whereas increasingly negative values indicate that the association between more positive mood and more frequent switches in emotion regulation strategies is becoming stronger. CES-D = Center for Epidemiologic Studies Depression, measure of trait depression symptom severity.

6 Conclusion

Stability and spread appear unbiased, in the aggregate, when applied to data with common types of time interval misspecification. Stability and spread also demonstrate approximately 95% coverage when using between- and within-person random sampling schedules. However, coverage is poor when using an off/on burst sampling schedule. When applied to a previously collected EMA dataset measuring adults' naturalistic use of 20 ER strategies, depressive symptoms were differentially associated with the ways in which a person's state mood and ER strategy switching decisions were related in daily life. Taken together, stability and spread appear to be appropriate metrics to use with data that are collected along common unequal time spacing conditions. Further, researchers interested in transitions within high-dimensional binary ILD should prioritize sampling at a rate that is conceptually or empirically matched to the transition process of interest, rather than prioritize equal sampling at the expense of missing the suspected transition process. However, taking long breaks between sampling bursts, assuming transitions are expected to continue as usual, may increase the likelihood that specific stability and spread values are inaccurate.

7 Acknowledgments

Katharine E. Daniel (KED's efforts on this project were supported by a University of Virginia Jefferson Fellowship Grant, a P.E.O. Scholars Award, and University of Virginia Dissertation Completion Fellowship awarded to KED. Steven M. Boker (SMB)'s efforts on this project were supported by a fellowship from the Max Planck Institute for Human Development awarded to SMB. Matthew W. Southward's efforts on this project were partially supported by the National Institute of Mental Health under award number K23MH126211. The funding sources had no involvement in the conduct or preparation of the research, and the content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Aldao, A., & Nolen-Hoeksema, S. (2013). One versus many: Capturing the use of multiple emotion regulation strategies in response to an emotion-eliciting stimulus. *Cognition & emotion*, *27*(4), 753–760. doi: https://doi.org/10.1080/02699931.2012.739998

Boker, S. M., Tiberio, S. S., & Moulder, R. G. (2018). Robustness of time delay embedding to sampling interval misspecification. *Continuous time modeling in the behavioral and related sciences*, 239–258. doi: https://doi.org/10.1007/978-3-319-77219-6_10

Bonanno, G. A., & Burton, C. L. (2013). Regulatory flexibility: An individual differences perspective on coping and emotion regulation. *Perspectives on psychological science*, *8*(6), 591–612. doi: https://doi.org/10.1177/1745691613504116

Burke, L. E., Shiffman, S., Music, E., Styn, M. A., Kriska, A., Smailagic, A., . . . others (2017). Ecological momentary assessment in behavioral research: addressing technological and human participant challenges. *Journal of medical Internet research*, *19*(3), e77. doi: https://doi.org/10.2196/jmir.7138

Cheng, C. (2001). Assessing coping flexibility in real-life and laboratory settings: a multimethod approach. *Journal of personality and social psychology*, *80*(5), 814. doi: https://doi.org/10.1037//0022-3514.80.5.814

Coifman, K. G., & Bonanno, G. A. (2010). When distress does not become depression: emotion context sensitivity and adjustment to bereavement. *Journal of Abnormal Psychology*, *119*(3), 479. doi: https://doi.org/10.1037/a0020113

Daniel, K. E., & Moulder, R. G. (2020). transitionmetrics: R package to calculate characteristics of transitions in multivariate binary time series data [Computer software manual]. Retrieved from https://github.com/KatharineDaniel/transitionMetrics (Version 0.1.0)

Daniel, K. E., Moulder, R. G., Boker, S. M., & Teachman, B. A. (2024). Investigating switches in emotion regulation strategies in the daily lives of socially anxious people. *Clinical Psychological Science*. doi: https://doi.org/10.1177/21677026241249192

Daniel, K. E., Moulder, R. G., Teachman, B. A., & Boker, S. M. (2023). Stability and spread: A novel method for quantifying transitions within multivariate binary time series data. *Behavior Research Methods*, *55*(6), 2960–2978. doi: https://doi.org/10.3758/s13428-022-01942-0

Daros, A. R., Daniel, K. E., Meyer, M. J., Chow, P. I., Barnes, L. E., & Teachman, B. A. (2019). Impact of social anxiety and social context on college students' emotion regulation strategy use: An experience sampling study. *Motivation and emotion*, *43*, 844–855. doi: https://doi.org/10.1007/s11031-019-09773-x

Ebner-Priemer, U. W., & Sawitzki, G. (2007). Ambulatory assessment of affective instability in borderline personality disorder. *European Journal of Psychological Assessment*, *23*(4), 238–247. doi: https://doi.org/10.1027/1015-5759.23.4.238

Gross, J. J. (2015). The extended process model of emotion regulation: Elaborations, applications, and future directions. *Psychological Inquiry*, *26*(1), 130–137. doi: https://doi.org/10.1080/1047840X.2015.989751

Heiy, J. E., & Cheavens, J. S. (2014). Back to basics: a naturalistic assessment of the experience and regulation of emotion. *Emotion*, *14*(5), 878. doi: https://doi.org/10.1037/a0037231

Johnson, K. J., Zaback, M., Tokuno, C. D., Carpenter, M. G., & Adkin, A. L. (2019). Exploring the relationship between threat-related changes in anxiety, attention focus, and postural control. *Psychological research*, *83*, 445–458. doi: https://doi.org/10.1007/s00426-017-0940-0

Mehta, P., Bukov, M., Wang, C.-H., Day, A. G., Richardson, C., Fisher, C. K., & Schwab, D. J. (2019). A high-bias, low-variance introduction to machine learning for physicists. *Physics reports*, *810*, 1–124. doi: https://doi.org/10.1016/j.physrep.2019.03.001

Muthén, B., & Asparouhov, T. (2009). Multilevel regression mixture analysis. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *172*(3), 639–657. doi: https://doi.org/10.1111/j.1467-985x.2009.00589.x

Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: new insights and technical developments. *World Psychiatry*, *17*(2), 123–132. doi: https://doi.org/10.1002/wps.20513

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., ... Boker, S. M. (2016). Openmx 2.0: Extended structural equation and statistical modeling. *Psychometrika*, *81*, 535–549. doi: https://doi.org/10.1007/s11336-014-9435-8

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fmri: a tutorial overview. *Neuroimage*, *45*(1), S199–S209. doi: https://doi.org/10.1016/j.neuroimage.2008.11.007

Plummer, M., Stukalov, A., & Denwood, M. (2023). Bayesian graphical models using mcmc [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=rjags` (Version 4-14)

R Core Team. (2022). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Radloff, L. S. (1977). A self-report depression scale for research in the general population. *Applied psychol Measurements*, *1*, 385–401. doi: https://doi.org/10.1177/014662167700100306

Rottenberg, J., Kasch, K. L., Gross, J. J., & Gotlib, I. H. (2002). Sadness and amusement reactivity differentially predict concurrent and prospective functioning in major depressive disorder. *Emotion*, *2*(2), 135. doi: https://doi.org/10.1037/1528-3542.2.2.135

Southward, M. W., Altenburger, E. M., Moss, S. A., Cregg, D. R., & Cheavens, J. S. (2018). Flexible, yet firm: A model of healthy emotion regulation. *Journal of Social and Clinical Psychology*, *37*(4), 231–251. doi: https://doi.org/10.1521/jscp.2018.37.4.231

Southward, M. W., & Cheavens, J. S. (2017). Assessing the relation between flexibility in emotional expression and symptoms of anxiety and depression: The roles of context sensitivity and feedback sensitivity. *Journal of social and clinical psychology*, *36*(2), 142–157. doi: https://doi.org/10.1521/jscp.2017.36.2.142

Southward, M. W., & Cheavens, J. S. (2020). More (of the right strategies) is better: Disaggregating the naturalistic between-and within-person structure and effects of emotion regulation strategies. *Cognition and Emotion*, *34*(8), 1729–1736. doi: https://doi.org/10.1080/02699931.2020.1797637

Southward, M. W., Eberle, J. W., & Neacsiu, A. D. (2022). Multilevel associations of daily skill use and effectiveness with anxiety, depression, and stress in a transdiagnostic sample undergoing dialectical behavior therapy skills training. *Cognitive behaviour therapy*, *51*(2), 114–129. doi: https://doi.org/10.1080/16506073.2021.1907614

Stawski, R. S., MacDonald, S. W., & Sliwinski, M. J. (2015). Measurement burst design. *The encyclopedia of adulthood and aging*, 1–5. doi: https://doi.org/10.1002/9781118521373.wbeaa313

Steinwart, I., & Christmann, A. (2008). *Support vector machines*. Springer Science & Business Media. doi: https://doi.org/10.1007/978-0-387-77242-4

Stone, A. A., & Shiffman, S. (1994). Ecological momentary assessment (ema) in behavorial medicine. *Annals of behavioral medicine*. doi: https://doi.org/doi.org/10.1093/abm/16.3.199

Wen, A., Quigley, L., Yoon, K. L., & Dobson, K. S. (2021). Emotion regulation diversity in current and remitted depression. *Clinical Psychological Science*, *9*(4), 563–578. doi: https://doi.org/10.1177/2167702620978616

Wrzus, C., & Neubauer, A. B. (2023). Ecological momentary assessment: A meta-analysis on designs, samples, and compliance across research fields. *Assessment*, *30*(3), 825–846. doi: https://doi.org/10.1177/10731911211067538