Word Count: 5,130
Figures: 7

Mapping the Timescale of Suicidal Thinking

Daniel D.L. Coppersmith,[1] Oisín Ryan,[2] Rebecca G. Fortgang,[1] Alexander J. Millner,[1, 3] Evan M. Kleiman,[4] and Matthew K. Nock[1,3,5]

[1] Harvard University, Department of Psychology

[2] Utrecht University, Methods and Statistics Department

[3] Franciscan Children's, Mental Health Research

[4] Rutgers University, Department of Psychology

[5] Massachusetts General Hospital, Department of Psychiatry

Correspondence to: Daniel Coppersmith, Harvard University, 33 Kirkland Street, Cambridge, MA 02138; dcoppersmith@g.harvard.edu.

Abstract

Suicide is one of the most devastating aspects of human nature and has puzzled scholars for thousands of years. Most suicide research to date has focused on establishing the prevalence and predictors of the presence or severity of suicidal thoughts/behaviors. Surprisingly little research has documented the fundamental properties of suicidal thoughts/behaviors, such as: when someone has a suicidal thought, how long do such thoughts last? Documenting the basic properties of a phenomenon is necessary to understand, study, and treat it. This study aims to identify the timescale of suicidal thinking, leveraging novel real-time monitoring data and a number of different novel analytic approaches. Participants were 105 adults with past week suicidal thoughts who completed a 42-day real-time monitoring study (total number of observations=20,255). Participants completed two forms of real time assessments: traditional real-time assessments (spaced hours apart each day) and high-frequency assessments (spaced 10 minutes apart over one hour). We found that suicidal thinking changes rapidly. Both descriptive statistics and Markov-Switching models indicated that that elevated states of suicidal thinking lasted on average 1 to 3 hours. Individuals exhibited considerable heterogeneity in how often and for how long they reported elevated suicidal thinking, and our analyses suggest that different aspects of suicidal thinking operated on different timescales. Continuous-time autoregressive models suggest that current suicidal intent is predictive of future intent levels for 2 to 3 hours, while current suicidal desire predictive of future suicidal desire levels for 20 hours. Multiple models found that elevated suicidal intent has on average shorter duration than elevated suicidal desire. Finally, our ability to capture within-person dynamics of suicidal thinking was improved using high-frequency sampling. For example, traditional real-time assessments alone estimated the duration of severe suicidal states of suicidal desire as 9.5 hours, whereas, the high-frequency

assessments shifted the estimated duration to 1.4 hours. The high-frequency assessments identified 19% more participants with a high-risk response than the traditional real-time assessment, and high frequency measurements were shown to capture considerable levels of variation across consecutive measurement occasions. These results provide the most detailed characterization to date of the temporal dynamics of suicidal thinking. Furthermore, these findings highlight the importance of sampling frequency in capturing the dynamics of a phenomenon.

Suicide is a leading cause of death (1). Despite an increase in suicide research over the past few decades (2), there have not been improvements in predicting (2) and preventing (3) suicide. Thus, there is an urgent need for new approaches to understanding suicide (4).

Much of the empirical research on suicide to date has focused on cross-sectional or retrospective studies of the presence or severity of suicidal thoughts/behavior, focusing on the between person characteristics that distinguish those who do vs. do not experience these outcomes (1, 4, 5). Increasingly, however, researchers have argued that suicidal thoughts/behaviors can best be understood as a *process* that evolves over time within an individual person (4, 6, 7). By understanding the *dynamics* of suicidal thinking, that is, how thoughts change over time, we might gain new insights into the mechanisms through which suicidal behavior develops. The recent widespread availability of smartphones has created an opportunity for researchers to begin collecting real-time data on suicidal thoughts (8, 9), providing new insights into the dynamics of suicidal thinking.

One foundational question for understanding the dynamics of suicidal thinking is over what timescale suicidal thoughts evolve (i.e., duration from onset to offset of a given episode of suicidal thinking). The timescale of a process is the cornerstone of any empirical or theoretical approach to understanding it: it determines how researchers should measure the constructs or system of interest and it is the starting point for any mathematical or computational model of underlying mechanisms of action. If the timescale of a process is slow, it is not necessary to measure it as frequently as a faster process (11): it would be sufficient to measure the physical growth of a child once a day, but to capture the reaction of heart rate to caffeine intake, a one-day gap in measurements would be insufficient. Unfortunately, very little is known about the timescale of suicidal processes because we have not studied how suicidal thoughts or behaviors change over minutes, hours, or days (10). As a result, researchers and clinicians have little guidance on how often to measure suicidal thinking or how to appropriately formalize theories on suicidal behavior.

In this paper, we aim to gain insights into the timescale of suicidal thoughts using a novel ecological momentary assessment dataset consisting of a mix of (a) observations taken at intervals of several hours apart and (b) burst observations spaced at a higher frequency of ten minutes apart. We quantify the timescale of suicidal thoughts using three approaches. First, we use descriptive statistics to quantify the rate at which we observe changes in suicidal desire and intent in our dataset. These descriptive statistics allow us to gain direct insight into how long states of elevated desire and intent last within and between individuals. Second, we examine if higher-frequency measurements captured additional occurrences of rare events (i.e., high suicidal intent). Third, we attempt to model the dynamics of suicidal thinking using continuous time models, and explore if high frequency observations contain unique information about the dynamics of suicidal thinking.

Methods

Participants

Participants were 105 adults recruited through online advertisements (see supplemental materials for more information on recruitment and screening). The average age of participants in this sample was 29.22 years (*SD* = 9.10, range = 18-60). Most participants (68.57%) reported being assigned female sex at birth (*n* = 72), 30.48% reported being assigned male sex at birth (*n* = 32), and 0.95% reported not knowing their assigned sex at birth (*n* =1). For gender, 56.19% identified as cisgender female (*n*=59), 33.33% identified as cisgender male (*n* = 35), 1.90% identified as transgender (*n* = 2), 7.62% identified as non-binary (*n* = 8) and 0.95% did not report (*n* = 1). Regarding racial identity, 21.90% of participants identified as multi-racial (*n* = 23), 8.57% as Black (*n* = 9), 4.76% as Asian (*n* = 5), 57.14% as White (*n* = 60), and 1% as Middle Eastern (*n* = 1). Race was unknown for 6.66% of participants (*n* = 7). For ethnicity, 15.24% of participants identified Latinx (*n* = 16). The median number of lifetime days with suicidal thoughts was 1,825 (range = 30 - 8000 days). More than two-thirds of participants (65.74%) reported a prior suicide attempt (*n* = 71).

Procedure

The study procedure consisted of a baseline survey, a 42-day real-time monitoring period, and 4-week follow-up survey. Participants were compensated with Amazon gift cards. Participants were paid $5 for the baseline survey and $5 for the follow-up survey. Participants were paid $0.25 for each completed survey and a $1 bonus per for any day where they completed at least 5 surveys. The maximum amount participants could earn in the study was $190. All study procedures were approved by the Harvard University-Area Institutional Review Board (IRB# 19-1819; "High-Resolution Real-Time Capture of Suicidal Thoughts and Urges"). Informed consent was obtained from all participants.

Real-Time Monitoring.

We used Metricwire to collect real-timed monitoring data. For the real-time monitoring period, participants downloaded the Metricwire app that sent them three types of surveys over a 6-week period daily surveys (1 time per day), EMA surveys (5 times per day), and burst surveys (6 times/hour, 2 per day, 4 days/week). In this paper, we focus on the EMA surveys and burst surveys (Figure 1). EMA surveys were sent 5 times per day at least 90 minutes apart between participant custom wake/sleep times. EMA surveys stayed open for 1 hour. Burst surveys were sent 6 times per hour (episode) with the maximum of 2 episodes per day and 4 days per week.

The burst surveys were sent immediately following the completion of a longer EMA (**Figure 1**). Each burst survey stayed open for 5 minutes. In each EMA survey, participants were provided with resources for treatment and safety (e.g. suicide prevention hotlines). Participants were paid every 2 weeks during the 42-day real-time monitoring period to increase compliance.

Real-Time Monitoring Measures

Suicidal thoughts were measured with two items: desire to kill self and intent to kill self "right now" (see Figure 1b). Both items were rated on a 0 (not at all) to 10 (very strong scale). Desire was defined for participants as how much do you want to kill yourself. Intent was defined as to what extent are you actually going to kill yourself. These items were selected based on the Beck Scale for Suicidal Ideation (11), a widely-used measure of suicidal thinking. Questions pertaining to desire and intent are similar to momentary measures of suicidal thinking used in other real-time monitoring studies (9, 12) and have shown predictive validity of suicidal behavior (13).

Analytic Approach

All analyses were conducted in R version 4.10. All data analysis code is available at https://github.com/ryanoisin/TimescaleSuicidalThinking. The data and materials that support the findings of this study are available from the corresponding author on request.

We used three different strategies to leverage our unique sampling design to gain insight into the timescale of change of suicidal thinking. First, we examined descriptive and time-invariant properties of our variables, such as the within-person mean, variance and mode, and the number of high risk responses captured, with the latter defined as ratings of 8, 9, or 10 on the suicidal intent item based on previous research (15). We assessed whether measurements taken at different frequencies capture fundamentally different processes on the aggregate level by examining whether these characteristics differed between high-frequency (from the burst design, every 10 minutes) and low-frequency (from the EMA design, approximately 3 to 12 hours apart) measures.

Second, we assessed measurement-to-measurement variability in self-reported desire to kill self and intent to kill self both across and within individuals as a function of the sampling frequency. This allowed us to gain direct insight into the timescale of change of these variables: For instance, if desire only changes appreciably on a timescale of hours, then we would expect to see little to no variance in measurements taken at 10-minute time intervals. The unique sampling scheme of the empirical data allowed us to probe the degree of variation at different timescales directly.

Finally, we used continuous-time (CT) models (14) to estimate the moment-to-moment dynamics of suicidal desire and intent. CT approaches allow users to estimate models from data collected at varying intervals by explicitly modeling time-forward relationships as a function of the time-interval between measurement occasions. In this way, CT models can be seen as a more appropriate alterative to discrete-time models such as vector auto-regressive or Markov-switching models more frequently used in psychological settings (15, 16), since the latter assume evenly spaced data, an assumption which is almost always violated in real-time monitoring settings (18). This feature also means the CT models can be used in a straightforward way to compare model parameters from datasets which contain different sampling schemes (17).

In the current paper, we fit CT versions of both the vector auto regressive model (CT-VAR; 17) and the Markov multi-state models (CT-Markov; 18) to the empirical data. The CT-VAR models desire and intent as continuous-valued processes, which influence each other over time through a linear system of first-order differential equations. From a qualitative perspective, they model processes which fluctuate around a stable equilibrium: external shocks (random noise) push the processes away from equilibrium, and the underlying lagged dependencies pull the process back to equilibrium over time. The CT-Markov model, in contrast, treats these processes as consisting of discrete states, modeling the probability of changing from one state to another at the next measurement occasion given the current state. To enable us to use this model we recoded the 11-point scales into three states: low (0 to 3), mild (4 to 7), severe (8 to 10).

Both model types use information about the time-interval between measurement occasions in order to estimate *moment-to-moment* dynamic relationships (known as the *drift matrix* for the CT-VAR and *intensity matrix* for the CT-Markov). These in turn can be used to model how lagged relationships and state-switching probabilities, respectively, depend in a non-linear way on the time-interval between measurements. In this way both models yield different but potentially complimentary models for how suicidal desire and intent evolve and vary at different timescales. We fit a bivariate hierarchical CT-VAR model with the ctsem R package (18) and a fixed-effects CT-Markov model desire and intent separately with the msm package (17). As a secondary analysis, we examined whether fitting both sets of CT models separately on high-frequency (pairs of observations spaced < 1.5 hours apart) and low-frequency (pairs spaced >1.5 hours apart) subsets of the data (see **Figure 1** panels b and c) yielded different conclusions about the underlying dynamics. This allows us to assess whether high frequency measurements contained unique information about the dynamics underlying suicidal thinking when compared to more typical low frequency EMA measurements.

Results

Descriptive Statistics

The 105 participants of the study completed a total of 20,255 surveys. The number of observations per person ranged from 5 to 456 (mean = 192.9, IQR = 58 to 296). The mean length of the observation period, that is, the time elapsed between the first and last observations, was 33.64 days (min = 2, max =41.62, IQR = 30.03 to 41.08). We assess whether measurements taken at different frequencies (burst vs EMA) are capturing the same underlying processes by examining the (within-person, over time) means and variances of desire and intent. **Figure 2** panels a and b show that the distributions of mean within-person desire and intent are highly similar in both the burst and EMA measurements. In panel c we see that individual's EMA-mean and burst-mean are highly correlated (r = 0.94 for desire and r = 0.97 for intent), and in panel c that the within-person standard deviation for both measurement types are highly correlated (r = 0.74 desire and r = 0.88 for intent). The similarity of these characteristics across measurement types indicates that participants do not appear to be responding to EMA and burst measurements in a systematically different manner. We recorded 1,213 high-risk observations (6% of all observations), spread across 31 participants. 544 of these high-risk observations were captured by traditional EMA samples (544/9755 = 5.6% of all EMA measurements were high risk). The remaining 669 were captured by burst surveys (669/10500 = 6.4% of all burst measurements were high risk). This suggests that high risk moments were slightly more likely to be observed with burst measurements than with traditional EMA. To account for the fact that high-risk states may last relatively long, we computed the number of unique high-risk moments captured by burst measurements (high risk moments reported during burst in which the preceding EMA measurement was not high risk). In total, 74 unique high-risk observations were captured by burst measurements. In terms of individuals, in total 6 individuals recorded high-risk moments in burst measurements while never recording a high-risk moment in an EMA measurement. In other words, with burst measurements we identified 19% more participants with a high-risk moment.

For the majority of participants, the most common response to the self-report items, i.e. the *modal response*, was 0 (not at all): (59/105) 56.2% of participants had a modal response of 0 to the desire item and (78/105) 74.3% had a modal response of 0 to the intent item. We can consider the most common response to be an indicator of the resting state of desire and intent for a given individual (19). All participants who had a mode of 0 on desire also had a mode of 0 on intent.

The degree of variability in responses also differed across individuals and items. Since the items are measured using a likert scale with limited answer categories, we can understand variability in terms of how often a participant reacts to a measurement prompt by using a response category which is equal to their resting state value, with p_{mode} = 1 indicating that an individual responds with the same answer category to every prompt, and therefor shows no variability. On average, around 50 percent of responses to the desire item were different from the individual's mode, but we see a large degree of variability across individuals (median p_{mode} = 50, IQR = 0.3 – 0.83), with one individual reporting no variability at all. Variability on the intent item was even lower (median p_{mode} = .91, IQR = .99 - .49), with 15 individuals showing no variability at all in reported suicidal intent. All individuals who exhibited no variability in responses had a resting state of 0, and there was a high positive correlation between the variability metrics of desire and intent (r = .621).

Variability as a Function of Time-Interval

Based on the large between-person differences we observe in the variability of desire and intent, we can create a distinction between three different groups of participants in our dataset: those for whom the experience of both Desire and Intent was a rare occurrence (*n* = 10, characterized by a p_{mode} >= .95 for both variables); those for whom the experience of intent was a rare occurrence, but who felt desire more frequently *(n* = 31, characterized by a p_{mode} >= .95 for intent only); and those who experience a reasonable degree of variation in both desire and intent (*n* = 64, all other participants). These qualitative differences in response patterns lead to different analysis strategies through which we might understand the timescale of change of suicidal thinking.

Participants with low suicidal desire and/or intent. For those participants who experienced desire and/or intent as rare events (*n* = 10 and *n* = 31 respectively) we can simply examine how often these events occur, and how long elevated (i.e., non-modal) levels of desire and intent last. We do this by counting sets of consecutive observations that show elevated desire or intent per-person. We calculate the duration of elevated desire/intent by observing how much time elapses before the next occasion where desire or intent are back to their resting state. As such, we can consider the duration estimated in this way as an upper bound (as participants may have returned to normal before the next observation). In our analysis, we omit duration estimates which include night gaps (observations which are elevated before bed, but the next day back to normal). We counted 43 episodes of elevated desire across 9 participants with a median estimated duration of 1.2 hours (IQR = .46 to 2.47, mean = 1.66). We count 67 episodes of elevated Intent events spread across 26 individuals, with a median duration of 0.82 hours (IQR = .17 to 2.55, mean = 2.2). This shows that for participants for whom desire and intent is a rare event on a 42-day timescale, the event itself is typically quite short lived – lasting from a number of minutes up to typically around 2 hours, with only very rare exceptions of longer lasting episodes.

Participants with higher variation in suicidal desire and intent. For those individuals for whom desire or intent are not "rare" events, but which vary more frequently over time, it is less feasible to distinguish distinct episodes of elevated intent. Instead, we take the following approach. First, for each individual, we bin pairs of consecutive observations according to how far apart in time they are spaced. Second, we calculate the degree of variation at different timescales by calculating how often these pairs of differ in value. To facilitate this analysis, we create three-timescales bins: less than 45 minutes, 45 minutes to 3 hours, and greater than 3 hours. These bins were chosen such that, as much as possible, all individuals have at least two pairs of observations within each bin. The results for desire and intent are shown in **Figure 3**. For desire, we used data from n = 85 individuals and for intent n = 58 individuals (see supplemental material for more details on data filtering). The proportion of change increases as the duration of timescale increases, which is to be expected. There was substantial variation (on average 36.5% for desire, 33% for intent) in the shortest timescale (less than 45 minutes). It is important to note that **Figure 3** highlights individual differences in both the amount of variation and how variation changes across timescales.

Dynamics of Suicidal Thinking

Figure 4 panel (a) shows the CT-VAR model fixed effects (drift matrix) estimates along with their 95% credible intervals. The parameters should be interpreted as estimates of the average moment-to-moment dynamics across individuals. We see that the model estimates a positive (i.e. activating) effect from desire to intent but no reciprocal effect from intent to desire. The interpretation of these parameters is that elevations in levels of Desire are predicted to co-occur with future elevations in Intent, but that elevations in Intent do not predict changes in the level of Desire. The self-loops depicted in the graph represent the estimated effect of each variable on its own rate of change. These parameters are negative, and so can be interpreted as self- regulating effects or dampening effects of desire and intent on themselves over time: the further away from equilibrium the process finds itself, the quicker the process moves towards equilibrium.[1] The more strongly negative self-effect of Intent should be interpreted as reflecting that elevated levels of intent are more quickly regulated away than elevated levels of desire.

Panels (b) and (c) of **Figure 4** show how the moment-to-moment dynamics can be used to understand the implied timescale of the process. Panel (b) shows how, according to the parameter estimates, current observations of Desire and Intent are expected to be predictive of each other's future values, as a function of the time-interval between observations. We see that Desire is predictive of itself and Intent for a period of around 20 hours; the relationship between current Desire and future Intent is strongest at a time-interval of around 2-3 hours. We can also see that Intent is weakly predictive of Desire over any time-interval, and current levels of intent no longer have any predictive value for future Intent levels after 2-3 hours. Another way to understand the timescale of change implied by the model parameters is the impulse response function (IRF), that is, the predicted trajectory of the system over time given an impulse. Panel (c) shows the predicted IRF given an impulse to increase levels of Desire. Again, we see the model predicts that Intent will increase over a period of 2-3 hours before both Desire and Intent return to baseline over a period of 20 hours.

Figure 5 panels (a) and (b) show the CT Markov switching model estimates for the Desire and Intent variables respectively. For the sake of interpretability, we use the estimated *intensity matrix* (see S.2) to derive, on the left-hand-side of each figure, the estimated transition probability matrix over a time-interval of one hour, that is, the probability of either staying in the same state (self-loops) or transitioning to a different state at an assessment taken one-hour later.

[1] The self-loops depict the diagonal elements of the drift matrix, sometimes referred to as auto-effects. Discrete-time auto-regressive effects between 0 and 1 often interpreted as inertia parameters, can be interpreted in the same way as continuous-time auto-effects running from 0 to negative infinity.

For both desire and intent, we observe that, at one hour intervals, participants are most likely to stay in their current state, and that the probability of staying in the low state is largest. This reflects that most participants frequently endorse low feelings of Desire and Intent through the observation period. On the right-hand-side of each panel we show the estimated sojourn times for each state, that is, how long on average a participant is predicted to stay in one state before transitioning to another. The sojourn times give us a direct estimate of the time-scale of suicidal desire and intent in terms of estimated state durations. We see that on average, Low levels of Desire are estimated to last 19.42 hours [CI: 17.19, 21.95], moderate levels 1.89 hours [CI: 1.70, 2.10] and severe levels 2.97 hours [CI: 2.55, 3.46]. We see a similar pattern for Intent, with low levels estimated to last 17.43 hours [CI: 15.1, 20.12], moderate levels 1.15 hours [CI: 1.02, 1.30], and high levels 3.03 hours [CI: 2.43, 3.79].

As a final analysis, we estimated both the CT-VAR and CT-Markov model on subsets of our data consisting of only regular EMA measurements and short time-interval burst measurements respectively. **Figure 6** shows the CT-VAR parameter estimates across both datasets. Comparing panels (a) and (b) we see many qualitative similarities in parameter estimates. The self-effect of intent in both cases is higher than that of desire; desire has a positive cross-effect on Intent, although the weakly negative effect of Intent on desire is considered “significant” in the EMA-dataset, in that the upper end of the credible interval does not cross zero. The parameter estimates themselves, however, are quite different, and yield somewhat different interpretations of the timescale of suicidal thinking. Based only on the EMA measurements, we would derive that elevated levels of Intent return to baseline much more slowly than we would conclude based on the full dataset or the short-time-interval subset. This can be seen in panels (c) where the self-effect of intent reaches zero after 8 hours, and in panel

(e) where, according to the IRF, the effect of an impulse on Intent is still present after 24 hours. **Figure 7** shows the CT-Markov parameter estimates across both datasets. We again see a qualitatively similar pattern of results, but that both datasets yield different quantitative estimates of the timescale of change of suicidal thinking. Across both desire and intent, the probability of transitioning was higher in the high frequency dataset. For example, the probability of transitioning from a mild state to a severe state for desire in the EMA dataset is 9.2%, whereas in the high frequency dataset the probability for desire is 29%. For both desire and intent, the average sojourn time for mild and severe states were shorter in the high frequency dataset than the EMA dataset. In the EMA datasets, the average sojourn time for mild and severe states were 5.6 and 9.5 hours (Desire) and 4.0 and 9.4 hours (Intent). In the high-frequency datasets, the average sojourn time for mild and severe states were merely 0.84 and 1.4 hours (desire) and 0.49 and 1.2 hours (intent).

Discussion

This study represents the highest resolution examination of the temporal dynamics of suicidal thinking to date. There were three key findings. First, suicidal thinking changes rapidly. Second, different aspects of suicidal thinking (i.e., suicidal desire and intent) operate on different timescales. Third, capturing within-person dynamics of suicidal thinking is improved using high-frequency sampling. Each of these findings warrants additional comment.

Across descriptive statistics and multiple statistical models, we found evidence that suicidal thinking changes rapidly. Suicidal thinking returned to baseline within several hours and elevated suicidal desire led to elevated suicidal intent within minutes to hours. These findings provide empirical support to theoretical work that conceptualizes of suicide as a dynamic system (6) and preliminary descriptive research on the dynamics of suicidal thinking (12). Converging lines of research (13, 20, 21), support conceptualizing of suicide risk as a dynamical system that unfolds over short timescales. This dynamic conceptualization of suicide, should inform future theories (4), assessment (13), and interventions (22).

From this study, we learned two new things about different aspects of suicidal thinking. First, suicidal desire lasts longer than suicidal intent. Second, suicidal desire precedes suicidal intent. Prior research has often either measured suicidal thinking with a single item assessment (23, 24) or collapsed multiple aspects into a single sum score of suicidal thinking (9, 25). Our findings suggest taking such an approach would be problematic because different dimensions of suicidal thinking have unique dynamics. Through providing greater specificity in characterizing different dimensions of suicidal thinking, our findings support conceptualizing of suicidal thinking as a multidimensional phenomenon. This more precise understanding of the different dimensions has implications for both assessment and intervention. For example, we characterized the duration from an instance of elevated suicidal desire to suicidal intent. This time between desire and intent represents a potential window for intervention (22) to reduce risk.

Finally, we showed how sampling frequency matters for characterizing suicidal thinking. If one were only interested in extracting features (e.g., the mean) from the time-series of suicidal thinking (13), lower-frequency sampling would show no difference from higher-frequency sampling. The proposed promise, however, of real-time monitoring research has been in capturing dynamics as they unfold (8, 26). Our results suggest that accurately characterizing within-person dynamics of suicidal thinking requires higher-frequency sampling (i.e. minutes apart). Statistical models cannot make up for under-sampled data (27). This is evidenced by our analysis of subsets of the empirical data, which showed that, even when using continuous-time models to correct for the differences in time-intervals, high-frequency measurements yielded different quantitative estimates of the timescale of suicidal thinking than lower-frequency EMA measurements. The risk of using classic real-time monitoring designs (i.e., hours apart; 29) is that one may underestimate the speed of the system because one cannot infer shorter lags (e.g. 30 minutes) from longer lags (e.g., 6 hours). Our findings have implications for the measurement of all psychological constructs and provide empirical support to speculative discussions on the importance of sampling frequency (29, 30). The implication of these findings is not that all psychological constructs need to be measured in brief (e.g., 10 minute) intervals, but rather highlight the crucial importance of sampling frequency as a design decision. Specifically, psychological constructs need to be measured at a high-frequency to understand the rate at which they change.

While the current study has several strengths such as the novel sampling design, there are multiple limitations that require discussion. First, a concern of high frequency assessments is reactivity to the assessments, such as that monitoring changes the process under observation. While analyses from this overall project (31) and other projects (32) suggest that frequently assessing suicidal thinking does not increase suicidal thinking, it could have impacted the data in some other way. Second, for continuous-time models we were only able to use the subset of the participants with high compliance and some variability in their suicidal thinking. This limits the generality of the findings. Third, for the Markov models we arbitrarily categorized suicidal thinking into three states, but we don't know if these are the exact states of interest. In future work, hidden Markov models (33) could be applied for a data-driven approach to identifying states of interest. Fourth, the true data generating model of suicidal thinking is unknown and therefore we may have picked inappropriate models for the dynamics of suicidal thinking. While we used two of the most widely studied models for characterizing dynamics of systems over time, model misspecification is probable, in the sense that the dynamics governing intent and desire may not be well approximated by linear and first-order dynamic models. Although the Markov models and the mode-based descriptive statistics took different approaches to characterizing the timescale of suicidal thinking, both approaches can be interpreted as indicating that elevated levels of suicidal thinking last on average 1-3 hours. More exploratory tools (34) are needed for data sampled at uneven time intervals (27). Finally, the current study only measured two aspects of suicidal thinking and did not measure suicidal behavior. It is currently unknown how these very short dynamics relate to suicidal behavior.

There are several future directions that build upon the current study. First, one could use higher density sampling over a longer interval than one hour. For example, one could try to sample every 30 minutes over the course of a day with a small number of questions of suicidal thinking. This would allow for building a continuous time model for a 24-hour period. Second, one could track suicidal behavior after the real-time monitoring period to understand if higher sampling of suicidal thinking improves the prediction of future suicidal behavior (13). This would be insightful for further understanding the value gained by more frequent assessments. Finally, the ultimate aim of the current study was to richly characterize the dynamics of the phenomenon of suicidal thinking. Future theoretical work could attempt to integrate these dynamics into a formal theory of suicide (4, 35). A valuable test of a formal theory of suicide would be evaluating if it could recover the dynamics identified in the current study.

Suicide has historically represented one of the most difficult topics to study. The current study highlights how we can use new forms of data collection and statistical models to zoom the microscope in on this perplexing phenomenon. We have provided new insights into suicidal thinking and hope this work catalyzes conceptualizing of suicide risk as a dynamical process that unfolds over short windows of time.

Funding: This research was supported by the Pershing Square Venture Fund for Research on the Foundations of Human Behavior, the Sydney DeYoung Foundation, and the Knox Fund at Harvard University. DDLC is supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1745303 and the National Institute of Mental Health under Grant F31MH130055. RGF is supported by a NIMH-funded postdoctoral fellowship [T32MH017119]. This content is solely the responsibility of the authors and does not necessarily represent the official views of the National Science Foundation or the National Institute of Mental Health. This work was also supported by the Fuss Family Research Fund at Harvard University to MKN.

Disclosures: Dr. Nock receives publication royalties from Macmillan, Pearson, and UpToDate. He has been a paid consultant in the past year for Microsoft Corporation, the Veterans Health Administration, and for a legal case regarding a death by suicide. He holds stock options in Cerebral Inc and is an unpaid scientific advisor for Empatica, Koko, and TalkLife.

References

1. S. Fazel, B. Runeson, Suicide. *New England Journal of Medicine* **382**, 266–274 (2020).

2. J. C. Franklin, *et al.*, Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological Bulletin* **143**, 187–232 (2017).

3. K. R. Fox, *et al.*, Interventions for suicide and self-injury: A meta-analysis of randomized controlled trials across nearly 50 years of research. *Psychol Bull* (2020) https:/doi.org/10.1037/bul0000305.

4. A. J. Millner, D. J. Robinaugh, M. K. Nock, Advancing the Understanding of Suicide: The Need for Formal Theory and Rigorous Descriptive Research. *Trends in Cognitive Sciences* **24**, 704–716 (2020).

5. G. Turecki, *et al.*, Suicide and suicide risk. *Nature Reviews Disease Primers* **5**, 1–22 (2019).

6. C. J. Bryan, *et al.*, Nonlinear change processes and the emergence of suicidal behavior: A conceptual model based on the fluid vulnerability theory of suicide. *New Ideas in Psychology* **57**, 100758 (2020).

7. D. De Beurs, *et al.*, A network perspective on suicidal behavior: understanding suicidality as a complex system. *Suicide and Life-Threatening Behavior* (2020).

8. E. M. Kleiman, M. K. Nock, Real-time assessment of suicidal thoughts and behaviors. *Current Opinion in Psychology* **22**, 33–37 (2018).

9. B. L. Gee, J. Han, H. Benassi, P. J. Batterham, Suicidal thoughts, suicidal behaviours and self-harm in daily life: A systematic review of ecological momentary assessment studies. *Digit Health* **6**, 2055207620963958 (2020).

10. C. J. R. Sewall, A. G. C. Wright, Personalizing Suicidology. *Crisis* **42**, 405–410 (2021).

11. A. T. Beck, M. Kovacs, A. Weissman, Assessment of suicidal intention: The Scale for Suicide Ideation. *Journal of Consulting and Clinical Psychology* **47**, 343–352 (1979).

12. E. M. Kleiman, *et al.*, Examination of real-time fluctuations in suicidal ideation and its risk factors: Results from two ecological momentary assessment studies. *Journal of Abnormal Psychology* **126**, 726–738 (2017).

13. S. B. Wang, *et al.*, A Pilot Study Using Frequent Inpatient Assessments of Suicidal Thinking to Predict Short-Term Postdischarge Suicidal Behavior. *JAMA Network Open* **4**, e210591–e210591 (2021).

14. O. Ryan, R. M. Kuiper, E. L. Hamaker, "A Continuous-Time Approach to Intensive Longitudinal Data: What, Why, and How?" in *Continuous Time Modeling in the Behavioral and Related Sciences*, K. van Montfort, J. H. L. Oud, M. C. Voelkle, Eds. (Springer International Publishing, 2018), pp. 27–54.

15. D. G. Jordan, E. S. Winer, T. Salem, The current status of temporal network analysis for clinical science: Considerations as the paradigm shifts? *Journal of Clinical Psychology* **76**, 1591–1612 (2020).

16. S. de Haan-Rietdijk, *et al.*, On the Use of Mixed Markov Models for Intensive Longitudinal Data. *Multivariate Behavioral Research* **52**, 747–767 (2017).

17. C. C. Driver, M. C. Voelkle, Hierarchical Bayesian continuous time dynamic modeling. *Psychol Methods* **23**, 774–799 (2018).

18. C. Jackson, Multi-State Models for Panel Data: The msm Package for R. *Journal of Statistical Software* **038** (2011).

19. W. R. Ringwald, A. G. C. Wright, Overcoming the Confound of Means and Variability for Measuring Everyday Emotion Dynamics Related to Neuroticism (2022) https:/doi.org/10.31234/osf.io/nxbyd (May 4, 2022).

20. C. J. Bryan, M. D. Rudd, The Importance of Temporal Dynamics in the Transition From Suicidal Thought to Behavior. *Clinical Psychology: Science and Practice* **23**, 21–25 (2016).

21. A. J. Millner, M. D. Lee, M. K. Nock, Describing and Measuring the Pathway to Suicide Attempts: A Preliminary Study. *Suicide and Life-Threatening Behavior* **47**, 353–369 (2017).

22. I. Nahum-Shani, *et al.*, Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. *Annals of Behavioral Medicine* **52**, 446–462 (2018).

23. P. J. Batterham, *et al.*, A systematic review and evaluation of measures for suicidal ideation and behaviors in population-based research. *Psychological Assessment* **27**, 501–512 (2015).

24. A. J. Millner, M. D. Lee, M. K. Nock, Single-Item Measurement of Suicidal Behaviors: Validity and Consequences of Misclassification. *PLOS ONE* **10**, e0141606 (2015).

25. D. D. L. Coppersmith, E. M. Kleiman, C. R. Glenn, A. J. Millner, M. K. Nock, The dynamics of social support among suicide attempters: A smartphone-based daily diary study. *Behaviour Research and Therapy* **120**, 103348 (2019).

26. C. L. Davidson, M. D. Anestis, P. M. Gutierrez, Ecological Momentary Assessment is a Neglected Methodology in Suicidology. *Archives of Suicide Research* **21**, 1–11 (2017).

27. J. M. B. Haslbeck, O. Ryan, Recovering Within-Person Dynamics from Psychological Time Series. *Multivariate Behavioral Research* **0**, 1–32 (2021).

28. S. Shiffman, A. A. Stone, M. R. Hufford, Ecological Momentary Assessment. *Annual Review of Clinical Psychology* **4**, 1–32 (2008).

29. C. J. Hopwood, W. Bleidorn, A. G. C. Wright, Connecting Theory to Methods in Longitudinal Research. *Perspect Psychol Sci*, 17456916211008408 (2021).

30. N. Ram, M. Brinberg, A. L. Pincus, D. E. Conroy, The Questionable Ecological Validity of Ecological Momentary Assessment: Considerations for Design and Analysis. *Research in Human Development* **14**, 253–270 (2017).

31. D. D. L. Coppersmith, *et al.*, Effect of frequent assessment of suicidal thinking on its incidence and severity: high-resolution real-time monitoring study. *The British Journal of Psychiatry*, 1–3 (undefined/ed).

32. M. K. Law, *et al.*, Does assessing suicidality frequently and repeatedly cause harm? A randomized control study. *Psychological Assessment* **27**, 1171–1181 (2015).

33. R. M. Altman, Mixed Hidden Markov Models. *Journal of the American Statistical Association* **102**, 201–210 (2007).

34. N. C. Jacobson, S.-M. Chow, M. G. Newman, The Differential Time-Varying Effect Model (DTVEM): A tool for diagnosing and modeling time lags in intensive longitudinal data. *Behav Res* **51**, 295–315 (2019).

35. J. M. B. Haslbeck, O. Ryan, D. J. Robinaugh, L. J. Waldorp, D. Borsboom, Modeling psychopathology: From data models to formal theories. *Psychol Methods* (2021) https:/doi.org/10.1037/met0000303.

Figure 1. Overview of sampling design. (a) combined EMA and burst measurements; (b) EMA only dataset; (c) Burst-only dataset; and (d) Questions asked about suicidal thinking during real-time measurements

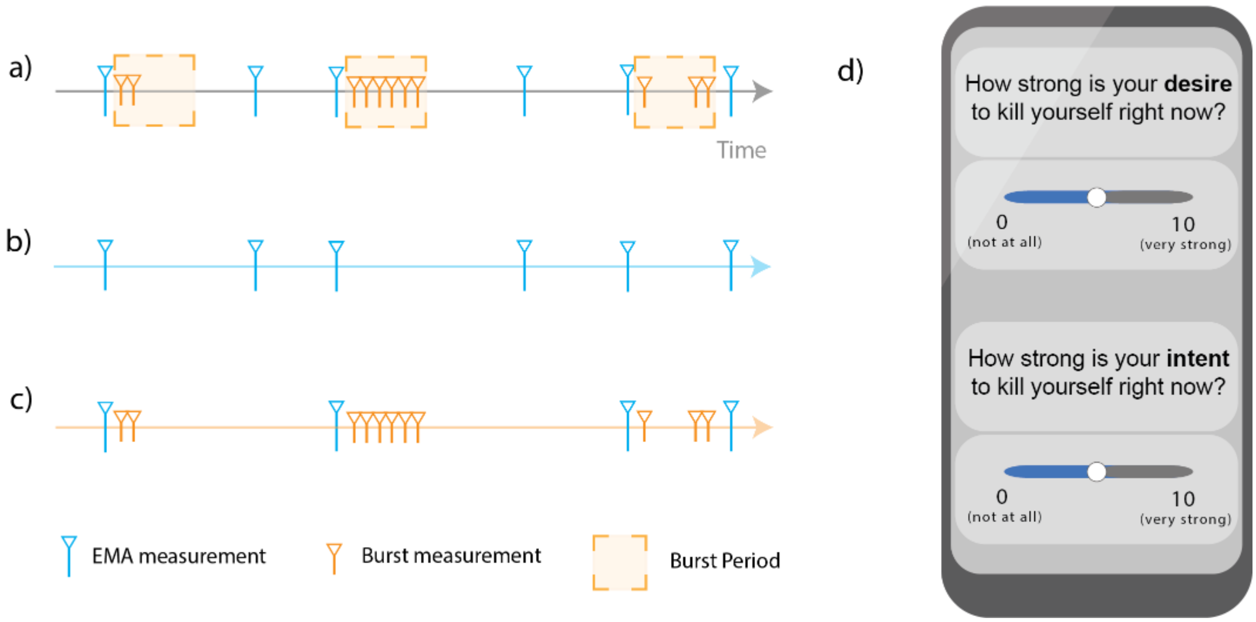


Figure 2. Descriptive Statistics across Different Types of Real-Time Measures of Suicidal Thinking

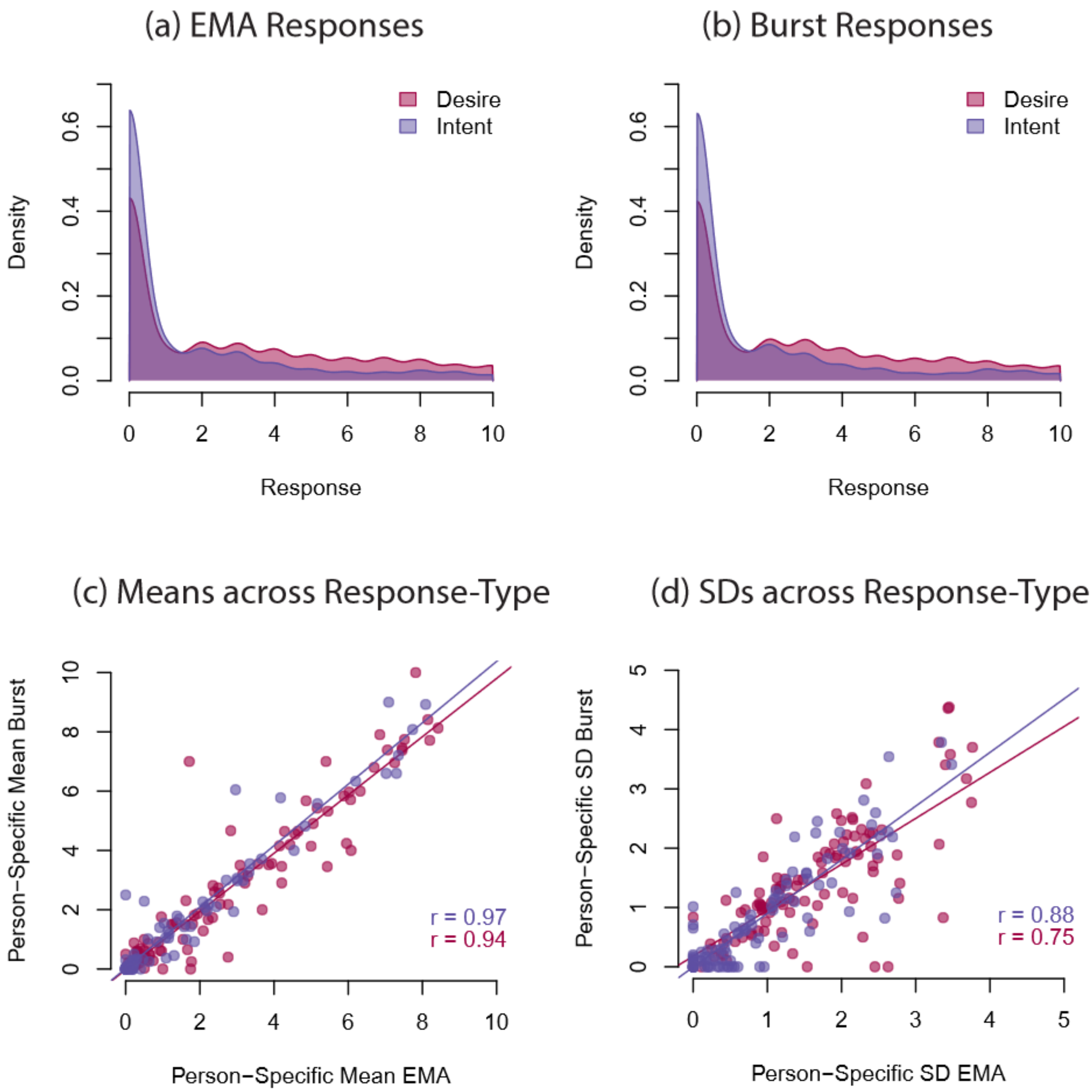


Figure 3. Proportion of Consecutive Observations that Show Variation as a Function of Time Between Observations for Desire (panel a) and Intent (panel b). Each red dot represents an individual participant, with the light grey lines connecting values of the same participant across timescales. The grey diamond represents the mean proportion in a given timescale (Desire = {.365, .473, .545}, Intent = {.33, .40, .48})

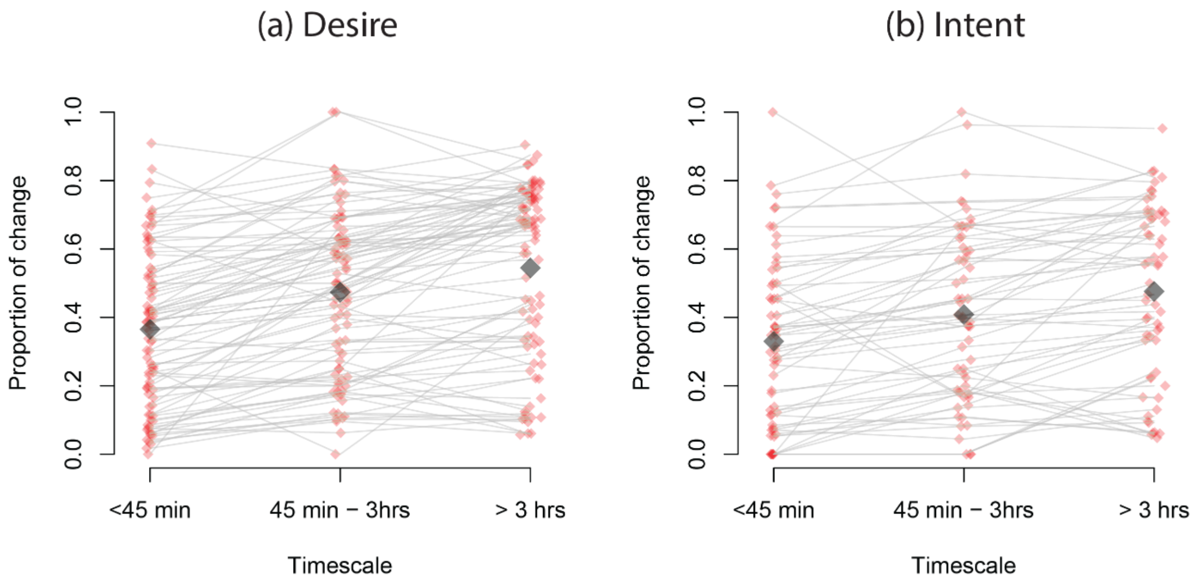


Figure 4. Continuous-time Vector Autoregression Results. Panel (a) depicts the estimated drift matrix fixed effects, with CIs, as a network. Panel (b) shows the model-implied lagged regression coefficients, as a function of the time-interval. Φ_{ij} represents the effect of X_i now on X_j some time-interval later, with X_1 = Desire and X_2 = Intent. Shaded lines represent 95% credible intervals. Panel (c) shows the model-implied Impulse Response Function, with the impulse values indicated by the filled diamond.

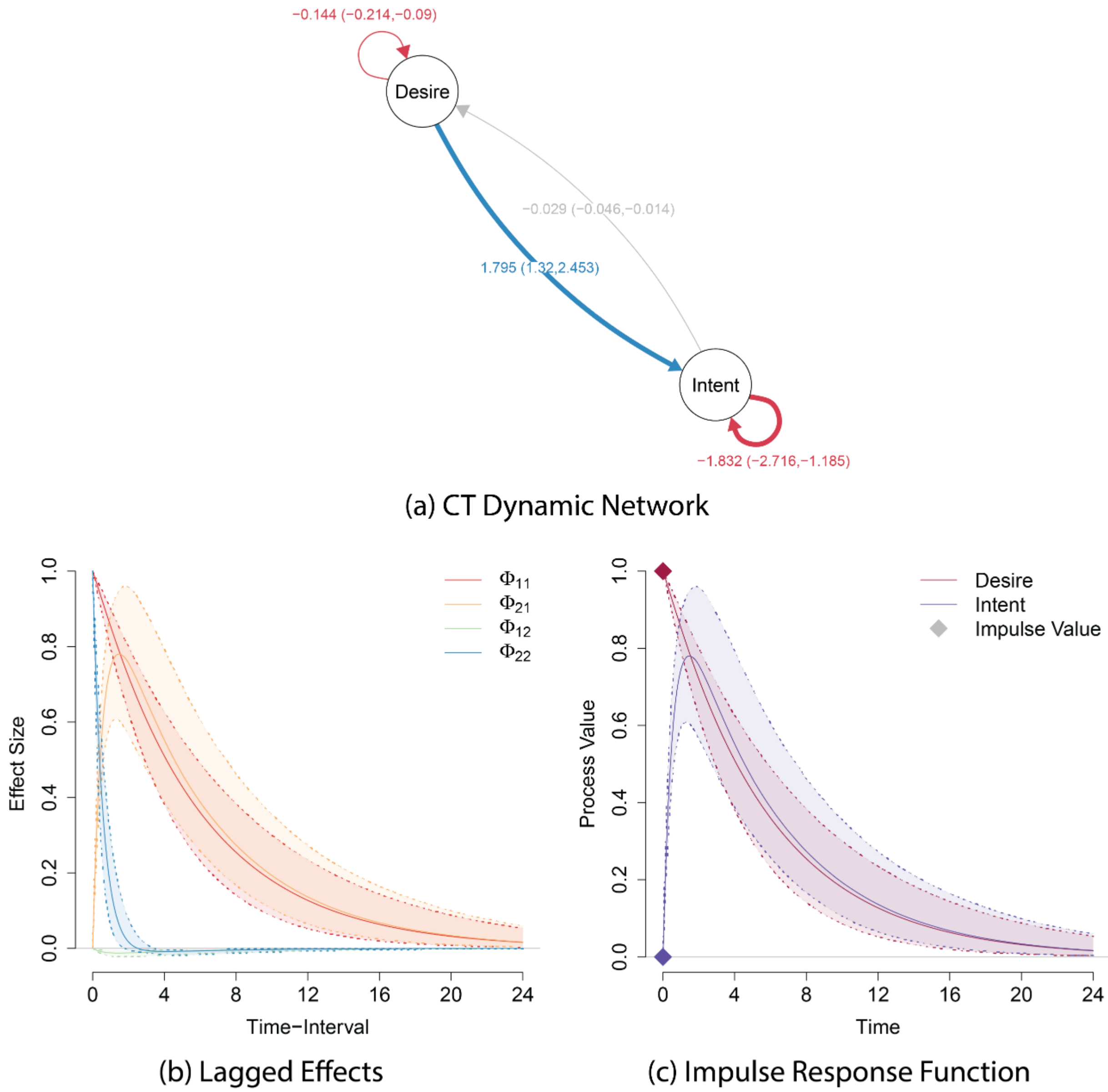


Figure 5. Transition Probabilities (at a one-hour interval, left) and Sojurn Times (right) from the Continuous Time Markov Models. Clocks represent sojurn time duration in blocks and fractions of 12 hours.

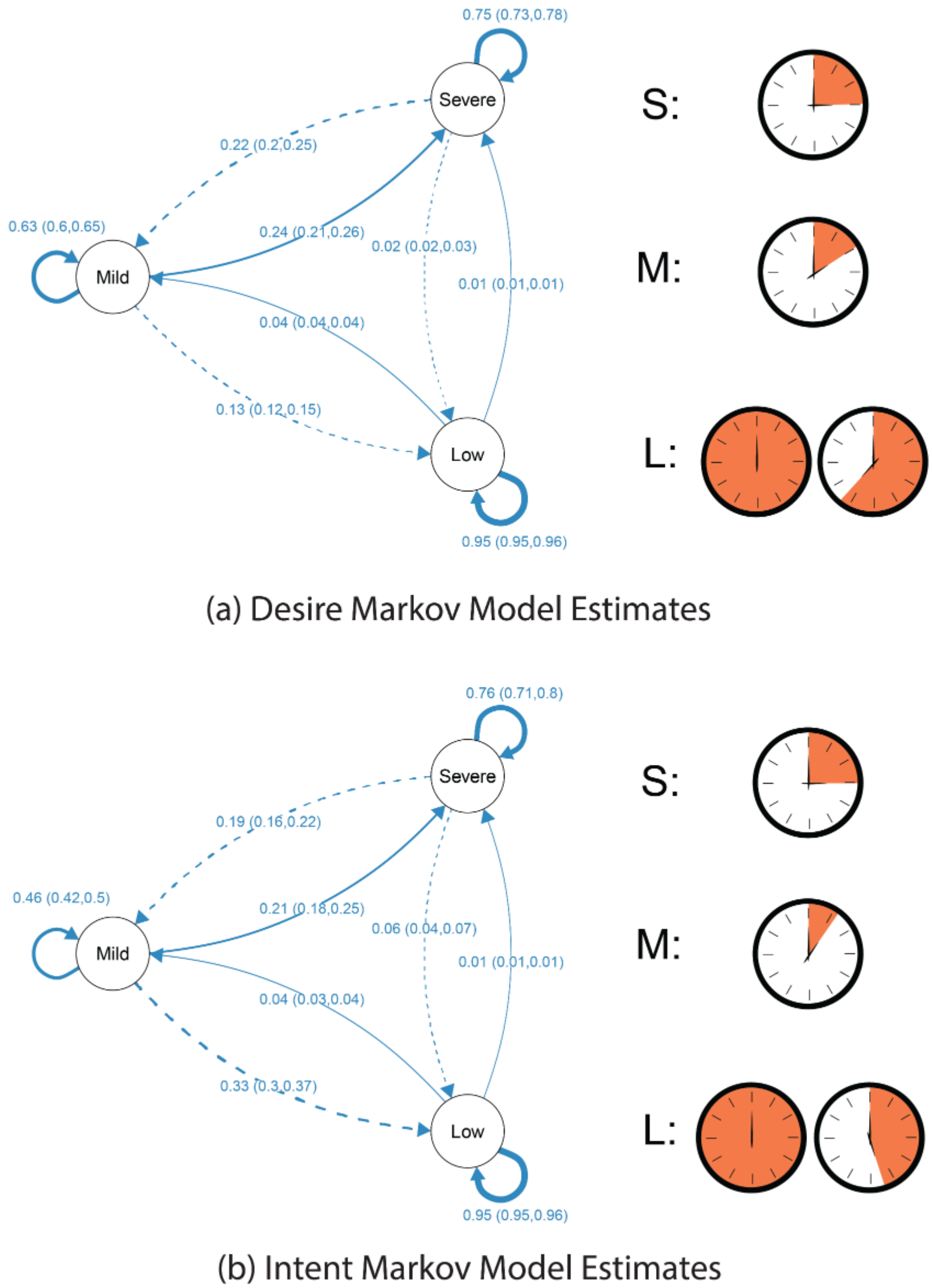


Figure 6. Continuous-time Vector Autoregression Results across EMA and Short Datasets.

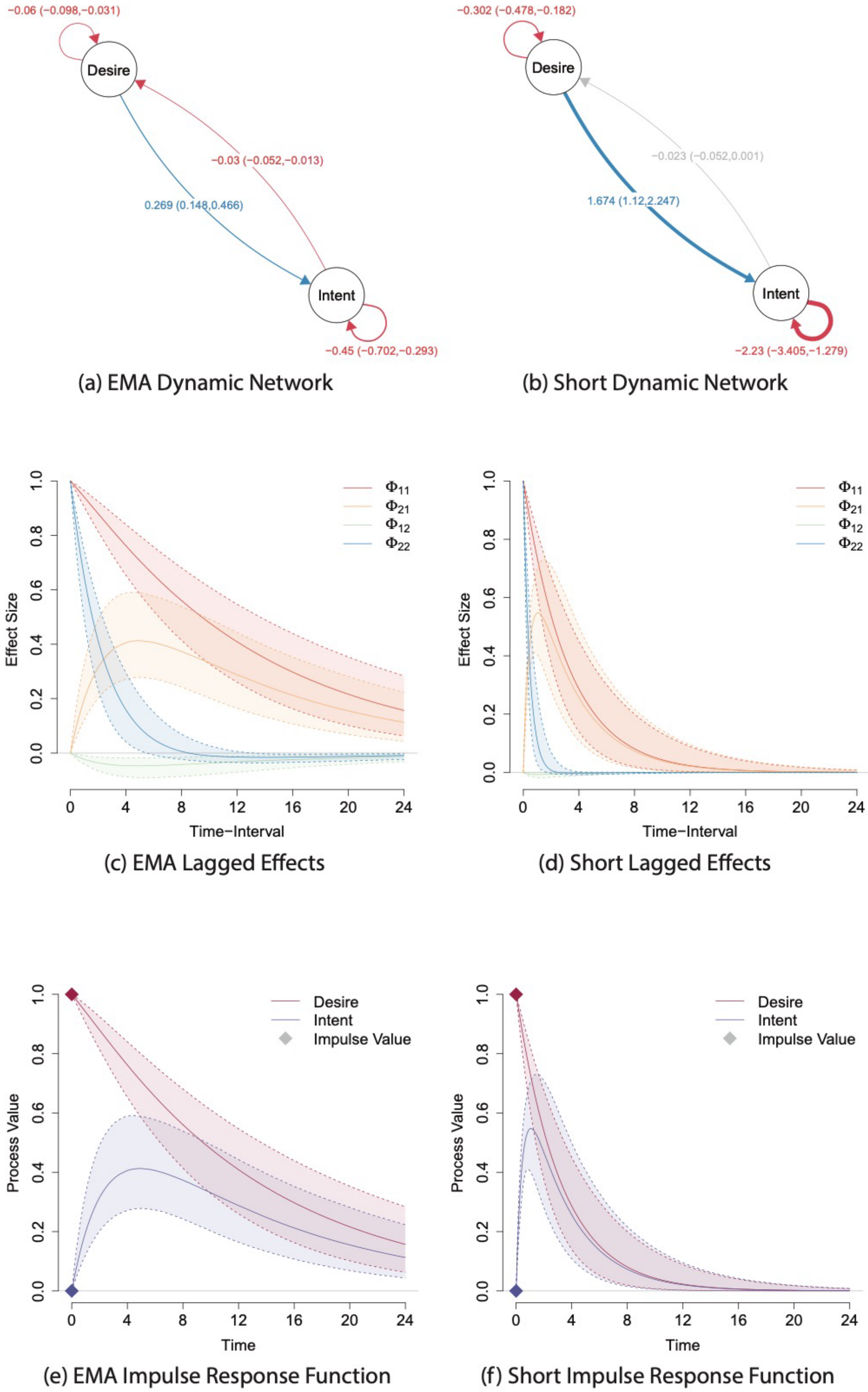
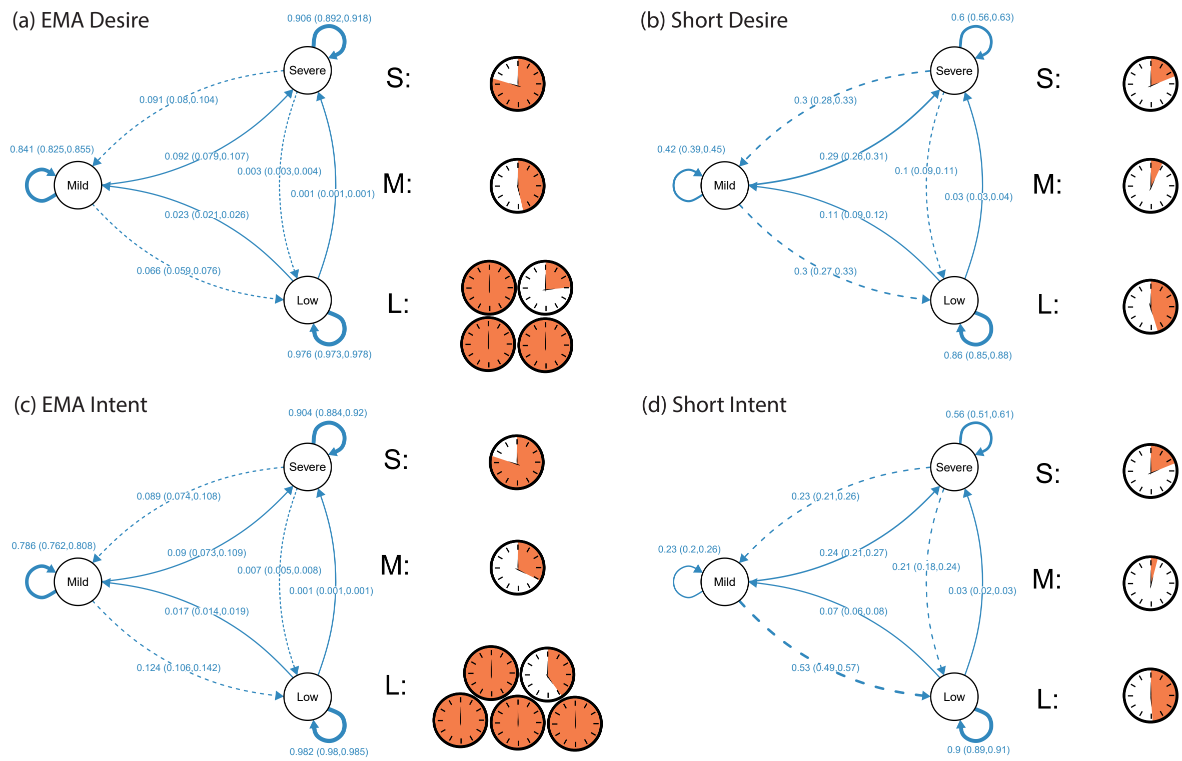


Figure 7. Transition Probabilities (at a one-hour interval) and Sojurn Times from Continuous Time Markov Models across EMA and Short Datasets. Note: clocks represent sojurn time duration.



Supplemental Materials

Participant Recruitment

Participants were recruited online (e.g., Reddit, Craigslist), with postings seeking "people who have experienced difficult emotions." Participants completed an eligibility screener that assessed self-injurious thoughts and behaviors. Inclusion criteria were active suicidal thoughts in the past week, fluency in English, >17 years of age, and regular access to a smartphone. An exclusion criterion was living in Europe due to GDPR restrictions, failing a CAPTCHA test or providing inconsistent or illogical responses.

A total of 8,035 individuals completed the recruitment screener; 279 qualified and were emailed a baseline assessment; 161 completed it but 30 were removed for denying active suicidal thoughts or responding inconsistently (e.g., several past-year suicide attempts but no lifetime attempts); 131 qualified for the study; and 115 downloaded the application. The first 5 participants were pilot subjects for feasibility testing and were excluded from analyses and another 5 failed to complete a single momentary (n = 3) or responded inconsistently (n = 2). A total of 105 participants were included in the final sample.

Distribution of time-intervals between consecutive measurements

Figure S1. shows the distribution of time-intervals between consecutive measurement occasions for 97.5% of the observations in the dataset (i.e. across all observations in all individuals). As we might expect, there is a strong clustering of observations between 0 and 2 hours, reflecting burst observations, and from 2 to 6 hours, reflecting the regular ESM measurements. Due to participants skipping observation waves and gaps for night we have a second cluster of observations around the 12 to 24 hour range. For the purpose of the visualization we omit the highest 2.75% of time-intervals since these range from 22.8 hrs to more than 19 days.

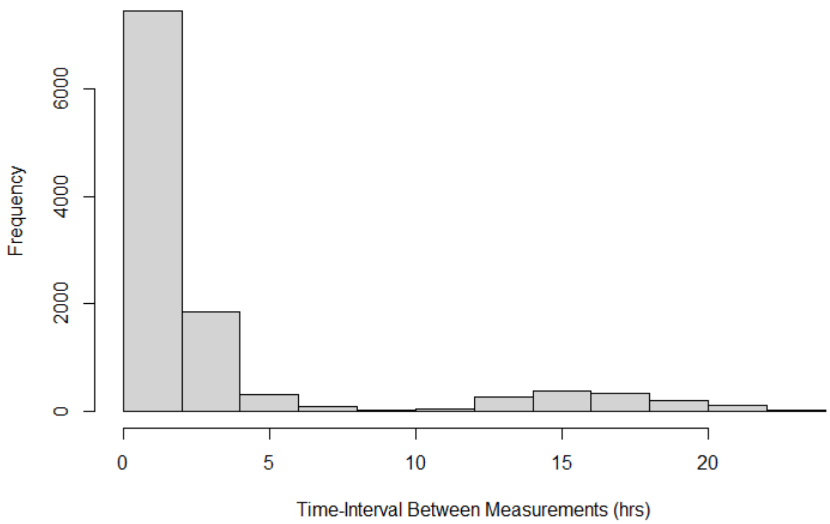
Data processing for timescale descriptive statistics

In Figure 2 we analyze the variation across different time-scale bins within individuals separately for Desire and Intent. For the Desire analysis, we omit the 10 individuals who exhibited low variance (<5% non-modal response). We additionally omit 9 individuals with very short time-series (< 14 observations) and one individual who showed a highly unusual response pattern (only responses very closely or very distantly spaced in time). This allows us to define timescale bins which are substantively interesting and for which all individuals have at least two observations in each bin. In total this yielded 85 individuals in the Desire analysis (20 subtracted from the total of 105). For the Intent we performed a similar procedure, with dropping the 42 individuals who had low variation in Intent. Omitting individuals with short time-series resulted in dropping an additional 5 participants from the analysis, giving us a total of 58 participants (47 subtracted from the total of 105). The relative low variation in Intent necessitates that a different number of individuals are used for the analysis of Desire and Intent separately.

Data processing for continuous-time models

For the continuous time models, we chose to use a subset of individuals who exhibited substantial variation in both Desire and Intent as discussed above and in the main text. We also omitted individuals who have a very short time-series, in this instance using a less strict cut-off of less than 10 total observations (omitting 8 individuals). We also identified an additional participant with an unusual response style who was omitted from the analysis: This participant was just above the mode threshold described above (with values ranging from 6 to 9% percent) and responded 258 times. However, after the first 42 observation points, the participants answered zero on both items for the remaining measurements. This led to a dataset with a total of 49 participants which we used to estimated the continuous time models in the main text.

S1. Distribution of time-intervals



S2. CT-MSM estimates

	low	**mild**	**severe**
low	**-0.052 (-0.06,-0.04)**	**0.052 (0.043, 0.062)**	**0**
mild	**0.175 (0.145, 0.213)**	**-0.530(-0.632,-0.454)**	**0.355 (0.281, 0.460)**
severe	**0**	**0.337 (0.265, 0.435)**	**-0.337 (-0.435,-0.265)**

Estimated intensity matrix with 95% CIs for the Desire variable

	low	**mild**	**severe**
low	**-0.057 (-0.072,-0.047)**	**0.057 (0.047, 0.072)**	**0**
mild	**0.501 (0.404, 0.620)**	**-0.869(-1.087,-0.721)**	**0.368 (0.250, 0.575)**
severe	**0**	**0.330 (0.223, 0.524)**	**-0.330 (-0.524,-0.223)**

Estimated intensity matrix with 95% CIs for the Intent variable