Multilevel Survival Analysis:

Studying the Timing of Children's Recurring Behaviors

Jessica P. Lougheed[1], Lizbeth Benson[2], Pamela M. Cole[2], and Nilam Ram[2,2]

[1]Purdue University

[2]The Pennsylvania State University

[3]German Institute for Economic Research (DIW), Berlin

Author Note

Jessica P. Lougheed, Department of Psychology, The Pennsylvania State University; Lizbeth Benson, Department of Human Development and Family Studies, The Pennsylvania State University; Pamela M. Cole, Department of Psychology, The Pennsylvania State University; Nilam Ram, Department of Human Development and Family Studies, The Pennsylvania State University.

We gratefully acknowledge support provided by a Banting postdoctoral fellowship from the Social Sciences and Humanities Research Council of Canada (Lougheed) and the NICHD (R01-HD076994; Cole and Ram).

Correspondence concerning this article should be addressed to Jessica P. Lougheed, Department of Human Development and Family Studies, Purdue University, 202 Fowler Memorial House, West Lafayette, Indiana, 47907. E-mail: jlougheed@purdue.edu

Abstract

The timing of events (e.g., how long it takes a child to exhibit a particular behavior) is often of interest in developmental science. Multilevel survival analysis (MSA) is useful for examining behavioral timing in observational studies (i.e., video recordings) of children's behavior. We illustrate how MSA can be used to answer two types of research questions. Specifically, using data from a study of 117 36-month-old (*SD* = .38) children during a frustration task, we examine the timing of 36-month-olds recurring anger expressions, and how it is related to: (1) negative affectivity, a dimension of temperament related to the ability to regulate emotions; and (2) children's strategy use (distraction, bids to their mother). Contrary to expectations, negative affectivity was not associated with the timing of children's recurring anger expressions. As expected, children's recurring anger expressions were less likely to occur in the seconds when children were using a distraction strategy, whereas they were more likely when children made bids to their mother. MSA is a flexible analytic technique that, when applied to observational data, can yield valuable insights into the dynamics of children's behaviors.

Keywords: survival analysis, observational data, emotion regulation

Multilevel Survival Analysis:

Studying the Timing of Children's Recurring Behaviors

The timing of behavior is of interest in many areas of developmental science. For example, emotion regulation has been conceptualized as a process of modulating emotion, including its timing (Thompson, 1994). Relatively stable between-child differences, such as temperament, and within-child dynamics, such as behavioral strategies, are both thought to contribute to children's emotion regulation. Negative affectivity (Rothbart & Posner, 2006) is a dimension of temperament associated with difficulties regulating anger during delayed reward situations in early childhood (Santucci et al., 2008; Tan, Armstrong, & Cole, 2013). In contrast, the use of distraction strategies should increase young children's ability to forestall the anger known to arise in delay task situations, i.e., influence the timing of anger expressions (Kopp, 1989). When confronted with a situation in which a goal is blocked, such as being required to wait for a desired object, distraction redirects attention from the blocked goal and should delay anger expressions (Gilliom, Shaw, Beck, Schonberg, & Lukon, 2002; Sethi, Mischel, Aber, Shoda, & Rodriguez, 2000). Other strategies, such as bidding to an adult about the problem, may keep children's attention focused on the blocked goal and actually hasten anger expressions (Sethi et al., 2000).

Investigating the timing of events can yield a description of emotion regulation as a within-child process and not merely as between-child differences in ability (Diaz & Eisenberg, 2015). Despite calls for investigations of children's emotion dynamics (Cole, Martin, & Dennis, 2004), most studies have not used statistical approaches that quantify emotion timing. Moment-to-moment changes in children's behavior are often coded in observational studies in ways that can capture how behaviors unfold over time. Often, however, these data are collapsed across time, creating summary variables (e.g., frequencies, averages) and using static statistical approaches (e.g., linear regression) that mask dynamic processes. Thus, there is a gap between conceptual questions about emotion regulation and the methods used to investigate them.

One such gap involves questions about how specific between-child or within-child factors relate to the timing of children's emotion expressions. Given the social significance of tolerating frustration, are some children disposed to express anger more readily than others? And, what types of strategies forestall children's anger expressions? Survival analysis, a technique for examining event timing (Allison, 1984), can be fruitfully applied to temporal data, namely micro-coded observational data—the products of quantifying *in situ* child behavior from video recordings—that are ubiquitous in developmental research. Survival analysis is well established in other fields of study but has not been used much in analysis of micro-coded behavioral observations (Fairbairn, 2016; Stoolmiller, 2016). Although we concentrate on this type of data here, researchers working with other types of data (e.g., questionnaire data concerning event timing) may also find survival analysis, and our introduction to it, useful.

We provide a practical introduction to multilevel survival analysis (MSA) for observational data using data from a study of emotion regulation in early childhood. In this study, 36-month-olds were observed while being required to wait for a desired object, a task often used to elicit children's anger (Cole, Teti, & Zahn-Waxler, 2003; Gilliom et al., 2002). We discuss how research questions about moment-to-moment behavioral processes can be tested using MSA, provide a general overview of survival analysis, and then illustrate how MSA is applied to observational data.

Mapping Developmental Theory to Method with Multilevel Survival Analysis

Developmental theory considers both between-child differences and within-child processes. Temperament refers to constitutionally-based dispositions – relatively stable, *between-child* differences – that are believed to underlie children's socioemotional development (Rothbart & Posner, 2006). Negative affectivity, one dimension of temperament, is related to the frequency and intensity of children's reactions to environmental changes. Children high in negative affectivity tend to experience negative emotions (e.g., frustration, anger, distress) and have difficulty regulating them (Calkins & Johnson, 1998; Rothbart & Posner, 2006). Research on young children during delay tasks has shown that negative affectivity is related to the frequency of children's anger expressions (e.g., Santucci et al., 2008). In line with dynamic perspectives of emotion regulation (Cole et al., 2004), high negative affectivity may be associated with recurring anger expressions during a frustrating situation.

Theories of emotion regulation typically adopt a dynamic perspective that prioritizes consideration of within-child processes (Cole, 2014; Thompson, 1994). In conceptualizing emotion regulation as a process, Thompson (1994) asserted that the latency to an emotion is one index of emotion regulation dynamics. Studies using such measures have shown that longer latencies until young children's distress during frustrating situations are associated with longer durations of regulatory behaviors (e.g., Calkins & Johnson, 1998). In addition, the latency to young children's anger expressions increases as children get older, whereas the durations of their anger expressions decrease with age (Cole et al., 2011). However, the use of these temporal variables does not reveal critical information, such as whether and how strategy use influences an emotion. Most models of emotion regulation have an implicit assumption that a strategy influences an emotion's "survival". That is, it is assumed, but as of yet untested, whether strategic efforts actually alter the lifespan of emotions. Kopp (1989) posited that children begin to use strategies during the third year of life. However, it is unclear that strategy use at this early age is actually effective (Buss & Goldsmith, 1998; Cole, Bendezú, Ram, & Chow, 2017).

Consider the frustration of waiting for something desirable that one cannot yet have. Young children may bid to a parent, e.g., verbalize about the situational demands (e.g., “I can open the surprise when you’re done, right?”). However, such statements can be said angrily and maintain focus on the frustrating aspects of the situation (Sethi et al., 2000) rather than forestall frustration (Cole et al., 2011). Around 36 months of age, children gain the ability to volitionally direct their attention away from a restricted item and become highly absorbed, at least temporarily, in a substitute activity. In general such attention shifting—strategic distraction—should reduce focus on the restricted item and the likelihood of expressing anger (Sethi et al., 2000). Alternatively, a child may seek adult support, bidding to a parent about the demands of waiting. The temporal relation between strategy use and emotion describes a *within-child* process—how the likelihood of expressing anger at a given moment is modified by the use of strategies such as bidding and distraction.

We focus our demonstration on how negative affectivity, and each of two putative strategies (bidding and focused distraction), are associated with the timing of children’s recurring anger expressions during a frustrating wait task. Looking across children, we examine whether negative affectivity is associated with a greater likelihood of recurring anger expressions. Looking within child, we examine whether bidding is associated with earlier anger expressions and distraction is associated with forestalled anger expressions. Before presenting the empirical example, we introduce survival analysis and its multilevel extension, including key concepts such as events; time; censoring; the hazard and survival functions; parametric, non-parametric and semi-parametric approaches; and the proportional hazards assumption.

Introduction to Survival Analysis

Survival analysis is a family of statistical techniques in which the outcome variable is the timing of an event—the time from a specific starting point such as the start of a task until a defined event occurs (Allison, 1984). The objective is to examine what predicts the timing of events. The terminology associated with survival analysis comes from its legacy in epidemiology, and research questions about the timing of events such as death. Individuals who experience an event during the observation period experience a “death,” whereas individuals who do not die “survive.” In the parlance of survival analysis, an individual’s likelihood of survival (i.e., not dying) may be associated with different predictor types, both time-invariant (e.g., gender) and time-varying (e.g., weekly alcohol consumption). In our empirical example, anger expressions are the event of interest. We examine the timing of children’s anger expressions with a multilevel approach, and how (time-invariant) negative affectivity and (time-varying) strategy use are related to it.

Key concepts. The outcome of interest in survival analysis is the timing of an *event*. An event is defined as a qualitative change in an individual’s state (e.g., from alive to dead) at an observable point in time (Allison, 1984). Event variables are often easily extracted from observational data (video recordings of participant behavior) collected in many developmental studies. For example, events can be defined as the onset or offset of specific behaviors or emotions. In our empirical example, wherein children’s anger expressions were coded as present/not present during each second of a 480-second task, we define events as the qualitative changes in a child’s emotion expression from a non-anger state to an anger state.

Events occur at specific points in *time*, with time operationalized as either a *discrete* or *continuous* variable. The specific operationalization depends on the sampling rate at which data are collected and/or coded (Singer & Willett, 2003). Observational data that are coded in discrete time intervals (e.g., 10- or 30-second epochs) imply implementation of a discrete time survival analysis. In contrast, data that are coded continuously with an event-based coding scheme (e.g., events could occur at any point in time, not just within discrete windows of time; see Stoolmiller & Snyder, 2006 for more details) imply implementation of a continuous time survival analysis. The distinctions between discrete and continuous time are not always clear, and the results of discrete- and continuous-time models will converge as the discrete-time sampling rate approaches a continuous scale (Efron, 1988; Singer & Willett, 2003). In our empirical example, children's behaviors were coded in 1-second epochs, which could be conceptualized as either a discrete or continuous time scale because the sampling rate consists of discrete, equally-spaced intervals but also provides a continuous record of observations (e.g., behaviors of interest typically last longer than the time unit, see Ram & Reeves, 2018). Our examples use time as continuous, but the procedures described here will be similar to other common coding schema (e.g., global coding of 10- or 30-second epochs).

Censoring is a common feature of time-to-event data and refers to cases in which event times are unknown because they did not occur within the observation period (Allison, 1984). *Right censoring* refers to cases when the event did not occur within the observation period; but may have occurred after the observation period ended. In our example, a child who did not express anger during the task would be right censored (even if the child expressed anger after the observation period ended). Right censoring is a common feature of time-to-event data and is not necessarily problematic in survival analysis given that the technique was developed in part to handle right-censored cases (Singer & Willett, 2003). As such, survival analysis is preferable to using conventional linear models predicting durations, which would not account for censoring, return biased estimates for event times, and predict impossible values such as negative durations.

Non-informative censoring refers to when censored and non-censored cases differ from each other only on the predictor variables included in the model, and is not considered problematic (Singer & Willett, 2003). Other types of censoring can, however, be problematic. Censoring is *informative* when censored cases differ systematically from non-censored cases on the risk of the event or on variables not included in the model. As a hypothetical example, we could imagine informative censoring if right-censored children (never expressed anger within the observation period) withdrew from the study because they experienced an extreme level of sadness. If this sadness was related to anger tendencies, the censoring would not be randomly distributed. Informative right censoring is considered problematic and will lead to biased analyses. Unfortunately, testing whether the censoring is informative is not possible if additional if predictors are not available, and little can be done to correct it when it occurs (Allison, 2010). *Left censoring* occurs when the start time of an event is unknown, such as when an event has occurred prior to or coincident with the start of the observation period. In contrast to right censoring, left censoring can be problematic in survival analysis (Singer & Willett, 2003). We will discuss this issue in more detail when working through the empirical example.

Two important statistical terms are the *hazard function* and the *survival function*. The *hazard function* describes the likelihood an individual will experience the event in time (Mills, 2011). Formally, the hazard function is expressed as

$$h_{ij}(t) = \Pr\left(T_{ij} = t \,\middle|\, T_{ij} \geq t\right) \tag{1}$$

where $h_{ij}(t)$ is the rate of the *j*th episode occurring at time interval *t* for individual *i*, given that the event has not yet occurred, and T_{ij} is the observed event time (Mills, 2011).

The hazard function provides the basis for deriving many other statistical quantities of interest (Stoolmiller, 2016). In particular, hazard functions are often reframed as *survival* *functions* to facilitate interpretation. Formally, the *survival function* is expressed as

$$S_{ij}(t) = \Pr(T_{ij} \geq t) \tag{2}$$

where $S_{ij}(t)$ is the proportion of individuals, i, whose event time for episode j, T_{ij}, is greater than time t (i.e., those who are still alive). $S_{ij}(t)$ is a non-increasing function that describes how, as time progresses, fewer and fewer individuals survive (Kleinbaum & Klein, 2012). Thus, whereas the hazard reflects the probability of experiencing the event at a given time among those who are still considered at risk for the event, the survival function reflects the cumulative "loss" from all those who were initially observed.

Survival analysis centers on examining the shape of the hazard function, and whether this function differs systematically across persons or time in relation to other between-person differences or within-person changes. These analyses can be conducted using *parametric, semi-parametric*, and *non-parametric* approaches. Parametric approaches are used when researchers expect that the hazard function has a specific shape (e.g., Gompertz, Weibull; see Mills, 2011 for more information). In contrast, non-parametric approaches are used when the shape of the hazard function is unknown. These approaches are free of assumptions about shape, purely descriptive, and well-suited for comparing survival functions among a small number of groups (Mills, 2011).

When examining how the hazard or survival function is related to between-person difference variables, researchers in social science typically use a semi-parametric Cox regression model (Cox, 1972). The Cox regression model does not involve assumptions about the shape of the hazard function (similar to a non-parametric approach) but does rely on a *proportional hazards assumption* that the log hazard is a linear, time-invariant (parametric) function of the predictors. The Cox proportional hazards model is sometimes referred to as an extended Cox model when conducted with time-varying predictors and/ or random effects (Therneau &

Grambsch, 2013).

In many cases, events can recur, in that the qualitative change in state can occur more than once. In such situations, events define episodes and the event time outcome variable indicates the timing since the last occurrence. To accommodate the fact that recurring events are clustered within individuals, survival analysis is placed in multilevel framework (MSA) that accommodates the potential interdependence between events within persons.

Our examples consider recurrent anger events that are nested within children. For example, if a child were observed for 8 minutes and expressed anger twice, then from the beginning of the observation period until the first anger expression would be considered the first episode, the time beginning after the end of the first anger expression to the second expression would be considered the second episode, and the time after the second expression to the end of the recorded observation being the third episode. For these data we designate individuals using *i*, episode as *j*, and time as *t*, thus t_{ij} represents the timing of the *j*th anger episode for individual *i*. The possibility that some individuals may be more at risk for the event than others is accommodated by the inclusion of a random effect that allows the hazard function to vary across individuals. The multilevel Cox regression model is specified as

$$h_{ij}(t) = h_0(t)\exp(v_i)\ \exp(\beta_1 x_{1i} + \beta_2 x_{2ij} + \beta_3 x_{3ij}(t) + \cdots \beta_q x_{qi}) \quad (3)$$

where the hazard of the *j*th episode at time interval *t* in individual *i* is the product of the baseline hazard $h_0(t)$, an exponentiated random effect v_i for individuals, and an exponentiated linear function of *q* predictors that may be time-invariant (e.g., x_{1i}) or time-varying (e.g., x_{2ij}, $x_{3ij}(t)$). The baseline hazard function is the shape of the hazard function when there is no influence of predictors, as in an unconditional model or when predictors have been centered at 0 (Kleinbaum & Klein, 2012). In our multilevel Cox model, the random effect is assumed to follow a gamma distribution with a mean of 1 and variance of θ, although other options are available (see Austin, 2017 for more discussion; Mills, 2011). Estimated random effects can be inspected after model fitting to evaluate the choice of distribution. Associations between the predictors and the hazard of anger expressions will be represented by β coefficients. Specifically, β_1 represents a coefficient for a time-invariant predictor, and β_2 and β_3 represent coefficients for time-varying predictors. In our empirical example, we will test if differences in the hazard of anger are related to fixed effects of children's (1) negative affectivity, and (2) use of bids and focused distractions, accounting for individual differences (i.e., random effects) in children's risk of recurring anger.

One advantage of the Cox model is that it is considered robust—results from a Cox model will closely approximate the results of a correctly specified parametric model, with the additional flexibility of being applicable to data for which the shape of the hazard function is unknown. Although the Cox model can be limiting because of the cost in statistical power associated with non-parametric estimation, the model can be informative about how event times are distributed in empirical data. Knowledge gained about the shape of the hazard function can inform theory development regarding the functional forms that manifest in different contexts, and open opportunity to test parametric functions in future studies.

Illustration of Multilevel Survival Analysis with Observational Data

Our motivating research questions are about whether children's negative affectivity and strategy use are related to the timing of recurring anger expressions during a situation in which children are required to wait for something they want—a situation in which, according to Western social norms, expressing anger is not desirable (Kopp, 1989). The timing of children's anger expressions during this type of situation is considered an indicator of self-regulation (Cole et al., 2011; Gilliom et al., 2002; Sethi et al., 2000). Between-child, dispositional differences in the reactivity of negative emotions—negative affectivity—are believed to be related to difficulties regulating anger (Rothbart & Posner, 2006). Within-child use of attentional control to turn attention away from a restricted object, with a strategy such as distraction, is believed to forestall anger. In contrast, behaviors that keep children's attention focused on the blocked goal, such as bidding to a caregiver about the demands of the wait, may foster anger (Sethi et al., 2000). To date, most studies examining children's emotion regulation have aggregated observational data across events and time, and examined the frequencies or durations of behaviors (e.g., Calkins & Johnson, 1998; Grolnick, Bridges, & Connell, 1996; Liebermann, Giesbrecht, & Müller, 2007; Santucci et al., 2008; Sethi et al., 2000). In our previous work, we examined temporal features of anger by quantifying its latency and duration and examining how these quantities were associated with temperament and strategy use (Cole et al., 2011; Tan et al., 2013). Taking a dynamic perspective on emotion regulation (e.g., Cole et al., 2004; Thompson, 1994) that emphasizes the unfolding of emotions *in situ*, we use MSA to extend our previous findings—directly testing how children's (1) negative affectivity, and (2) strategy use are related to the timing of their recurring anger expressions.

Relations between the timing of recurring anger expressions and children's temperament and strategy use can be conceptualized in terms of between-child and within-child associations, respectively. In the context of anger regulation, examinations of between-child differences address research questions about whether children with a particular disposition (e.g., high negative affectivity) differ from other children on the timing of recurring anger expressions. Examinations of within-child differences address research questions about how children's strategy use influences the timing of anger when children use specific strategies compared to when they do not. One advantage of MSA over examination of correlations between temperament or strategies and anger is that MSA provides a statistically rigorous way to simultaneously examine between-child and within-child associations.

Survival analysis can be used to examine both single and recurring events. A *single episode* model can be used for events that can occur only once (e.g., task completion, giving up). A single episode conceptualization captures between-child differences in event timing. In our examples, we conceptualize anger as a recurring event, as children can express anger repeatedly during the observational period. Examinations of recurring episodes in observational data are uncommon (for exceptions, see Dagne & Snyder, 2011; Lougheed, Hollenstein, Lichtwarck-Aschoff, & Granic, 2015; Snyder, Stoolmiller, Wilson, & Yamamoto, 2003) but may be particularly useful for understanding the dynamics of regulatory or other processes. One reason recurring-episode models are less common is that estimation of model parameters is more difficult computationally than estimation of the parameters in a single-episode models, with some of the difficulties arising from the utility of software options (Stoolmiller, 2016). Current software options for estimating these types of models include the *survival* (Therneau, 2015a) and *coxme* (Therneau, 2015b) packages in R (R Core Team, 2016), and Mplus (Muthén & Muthén, 2012). Both programs have advantages and disadvantages. R has the advantages of being freely available and open source, but with the disadvantages that technical support comes from the community of R users (and may be sporadic) and that packages are not always well documented. Mplus has the advantages of reliable technical support but it is not open source or freely available. Interested readers are referred to Austin (2017) for a detailed comparison of several programs.

To illustrate research questions that MSA can answer, we present two example analyses. With Model 1, we considered between-child differences in children's temperament by examining if (1) the time until recurring anger expressions was shorter for children with higher negative affectivity. Then, in Model 2, we considered within-child changes in strategy use by examining if children's anger expressions were: (2a) more likely in the seconds they made bids compared to the seconds they did not, and (2b) less likely in the seconds they distracted themselves compared to the seconds they did not. There are many possibilities afforded by survival analysis beyond the two research questions examined here. For example, researchers can use a single-episode approach to examine time-invariant and time-varying predictors of single events. Although a comprehensive demonstration of all possibilities is beyond the scope of this paper, interested readers are referred to online materials for examples of other model types (predicting a single episode from time-invariant and time-varying predictors; https://quantdev.ssri.psu.edu/), in addition to the examples covered in this article.

Participants

Data for our empirical illustrations are drawn from a longitudinal study of children's emotion regulation, wherein children and their caregivers visited the laboratory when children were ages 18, 24, 36, and 48 months. We used data collected when the children were age 36 months, when children are believed to develop the ability self-regulate emotions (Kopp, 1989). The analysis sample consisted of 117 children (64 boys, 53 girls) described by their mothers as White (92%) or biracial (8%). Complete demographic information and description of the larger study can be found in Cole et al. (2011). All procedures for the Development of Toddlers Study were approved by the Pennsylvania State University's institutional review board, IRB protocol numbers 18993 and 45013.

Procedure

We examined the timing of children's recurring anger expressions in the context of *the* *wait task* (Vaughn, Kopp, & Krakow, 1984), a task often used to study children's self-regulation when faced with a blocked goal (frustration). The child and mother were seated in an observation room. A research assistant provided the child with a boring toy (a toy car with no wheels) and the mother with questionnaires to fill out. The research assistant then placed a shiny gift-wrapped bag on the table and indicated that the gift was a surprise for the child. When the research assistant left the room, the mother (as instructed by the research assistant) told the child to wait to open the gift until she finished her work. All mothers complied with this instruction. The child's behavior during the task was videotaped for 8 minutes. Then, the research assistant returned and prompted the mother to let the child open the gift. Data from this task have been examined previously using other analytic techniques (Cole et al., 2011, 2017).

Measures

Observational data were derived from the videotapes. Children's nonverbal emotion expressions and strategies were coded on a second-by-second basis by two independent coding teams. Coders were trained to at least 80% agreement with master coders on the second-by-second coding prior to coding videos. Inter-rater reliability checks were conducted on a randomly selected set of 15% of the children in the sample for each coding system.

Anger events. *Anger events* were coded using a system based on facial expressions and vocal quality (Cole, Zahn-Waxler, & Smith, 1994). Anger intensity was coded for each second on a 4-point scale (0 = *not present*, 1 = *low intensity*, 2 = *moderate intensity*, 3 = *high intensity*). Inter-rater reliability was good, Cohen's κ = .86. Conceptualized as an event, anger intensity was recoded as a dichotomous variable that indicated whether anger had not (0 = 0) or had (1, 2, or 3 = 1) occurred in each second (see Anger_{it} under Original Data Structure in Table 1).

Episode and event time variables. Several variables are required for MSA. *Episode* indicates the time span for each anger event (see Episode$_{ij}$ in Table 1). The episode variable begins at one for each child and increases by 1 for each recurrence of the anger event. Thus, right-censored children (for whom anger never occurs) will have episode equal to 1 for their entire time series. *Event time* indicates elapsed time within each episode (see Event time$_{ijt}$ in Table 1).

Strategy use. Two strategies were of interest, *bids* to the mother about the demands of the wait (e.g., asking how much longer the wait was) and *focused distraction* that was initiated by the child and not done in a disruptive manner (i.e., becoming absorbed in an alternate activity, such as playing with the boring toy). Inter-rater reliability for strategies was good, Cohen's κ = .84. Like anger, bids and focused distraction were coded as dichotomous variables that indicated whether the strategy had not (=0) or had (=1) occurred in each second (see Bids$_{it}$ and Distraction$_{it}$ under Original Data Structure in Table 1). To examine within-child differences (Model 2), the occasion-specific, time-varying *bids* and *distraction* variables were used in their original form (see Bids$_{ijt}$ and Distract$_{ijt}$ under Model 2 in Table 1).

Negative affectivity. Child negative affectivity was measured with the negative affectivity factor score (Cronbach's α = .86) of the 105-item toddler behavior assessment questionnaire-revised (TBAQ-R, Goldsmith, 1996). Mothers completed the TBAQ-R at a home visit when children were 30 months old as part of the broader longitudinal study. Mothers rated items describing their child's behavior in the past two weeks using a 7-point scale (1 = *extremely untrue*; 7 = *extremely true*). For convenience of interpretation, the child-level negative affectivity variable was centered at its sample mean.

Empirical Illustration: Five Steps for Multilevel Survival Analysis

We now illustrate how MSA can be used to ask two types of research questions. We present MSA as a five-step process: (1) preliminary considerations, (2) data preparation, (3) data description, (4) model building and assessment of fit to data, and (5) presentation and interpretation of results. Tutorials walking through R code for each model are available online (https://quantdev.ssri.psu.edu/resources/survival-analysis).

Step 1: Preliminary considerations. Preliminary considerations include selecting a model that is appropriately matched to research questions and available data. Several decisions must be made. First, it must be decided whether the data will be analyzed with respect to discrete or continuous time. The data for our examples were coded in 1-second epochs, and we selected a continuous time approach.

Second, it must be decided if the data should be analyzed using a non-parametric, semi-parametric, or parametric approach. We decided to use a semi-parametric approach because we did not have theoretically-driven knowledge about the distribution of event times and because the Cox approach (Cox, 1972) allowed us to examine how both continuous time-invariant predictors and categorical time-varying predictors were associated with the timing of anger expressions.

Third, it must be decided how the outcome variable is best conceptualized—whether the research questions require modeling of single or recurring episodes. The use of single or recurring episode models should be determined by theory and research questions. With observational data, some events may only occur once (e.g., time to completing a task). Other events, such as emotions, may occur multiple times. Survival analysis affords many options, and single and recurring episodes need not be mutually exclusive. If researchers are interested in recurring events but the timing of the first event is also of interest, researchers could incorporate a first-episode effect (via a dummy variable) into a recurring-episode model. In our case, children expressed anger multiple times during the wait task, so we used recurring-episode models.

The fourth decision is to determine whether testing hypotheses requires the use of time-invariant or time-varying predictors (or both). Some questions are inherently focused on examinations of between-person differences (e.g., gender, ethnicity, traits), and suggest the use of time-invariant predictors. In Model 1, we use the child negative affectivity variable as a time-invariant predictor to examine if differences in child temperament are related to differences in the hazard of recurring anger expressions. Other questions focus on within-child changes, which suggest the use of time-varying predictors. In Model 2, we use time-varying predictors to examine how children's second-by-second changes in strategy use (bids and focused distraction) are related to the hazard of recurring anger expressions.

For illustration, we combine different decisions together to articulate two specific research questions. With Model 1, we examined if: (1) the time until recurring anger expressions was shorter for children with higher negative affectivity. Model 1, with episode *j* clustered within individual *i* is

$$h_{ij}(t) = h_0(t)\, exp(v_i) exp(\beta_1\, Negative\ Affectivity_i) \tag{4}$$

This equation states that the hazard of transitioning to anger for episode *j* within each child *i*, $h_{ij}(t)$, is a product of the baseline hazard function $h_0(t)$ for each episode; an exponentiated child-specific random effect (or "frailty") that indicates the extent to which children differ from each other on baseline risk of anger, v_i; and the exponentiated linear function of the time-invariant predictor, *Negative Affectivity*$_i$. Of greatest interest is the test of whether the β_1 parameter is different than 0 (i.e., that there is an association between negative affectivity and the hazard of recurring anger expressions).

With Model 2, we examined if children's recurring anger expressions were (2a) more likely in the seconds they made bids compared to the seconds they did not, and (2b) less likely in the seconds they distracted themselves compared to the seconds they did not. Model 2 tests the likelihood that children transition to anger in the same seconds that they make bids or engage in focused distraction,

$$h_{ij}(t) = h_0(t)\, exp(v_i) \exp(\beta_1 Bids_{ij}(t) + \beta_2\, Distract_{ij}(t)) \quad (5)$$

The hazard of transitioning to anger at time (t) is a product of the baseline hazard function $h_0(t)$ for each episode *j* within each child *i*; a child-specific random effect, v_i; and the exponentiated linear function of the two time-varying predictors ($Bids_{ij}$ and $Distract_{ij}$). Of primary interest to our research questions, the β_1 and β_2 parameters indicate the prototypical associations between time-varying use of bids and distraction, respectively, and the hazard of anger. Of note, these two time-varying behaviors were unlikely to co-occur and only did so less than .01% of the time. For this reason, we did not examine the interaction between bids and distraction, but it may be fruitful to test interactions among time-varying predictors when co-occurrences are present and of theoretical interest.

Step 2: Data preparation. Table 1 shows a hypothetical original data structure (raw data in long format) and data structures for each of the models. The original data structure consists of one row for each second of the wait task (t = 1 to 480) for each participant in the sample (N = 117). *ID* is the participant identification variable. *Second* indicates the temporal location in the task. *Anger*, *Bids*, and *Distraction* are binary variables that indicate whether the behavior was present or not in each second of the task.

Step 3: Data description. It is useful to examine some characteristics of the data before fitting survival models. For example, the data should be checked for right censoring—

participants who did not express anger before the end of the observation. In our sample, 12 of the 117 individuals (10% of the sample) were right censored. The low number of censored cases precluded comparisons of censored cases to non-censored cases on event risk, but this step should be performed in the presence of high numbers of censored cases to assess whether the censoring is informative and problematic. The data should also be checked for left censoring. We identified 7 left-censored children who were expressing anger when the observation period started. In the context of the study design, children who were expressing anger while the task started were likely not expressing anger because of the task. To overcome the presence of problematic left censoring, we defined the start time for all participants' observations as the first second at which anger was not occurring during the task, so that all children's first episode was defined as the first anger expression during the task. Another option is to remove left-censored cases from the data set (Singer & Willett, 2003). Researchers should also note that left censoring is less problematic in parametric models than it is in the semi-parametric Cox approach used here (Allison, 2010).

It is also important to examine plots of raw data and descriptive statistics. Figure 1 shows survival times for each child (length of horizontal lines) grouped by anger episode (different colors). The number of survival times (horizontal lines) per child is equal to their number of anger episodes. As seen by looking vertically up the graph, the time until each anger event appears to decrease, with fewer children represented at the later episodes. Descriptive statistics for all the variables used in Models 1 and 2 are shown in Table 2.

Step 4: Model building and assessment of fit to data. Model building often begins with fitting an unconditional model (no predictors included) to obtain the baseline hazard function. Then, predictors are added to test hypotheses, with model fit and diagnostics examined before interpreting the model results in Step 5.

We used the *coxph*() function of the *survival* package (Therneau, 2015a) in R to fit the models, specifying a random effect for children with a frailty term, and using the Efron approximation of the partial likelihood algorithm for estimation. The Efron algorithm was used because it is appropriate when there are multiple survival times with the same value (“ties”) in the data (Mills, 2011). First, an unconditional (baseline hazard) model was fit to the data to examine the baseline survival and cumulative hazard functions (see Figure 2). Then, we added negative affectivity into the model as a predictor to examine the proportional hazards assumption for Model 1. The proportional hazards assumption can be examined by obtaining the Schoenfeld residuals, which are the observed values of the predictors minus their predicted values at each event time (Mills, 2011). We then tested the assumption statistically using the *cox.zph* function, which creates an interaction between each predictor in the model and log-transformed time (Therneau, 2015a). Significant deviation of observed values from expected values of the predictors indicates non-proportional hazards (Mills, 2011). We concluded that the data for Model 1 did not violate the proportional hazards assumption (p = .78). If this assumption had been violated, there are several options. One option is to restrict the model to the range in which the assumption is not violated (Singer & Willett, 2003). A second option is to use time-varying predictors in place of the predictors that violate the assumption. This can be done either by creating interaction terms between the non-proportional predictors and time (Singer & Willett, 2003), or by using predictors that are truly time-varying. A third option is to break up the observation into periods of time based on when the hazard differs, so that the time-varying effects of the predictor can be examined for each time period (Muthén, Asparouhov, Boye, Hackshaw, & Naegeli, 2009).

Goodness of model fit can be assessed, for nested models, using a likelihood ratio test that compares fit of one model to another. This approach allows us to assess the explanatory power of predictors by adding fixed and random components to the model and comparing the model fit with this test (see Mills, 2011 for a full description of model building steps). In our examples, we compared the model including the predictors to the unconditional model, which is a joint test of the random effect variance and the fixed effect of negative affectivity. The likelihood ratio test was significant for Model 1, X^2 (90.14) = 414.70, $p < .001$, indicating that Model 1 fit the data better than the unconditional model. The likelihood ratio test comparing Model 2 and the unconditional model was also significant, X^2 (92.65) = 612.20, $p < .001$, indicating that Model 2 also fit the data better than the unconditional model. Relative fit of nested or non-nested models can be assessed using Akaike and Bayesian information criterion (Mills, 2011; Singer & Willett, 2003).

Step 5: Presentation and interpretation of results. We report our model results in the manner they might be presented within an empirical journal article. Results tables are formulated as typical regression tables, but often include an additional column where parameter estimates are transformed into the more easily interpreted hazard ratio metric (HR = exp[β]), which is, similar to the relative risk, an indication of effect size (Ressing, Blettner, & Klug, 2010). Specifically, hazard ratios can be interpreted as the magnitude of difference in the risk of event occurrence between two groups being compared. Similar to odds ratios: HR = 1.00 indicates no association between the predictor and outcome variable, HR > 1 indicates higher hazard of event occurrence for higher values of the predictor, and HR < 1 indicates lower hazard of event occurrence for higher values of the predictor (Mills, 2011). HRs can also interpreted as a percent change in hazard as 100 × [HR -1] (Mills, 2011).

Model 1: Recurring Episode Model with Time-Invariant Predictors. In Model 1, we tested the hypothesis that children with higher negative affectivity have shorter time until recurring anger expressions. The results of Model 1 are shown in Table 3. Contrary to expectations, differences in children's negative affectivity were not related to differences in the timing of children's recurring anger expressions ($\beta_1 = 0.01$, $p = .96$, HR = 1.01).

Model 2: Recurring Episode Model with Time-Varying Predictors. In Model 2, we tested if children's recurring anger expressions were (a) more likely in the seconds they made bids compared to the seconds they did not make bids, and (b) less likely in the seconds they distracted themselves compared to the seconds they did not distract themselves. The results of Model 2 are shown in Table 3. In line with expectations, children were 2.53 times (HR = 2.53) more likely (percent change = 100 × [2.53 - 1.00] = 153%) to express anger in the seconds that they bid compared to the seconds they did not. Also in line with expectations, children were 0.27 times as likely, or 73% less likely (percent change = 100 × [0.27 - 1.00] = -73%), to express anger in the seconds that they were engaged in focused distraction compared to the seconds they were not (HR = 0.27). Random effects can be interpreted by exponentiating their standard deviation, which indicates the relative risk of a child who is more (or less) "frail" than the prototypical child (e.g., at 1 *SD* above the sample mean; Therneau, 2015b). The variance of the random effect was 0.65 (see Table 3) and taking the square root of this value gives the standard deviation (*SD* = 0.81). This indicated that a "frail" child (+ 1 *SD*) above the sample mean on risk of anger had 2.25 = exp[0.81] times (95% CI [1.94, 2.60]) the risk of expressing anger than the average child. The extent of difference indicated a significant degree of variation in children's hazard of anger.

Discussion

MSA is a useful method for examining the timing of events. This method can be especially useful in testing theoretical perspectives of dynamic processes (e.g., Cole et al., 2004; Thompson, 1994) using time series data derived from video-recorded observations. To illustrate the utility of MSA, we used it to test predictions that 36-month-olds’ temperament and strategy use influence the timing of their anger expressions as they tolerate a frustrating wait for a desired gift.

This study is the first to show that 36-month-olds’ use of distraction, a strategy that is generally believed to be ideal when young children are faced with blocked or delayed rewards (Cole et al., 2011; Grolnick et al., 1996; Sethi et al., 2000), is associated with a decreased likelihood of anger expressions in the seconds that children engage in this strategy. In addition, we found that children’s use of bids, which is a strategy that keeps children’s attention on the blocked reward, increases the likelihood of anger expressions in the seconds that children engage in this strategy, which may be why anger is highly probable when this strategy is used. That is, the use of MSA provides the first evidence for a long-held view that attentional control is a central feature of early childhood emotion regulation, particularly in contexts involving blocked or delayed rewards (Kopp, 1989; Sethi et al., 2000), and that strategy use influences the “lifetime” of emotional expressions. This method builds on prior evidence about the latency and duration of young children’s anger expressions and use of distraction and bids (Cole et al., 2011), showing that strategy use influences the timing of anger expressions.

Contrary to expectations, we found that children’s negative affectivity was not associated with the timing of recurring anger expressions. We speculate that this null result could be for a number of reasons. One is that, although negative affectivity is considered to be a relatively stable between-child difference (Rothbart & Posner, 2006), there is some evidence of within-child changes during early childhood (Tan et al., 2013). Thus, the null result could be related to the 6-month difference in the timing of the temperament and observed anger measures. It could also be a Type II error due to lack of precision of estimated child anger hazard rates due to a short observation task (480 seconds), due to lack of re-test reliability for the child anger hazard rates, or both; or a true lack of effect.

Advantages of survival analysis. Survival analysis is well-suited to testing theoretical propositions about children's emotion regulation. MSA will enable researchers to go beyond linking behavioral tendencies (aggregated counts of observed strategy use and emotion expressions) with traditional approaches such as linear regression to examining the temporal influence of children's strategy use on their emotion expressions—one dynamic at the core of emotion regulation (Cole et al., 2004; Kopp, 1989). Example research questions that illustrate insights to be gained from MSA include: (1) Do the dynamics of children's strategy use and emotion expressions differ by child temperament? A cross-level interaction can be incorporated into MSA to test between-child differences in within-child processes. (2) Is the use of some strategies more effective for modulating emotion expressions than others, and do these associations vary by context? The temporal associations between strategies and emotion expressions can be contrasted with observations of children's behaviors in different contexts. (3) How does the effectiveness of children's strategy use develop? MSA can be extended to incorporate multiple time scales with repeated assessments of child behavior over years to examine how children develop the ability to regulate emotion expressions with strategies. MSA can also be used to move forward research on other topics for which behavioral timing is theoretically important, such as dyadic (e.g., parent-child) coercive processes (Granic & Lougheed, 2016; Lunkenheimer, Lichtwarck-Aschoff, Hollenstein, Kemp, & Granic, 2016), interpersonal anger regulation (Snyder et al., 2003), and parental scaffolding of adolescents' emotions (Lougheed, Craig, et al., 2016; Lougheed, Hollenstein, & Lewis, 2016).

The methodological advantages of survival analysis pertain to it being a flexible method. For example, it is flexible with respect to the time scale of event occurrence. We have demonstrated how MSA can be used at a short time scale (seconds), but it could also be used to examine development (e.g., the attainment of cognitive milestones or puberty) at longer time scales (e.g., months, years) using other kinds of data. When tasks are repeated at multiple ages, MSA can be used to examine processes that manifest at multiple time scales (Ram & Diehl, 2015). For example, Stoolmiller (2016) has used MSA to examine the interplay between short-term parent-child dynamics and the longer-term development of externalizing problems (Stoolmiller, 2016). Recently MSA has been embedded within broader structural equation models (McCurdy, Molinaro, & Pachter, 2017; Stoolmiller & Snyder, 2014; Wong, Zeng, & Lin, 2017). This extension allows the temporal associations between behaviors—hazards—to be used as both predictors and outcomes in longitudinal studies.

MSA can also be used to examine between-child differences in survival time in different ways. For example, researchers can examine between-child differences in event timing with single-episode models. Researchers can also compare groups on survival time (e.g., Morack, Ram, Fauth, & Gerstorf, 2013). In addition, between-child differences can be examined in terms of unmeasured, categorical differences within a mixture modeling framework if researchers expect meaningful classifications of participants based on survival time (Masyn, 2009).

Survival analysis is flexible regarding the number of event types that are examined. We examined the timing of one event—onset of anger—but the timing of multiple types of events (e.g., anger versus sadness) can be compared with competing hazard models (Stoolmiller &

Snyder, 2006). In competing hazards models, different event types (e.g., anger and sadness) each have their own hazard function.

Cautionary notes. A few issues with MSA should be noted. One issue is informative censoring, in which right-censored cases drop out before the end of the observation period in a non-random way that is related to the risk of the event or to predictors not included in the model. One recommendation to prevent informative censoring is to design studies to minimize all right censoring (Allison, 2010). Tasks could be designed so that their level of difficulty (e.g., in terms of length, emotional challenge) is sufficient to capture events but not so difficult that participants are likely to drop out.

Other challenges relate to the design of observational studies for survival analysis. Although there are no clear guidelines regarding the minimum base rates of events for survival analysis, observation periods need to be sufficiently long for the event to occur multiple times for at least some participants. Determining the required observation period will involve design issues such as how to elicit the event of interest—the number of events observed may be related to how evocative the task is in addition to its length. To assess whether the number of events is sufficient, researchers can compute the finite sample reliability of the individual-level hazard rates (see Stoolmiller, 2016). If the variance in the hazard rates at the individual level is significant and substantial, and if the reliabilities of the estimated random effect hazard rates at the individual level is high across all participants (e.g., .80 or higher), then it is a good indication that the task is revealing reliable individual differences in events. The issue of finite sampling reliability may be more of a concern in study designs using unstructured or naturalistic observations compared to structured lab tasks, because the observation length might be more likely to influence the number of observed events in unstructured observations.

Another issue related to study design is the consideration of re-test reliability of the behaviors being observed. Whenever possible, researchers considering using survival analysis with observational data should obtain a second, re-test observation of the same design several days after the initial observation. A thorough discussion of this topic can be found in Stoolmiller (2016), but a brief description of the issue is that participants may show low re-test reliability on their behaviors, which would influence how trustworthy the inferences to be drawn from model results are when relating hazards to child-level predictors or outcomes.

Modeling issues also include the assumption of proportional hazards for time-invariant predictors when survival analysis is conducted in the Cox framework (or with any linear model that uses the log hazard). With the uptake of MSA for observations of child behavior in the field using semi-parametric (e.g., Cox) and non-parametric approaches, we will be able to make informed decisions in applying parametric approaches to such data. Doing so will, in turn, inform theoretical perspectives on the functional form of how event times are distributed for different types of behaviors in different contexts. Parametric approaches are also advantageous because of the statistical power gained from using a parametric form of the baseline hazard function. We can make use of non-parametric and Cox regression models to explore what the distributions are while we move toward greater theoretical precision with the use of parametric approaches.

Conclusion

Developmental scientists often have research questions about the timing of events. We illustrated how MSA can be used to examine between-child (negative affectivity) and within-child (strategy use) predictors of recurring events (anger expressions). Our field is rich in observational data containing nuanced temporal information, and developmental theories are increasingly emphasizing dynamic processes in which timing plays a central role. MSA is a useful analytical technique to add to our research toolkit.

References

Allison, P. D. (1984). *Event history analysis: Regression for longitudinal event data*. Newbury Park, CA: SAGE.

Allison, P. D. (2010). Survival analysis. In G. R. Hancock & R. O. Mueller (Eds.), *The Reviewer's Guide to Quantitative Methods in the Social Sciences* (pp. 413–425). New York: Routledge.

Austin, P. C. (2017). A tutorial on multilevel survival analysis: Methods, models and applications. *International Statistical Review*, *85*(2), 185–203. https://doi.org/10.1111/insr.12214

Buss, K. A., & Goldsmith, H. H. (1998). Fear and anger regulation in infancy: Effects on the temporal dynamics of affective expression. *Child Development*, *69*(2), 359–374. https://doi.org/10.1111/j.1467-8624.1998.tb06195.x

Calkins, S. D., & Johnson, M. C. (1998). Toddler regulation of distress to frustrating events: Temperamental and maternal correlates. *Infant Behavior and Development*, *21*(3), 379–395. http://dx.doi.org/10.1016/S0163-6383(98)90015-7

Cole, P. M. (2014). Moving ahead in the study of the development of emotion regulation. *International Journal of Behavioral Development*, *38*(2), 203–207. https://doi.org/10.1177/0165025414522170

Cole, P. M., Bendezú, J. J., Ram, N., & Chow, S.-M. (2017). Dynamical systems modeling of early childhood self-regulation. *Emotion*, *17*(4), 684–699. https://doi.org/10.1037/emo0000268

Cole, P. M., LeDonne, E. N., & Tan, P. Z. (2013). A longitudinal examination of maternal emotions in relation to young children's developing self-regulation. *Parenting: Science* *and Practice*, *13*(2), 113–132. https://doi.org/10.1080/15295192.2012.709152

Cole, P. M., Martin, S. E., & Dennis, T. A. (2004). Emotion regulation as a scientific construct: Methodological challenges and directions for child development research. *Child Development*, *75*(2), 317–333. https://doi.org/10.1111/j.1467-8624.2004.00673.x

Cole, P. M., Tan, P. Z., Hall, S. E., Zhang, Y., Crnic, K. A., Blair, C. B., & Li, R. (2011). Developmental changes in anger expression and attention focus: Learning to wait. *Developmental Psychology*, *47*(4), 1078–1089. https://doi.org/10.1037/a0023813

Cole, P. M., Teti, L. O., & Zahn-Waxler, C. (2003). Mutual emotion regulation and the stability of conduct problems between preschool and early school age. *Development and Psychopathology*, *15*(1), 1–18. https://doi.org/10.1017/S0954579403000014

Cole, P. M., Zahn-Waxler, C., & Smith, D. K. (1994). Expressive control during a disappointment: Variations related to preschoolers' behavior problems. *Developmental Psychology*, *30*(6), 835–846. https://doi.org/10.1037/0012-1649.30.6.835

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *34*(2), 187–220.

Dagne, G. A., & Snyder, J. (2011). Relationship of maternal negative moods to child emotion regulation during family interaction. *Development and Psychopathology*, *23*(1), 211–223. https://doi.org/10.1017/S095457941000074X

Diaz, A., & Eisenberg, N. (2015). The process of emotion regulation is different from individual differences in emotion regulation: Conceptual arguments and a focus on individual differences. *Psychological Inquiry*, *26*(1), 37–47. https://doi.org/10.1080/1047840X.2015.959094

Efron, B. (1988). Logistic regression, survival analysis, and the Kaplan-Meier curve. *Journal of* *the American Statistical Association*, *83*(402), 414–425. https://doi.org/10.2307/2288857

Fairbairn, C. E. (2016). A nested frailty survival approach for analyzing small group behavioral observation data. *Small Group Research*, *47*(3), 303–332. https://doi.org/10.1177/1046496416648778

Gilliom, M., Shaw, D. S., Beck, J. E., Schonberg, M. A., & Lukon, J. L. (2002). Anger regulation in disadvantaged preschool boys: Strategies, antecedents, and the development of self-control. *Developmental Psychology*, *38*(2), 222–235. https://doi.org/10.1037//0012-1649.38.2.222

Goldsmith, H. H. (1996). Studying temperament via construction of the toddler behavior assessment questionnaire. *Child Development*, *67*(1), 218–235. https://doi.org/10.1111/j.1467-8624.1996.tb01730.x

Granic, I., & Lougheed, J. P. (2016). The role of anxiety in coercive family processes with aggressive children. In T. J. Dishion & J. Snyder (Eds.), *The Oxford Handbook of Coercive Relationship Dynamics* (pp. 231–248). Oxford, England: Oxford University Press.

Grolnick, W. S., Bridges, L. J., & Connell, J. P. (1996). Emotion regulation in two-year-olds: Strategies and emotional expression in four contexts. *Child Development*, *67*(3), 928–941. https://doi.org/10.1111/j.1467-8624.1996.tb01774.x

Kleinbaum, D. G., & Klein, M. (2012). *Survival analysis: A self-learning text*. Springer Science & Business Media.

Kopp, C. B. (1989). Regulation of distress and negative emotions: A developmental view. *Developmental Psychology*, *25*(3), 343–354. https://doi.org/10.1037/0012-1649.25.3.343

Liebermann, D., Giesbrecht, G. F., & Müller, U. (2007). Cognitive and emotional aspects of self-regulation in preschoolers. *Cognitive Development*, *22*(4), 511–529. https://doi.org/10.1016/j.cogdev.2007.08.005

Lougheed, J. P., Craig, W. M., Pepler, D., Connolly, J., O'Hara, A., Granic, I., & Hollenstein, T. (2016). Maternal and peer regulation of adolescent emotion: Associations with depressive symptoms. *Journal of Abnormal Child Psychology*, *44*(5), 963–974. https://doi.org/10.1007/s10802-015-0084-x

Lougheed, J. P., Hollenstein, T., & Lewis, M. D. (2016). Maternal regulation of daughters' emotion during conflicts from early to mid-adolescence. *Journal of Research on Adolescence*, *26*(3), 610–616. https://doi.org/10.1111/jora.12211

Lougheed, J. P., Hollenstein, T., Lichtwarck-Aschoff, A., & Granic, I. (2015). Maternal regulation of child affect in externalizing and typically-developing children. *Journal of Family Psychology*, *29*(1), 10–19. https://doi.org/10.1037/a0038429

Lunkenheimer, E. S., Lichtwarck-Aschoff, A., Hollenstein, T., Kemp, C. J., & Granic, I. (2016). Breaking down the coercive cycle: How parent and child risk factors influence real-time variability in parental responses to child misbehavior. *Parenting: Science and Practice*, *16*(4), 237–256. https://doi.org/10.1080/15295192.2016.1184925

Masyn, K. (2009). Discrete-time survival factor mixture analysis for low-frequency recurrent event histories. *Research in Human Development*, *6*(2–3), 165–194. https://doi.org/10.1080/15427600902911270

McCurdy, S. R., Molinaro, A., & Pachter, L. (2017). A latent variable model for survival time prediction with censoring and diverse covariates. *ArXiv:1706.06995 [Stat]*. Retrieved from http://arxiv.org/abs/1706.06995

Mills, M. (2011). *Introducing survival and event history analysis*. London: Sage Publications

Ltd.

Morack, J., Ram, N., Fauth, E. B., & Gerstorf, D. (2013). Multidomain trajectories of psychological functioning in old age: A longitudinal perspective on (uneven) successful aging. *Developmental Psychology*, *49*(12), 2309.

Muthén, B. O., Asparouhov, T., Boye, M., Hackshaw, M., & Naegeli, A. (2009). *Applications of continuous-time survival in latent variable models for the analysis of oncology randomized clinical trial data using Mplus*. Retrieved from https://www.statmodel.com/download/lilyFinalReportV6.pdf

Muthén, L. K., & Muthén, B. O. (2012). Mplus user's guide (Version 7). Los Angeles: Muthén & Muthén.

R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Ram, N., & Diehl, M. (2015). Multiple time-scale design and analysis: Pushing towards real-time modeling of complex developmental processes. In M. Diehl, K. Hooker, & M. J. Sliwinski (Eds.), *Handbook of intraindividual variability across the lifespan* (pp. 308–323). New York: Routledge.

Ram, N., & Reeves, B. (2018). Time sampling. In M. H. Bornstein, M. E. Arterberry, K. L. Fingerman, & J. E. Lansford (Eds.), *Encyclopedia of Lifespan Human Development* (pp. 2247–2258). Thousand Oaks, CA: Sage Publications Ltd.

Ressing, M., Blettner, M., & Klug, S. J. (2010). Data analysis of epidemiological studies. *Deutsches Arzteblatt International*, *107*(11), 187–192. https://doi.org/10.3238/arztebl.2010.0187

Rothbart, M. K., & Posner, M. I. (2006). Temperament, attention, and developmental psychopathology. In D. Cicchetti & D. J. Cohen (Eds.), *Handbook of Developmental psychopathology* (Vol. 2, pp. 465–501).

Santucci, A. K., Silk, J. S., Shaw, D. S., Gentzler, A., Fox, N. A., & Kovacs, M. (2008). Vagal tone and temperament as predictors of emotion regulation strategies in young children. *Developmental Psychobiology*, *50*(3), 205–216. https://doi.org/10.1002/dev.20283

Sethi, A., Mischel, W., Aber, J. L., Shoda, Y., & Rodriguez, M. L. (2000). The role of strategic attention deployment in development of self-regulation: Predicting preschoolers' delay of gratification from mother-toddler interactions. *Developmental Psychology*, *36*(6), 767–777.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford: Oxford University Press.

Snyder, J., Stoolmiller, M., Wilson, M., & Yamamoto, M. (2003). Child anger regulation, parental responses to children's anger displays, and early child antisocial behavior. *Social Development*, *12*(3), 335–360. https://doi.org/10.1111/1467-9507.00237

Stoolmiller, M. (2016). An introduction to using multivariate multilevel survival analysis to study coercive family process. In T. J. Dishion & J. J. Snyder (Eds.), *The Oxford handbook of coercive relationship dynamics* (pp. 363–378). New York: Oxford University Press.

Stoolmiller, M., & Snyder, J. (2006). Modeling heterogeneity in social interaction processes using multilevel survival analysis. *Psychological Methods*, *11*(2), 164–177. https://doi.org/10.1037/1082-989X.11.2.164

Stoolmiller, M., & Snyder, J. (2014). Embedding multilevel survival analysis of dyadic social interaction in structural equation models: Hazard rates as both outcomes and predictors. *Journal of Pediatric Psychology*, *39*(2), 222–232. https://doi.org/10.1093/jpepsy/jst076

Tan, P. Z., Armstrong, L. M., & Cole, P. M. (2013). Relations between temperament and anger regulation over early childhood. *Social Development*, *22*(4), 755–772. https://doi.org/10.1111/j.1467-9507.2012.00674.x

Therneau, T. M. (2015a). A package for survival analysis in S (Version 2.38). Retrieved from http://CRAN.R-project.org/package=survival

Therneau, T. M. (2015b). coxme: Mixed effects Cox models (Version 2.2-5). Retrieved from https://CRAN.R-project.org/package=coxme

Therneau, T. M., & Grambsch, P. M. (2013). *Modeling Survival Data: Extending the Cox Model*. Springer Science & Business Media.

Thompson, R. A. (1994). Emotion regulation: A theme in search of definition. *Monographs of the Society for Research in Child Development*, *59*(2–3), 25–52. https://doi.org/10.2307/1166137

Vaughn, B. E., Kopp, C. B., & Krakow, J. B. (1984). The emergence and consolidation of self-control from eighteen to thirty months of age: Normative trends and individual differences. *Child Development*, *55*(3), 990–1004. https://doi.org/10.2307/1130151

Wong, K. Y., Zeng, D., & Lin, D. Y. (2017). Efficient estimation for semiparametric structural equation models with censored data. *Journal of the American Statistical Association*, 0–0. https://doi.org/10.1080/01621459.2017.1299626

Table 1

Data Structure Examples

Original Data Structure

ID_i	$Second_{ijt}$	$Anger_{ijt}$	$Bids_{ijt}$	$Distraction_{ijt}$	Negative $Affectivity_i$
1	1	0	0	1	4.37
1	2	0	1	1	4.37
1	3	1	1	0	4.37
1	4	0	0	0	4.37
1	5	0	0	0	4.37
1	6	0	0	1	4.37
1	7	0	0	1	4.37
1	8	0	0	1	4.37
1	9	0	1	0	4.37
1	10	1	1	0	4.37
…	...	...	...	...	…
1	480	0	0	0	4.37
2	1	0	0	0	5.16
2	2	0	1	0	5.16
…	…	…	…	…	…
117	480	1	0	0	3.98

Model 1: Recurring Episode Model with Time-Invariant Predictor

ID_i	$Second_{ij}$	Anger $Event_{ij}$	Negative $Affectivity_i$	$Episode_{ij}$	Event $Time_{ijt}$
1	3	1	4.37	1	3
1	10	1	4.37	2	7
1	480	0	4.37	3	470
…	…	…	…	…	…
117	480	1	3.98	4	50

Model 2: Recurring Episode Model with Time-Varying Predictors

ID_i	$Second_{ijt}$	Anger $Event_{ijt}$	$Bids_{ijt}$	$Distract_{ijt}$	$Start_{ijt}$	$Stop_{ijt}$
1	1	0	0	1	0	1
1	2	0	1	1	1	2
1	3	1	1	0	2	3
1	4	0	1	0	3	4
…	…	…	…	…	…	…
117	480	1	0	0	479	480

Note. Subscripts refer to variation over individuals (*i*), episodes (*j*), and time periods (*t*).

Table 2
Descriptive Statistics for both Models

Predictor Variables	Mean (*SD*)	Minimum	Maximum	Outcome Variables	Number of Episodes		
					Mean (*SD*)	Minimum	Maximum
Model 1: Recurring Episode Model with Time-Invariant Predictor							
Negative Affectivity (1 to 7)	3.60 (.58)	2.27	4.96	Anger (recurring)	8.26 (8.38)	1	43
Model 2: Recurring Episode Model with Time-Varying Predictors							
Bids	34.74 (24.43)	0	157	Anger (recurring)	8.26 (8.38)	1	43
Focused distraction	101.75 (84.41)	0	321				

Note. N = 117. *Bids* and *Focused distraction* are represented as the number of seconds the behavior was present during the 480-second task.

Table 3
Results from Model 1 and Model 2

Predictor	Estimate	Standard Error	p	Hazard Ratio	95% Confidence Interval of Hazard Ratio
Model 1: Recurring Episode Model with Time-Invariant Predictor					
Negative Affectivity, β_1	0.01	0.18	.96	1.01	[0.71, 1.43]
Child-level random effect variance, v_i	0.80				
Log-Likelihood (Fitted)	-4767.56				
Log-Likelihood (Unconditional)	-4974.91				
AIC	9715.39				
Model 2: Recurring Episode Model with Time-Varying Predictors					
Bids, β_1	0.93	0.09	<.001	2.53	[2.14, 2.99]
Distractions, β_2	-1.32	0.17	<.001	0.27	[0.19, 0.37]
Child-level random effect variance, v_i	0.65				
Log-Likelihood (Fitted)	-4877.07				
Log-Likelihood (Unconditional)	-5183.16				
AIC	9939.44				

Note. AIC= Akaike Information Criterion. Model 1 N = 117 persons. Model 1 Likelihood ratio test: X^2 (90.14) = 414.70, p < .001. Model 2 N = 480 seconds nested within 117 persons. Model 2 Likelihood ratio test: X^2 (92.65) = 612.20, p < .001.

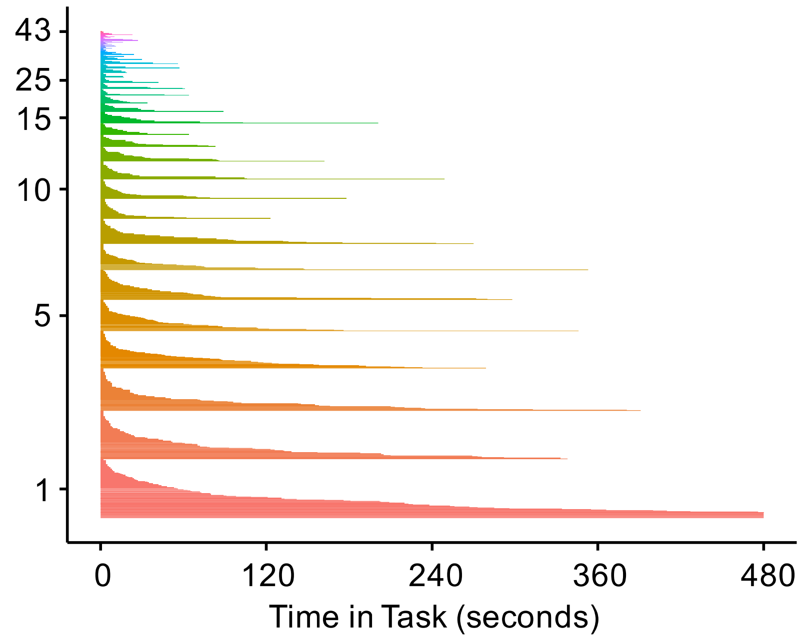

Figure 1. Survival times for each child for each episode of anger.

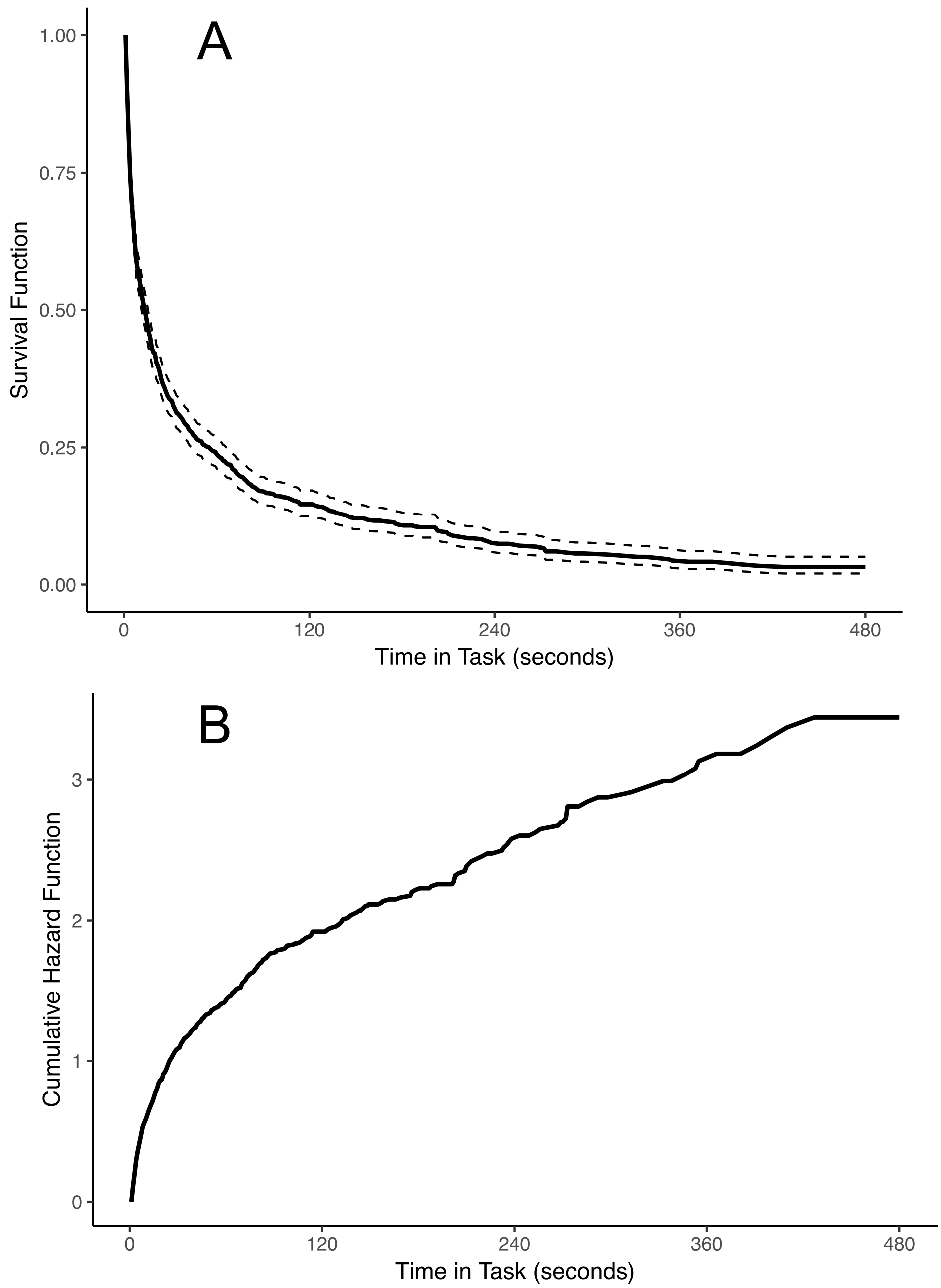

Figure 2. (A) Baseline survival function and (B) cumulative hazard function.