Journal of Behavioral Education, Vol. 13, No. 4, December 2004 (© 2004), pp. 227–246

Empirical Guidance for Time-Window Sequential Analysis of Single Cases

Paul J. Yoder, PhD,[1,2] and Jon Tapp[1]

Time-window sequential analyses test whether a target behavior occurs within a temporal window (e.g., within 2 seconds) after an antecedent behavior more than is expected by chance. This type of question is common when we need to know how one person or event may immediately affect another event or person in the natural environment. Theoretically, the significance of sequential associations from time-window analysis can be tested on the single subject level (Bakeman & Quera, 1995). The present Monte Carlo study was conducted to test the Type I error rates and the difference in sequential associations derived from four methods of time-window sequential analysis. The four methods vary according to whether they analyze the duration of antecedent and target behaviors. The results indicate that time-window sequential analysis method is generally valid. The results were most accurate when antecedent duration and target onset was analyzed. Although analyzing duration of the antecedent did affect the results, the effect size for the difference in results due to presence or absence of measuring duration of the antecedent was extremely small. Time-window analysis results appear unaffected by the decision to analyze the duration of the target event.

KEY WORDS: sequential analysis, observational measurement, time-window analysis.

This paper will give the rationale for, the description of, and a test of the Type I error for time-window sequential analysis (Bakeman & Quera, 1995). This type of sequential analysis addresses questions such as "Does teacher instruction increase the probability that student self-injury will occur within 2 seconds after teacher instruction?" We will present the results of a Monte Carlo study that estimates the accuracy of time-window sequential analyses when research questions are

[1]Vanderbilt University, Nashville, Tennessee.
[2]Correspondence should be addressed to Paul J. Yoder, PhD, Department of Special Education, Box 328 Peabody College, Vanderbilt University, Nashville, Tennessee 37203; e-mail: paul.yoder@vanderbilt.edu.

addressed at the single subject level. Additionally, this simulation study examines if results vary by whether duration of the antecedent and target events are analyzed.

Sequential analysis is the examination of whether one behavior (e.g., teacher instruction) increases or decreases the probability of another behavior (e.g., self-injury) occurring within a specified number of coded behaviors or time units. In the sequential analysis literature we use the term "antecedent" to refer to the "hypothesized causal behavior" and "target" to refer to the "hypothesized affected behavior." However, we do not mean to imply that a relation between the two behaviors is assumed *a priori*. The association between antecedent and target behavior is what is being tested in sequential analysis.

Because sequential analysis requires the investigator to be so specific about *when* the target event is expected to occur after the antecedent, the number of alternative explanations for the association of interest is smaller than with summary-level associations. An example of a summary-level association addressing a similar research question to our example is the association between the frequency teacher instruction and the frequency of student self-injury. This ruling out of alternative explanations for the association of interest is what makes sequential analysis a useful tool in identifying potential causal or functional relationships. However, only by altering the frequency of antecedents in the context of internally valid experiments can we infer that antecedents cause targets to occur more (or less) often. Therefore, sequential analyses are best used to identify potential causal behaviors that are subsequently tested in the context of internally valid experiments.

The aspect of our theory that predicts that the antecedent has an *immediate effect* on the target is what motivates our choice to use sequential analysis. A critical attribute of sequential analysis is that we *specify* what we mean by "immediate." Immediacy is defined by the expected number of (a) time units (e.g., 1 second later), (b) coded behaviors (e.g., the next coded behavior), or (c) window of time units (e.g., within the next 2 seconds) between the antecedent behavior and target behaviors. In fact, the three major types of sequential analysis are distinguished by whether they count time units or coded events between antecedent and target behaviors and by how specific they are about how many of these units are between the two behaviors. For example, "event sequential analysis" (using lag 1) would frame the example research question as "Does teacher instruction increase the probability that self-injury will occur as *the next coded behavior*?" "Timed lag sequential analysis" (using lag 1) would frame the example research question as "Does teacher instruction increase the probability that self-injury will occur in *the next second*?" "Time-window sequential analysis" (using a 2 second window) would frame the question, "Does teacher instruction increase the probability that self-injury will occur *sometime in the next two seconds*?" Detailed explanations of event-lag sequential analysis and timed lag sequential analysis are available elsewhere (Bakeman & Gottman, 1997; Yoder, Short-Meyerson, & Tapp, 2004). Because, time-window sequential analysis is relatively new to the literature and because no empirical work has been directed at guiding decisions regarding this approach, the remainder of the paper will focus on time-window sequential analysis.

Time-window sequential analysis is a relative new-comer to the scene. It was first described in Bakeman and Quera (1995). It has some potential advantages over event-lag and timed-lag sequential analyses. For example, time-window sequential analysis allows less precision than the timed- or event- lag sequential analysis in the prediction of the exact time or number of events that the target (e.g., self-injury) is expected to occur after the antecedent (e.g., teacher instruction). Therefore, time-window sequential analysis frequently matches the complexity of human behavior and our limited state of knowledge of human interactions better than timed- or event- lag sequential analysis. Second, the time-window analysis does not require that the investigator "correctly" identify all of the behaviors that are relevant to code to correctly conduct the analysis. Event-lag sequential analysis does require making this type of decision. To read more about this issue, the reader is referred to Yoder, et al. (2004).

Because all sequential analyses, including time-window sequential analysis, are correlational techniques, it is helpful to state our research questions in correlational terms. For example, we might ask whether there is a sequential association between teacher instruction (i.e., the antecedent) and child self-injury (i.e., the target). When we say "sequential association," we mean that the target follows the antecedent more than one would expect by a random sequencing of behaviors or time units (i.e., more than chance).

Accounting for Chance Occurrences of the Sequence of Interest

It is important compare the *observed* number of times a target follows an antecedent with the *expected* number of times the target follows the antecedent because we expect a certain number of such sequences to occur by random processes. The more instances of the antecedent and target behaviors, the more instances of the sequence of interest we expect by chance.

Two variables that are derived from sequential analysis of behavior do not control for chance. These two indices are sequential frequency and transitional probabilities. The sequential frequency is the number of times the target occurs within a time window (e.g., 2 seconds) of the antecedent. In our example, the sequential frequency is the number of times the child's self-injury occurred with 2 seconds of the teacher's instructions. The sequential frequency divided by the total time units within antecedent time windows is the "transitional probability". In our example, the transitional probability is the sequential frequency of self-injurious behavior within two seconds of the teacher's instructions divided by the total number of seconds in 2-second antecedent time windows.

An example illustrates why sequential frequency and transitional probabilities are not good indices of sequential association in most situations. Suppose that two children, Joe and Lisa, injure themselves within two seconds after adult instructions 5 times and 10 times, respectively. That is, their sequential frequencies are 5 and 10. It is clear that these sequential frequencies do not reflect the sequential association of self-injury after teacher instruction when one hears that Joe's adult used half as many instructions as Lisa's adult. This different frequency of instruction-giving resulted in their being 10 and 20 seconds that are within the 2-second window after adult instructions in Joe's and Lisa's data, respectively. That is, even though Lisa's sequential frequency is twice that of Joe's, the transitional probability of interest is the same for Joe and Lisa (i.e., 5/10, 10/20, respectively). However, it is clear that transitional probabilities are inadequate as indices of sequential association when one hears the new information that 5/6 of Joe's self injury instances occurred after teacher instruction; while, only 1/8 (i.e., 10/80) of Lisa's self injury occurred after teacher instruction. Given that Lisa and Joe's transitional probabilities of self-injury after teacher instruction are equal and that Lisa's self-injury is 13 (80/6 = 13) times more frequent than Joe's, it is clearly more likely that teacher instructions elicit self-injury in Joe than in Lisa. Sequential frequency and transitional probabilities are not adequate indices of sequential association because they do not control for the base rates of the antecedent and target behaviors (Bakeman & Gottman, 1997; Yoder & Feurer, 2000). To help organize our discussion and thinking about controlling for the base rates of antecedent and target behaviors, it is useful to introduce the concept of 2 × 2 contingency tables.

Contingency Tables

To see why contingency tables are so useful, it will be helpful to illustrate a short segment of coded data and to indicate how these data are organized into a 2 × 2 contingency table. The reader is referred to Table I to illustrate one way that

Table I. Timed-Event Data With Only Onset Coded

Time	Teacher behavior	Student Behavior
00:01	Other	Other
00:02	Instruction Onset	Other
00:03	Other	Other
00:04	Other	Self-Injury Onset
00:05	Other	Other
00:06	Other	Other
00:07	Other	Other
00:08	Other	Self-Injury Onset
00:09	Instruction Onset	Other
00:10	Other	Other
00:11	Other	Other
00:12	Other	Self-Injury Onset
00:13	Other	Other

Table II. Timed-Event Data for Behavior Onsets, 2-Second Time Window and Cell Addresses Using the Onset-Onset Method Indicated

Time	Teacher behavior	Student Behavior	Cell Address
00:01	Other	Other	D
00:02	Instruction Onset	Other	B
00:03	Other (Window)	Other	B
00:04	Other (Window)	Self-Injury Onset	A
00:05	Other	Other	D
00:06	Other	Other	D
00:07	Other	Other	D
00:08	Other	Self-Injury Onset	C
00:09	Instruction Onset	Other	B
00:10	Other (Window)	Other	B
00:11	Other (Window)	Other	B
00:12	Other	Self-Injury Onset	C
00:13	Other	Other	D

data that is appropriate for time-window sequential analysis may look. These data are called "timed-event data" because they include not only a code representing what type of behavior that occurred (i.e., the event) but also the event's time of occurrence (Bakeman & Gottman, 1997). For simplicity's sake, the data in Table I only includes the onset of the antecedent and target behaviors. Duration of these behaviors is ignored for the time being. We call this method the "Onset-Onset" method of analysis. We will return to the issue of whether duration *should* be considered later in this paper.

Table II indicates how these timed-event data in Table I are recoded for the Onset-Onset method of time-window sequential analysis. Time-window sequential analysis requires some operation that is *equivalent* to recoding or actual recoding so that all time units are analyzed once and only once. The math of the analysis requires this condition to be met.

Table III indicates how the recoded data is tallied into a 2 × 2 table using the Onset-Onset method of time-window sequential analysis. A convention used in the sequential analysis literature is to use the rows to classify presence or absence of antecedents and to use columns to indicate presence or absence of targets. Applied to time-window analysis, rows are used to categorize seconds coded for the antecedent actor (e.g., teacher) and columns are used to categorize seconds coded for the target actor (e.g., student).

Table III. Onset-Onset Method Contingency Table From the Data in Table II

	Student			
Teacher	Student self-injury		Student other	
Inside of the 2 second window after teacher instruction onset	1 second	A	B	5 seconds
Outside of the 2 second window after teacher instruction onset	2 seconds	C	D	5 seconds

The 1st row is for seconds coded for the antecedent actor that are within the time window after the antecedent (e.g., 2 seconds after the onset of teacher instructions). In time-window sequential analysis, it should be noted that the base rate of the antecedent is the number of time units *within the specified time window*, not the number of antecedent events. The 1st column in time window 2×2 contingency tables is for seconds coded for the target actor in which the target behavior occurred (e.g., student self-injury). The base rate of the target event is the total number of time units in which a target event occurred. The 2nd row is for seconds coded for the antecedent actor that are not within the antecedent time window (i.e., seconds that are more than 2 seconds after the most recent antecedent). The 2nd column is for seconds coded for the target actor that are not the target behavior (e.g., student other).

The cell labels (A–D) in Table II are for the four cells in the 2×2 table in Table III. The cell label indicated in Table II is for the recoded data represented in that figure. The A cell indicates the sequence of interest (e.g., the number of seconds in which student self injury occurred within 2 seconds of teacher instruction). The B cell indicates all seconds that are within 2 seconds of the antecedent at which the target does not occur. The C cell indicates all seconds at which the target occurs outside of the 2-second antecedent time. The D cell indicates all seconds at which neither the 2-second antecedent time window or target behavior occurs.

Each of these cells can be related to the various concepts we have discussed so far. The sequential frequency is the value in cell A. The base rate of the antecedent time window is the sum of cells A and B. The transitional probability is (cell A/(cell A + cell B)). The base rate of the target event is the sum of cells A and C. Finally, the total observation time is the sum of all four cells.

Should Duration of the Antecedent or Target Event be Analyzed?

A critical question in time-window sequential analysis is whether to collect and analyze duration information for the antecedent and target events. To date, published literature does not inform us whether analyzing only the onset of the antecedent and/or target behaviors bias the results. Because we have no empirical guidance on this question yet, it is tempting to say that the way timed-event data are analyzed should depend on whether the motivating theory behind the research question considers the duration information important to addressing the research question. More specifically, one might say that if the theory indicates that the duration of the antecedent influences the probability of the target occurring, then duration of the antecedent information should be analyze. Similarly, one might say that if the theory indicates that the duration of the target influences the probability that the target will occur after the antecedent, then duration of the target information should be analyzed. However, using theory in this way may not be wise.

For one thing, our theories are often not sufficiently specific to offer clear guidance on this matter. Because we lack sufficient information, often we can derive convincing theories that justify both analyzing and not analyzing duration of the antecedent and target behavior. For example, assume the investigator thinks that teacher instructions are antecedents for student self-injury because self-injury helps the child escape teacher demands. We might consider duration of the teacher instruction important to analyze if we think that the longer the instruction the more likely the child will experience the social pressure of the demand (or the boredom of being lectured) and thus be more motivated to escape it. If such is the case, it would be sensible to analyze the duration of the antecedent. Alternatively, we might think that the child uses self-injury after teacher instruction because he does not know how to comply with the teacher's instruction. In such a case, no matter how long the teacher instructs, the student still does not know how to comply. And thus longer instructions do not increase the probability that self-injury will occur. Therefore, it could also be argued that only onset of the antecedent should be analyzed. Turning to whether *target* behavior duration should be analyzed, one might speculate that longer self-injurious behavior might increase the success of escaping the social pressure of the teacher's instruction. In such cases, it would seem that target duration should be analyzed. Alternatively, one might consider target-behavior duration irrelevant because once the child begins self-injury, the teacher immediately ceases to give instructions, regardless of how long the self-injury lasts.

Different Ways to Analyze Timed-Event Data

Assuming durations of the antecedent and target behaviors are coded, one has four options of how to analyze timed-event data for sequential associations. Table IV adds the duration information to the timed-event data that was presented in Table I. The four analysis methods vary according to whether duration information of the antecedent and target events is included in the analysis. We have already introduced one of these methods (i.e., the Onset-Onset method illustrated in Table III). Table IV illustrates the other three ways that some or all of the duration information can be analyzed. In Table IV, we can see that the decision to analyze the duration information has an impact on how several of the seconds are treated in three analysis methods. Seconds 00:05, 00:06, 00:09, 00:11, 00:12, and 00:13 are analyzed in a different way depending on whether duration of the antecedent or target behavior is analyzed.

To further understand why analyzing duration of the antecedent and target behaviors can produce different results than just analyzing onset of these behaviors, it is useful to note that the definition of what constitutes the antecedent time window differs depending on whether duration or only onset of the antecedent is analyzed. When antecedent duration information is analyzed in time-window

Table IV. Timed-Event Data With Duration and Cell Addresses for 3 Ways to Analyze the Duration Information in Time-Window Sequential Analyses

			Analysis approach to dealing with duration information		
Time	Teacher behavior	Student Behavior	Cell address Duration-Duration method[a]	Cell address Duration-Onset method[b]	Cell address Onset-Duration method[c]
00:01	Other	Other	D	D	D
00:02	Instruction	Other	B	B	B
00:03	Instruction	Other	B	B	B
00:04	Instruction	Self-Injury	A	A	A
00:05	Other	Self-Injury	A	B	A
00:06	Other	Other	B	B	D
00:07	Other	Other	D	D	D
00:08	Other	Self-Injury	C	C	C
00:09	Instruction	Self-Injury	A	B	A
00:10	Instruction	Other	B	B	B
00:11	Other	Other	B	B	D
00:12	Other	Self-Injury	A	A	C
00:13	Other	Self-Injury	C	D	C

[a]Duration-Duration refers to 2 × 2 tables that analyze duration for antecedent and target behaviors.
[b]Duration-Onset refers to 2 × 2 tables that analyze duration for only antecedents.
[c]Onset-Duration refers to 2 × 2 tables that analyze duration for only targets.

sequential analysis, the antecedent time window count begins after the *offset* of the antecedent behavior. In contrast, when only antecedent onset information is analyzed, the antecedent time window begins after the *onset* of the antecedent. For example, in Table IV, the first antecedent time window that consider duration of the antecedent begins at second 00:02 and ends at 00:06. In contrast, if antecedent duration is not analyzed, the first antecedent time window begins at second 00:02 and ends at 00:04. This different definition of the antecedent time window has implications for the base rate of the antecedent time window. In Table IV, the base rate for the antecedent time window is nine seconds if duration is analyzed, but is only six seconds if duration of the antecedent is not analyzed. Similarly, if target duration information is analyzed, the target base rate in the data in Table IV is six seconds. However, if only target onset is analyzed, the target base rate in the data in Table IV is only three seconds.

A Return to Estimating Chance Occurrence of the Sequence of Interest

At this point, it is useful to note that the expected frequency of the sequence of interest (e.g., frequency of self-injury following teacher instruction) by chance processes is a function of the base rates of the antecedent time window and the target behavior. Bakeman and Gottman (1997) indicate that one mathematical estimate of the sequential frequency that is expected by random processes is the product of the base rate of the target multiplied by the simple probability of antecedent. We

Table V. Comparisons of the Expected and Observed Sequential Frequencies Depending on Whether Durations of Antecedent and Target Behaviors Are Analyzed

Analysis method	Simple probability of antecedent time window	Base rate of target behavior	Expected frequency of the sequence of interest	Observed frequency of the sequence of interest (A cell)
Onset-onset	.46	3	1.38	1
Duration-duration	.69	6	4.14	4
Duration-onset	.69	3	2.07	2
Onset-duration	.46	6	2.76	3

refer to this product as the "expected sequential frequency." A "simple probability of the antecedent" in time-window analysis is the base rate for the antecedent time windows divided by the total number of coded seconds in the behavior sample. Using the contingency table cell labels, this is $(A + B)/(A + B + C + D)$. Since the base rates of the antecedent time window and base rates of the target behavior vary across the four contingency tables, the expected sequential frequency will also vary across the four contingency tables. Additionally, the four A cell counts vary across the four contingency tables.

Table V illustrates these differences. It should be noted that these four ways to analyze timed-event data for time-window sequential analysis came from the same observational data. The difference score between observed and expected sequential frequencies varies by whether the durations of the antecedent and target behaviors are analyzed. In fact, in the first three tables, the observed sequential frequency is *less* than expected by chance. In the fourth table, the observed sequential frequency is *greater* than expected by chance.

An Index of Sequential Association: Yule's Q

Regardless of what 2×2 table we use to organize sequential data, we need an index of sequential association that controls for (a) the base rates of the antecedent time window and target behavior and (b) the total number of coded time units in the behavior sample. The odds ratio is one such index of sequential association (Bakeman & Gottman, 1997; Bakeman, McArthur, & Quera, 1996). The possible range of the odds ratio is 0 to infinity, with 1.0 representing a null association. Using the cell labels for the 2×2 table in Table III, the odds ratio is as follows:

$$\text{odds ratio} = (A \times D)/(B \times C)$$

In the context of sequential analysis, the meaning of an odds ratio above 1.0 is that the target is more likely to occur within the antecedent time window than outside the antecedent time window. An odds ratio below 1.0 indicates that the target is less likely to occur within the antecedent time window than outside of the antecedent time window. Audiences without training in the statistics that educators and psychologists use may relate to the odds ratio. For audiences that are familiar with Pearson's correlation coefficient, the asymmetry around 1 as an indication of statistical independence (i.e., no association) may result in confusion. Fortunately, the odds ratio can been transformed to Yule's Q without loss of information (Reynolds, 1984). Yule's Q has a possible minimum of -1.0 and a possible maximum of 1.0. A Yule's Q of 0 represents the null relationship between the antecedent and the target behaviors. A negative Yule's Q means that the target occurs within the antecedent time window less than it occurs outside the antecedent time window. A positive Yule's Q means that the target occurs within the antecedent time window more than it occurs outside the antecedent time window. Using the cell labels for the 2×2 table in Table III, Yule's Q is computed as follows:

$$\text{Yule's } Q = ((\mathrm{A} \times \mathrm{D}) - (\mathrm{B} \times \mathrm{C}))/((\mathrm{A} \times \mathrm{D}) + (\mathrm{B} \times \mathrm{C})).$$

A Monte Carlo study has been conducted (Yoder et al., 2004) to illustrate that Yule's Q is only minimally affected, whereas transitional probabilities are greatly affected, by the simple probabilities of the target behaviors. Neither is affect by the base rate of the antecedent behavior. In the Yoder, et al. (2004) Monte Carlo, the correlation between the Yule's Q scores and the corresponding simple probabilities of the target and antecedent behaviors were $-.02$ and .08, respectively. The correlation between the transitional probabilities and the corresponding simple probabilities for the target and antecedent behaviors were .70 and $-.04$, respectively. Because we want an index of sequential association that is not influenced by individual differences in the occurrence of the antecedent behavior, Yule's Q is a more informative index of sequential association than transitional probabilities.

Test of the Significance of Sequential Associations at the Single Subject Level

Because some difference between observed and expected sequential frequencies can occur by chance (i.e., sampling a particularly favorable or unfavorable behavior sample), we need to test the statistical significance this difference score. One can do this at the single subject level of analysis. Bakeman, Robinson, and Quera (1996) and Yoder, Bruce and Tapp (2001) have demonstrated through Monte Carlo studies that the adjusted residual in log linear analysis or its equivalent from 2×2 tables (i.e., the z score) produce an acceptably accurate estimate of the p value for the difference between observed and expected sequential frequency.

Bakeman and Gottman (1997) present one formula for the z score that makes it clear why it is a reasonable way to test significance of this difference. This formula is as follows:

(Observed sequential frequency–expected sequential frequency)/SD of this difference. From this formula it should be clear that this is simply the standardized difference between the observed and expected sequential frequency. This formula is appropriate when analyzed time units can follow other time units that are coded in the same category (Bakeman & Gottman, 1997). The standard deviation of the difference score is computed as follows:

$$\begin{aligned}&\text{Square root of [expected sequential frequency}\\&\times (1 - \text{simple probability of antecedent})\\&\times (1 - \text{simple probability of the target})]\end{aligned}$$

The adjusted residual or z score is normally distributed. Therefore, a $|z \text{ score}|$ that is greater than $|1.96|$ is "statistically significant" when a 2-tailed significance test and an alpha of .05 is used.

Although z scores based on such a small behavior sample are uninterpretable (Yoder et al., 2004), the above formula is applied here to the data in Table III for illustrative purposes. The expected sequential frequency (i.e., expected value of the A cell) of these data is 1.38 ($.46 \times 3$). The denominator of the z score is .76 (i.e., sqrt of $[1.38 \times (1 - .46) \times (1 - .23)]$. Therefore, the z for the 2×2 table in Table III is $-.5$ (i.e., $[1 - 1.38]/.76$). To date, no simulations have tested the Type I error rate of the four types of time-window sequential.

Purpose of Present Simulation Study

This study tested the Type I error rate of the four methods of time-window sequential analysis. We tested whether the Q and z scores differed across the four time-window sequential analysis methods. Finally, we tested whether Yule's Q and the z scores were related to the simple probabilities of the target and antecedent events. We expected that the Type I error rate of the Duration-Duration time-window analysis to be the closest to our alpha of .05 because recoding was restricted only to adding the time window (i.e., all duration information was analyzed). We expected that the decision regarding the analysis of duration would affect the results but we were not sure which decision would produce the most accurate results (i.e., produce Type I error rates closest to .05). We expected associations between simple probabilities of antecedent and target behaviors with Q and z to be a chance levels.

We also conducted a Concurrent time analysis on the simulated data. Differences in the results of time-window analysis methods versus Concurrent time analysis indicate how recoding the data affects the results. Concurrent time analysis does not add a time window onto the observed antecedent events, and it analyzes all duration information. Concurrent time analysis logic is identical to that used for event-lag sequential analysis, except that concurrent, instead of sequential, occurrences of the antecedent and target behaviors are analyzed. Time-window analysis uses Concurrent time analysis logic also, but instead of applying such logic to observed data, time-window analysis applies it to recoded data to account for the time window and decisions regarding the analysis of duration of the antecedent and target behaviors. The Type I error rate for the Concurrent time analysis was expected to be .05.

METHOD

Generation of the Simulated Pairs of Data Streams

To determine whether *z* scores were effective in controlling Type I error rates when applied to time-window sequential analysis and Concurrent time analysis, we conducted a simulation in which we generated 10,000 pairs of data streams, each with 3600 seconds (1 hour of observation time). One hour was selected as the session length because we considered this on the high side of typical session lengths for behavioral observational studies. Each data stream within the pair was "coded" for either antecedent or target actor. Within each data stream 1s or 0s represented presence or absence of the critical behavior, respectively. Duration information was represented in the data through consecutive 1s. For example, if three consecutive antecedent codes (e.g., "1") occurred in a data stream, it meant that this antecedent event lasted 3 seconds. In this example, the first 1 represents the onset of the behavior and the other two 1s represent continuations of the same instance of the behavior. The simple probability of 1s in each stream was randomly selected between .20 and .80 so that simple probability of events would not systematically affect the results and so results could be generalized across a wide range of base rates for antecedent and target events. Data in each stream was generated independently from the data in other streams to model a null sequential association in the population of simulated pairs of data streams.

Tallying the Data for the Concurrent Time Analysis

Because no time window was used and all duration information was analyzed for the Concurrent time analyses, no recoding of the data was necessary. Before computing the Yule's Q for these analyses, data was sorted into the four cells of the 2×2 tables. Then Yule's Q and z scores were computed.

Recoding and Tallying for the Four Time-Window Sequential Analyses

Because a time window is used and because duration was not always analyzed, we recoded the data before tallying the data into the 2×2 tables for the four time-window sequential analysis methods. Each method required a different recoding of the simulated data streams. Data streams were recoded for the type of analysis used before being recoded for a 2-second time window. When a time-window analysis method only analyzed the onset of a behavior (e.g., Onset-Onset), the simulated data stream that represents the onset-only behavior is recoded by changing 1s to 0s when the 1 immediately follows another 1. This type of recoding was not necessary for the Duration-Duration method. After recoding the simulated data streams according to whether duration was analyzed, the recoded data streams were recoded again to represent the 2-second antecedent time window. This 2nd recoding involved changing from 0 to 1 the code representing the two seconds following the antecedent. The time window began immediately following the *offset* of antecedents for the Duration-Duration and Duration-Onset methods and immediately following the *onset* of antecedents for the Onset-Onset and Onset-Duration methods. After recoding was complete, appropriate 2×2 tables were constructed and Yule's Q and z scores were computed.

Estimation of Type I Error Rates

Two-tailed significance tests were implemented. Negative z scores were tested against the lower tail of the probability distribution. Positive z scores were tested against the upper tail of the probability distribution. Alpha was set at .05. In simulation studies, the Type I error rate is the proportion of significant simulated findings when it is known that the effect being tested is zero in the population. In this study, the Type I error rate was number of z scores whose absolute value was greater than 1.96 divided by 10,000.

RESULTS

Description of Simulated Data Streams

To estimate Type I error rate, the data we generated must represent a null sequential association in the "population." To determine whether we were successful in generating such data, we examined the distributions of the Q scores. If the sequential association was null in the population, the mean Q should be zero. Table VI indicates the means and standard deviations of the Yule's Q scores for the 10,000 timed-event data stream pairs as computed by the five analysis methods. The results indicate that the population of the simulated data streams had a null sequential relationship between the antecedent and target behaviors, regardless of how the timed-event data were analyzed.

To help us determine the types of data streams to which the results can be generalized, it is useful to note the average duration and frequency of antecedent

Table VI. Mean and Standard Deviations of the Simulated Yule's Q Scores for Five Ways to Analyze Timed-Event Sequential Data

	Simple probability of the antecedent		Simple probability of the target		Yule's Q			
Analysis method	Mean	*SD*	Mean	*SD*	Mean	*SD*	min	max
Concurrent time analysis	.50	.17	.50	.17	.0002	.04	−.19	.16
Duration-duration	.83	.15	.50	.17	.001	.08	−.67	.58
Duration-onset	.83	.15	.22	.03	.003	.09	−.45	1.0
Onset-duration	.61	.07	.50	.17	.0007	.04	−.13	.13
Onset-onset	.61	.07	.22	.03	−.0004	.04	−.14	.14

and target events. The grand average (average of the 10,000 averages) for the duration of the generated antecedent event without the time window was 1.37 seconds ($SD = .28$; minimum = 1.05–2.32); thus indicating that most of the antecedent events were brief. The grand average for the duration of the generated target event was 2.3 seconds ($SD = .96$; minimum = 1.20; maximum = 5.39); thus indicating that targets were of greater duration than antecedents. This difference was significant ($t = 93.3$; $p = .000$), and the effect size was large (mean difference = .93 seconds; *SD* of difference = 1.0; $d = .93$). The average number of antecedent events was 361 ($SD = 140$; minimum = 110; maximum = 621).The average number of generated target events was 599 ($SD = 207$; minimum = 223; maximum = 976).

Table VI also indicates the simple probabilities of the antecedent time windows and target event for each of the 4 methods of time-window sequential analysis and for the generated data streams, as indicated by the Concurrent time analysis. The *SD* and range of the Q scores is approximately twice as high for time-window analysis methods that analyze the duration of the antecedent behavior (i.e., Duration-Duration and Duration-Onset) than for analysis methods that analyze the onset of the antecedent behavior (i.e., Onset-Duration and Onset-Onset). Additionally, all time window analysis methods have larger simple probabilities of the antecedent behavior than the generated data because the time window analysis methods recoded the two seconds after the antecedent. Those time-window analysis methods that analyze duration of the antecedent have greater simple probabilities of the antecedent than the generated data streams and than the methods that analyze onset of the antecedents. Finally, the time-window analysis methods that analyze duration of the target have greater simple probabilities of the target than the methods that analyze onset of the target. These differences are reflected in the Type I error rates.

Type I Error Rates

Table VII indicates the Type I error rate of each of the five ways of analyzing the simulated timed-event data. The results for the Concurrent time analysis

Table VII. Type I Error Rate for Four Time-Window Sequential Analysis and Concurrent Time Analysis

Analysis method	Type I error rate
Concurrent time analysis	.05
Duration-duration	.06
Onset-duration	.05
Duration-onset	.03
Onset-onset	.02

indicate that the z score method of testing significance of the sequential association produced accurate Type I error rates (i.e., .05). Similarly, Onset-Duration time-window sequential analysis also produced accurate Type I error rates. However, Duration-Duration time-window sequential analysis produced slightly liberal Type I error rates. The two time-window analysis methods that analyze onset of the target behavior produced Type I error rates that were too conservative.

Differences in Q and z Across the Five Analysis Methods

The results presented in Table VI illustrate that the variance in Q scores is greater for time-window analysis methods that analyze duration of the antecedent than methods that analyze onset of the antecedent. Additionally, the results in Table VII suggest that z scores may vary by whether duration information is analyzed, particularly target duration. To test whether the differences among groups on Q and z scores varies as a function of the duration of target and antecedent events, we tested (a) the statistical interaction between analysis method and duration of the target on Q and z and (b) the interaction between analysis method and duration of the antecedent on Q and z. These interactions were tested using a repeated measures MANOVA with analysis method (5 levels) as a within subjects factor and duration of the target and antecedent as continuous covariates. The z scores did not vary significantly among analysis method, regardless of the duration of the target and antecedent behaviors (all p values $> .07$). However, the statistical interaction of analysis method and duration of the target was significant for the Q scores (Hostelling's $T(4, 9994) = 4.5$; $p = .001$). Similarly, the interaction of analysis method and duration of the antecedent for the Q scores was also significant (Hostelling's $T(4, 9994) = 9.2$; $p < .001$).

We have illustrated these interactions in Figs. 1 and 2. The mean Q scores are presented by analysis method for low and high duration of the antecedent and target events, respectively. Duration was dichotomized using the median split method for illustrative purposes only. The presence of these interactions was tested using the duration as a continuous variable. These figures indicate that when the average duration of the antecedent was over 1.27 seconds (the median average duration), the sequential association was largest when duration of

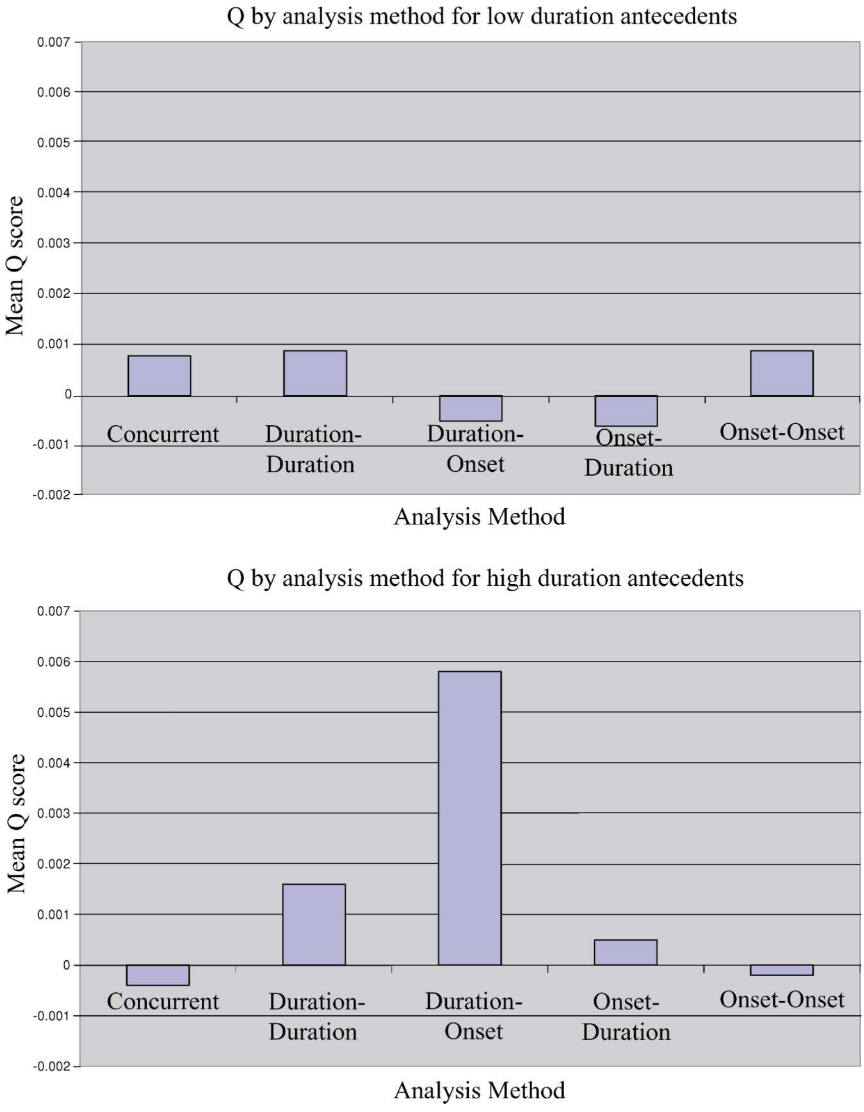

Fig. 1. Mean Q scores as a function of the duration of the antecedent behavior and analysis method.

the antecedent and the onset of the target (i.e., Duration-Onset method) are analyzed. Under these long-antecedent conditions, the average sequential association differed between Concurrent time analysis and Duration-Onset (paired $t = 3.6$; $p = .000$; $d = .05$). When the target duration is under 2 seconds (i.e., the median of the average duration), sequential association was largest when duration of

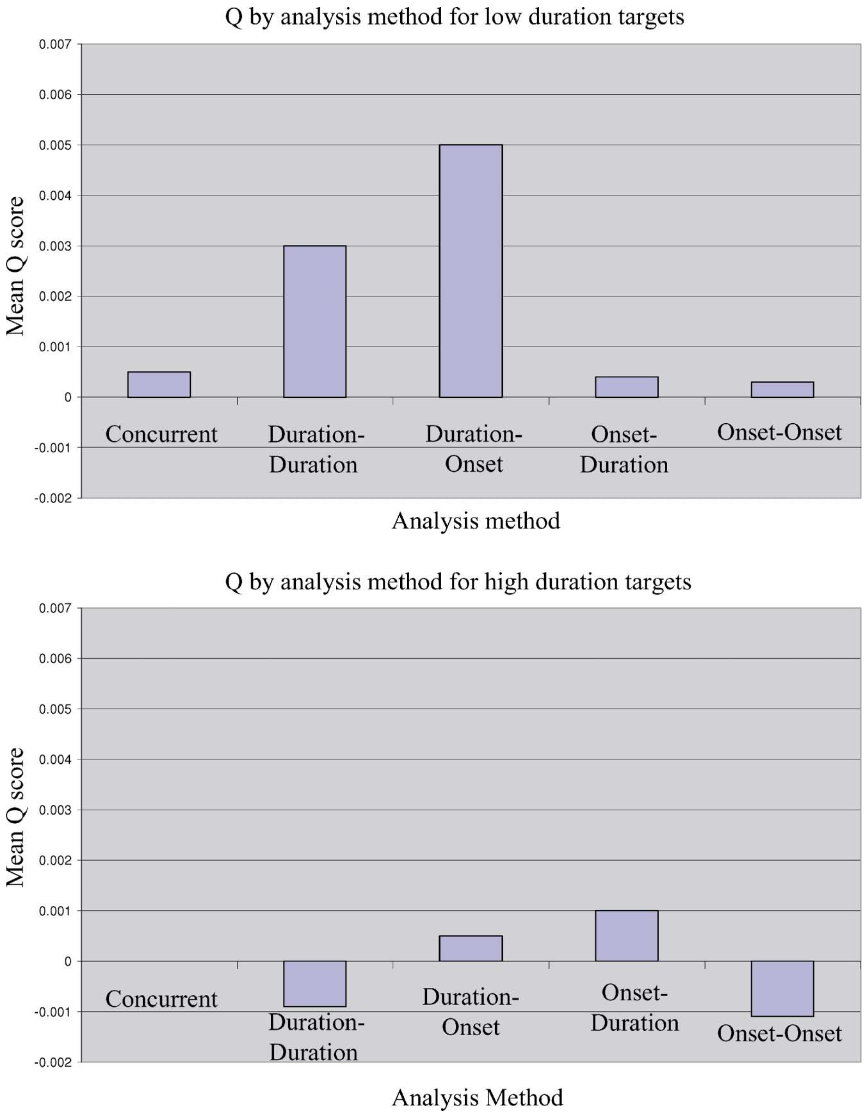

Fig. 2. Mean Q scores as a function of the duration of the target behavior and analysis method.

the antecedent was analyzed (i.e., Duration-Duration and Duration-Onset methods). Under these short-target conditions, the average sequential association as estimated by the Duration-Duration and Duration-Onset methods differed from the Concurrent time analysis (paired $t = 2.5, 3.5$; $p < .01$, $d = .035, .05$, respectively). However, it should be noted that all of these effect sizes are trivial.

Correlations of the *Q* Scores and *z* Scores With Simple Probabilities of the Antecedent Time Windows and Target Events

The *Q* scores from the various analysis methods were unrelated to their simple probabilities of the antecedent time windows and target events (r values ranged from .04 to −.03). The z scores for most of the analysis methods were also unrelated to their simple probabilities of the antecedent time windows and target events (r values ranged from .006 to −.005), with one exception. The z scores from the Onset-Onset method had small associations with that method's simple probabilities of the antecedent and target behaviors ($r = -.17$ and .12), respectively. Although these correlations were significant with such a large sample size (i.e., 10,000), the simple probability of the antecedent time window and target behaviors accounted for only 3% and 1% of the variance in the z scores, respectively. Cohen (1988) considers these effect sizes trivial.

DISCUSSION

This study was conducted to determine whether time-window sequential analysis is an accurate way to address research questions concerning whether one behavior occurs within a specified time after another behavior more than is expected by chance. To determine whether recoding for a time window affected the results, we also analyzed the data using a Concurrent time analysis, which does not require recoding the data and uses an already-tested analysis method. We judged the accuracy of Type I error rates by their discrepancy from our alpha rate of .05. To estimate Type I error rate, the simulated-data mean of the sequential association between antecedent and target behaviors had to be zero. Regardless of the analysis method used, the simulated data mean for our index of sequential association (i.e., Yule's Q) was zero when rounded off to two decimal places. The results confirmed that the Concurrent time analysis produced accurate Type I error rates (.05), thus validating the z score as a significance test in this context. The most accurate method of time-window sequential analysis was the Onset-Duration method (.05). The Duration-Duration method was slightly too liberal (.06). The methods that analyzed onset for target behavior were too conservative (under .05). The tests of significance of the sequential association (i.e., z scores) were not affected by whether durations of antecedent and target events were analyzed. Additionally, the test statistic (i.e., z score) and the sequential association index (i.e., Q score) were generally uncorrelated with or had trivial correlations with the simple probabilities of the antecedent and target behaviors. This latter point is a desirable characteristic because the concept of sequential association and its significance should be independent of base rates (Bakeman & Gottman, 1997).

The degree to which the sequential association index differed among analysis methods depended on the duration of the antecedent and target events. When the target duration was short or the antecedent duration was long, using the Duration-Onset method produced the strongest sequential associations when compared to other 4 analysis methods. This can be understood by noting that when the target duration is short, analyzing only the onset information for the target results in little loss of information. Similarly, when antecedent duration is long, we will lose information unless we analyze the duration of the antecedent. Under other duration conditions, the magnitude of the sequential association quantified by the four time-window sequential analysis methods was nonsignificantly different, even with a sample size of 10,000. It should be noted, however, that the effect size of analysis method on the sequential association was extremely small (i.e., $d = .035$ to .05). These effect sizes are significant only because the sample size was extremely large. It is possible that the effect size would be larger if the duration of the antecedent and target events was longer in the simulated data. In the simulated data, what constituted "long" was actually quite brief for both target (average duration over 1.99 seconds) and antecedent events (average duration over 1.27 seconds). The question of "How long do the events need to be before duration matters?" remains. At this point in our knowledge development, it is "safest" to analyze the duration of the antecedent, if, on average, the average duration is over 5 seconds. The decision to measure the duration of the target appears less important. However, it should be said that this decision does not appear critical when durations of the antecedent and target are brief (i.e., under 5 seconds).

These results are supportive of the validity of time-window sequential analysis. Recoding the observed data stream does not appear to affect the accuracy of tests of significance of the sequential association very much. However, it should be noted that the time window used in this analysis was relatively short. Future research is needed to determine if length of the time window affects the accuracy of the Type I error rate and the importance of analyzing duration of the antecedent and target events when duration of these events is over 5 seconds.

Additional future research is needed to determine the relative statistical power of the four time-window analysis methods. To do this, one needs to simulate data with a known, nonzero sequential association. The current study cannot address this question because we simulated data with a known, zero sequential association. Logic tells us that the statistical power of detecting a significant sequential association is inversely related to the duration of the time-window. Therefore, duration of the time window should be a factor in this future research.

In summary, time-window sequential analysis has much promise. The basic logic of the analysis appears sound. The z scores produces accurate Type I error rates when recoding is not used and produces nearly accurate Type I error rates when recoding for the time window is necessary. Sequential associations derived from time-window analysis are independent of base rates of the antecedent and target events.

ACKNOWLEDGMENT

This article was written when the authors were partially funded by the NICHD core grant HD15052 (to John F. Kennedy Center, Vanderbilt University).

REFERENCES

Bakeman, R., & Gottman, J. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). New York: Cambridge University Press.

Bakeman, R., McArthur, D., & Quera, V. (1996). Detecting group differences in sequential association using sampled permutations: Log odds, kappa, and phi compared. *Behavior Research Methods, Instruments and Computers*, *28*(3), 446–457.

Bakeman, R., & Quera, V. (1995). *Analyzing interaction: Sequential analysis with SDIS & GSEQ*. New York: Cambridge University Press.

Bakeman, R., Robinson, B., & Quera, V. (1996). Testing sequential association: Estimating exact *p* values using sampled permutations. *Psychological Methods*, *1*(1), 4–15.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Reynolds, H. T. (1984). *Analysis of nominal data* (2nd ed.). Beverly Hills, CA: Sage.

Yoder, P. J., Bruce, P., & Tapp, J. (2001). Comparing sequential associations within a single dyad. *Behavior Research Methods, Instruments and Computers*, *33*(3), 331–338.

Yoder, P. J., & Feurer, I. D. (2000). Quantifying the magnitude of sequential association between events or behaviors. In T. Thompson & D. Felce (Eds.), *Behavioral observation: Technology and applications in developmental disabilities* (pp. 317–333). Baltimore, MD: Paul H. Brookes.

Yoder, P. J., Short-Meyerson, K., & Tapp, J. (2004). Measurement of behaviors with special emphasis on sequential analysis of behavior. In E. Emerson, C. Hatton, T. Thompson, & T. Parmenter (Eds.), *International handbook of research methods in intellectual disability* (pp. 179–202). New York: Wiley.