Multivariate Behavioral Research, 50:41–55, 2015

ISSN: 0027-3171 print / 1532-7906 online
DOI: 10.1080/00273171.2014.961056

The Importance of Temporal Design: How Do Measurement Intervals Affect the Accuracy and Efficiency of Parameter Estimates in Longitudinal Research?

Adela C. Timmons
Department of Psychology, University of Southern California

Kristopher J. Preacher
Department of Psychology and Human Development, Vanderbilt University

The timing (spacing) of assessments is an important component of longitudinal research. The purpose of the present study is to determine methods of timing the collection of longitudinal data that provide better parameter recovery in mixed effects nonlinear growth modeling. A simulation study was conducted, varying function type, as well as the number of measurement occasions, in order to examine the effect of timing on the accuracy and efficiency of parameter estimates. The number of measurement occasions was associated with greater efficiency for all functional forms and was associated with greater accuracy for the intrinsically nonlinear functions. In general, concentrating measurement occasions toward the left or at the extremes was associated with increased efficiency when estimating the intercepts of intrinsically linear functions, and concentrating values where the curvature of the function was greatest generally resulted in the best recovery for intrinsically nonlinear functions. Results from this study can be used in conjunction with theory to improve the design of longitudinal research studies. In addition, an R program is provided for researchers to run customized simulations to identify optimal sampling schedules for their own research.

In the process of conducting longitudinal research, investigators usually pay scrupulous attention to the design of their studies. But, despite careful attention to many aspects of research design, less attention is typically given to *temporal design*. Temporal design refers to the timing and spacing of occasions of measurement (Collins & Graham, 2002). One of the most important components of temporal design is the measurement interval. As discussed by Collins and Graham (2002), measurement interval (or *lag*) refers to the amount of time that elapses between occasions of measurement. Although it is often overlooked by researchers, the choice of measurement interval is a central component of longitudinal research design.

Correspondence concerning this article should be addressed to Adela C. Timmons, Department of Psychology, University of Southern California, SGM 921, 3620 South McClintock Avenue, Los Angeles, CA 90089-1061. E-mail: adelatim@usc.edu

In the psychological literature, researchers often report the measurement interval that they choose, but they usually do not justify their particular chosen interval. If justification is provided, reasons typically are related to logistical constraints, or more rarely to theoretical considerations. In one study, for example, Chaiton, Cohen, O'Loughlin, and Rehm (2010) studied smoking and depressive symptoms. The investigators surveyed participants at 3-month intervals for 5 years. Although they used a relatively small measurement interval, the researchers did not justify their choice. Similarly, Kyle and Harris (2010) studied predictors of reading development in hearing-impaired children. They sampled every 12 months for 3 years, but they did not provide justification for their decision. Finally, in a study examining the relationship between birth weight and cognitive development, investigators collected data every 2 years, but they provided no explanation for their choice of measurement interval (Cheadle & Goosby, 2010). This leaves open the possibility that had the researchers chosen different intervals, qualitatively different results may have been obtained.

The above studies represent a few examples of a more general trend in the psychological literature. Because there is little authoritative guidance from methodologists, researchers must rely on other methods for making decisions about measurement intervals. For now, considerations that inform these decisions are constrained to a number of distinct issues. First, researchers face logistical constraints. They are limited by the amount of time and money they can devote to a particular project. Still, most investigators are aware that collecting data at more frequent intervals often is advantageous, and they will typically collect more data if logistical constraints do not prevent them from doing so. Researchers also consider theory when making decisions about measurement intervals. Theory can give researchers an idea about how quickly change will occur and the time span over which the change will occur. Finally, researchers must consider other practical limitations associated with frequent sampling. As can occur with test–retest designs (McArdle & Woodcock, 1997), sampling too frequently can cause practice effects, lead to attrition, and make recruitment more difficult (Adolph, Robinson, Young, & Gill-Alvarez, 2008). Moreover, as Collins and Graham (2002) explain, it may make data management and analysis more cumbersome and can heighten sensitivity to error while obscuring more general patterns in the data.

The Importance of Temporal Design

Although investigators typically do not provide empirical justification for their choice of measurement interval, methodological work suggests that this decision is important and can have a large impact on statistical analyses, and therefore on the conclusions that researchers draw (Cohen, 1991; Collins, 1996a; Collins & Graham, 1991; Gollob & Reichardt, 1991; Nesselroade & Jones, 1991; Windle & Davies, 1999). First, measurement intervals can impact the strength of correlations and the magnitudes of regression coefficients. As Collins and Graham (2002) discuss, if change occurs before or after the occasion of measurement, then these correlations and slopes may be biased. This is because the size of an effect is likely to change over time. Although this difference could be in either direction, effects typically become weaker or disappear as time passes (Collins & Graham, 2002). Poorly chosen measurement intervals can cause an effect to be underestimated or overestimated, or to not be detected at all. For example, Collins and Graham (2002) demonstrated that smoking and peer smoking are correlated. However, the strength of the relationship decreased as the lag between measurements of the two variables increased, and after 5 years the effect was not significant. Therefore, suboptimal measurement intervals can lead researchers to make inaccurate conclusions about variables under study.

In addition to correlation and multiple regression, the choice of measurement interval has implications for other statistical techniques. For example, stage-sequential models, which are used to model transitions between distinct states, are particularly sensitive to changes in measurement interval. As Collins and Graham (2002) show, when measured once per month, it is seen that adolescents tend to progress through separate stages of trying alcohol, tobacco, and marijuana. However, when measured every 10 months, there is no evidence for a separate tobacco stage. Thus, if researchers collect data every 10 months, it appears that adolescents first try alcohol and tobacco at roughly the same time. Similarly, Adolph et al. (2008) showed how measurement intervals affect estimates of human development. These researchers measured infant motor skills daily. Then, they manipulated the measurement interval by selecting observation points at various intervals. They showed that the change trajectory was lost as interval size increased. Sampling once per day resulted in 15.7% of the motor skills showing an abrupt transition from absent to present. The majority of infants wavered in their acquisition of skills, sometimes demonstrating the skill and other times not. However, sampling once per month resulted in 92.7% of the motor skills showing abrupt transitions from one stage to another. Therefore, differences in measurement interval can yield dramatically different conclusions regarding developmental trajectories.

Well-selected measurement intervals are also essential in mediation analysis. Tests of mediation typically require longitudinal data because the researcher must demonstrate that the causal variable has sufficient time to influence the mediator, which in turn has sufficient time to influence the outcome (Cole & Maxwell, 2003; Collins, Graham, & Flaherty, 1998). However, suboptimal measurement intervals can obscure tests of mediation. For instance, if the length of time between an effect and measurement is too long, the strength of the relationship could decay, making it more difficult to detect a mediation effect (Shrout & Bolger 2002).

Finally, differences in measurement interval could affect growth curve models. If a trajectory follows a straight line, the timing and spacing of observations may not be particularly important. Although more occasions of measurement generally yield more accurate results, it is likely that changes in the timing and spacing of observations will not increase the accuracy of parameter estimates. But, if the growth curve is complex, then the timing and spacing of measurement occasions may impact the ability of a researcher to detect the effect. If measurement intervals are large, then changes in growth easily could be missed. In addition, changes in the timing and spacing of observations could yield more accurate or less accurate results, depending on what trajectory underlies the data (Collins & Graham, 2002).

Previous Methodological Guidance on Temporal Design

Given the importance of choosing optimal measurement intervals, it is surprising that researchers do not pay more attention to the temporal aspects of research design. Collins and Graham (2002) advocate employing "careful consideration" and "explicit justification" when choosing measurement intervals (p. 94), and other researchers have made similar suggestions (Collins, 1996b; Nesselroade, 1991; Nesselroade & Boker, 1994; Nesselroade & Jones, 1991; Ployhart & Vandenberg, 2010; Windle & Davies, 1999). However, this advice has not been widely heeded in the applied literature. One explanation for this is that there is little explicit justification for researchers to use when choosing measurement intervals. Because methodological guidelines for making these decisions are scarce, it is difficult for researchers to provide the explicit justification that methodologists ask of them.

Nevertheless, several methodologists have offered suggestions for researchers to use when making these decisions. Most often, methodologists recommend sampling more frequently and using theory to guide decisions about temporal design. The use of smaller measurement intervals is also a good rule of thumb. Smaller intervals often result in greater power to detect an effect, as well as increased accuracy and precision (Burchinal & Appelbaum, 1991; Collins, 2006; Hertzog & Nesselroade, 2003; Siegler, 2006; Thelen & Ulrich, 1991; Willett, 1989). However, even though smaller measurement intervals typically result in more accurate estimates of population parameters, this may not always be the case. It is possible that sampling at increasingly smaller intervals will yield more accurate results up to a point, but that there will be a "diminishing returns" effect. That is, it is possible that at some point adding more observations will not add meaningfully to the ability to detect an effect, even though it will cost the researcher extra time, money, and effort. Moreover, oversampling may result in retest effects, fatigue, and attrition. The advantages associated with small measurement intervals should be balanced with the high costs of frequent sampling. As with all aspects of research design, investigators must balance best-case scenarios with the real-world limitations of conducting research. Although some of the disadvantages associated with frequent sampling may be overcome by using a planned missing data design (see Graham, Taylor, & Cumsille, 2001; Graham, Taylor, Olchowski, & Cumsille, 2006), it is still important for researchers to seek ways to improve the design of longitudinal research.

Some researchers have advocated developing empirical methods for determining optimal measurement intervals. For example, Siegler (2006) stated that within a period of change, ". . .the density of observations [should be] high, relative to the rate of change" (p. 469). In addition, the Nyquist-Shannon sampling theorem often is used to determine the minimum number of measurement occasions required to detect waveforms (Nyquist, 1928; Shannon, 1949). The theorem states that the sampling frequency must be at least twice the bandwidth. However, this theorem is used primarily for physiological and psychophysiological variables. The theorem is inapplicable in many other cases because most psychological variables do not follow complex waveform functions. Other researchers have suggested hybrid models that measure over longer spans of time but also measure repeatedly in "bursts" at some occasions to capture both long-term and short-term change (e.g., Nesselroade, 1991). These models are useful for detecting change on multiple time scales but do not provide empirical guidelines for choosing measurement occasion spacing that maximizes parameter recovery. Other work suggests that to maximize power, researchers should sample more frequently early in a longitudinal study and less frequently later in the study (Rast & Hofer, 2014).

Apart from these recommendations, we have also heard informal advice from multiple colleagues, as well as suggestions in the literature (e.g., Ployhart & Vandenberg, 2010), that when fitting nonlinear functions, it is often advantageous to concentrate measurements more at times when the "greatest change" is expected to occur, or where the function is the "curviest," and less during periods of relatively static change. This advice may be sound accurate, but it is difficult to know with certainty. First, we could find no empirical evidence to support this claim. Second, precisely what is meant by "greatest change," "curvy," and "static" is ambiguous. Informal eyeball estimates of curvature do not necessarily agree with the formal mathematics of nonlinear change and curvature (see, e.g., Seber & Wild, 1989). For example, informally the point of greatest curvature in a quadratic function appears to be at the minimum or maximum, yet simple calculus proves that curvature is constant across the range of the function. Similarly, the inflection point of a logistic function (e.g., a child's time of largest gain in reading skill) may resonate with the lay interpretation of "greatest change" because the absolute level of the skill shows the greatest gains per unit of time during this period, yet the inflection point corresponds to what mathematicians would consider zero curvature (i.e., locally linear growth), so these concepts clearly are not identical. In the present study, we operationally define *change* as the absolute value of the first derivative of the function with respect to time (i.e., the slope of the tangent of the function), evaluated at a particular point in time. We operationally define *curvature* as the absolute value of the second derivative of the function with respect to time, evaluated at a particular point in time. For some functions the point of greatest change co-occurs with the point of greatest curvature; for others, they do not coincide.

Issues in the Application of Empirical Guidelines

In the process of developing empirical guidelines, methodologists often face a chicken-and-egg problem. Empirical guidelines suggest that, given that a specific function underlies a trend in repeated measures data, researchers should employ a certain technique or design that is well suited to estimate that trend. However, this information is not helpful if researchers do not have an *a priori* reason to expect to see a certain functional form. This situation produces a conundrum for researchers. But, even though researchers do not know beforehand what function exists, empirical guidelines about measurement intervals can still be of use. Collins and Graham (2002) discuss this point by comparing this problem to a power analysis:

> In both situations [power analysis and the use of empirical guidelines for measurement intervals] researchers must use prior research, pilot studies, and theory to come up with what amounts to an intelligent, informed guess. This process is difficult and fraught with uncertainty. However, as a body of knowledge accumulates in an area and more and more research reports include scientific justification of the choice of a temporal design, researchers will have an improved basis for making these decisions. This is analogous to what happened with respect to power analysis, which 40 years ago seemed strange to researchers who were first learning to think in those terms. The present authors believe that, just as today's biomedical and social scientists have come to appreciate the long range advantages of power analysis, they will also come to appreciate the value of basing temporal design considerations on explicit, justifiable, scientific criteria. (p. 94)

Thus, even though researchers may have only an "intelligent, informed guess" as to the shape of a function, guidelines for selecting optimal measurement intervals still may be useful.

Furthermore, the usefulness of empirical guidelines will depend on many factors, and such guidelines may be more useful in some situations than in others. In well-established areas of research, these rules could be applied easily. In less well-established areas, other methods could be employed. For example, researchers could initially use the microgenetic method of sampling. Microgenetic sampling entails sampling at very small intervals so that the shape of the underlying trend can be determined (Siegler, 2006). Then, once the underlying shape is determined, future research could use empirical guidelines, such as the ones discussed in the current article, to capture the same pattern more efficiently.

Alternatively, researchers may use theory and past research to make inferences about what shape to expect in the data, just as they might rely on past estimates of effect size to conduct a power analysis. In both cases, researchers recognize that they are proceeding with less than perfect information, but they also recognize that this information is better than none at all. In general, we argue that it is beneficial to approach the problem of temporal design armed both with prior theory and with empirical guidance from methodologists. Even if neither source provides clear, unambiguous guidance on temporal design, the two sources used in tandem can greatly strengthen the justification for design choices.

The Present Study

Purpose and General Procedure

The purpose of the present simulation study is to develop some basic empirical guidelines for researchers to use when choosing measurement intervals. Admittedly, the optimal measurement interval for each study is context-dependent and will vary according to a variety of factors. However, the present study is a preliminary step in improving decisions about measurement intervals. Past research dealing with this topic typically has used real data. However, simulated data are well suited for addressing this problem because with simulated data, the researcher knows *a priori* what functional form defines the trend underlying the data.

Because we view this study as a preliminary step, we limit our attention to linear and nonlinear mixed effects regression models describing change (Davidian & Giltinan, 1995). Such models are commonly used to model change over time in a sample of individuals (e.g., change in cortisol levels in individuals across time) or change in means (e.g., change in mean depression levels in a group of individuals undergoing treatment). It is expected that the general principles derived from the current study can be extended to more complex designs for investigating nonlinear change (e.g., nonlinear mixed models involving latent variables and/or individually varying occasions of measurement), although future work should examine these questions directly.

In the simulation, data were generated according to six specific functional forms: linear, quadratic, cubic, power, exponential, and Gompertz. These functions were chosen because they are observed commonly in social science research. For each type of function, population parameter values were chosen to be consistent with examples seen in the literature. The measurement intervals were manipulated by varying the number of occasions (6, 11, 16, or 21), as well as the spacing of occasions (spaced evenly, or concentrated toward the left, right, middle, or extremes).

After generating the data, the models were fit and parameter estimates were obtained. These estimates were compared to the population values that were set in the simulation. To evaluate how changes in measurement interval affected the parameter estimates, three outcome measures were evaluated: (1) percent relative bias (a measure of accuracy corresponding to how close an estimate is to the true value; PRB), (2) estimated standard error (SE), and (3) empirical standard deviation (the standard deviation of an estimate over repeated sampling; ESD). Smaller bias corresponds to greater accuracy on average, and less sampling error corresponds to greater efficiency. We examined the average estimated standard error because it is easily obtained in practice and commonly used in hypothesis testing and confidence interval construction, and the empirical standard deviation because it directly reflects observed variability in a parameter estimate over repeated sampling. Theoretically, these two measures should align closely.

Hypotheses

It was hypothesized that smaller measurement intervals (more measurement occasions) will be associated with lower PRB and less sampling error, all else being equal (Zhang & Wang, 2009). However, increasing the number of measurement occasions may create a "diminishing returns" effect, such that each time the interval length decreases, smaller gains in accuracy and efficiency will be observed. In this study, we differentiate between "intrinsically linear" and "intrinsically nonlinear" functions. "Intrinsically nonlinear" refers to a function that has at least one parameter that enters the function nonlinearly; a parameter is said to enter nonlinearly when the first partial derivative of the function with respect to the parameter contains that parameter. For example, $y = a_3 + b_3x + c_3x^2 + d_3x^3$ is intrinsically linear because the derivatives with respect to parameters a_3, b_3, c_3, and d_3 are, respectively, 1, x, x^2, and x^3, which do not contain the parameters in question. On the other hand, $y = a_5e^{b_5x}$ is intrinsically nonlinear because the derivative with respect to b_5, $a_5xe^{b_5x}$, contains b_5. For the intrinsically linear functions (i.e., the linear, quadratic, and cubic), it was hypothesized that the spacing of measurement occasions will not be associated with PRB but will be associated with efficiency. In the case of the slope of the linear function, it was hypothesized that the extreme spacing condition would be the most efficient, followed by even. We expected that left and right spacing would be equally efficient and that middle spacing would be least efficient. This hypothesis is justified in the following way (Willett, 1989).

Consider the fixed effects multiple linear regression model with one regressor, X:[1]

$$Y = B_0 + B_1X + e, \qquad e \sim N\left(0,\ \sigma_e^2\right) \tag{1}$$

The estimated asymptotic covariance matrix of the parameter estimates $\hat{B}_0$ and $\hat{B}_1$ is

$$\begin{aligned}
\sigma_e^2\left(X'X\right)^{-1} &= \sigma_e^2\begin{bmatrix} N & \Sigma X \\ \Sigma X & \Sigma X^2 \end{bmatrix}^{-1} \\
&= \frac{\sigma_e^2}{N\Sigma X^2 - (\Sigma X)^2}\begin{bmatrix} \Sigma X^2 & -\Sigma X \\ -\Sigma X & N \end{bmatrix} \\
&= \begin{bmatrix} \dfrac{\sigma_e^2 \Sigma X^2}{N\Sigma X^2 - (\Sigma X)^2} & \dfrac{\sigma_e^2 \Sigma X}{(\Sigma X)^2 - N\Sigma X^2} \\ \dfrac{\sigma_e^2 \Sigma X}{(\Sigma X)^2 - N\Sigma X^2} & \dfrac{N\sigma_e^2}{N\Sigma X^2 - (\Sigma X)^2} \end{bmatrix}.
\end{aligned} \tag{2}$$

$SE\left(\hat{B}_0\right)$ and $SE\left(\hat{B}_1\right)$ can be determined by calculating the square roots of the asymptotic variances:

$$\begin{aligned}
SE(\hat{B}_0) &= \sqrt{\frac{\sigma_e^2\Sigma X^2}{N\Sigma X^2 - (\Sigma X)^2}} & SE(\hat{B}_1) &= \sqrt{\frac{N\sigma_e^2}{N\Sigma X^2 - (\Sigma X)^2}} \\
&= \sqrt{\frac{\sigma_e^2\Sigma X^2}{\sigma_X^2 N(N-1)}} & &= \sqrt{\frac{\sigma_e^2}{\sigma_X^2(N-1)}} \\
&\propto \sqrt{\frac{\Sigma X^2}{\sigma_X^2}} & &\propto \sqrt{\frac{1}{\sigma_X^2}}
\end{aligned} \tag{3}$$

where $\propto$ denotes proportionality. The proportionality holds when σ_e^2 and sample size are held constant. Thus, $SE(\hat{B}_1)$ is inversely proportional to the variance of X. For the intercept of the linear function, we hypothesized that the left spacing condition would result in the smallest SE, followed by extreme, even/middle, and right. $SE(\hat{B}_0)$ is also inversely proportional to the variance of X. However, the presence of ΣX^2 in the numerator implies that $SE(\hat{B}_0)$ increases as the X values get further from 0. For the quadratic and cubic functions, we expect that similar processes may impact the efficiency of the slope and intercept terms. However, we expect that their implications will be more complex and more dependent on the span of time values considered.

For the intrinsically nonlinear functions (i.e., power, exponential, and Gompertz), we expect that sampling more frequently around the point in time when the absolute value of the second derivative is maximized will result in greater accuracy and efficiency. The location of greatest curvature will depend upon the type of function, as well as upon the parameter values used in the simulation. In this study, the locations of greatest curvature will be on the left side ($x = 0$; nearest to the origin) for the power function, on the right side ($x = 60$; farthest from the origin) for the exponential function, and in the middle ($x = 29.6$) for the Gompertz function. Based on the results, we will generate basic guidelines for choosing optimal measurement intervals and will discuss some implications for improving longitudinal research.

METHOD

Simulation Description

Programming Software and Generation of Data

To conduct the simulation, the programming language R was used (Version 3.0.2). The *rnorm* function in R was used to simulate random, normally distributed data. Using *rnorm*, the mean, standard deviation, and number of observations are set by the programmer. In the present study, the parameter values were set to generate data according to different functional forms, including linear, quadratic, cubic, power, exponential, and Gompertz functions. For each model, one parameter was treated as a random coefficient. The random seed for the simulation was set to 42596 using the *set.seed* command. Sample code from the study is included in an online appendix available at the second author's website.[2]

[1]The following expressions do not hold exactly in mixed effects models, but the basic logic is analogous.

[2]http://quantpsy.org/

TABLE 1
Summary of Simulation Conditions

Simulation Conditions	
Number of measurement occasions	6, 11, 16, and 21
Spacing of measurement occasions	Even, left, right, middle, and extreme
Functional form	Linear$_1$, quadratic$_2$, cubic$_3$, power$_4$, exponential$_5$, Gompertz$_6$
Parameters	a_1, b_1, a_2, b_2, c_2, a_3, b_3, c_3, d_3, a_4, b_4, a_5, b_5, a_6, b_6, c_6
Outcome Measures	
Percent relative bias	A measure of accuracy corresponding to how close an estimate is to the true value
Standard error	An estimate of the standard error calculated per parameter estimate per function per trial
Empirical standard deviation	The standard deviation of an estimate over repeated sampling

Description of Conditions

Three factors were manipulated in the study, including the function defining change over time, the number of measurement occasions, and the spacing of measurement occasions. As previously discussed, data were simulated for linear, quadratic, cubic, power, exponential, and Gompertz functions. These functions were chosen because they occur frequently in the social sciences (see Sit and Poulin-Costello, 1994, for a full description of these functions). Similarly, parameter values were chosen to reflect values that are plausible in social science research. Because parameter values vary according to scale and context, it is difficult to make meaningful comparisons of parameters across different areas of research. In addition, researchers typically do not report all estimates obtained in a study, so recreating exact functions is difficult. Thus, the parameters were not drawn directly from past research. Rather, values were chosen to approximate frequently occurring functions in the social sciences. To create functions that were plausible in social science research, it was necessary to use different parameters for different functions. Although this is a possible confound, the goal of the study was to determine how differences in measurement interval affect the accuracy and efficiency of parameter estimates. This question was investigated for a variety of function types, but this was not done with the intention of comparing the quality of estimates across functions. Thus, comparisons across functions should be made with caution.

A summary of the simulation factors is presented in Table 1, and the equations used to simulate the data, as well as the parameter values used, are included in Table 2. The coefficient labeled "a" in each model was treated as random. For all functions, data from $J = 50$ individuals were generated. The residual variances varied by functional form; values were chosen to approximate values plausible in social science research, as well as to minimize convergence errors while allowing for variability in parameter estimates. Interpretations of parameter values are included in Table 3, and graphical depictions of the functional forms defined by the fixed portion of each coefficient are presented in Figure 1. In addition to the functional form, the number of measurement occasions was manipulated. The total duration of measurement in all conditions was chosen to be 61 time units, beginning at 0 and terminating at 60. This is because 60 is the least common multiple of 5, 10, 15, and 20, which allowed us to use occasions differing by 5-point increments. For each condition, there were thus 6, 11, 16, or 21 occasions of measurement. The values on the y-axis of Figure 1 represent an arbitrary scale and should not influence the conclusions.

In the simulation, measurement occasion spacing was also manipulated. Occasions were spaced evenly, or were concentrated to the left, to the right, in the middle, or at the extremes. Even spacing was defined as spacing occasions of measurement at regular intervals. In the left-dense conditions, the occasions of measurement were concentrated to the left early in the study and became less concentrated across time. For the right-dense conditions, occasions of measurement were sparse at the beginning and became more concentrated over the measurement domain. Finally, for the middle and extreme spacing conditions, a similar pattern was used, except that the occasions of measurement were concentrated in the middle and at the extremes, respectively.

TABLE 2
Equations and Parameter Values Used to Simulate Data

Function	Equation	a_i	b_i	c_i	d_i	τ_{00}	σ_e^2
Linear	$y_{ij} = (a_1 + u_{1j}) + b_1 x_{ij} + e_{ij}$	3	5			225	600
Quadratic	$y_{ij} = (a_2 + u_{2j}) + b_2 x_{ij} + c_2 x_{ij}^2 + e_{ij}$	600	–40	2/3		800	1500
Cubic	$y_{ij} = (a_3 + u_{3j}) + b_3 x_{ij} + c_3 x_{ij}^2 + d_3 x_{ij}^3 + e_{ij}$	–600	100	–4	.0444	8000	2500
Power	$y_{ij} = (a_4 + u_{4j}) x_{ij}^{b_4} + e_{ij}$	10	.86			5	300
Exponential	$y_{ij} = (a_5 + u_{5j}) \exp(b_5 x_{ij}) + e_{ij}$	5	.08			2	475
Gompertz	$y_{ij} = (a_6 + u_{6j}) \exp(-\exp(b_6 - c_6 x_{ij})) + e_{ij}$	600	3	.15		1000	500

Note. τ_{00} refers to the population level-2 variance of the random a coefficient, and σ_e^2 refers to the population variance of the level-1 errors.

TABLE 3
Interpretation of Parameters

Function	Parameter	Interpretation
Linear	a_1	Intercept
	b_1	Linear slope
Quadratic	a_2	Intercept
	b_2	Linear slope where $x = 0$
	c_2	Controls the rate at which the linear slope changes; $\frac{1}{2}$ the amount by which the linear slope of x changes per unit change in x
Cubic	a_3	Intercept
	b_3	Linear slope where $x = 0$
	c_3	Controls the rate at which the linear slope changes where $x = 0$
	d_3	Controls the rate at which the quadratic slope changes; $^1/_6$ the amount by which the curvature changes per unit change in x
Power	a_4	The value of y when $x = 1$
	b_4	Controls the concavity of the function
Exponential	a_5	Intercept
	b_5	Controls the rate and direction at which the asymptote (0) is approached
Gompertz	a_6	Upper asymptote
	b_6	Shifts the curve horizontally
	c_6	Controls steepness of the curve

An exact formula was not used to determine the placement of intervals. Doing so would have resulted in non-integer values that would be unmanageable for the simulation. Rather, measurement occasions were spaced subjectively so that the amount of space between each occasion increased or decreased gradually across the measurement domain in accordance with the desired pattern. The exact spacing of intervals for all conditions is depicted in Figure 2.

Summary and Simulation Procedure

In summary, data were generated according to type of function (linear, quadratic, cubic, power, exponential, and Gompertz), the number of measurement occasions (6, 11, 16, and 21), and measurement occasion spacing (even, left, right, middle, and extremes). Not counting the residual variances, across the six function types, 16 parameters were estimated (a_1 and b_1 for the linear; a_2, b_2, and c_2 for the quadratic; a_3, b_3, c_3, and d_3 for the cubic; a_4 and b_4 for the power; a_5 and b_5 for the exponential; and a_6, b_6, and c_6 for the Gompertz). Therefore, there were 320 conditions (4 numbers of occasions $\times$ 5 measurement occasion spacings $\times$ 16 estimated parameters across the 6 functions). We fit the generated data using the *nlme* function in R, using population parameters as start values for each parameter. In each condition, the parameters were estimated using maximum likelihood, along with corresponding standard errors. This process was repeated 1,000 times for every cell of the design, creating a distribution of 1,000 values per parameter estimate per condition. The mean and standard deviation of the 1,000 parameter estimates were calculated in each cell of the design. As we discuss, the mean was used to calculate the PRB, and the SE and ESD are used as measures of (in)efficiency.

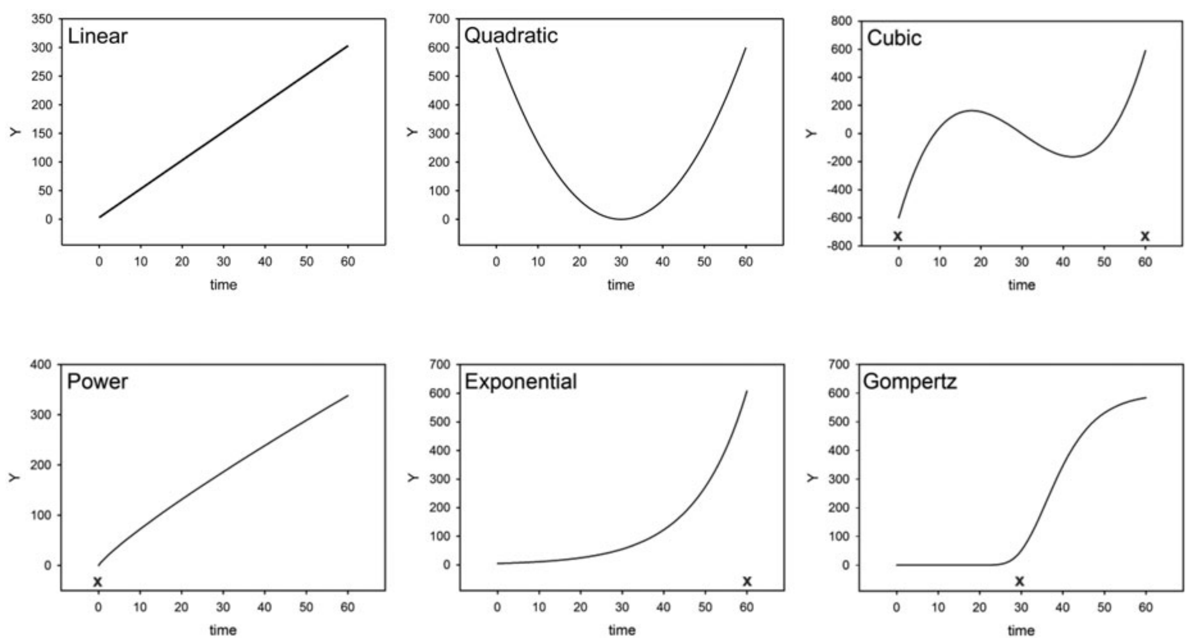

FIGURE 1 A graphical depiction of each functional form used in the simulation. **x** indicates the point(s) of greatest curvature for the cubic, power, exponential, and Gompertz functions (curvature is zero for the linear function and constant for the quadratic function).

	Even				Left				Right				Middle				Extreme			
	6	11	16	21	6	11	16	21	6	11	16	21	6	11	16	21	6	11	16	21
0	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
1								x										x	x	x
2							x													x
3				x		x		x											x	
4			x		x		x											x		x
5								x											x	
6		x		x		x	x													x
7								x									x	x	x	
8			x				x									x				x
9				x		x		x												
10							x								x					x
11								x								x			x	
12	x	x	x	x	x		x							x				x		
13						x		x					x							x
14							x								x	x				
15				x				x				x							x	
16			x																	x
17						x	x	x								x	x			
18		x		x								x		x	x					
19								x			x							x		x
20			x													x				
21				x	x		x	x				x							x	
22						x									x					
23										x	x			x						x
24	x	x	x	x				x				x				x				
25							x								x					
26													x			x				
27				x		x		x	x		x	x		x						
28			x							x					x	x				
29							x								x	x				
30		x		x				x				x		x	x	x				
31											x				x	x				
32			x			x										x				
33				x	x		x	x		x		x		x						
34													x		x	x				
35											x									
36	x	x	x	x				x				x				x				
37						x	x							x	x					
38										x										x
39				x				x	x		x	x				x				
40			x																x	
41							x					x			x					
42		x		x				x						x		x		x		x
43										x	x	x					x			
44			x																	
45				x				x				x			x	x			x	
46											x		x							x
47										x		x								
48	x	x	x	x					x		x			x		x				
49												x						x	x	x
50											x				x					
51				x						x		x								
52			x								x					x				x
53												x					x	x	x	
54		x		x						x	x									x
55												x							x	
56			x						x		x									x
57				x						x		x						x	x	
58											x									x
59												x							x	x
60	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

FIGURE 2 The exact interval spacing for all measurement occasion spacing conditions.

Outcome Measures

Percent Relative Bias (PRB)

To assess the accuracy of the point estimates, PRB was used. PRB is calculated by multiplying the ratio of bias to the population value by 100:

$$\mathrm{PRB} = \frac{\bar{\hat{\theta}} - \theta}{\theta} \times 100\% \tag{4}$$

In this equation, θ is the population value and $\bar{\hat{\theta}}$ is the mean estimated value. The mean estimated value $\bar{\hat{\theta}}$ was found by computing the mean of the 1,000 generated point estimates. Because different parameter values were used for different functions, PRB is a better indicator of the amount of bias than simply subtracting the population value from the mean estimated value because it takes into account the size of the original parameter value in its calculation. PRB estimates also indicate the direction of bias, with positive values indicating estimates greater than population values and negative values indicating estimates less than population values. Means per condition were used to generate plots. Individual PRB values, i.e., all 1,000 values for each parameter estimate per condition, were used to conduct analyses of variance (ANOVAs) across the conditions.

TABLE 4
ANOVA Results for the PRB of the Polynomial Functions

Parameter	Source	F	Partial η^2	p
a_1	(A) Spacing of occasions	0.38	.00	.82
	(B) Number of occasions	0.47	.00	.70
	(A × B) Interaction	0.76	.00	.69
b_1	(A) Spacing of occasions	0.94	.00	.44
	(B) Number of occasions	0.07	.00	.97
	(A × B) Interaction	0.64	.00	.81
a_2	(A) Spacing of occasions	2.26	.00	.06
	(B) Number of occasions	8.10	.00	<.001
	(A × B) Interaction	2.31	.00	<.01
b_2	(A) Spacing of occasions	2.92	.00	.02
	(B) Number of occasions	7.80	.00	<.001
	(A × B) Interaction	1.56	.00	.10
c_2	(A) Spacing of occasions	2.17	.00	.07
	(B) Number of occasions	6.31	.00	<.001
	(A × B) Interaction	1.21	.00	.27
a_3	(A) Spacing of occasions	0.36	.00	.83
	(B) Number of occasions	3.50	.00	.01
	(A × B) Interaction	0.60	.00	.84
b_3	(A) Spacing of occasions	0.13	.00	.97
	(B) Number of occasions	0.97	.00	.41
	(A × B) Interaction	0.65	.00	.80
c_3	(A) Spacing of occasions	0.27	.00	.90
	(B) Number of occasions	0.69	.00	.56
	(A × B) Interaction	0.71	.00	.74
d_3	(A) Spacing of occasions	0.54	.00	.71
	(B) Number of occasions	0.60	.00	.61
	(A × B) Interaction	0.77	.00	.68

Note. df between = 19; *df* within = 19,980; Significant if $p < .05$.

TABLE 5
ANOVA Results for the SE of the Polynomial Functions

Parameter	Source	F	Partial η^2	p
a_1	(A) Spacing of occasions	6697.72	.57	<.001
	(B) Number of occasions	10763.30	.62	<.001
	(A × B) Interaction	113.07	.06	<.001
b_1	(A) Spacing of occasions	47097.50	.90	<.001
	(B) Number of occasions	138329.41	.95	<.001
	(A × B) Interaction	1297.53	.44	<.001
a_2	(A) Spacing of occasions	11234.03	.69	<.001
	(B) Number of occasions	7433.06	.52	<.001
	(A × B) Interaction	297.98	.15	<.001
b_2	(A) Spacing of occasions	12024.98	.71	<.001
	(B) Number of occasions	125542.02	.95	<.001
	(A × B) Interaction	2848.17	.63	<.001
c_2	(A) Spacing of occasions	11247.21	.69	<.001
	(B) Number of occasions	125511.87	.95	<.001
	(A × B) Interaction	2527.14	.60	<.001
a_3	(A) Spacing of occasions	416.25	.08	<.001
	(B) Number of occasions	310.85	.04	<.001
	(A × B) Interaction	25.42	.02	<.001
b_3	(A) Spacing of occasions	15630.11	.76	<.001
	(B) Number of occasions	185486.48	.97	<.001
	(A × B) Interaction	5214.69	.76	<.001
c_3	(A) Spacing of occasions	21841.32	.81	<.001
	(B) Number of occasions	245365.62	.97	<.001
	(A × B) Interaction	8678.27	.84	<.001
d_3	(A) Spacing of occasions	22046.92	.82	<.001
	(B) Number of occasions	247164.39	.97	<.001
	(A × B) Interaction	9141.38	.85	<.001

Note. df between = 19; *df* within = 19,980; Significant if $p < .05$.

Standard Error (SE) and Empirical Standard Deviation (ESD)

The SE was used as a measure of efficiency, with smaller SE indicating better estimator performance. Averages across the 1,000 estimates were used to generate the plots, and individual SE values were used to conduct ANOVAs across the conditions. The ESD was obtained as a secondary statistic and was calculated by averaging the squared differences between the individual estimated values $\hat{\theta}_i$ and the mean estimate $\bar{\hat{\theta}}$ and taking the square root of the result. It is calculated by the equation:

$$\mathrm{ESD} = \sqrt{\frac{1}{T}\sum_i (\hat{\theta}_i - \bar{\hat{\theta}})^2} \quad (5)$$

where T is the number of repetitions used. Higher ESD indicates poorer estimator performance.

Comparing Results Across Conditions

A series of plots and ANOVAs were used to compare how the number and spacing of measurement occasions affected the accuracy and efficiency of the parameter estimates. Plots are labeled by the parameter estimate and corresponding function. The number of occasions was plotted on the horizontal axis, and the outcome measure was plotted on the vertical axis. Then, a line was plotted for each measurement occasion spacing condition. By using this system, a total of 48 plots were generated in R (16 parameter values and 3 outcome measures). These plots were used to detect general patterns. To conserve space, plots of significant interest are depicted here. Next, we conducted 32 ANOVAs (16 parameter values for the PRB and SEs) to determine which of the manipulated factors (in particular, measurement interval and measurement occasion spacing) affected key outcomes. Finally, important patterns and themes are emphasized in the discussion. Complete results are included in the online appendices.

RESULTS

Linear Function (First Degree Polynomial)

Because of the large number of analyses run in the simulation, ANOVA results and effect sizes for the polynomial functions (i.e., linear, quadratic, and cubic) are included in

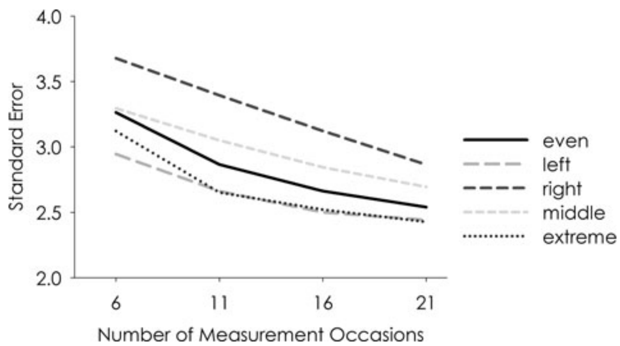

FIGURE 3 The SE for parameter a_1 of the linear function.

Table 4 (PRB) and Table 5 (SE).[3] Interaction effects are discussed generally and only for theoretically relevant findings. Refer to the online appendices for the complete ANOVA results. Results for the linear function indicated that neither the number of measurement occasions nor the spacing of measurement occasions was associated with PRB for either parameter. However, both the number and spacing of measurement occasions were associated with SE for both parameters. For parameter a_1, the intercept, increasing the number of measurement occasions was associated with increased efficiency for all conditions. The most efficient spacing condition was left, followed by extreme, even, middle, and finally right spacing. Left spacing was more efficient than extreme spacing in the condition with 6 measurement occasions but was no more efficient in the conditions with 11, 16, and 21 measurement occasions. For parameter b_1, which is the linear slope, increasing the number of measurement occasions was associated with smaller SE. In addition, measurement occasion spacing was associated with efficiency. The most efficient spacing condition was extreme, followed by even. Left and right spacing conditions were equally efficient, and the middle condition was the least efficient. The results for the SE of a_1 are depicted in Figure 3.

Quadratic Function (Second Degree Polynomial)

Parameter a_2 is the intercept, b_2 controls the linear slope where $x = 0$, and c_2 controls the rate at which the linear slope changes. Results indicated that the number of measurement occasions was positively associated with accuracy and efficiency for all parameters, although effect sizes for PRB were close to zero. Measurement occasion spacing was associated with PRB for parameter b_2 but was not significantly associated with PRB for parameters a_2 or c_2. Specifically for b_2, extreme spacing was more accurate than middle; however, the size of this effect was close to zero. In contrast, measurement occasion spacing was associated with SE for all parameters. For a_2, the most efficient spacing conditions were left, extreme, even, middle, and right spacing conditions, respectively. For b_2, left spacing was the most efficient, followed by even, middle, right, and extreme. Finally for c_2, right, left, and even were equal to each other and the most efficient, followed by middle and then extreme.

Cubic Function (Third Degree Polynomial)

Results from the simulation showed that the spacing of measurement occasions and the number of measurement occasions were not associated with PRB, with the exception that the number of measurement occasions was positively associated with accuracy for the intercept term a_3, though again the size of this effect was close to zero. Results also showed that increasing the number of measurement occasions was associated with greater efficiency for all parameters. For a_3, extreme spacing was the most efficient, followed by left, even, middle, and right spacing. Parameter b_3 indicates the slope where $x = 0$. For this parameter, the most efficient spacing conditions were extreme, even, middle, left, and then right. Similarly, for parameter c_3, which controls the rate at which the linear slope changes where $x = 0$, extreme was the most efficient, followed by even, middle, left, and right, Finally, for d_3, which controls the rate at which the quadratic slope changes, the most efficient conditions were extreme, even, and middle, respectively. Right and left performed the worst overall and equally well to each other.

Power Function

ANOVA results and effect sizes for the intrinsically nonlinear functions (i.e., power, exponential, and Gompertz) are included in Table 6 (PRB) and Table 7 (SE). In the power function, parameter a_4 represents the value of Y when $x = 1$, and parameter b_4 controls the concavity of the function. Increasing the number of measurement occasions was generally associated with lower PRB and lower SE for both parameters. For parameter a_4, the most accurate and efficient conditions were left, even, middle, extreme, and right, respectively. Left was more accurate than even with 6 and 11 measurement occasions but was not more accurate with 16 and 21 measurement occasions. For parameter b_4, the most accurate conditions were left, even, middle, extreme, and right, and the most efficient conditions were left, even, middle/extreme, and right. Results for the PRB of parameter a_4 are depicted in Figure 4.

Exponential Function

Parameter a_5 is the intercept of the exponential function, and parameter b_5 controls the rate at which the asymptote (0) is approached. Increasing the number of measurement occasions was associated with increased accuracy and efficiency for both parameters. For both a_5 and b_5, the most accurate conditions were right, even/middle, extreme, and then left.

[3]ESD and SE results aligned closely across nearly all of the conditions, and thus, results for SE are reported throughout. ESD values are available in the online appendices.

TABLE 6
ANOVA Results for the PRB of the Intrinsically Nonlinear Functions

Parameter	Source	F	Partial η^2	p
a_4	(A) Spacing of occasions	614.71	.11	<.001
	(B) Number of occasions	3850.21	.37	<.001
	(A × B) Interaction	131.71	.07	<.001
b_4	(A) Spacing of occasions	1004.57	.17	<.001
	(B) Number of occasions	6745.58	.50	<.001
	(A × B) Interaction	260.36	.14	<.001
a_5	(A) Spacing of occasions	277.53	.05	<.001
	(B) Number of occasions	1861.85	.22	<.001
	(A × B) Interaction	25.97	.02	<.001
b_5	(A) Spacing of occasions	455.52	.08	<.001
	(B) Number of occasions	3010.70	.31	<.001
	(A × B) Interaction	44.21	.03	<.001
a_6	(A) Spacing of occasions	6402.34	.56	<.001
	(B) Number of occasions	5658.48	.46	<.001
	(A × B) Interaction	6404.01	.79	<.001
b_6	(A) Spacing of occasions	3427.69	.40	<.001
	(B) Number of occasions	1429.08	.18	<.001
	(A × B) Interaction	3071.11	.65	<.001
c_6	(A) Spacing of occasions	4896.61	.50	<.001
	(B) Number of occasions	2531.91	.28	<.001
	(A × B) Interaction	4547.94	.74	<.001

Note. df between = 19; *df* within = 19,980; Significant if $p < .05$.

For a_5, right, middle, and even spacing were equally efficient, followed by extreme and then left. For b_5, the most efficient spacing conditions were right, middle, even, extreme, and left, respectively. For even spacing and parameter b_5, 11 measurement occasions were more accurate than 6, and 16 measurement occasions were more accurate than 11; however, there was not a significant difference between 16 and 21 measurement occasions. Results for the PRB of parameter b_5 are included in Figure 5.

TABLE 7
ANOVA Results for the SE of the Intrinsically Nonlinear Functions

Parameter	Source	F	Partial η^2	p
a_4	(A) Spacing of occasions	1463.95	.23	<.001
	(B) Number of occasions	15097.05	.69	<.001
	(A × B) Interaction	88.30	.05	<.001
b_4	(A) Spacing of occasions	10665.46	.68	<.001
	(B) Number of occasions	113319.18	.94	<.001
	(A × B) Interaction	1483.62	.47	<.001
a_5	(A) Spacing of occasions	1247.12	.20	<.001
	(B) Number of occasions	1810.64	.21	<.001
	(A × B) Interaction	958.13	.37	<.001
b_5	(A) Spacing of occasions	2086.83	.29	<.001
	(B) Number of occasions	4140.88	.38	<.001
	(A × B) Interaction	1305.42	.44	<.001
a_6	(A) Spacing of occasions	217032.91	.98	<.001
	(B) Number of occasions	289380.82	.98	<.001
	(A × B) Interaction	193039.32	.99	<.001
b_6	(A) Spacing of occasions	52102.78	.91	<.001
	(B) Number of occasions	51266.81	.89	<.001
	(A × B) Interaction	8994.44	.84	<.001
c_6	(A) Spacing of occasions	44608.85	.90	<.001
	(B) Number of occasions	57946.02	.90	<.001
	(A × B) Interaction	9272.27	.85	<.001

Note. df between = 19; *df* within = 19,980; Significant if $p < .05$.

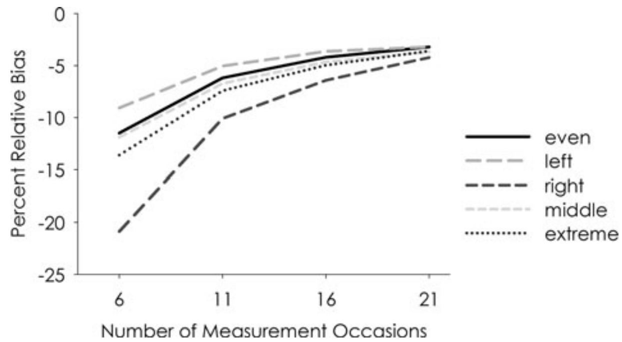

FIGURE 4 The PRB for parameter a_4 of the power function.

Gompertz Function

In the Gompertz function, a_6 controls the upper asymptote, b_6 shifts the curve horizontally, and c_6 controls the steepness of the curve. The number of measurement occasions was positively associated with accuracy and efficiency for a_6, b_6, and c_6. For all parameters, middle, right, and even spacing were equally accurate, followed by extreme and then left. For parameter a_6, the most efficient conditions were right, middle, even, extreme and then left. For parameters b_6 and c_6, the most efficient conditions were middle, right, even, left, and extreme, respectively. For the efficiency of parameter c_6, differences between middle and right were significant with 6 measurement occasions but were not significant with 11, 16, or 21 measurement occasions. SE results for b_6 are depicted in Figure 6.

Summary of Results

Increasing the number of measurement occasions was associated with increased efficiency for all parameters and was associated with increased accuracy for the parameters of the

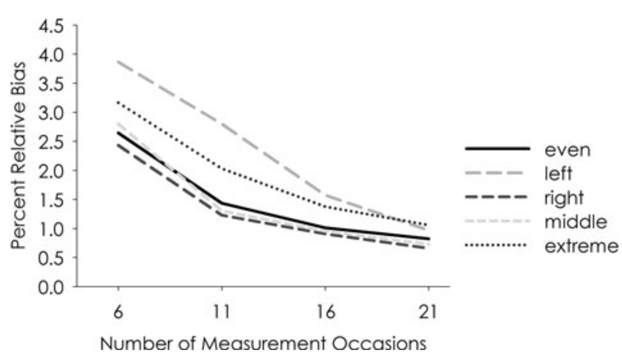

FIGURE 5 The PRB of parameter b_5 of the exponential function.

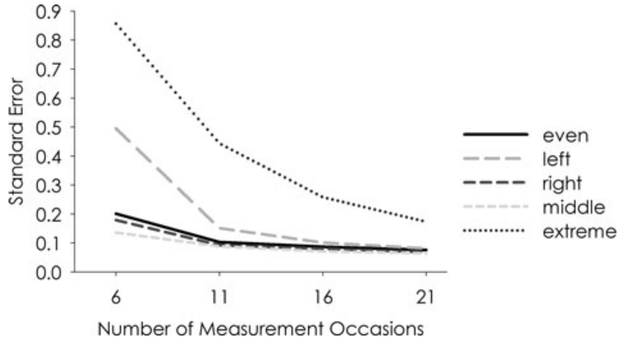

FIGURE 6 The SE of parameter b_6 of the Gompertz function.

intrinsically nonlinear functions. Although the number of measurement occasions was associated with PRB for some parameters of the intrinsically linear functions, the sizes of these effects were close to zero. Results for the intercept (a_1) and slope (b_1) terms of the linear function were generally consistent with our hypotheses; for a_1, left spacing was the most efficient, followed by extreme, even, middle, and right, and for b_1, extreme spacing was the most efficient, followed by even, left/right, and middle. Patterns for the quadratic and cubic functions were more complex. Results did indicate, however, that for intercepts, the extreme and left conditions were generally more efficient than the other spacing conditions.

For the intrinsically nonlinear functions, results were consistent with our hypothesis in that concentrating occasions of measurement where the curvature of the function was greatest tended to result in the best recovery of population parameters, although this was not uniformly true. For the power function, curvature was maximized on the left, and left outperformed the other spacing conditions. For the exponential function, curvature was maximized on the right, and right spacing was generally the most accurate and efficient. For Gompertz, curvature was maximized in the middle; results showed that for all parameters, middle, even, and right spacing were equally accurate and that for parameters b_6 and c_6, middle was the most efficient. Results for parameter a_6 showed that right spacing was the most efficient and that middle was the second most efficient.

DISCUSSION

Number of Measurement Occasions

The finding that increasing the number of measurement occasions is associated lower SE and PRB is not surprising. In most contexts, as researchers add observations, they can expect to obtain more efficient and accurate estimates of population parameters. However, it is important to note that adding more measurement occasions will not always add meaningfully to the accuracy or efficiency of an estimate. For example, in the case of even spacing for b_5, there was a “diminishing returns” effect. That is, each time the number of occasions was increased by a 5-point increment, the associated increase in accuracy decreased. Differences in accuracy were significant from 6 to 11 measurement occasions and from 11 to 16; however, differences from 16 to 21 were not significant (see Figure 5). Thus, there is a point beyond which adding more measurement occasions may no longer usefully improve the parameter estimate.

Given the cost associated with increasing the number of measurement occasions, it is important that researchers not add measurement occasions to their studies arbitrarily. Instead, researchers should determine *a priori* whether adding more measurement occasions would add meaningfully to their ability to detect or accurately estimate an effect. Determining when to add more occasions is difficult and depends upon a variety of factors, including the functional form, the parameter values, and the amount of error. To address this problem, R code adapted from our simulation study is provided in the online appendices. The use of this code will be discussed in the implications section.

Spacing of Measurement Occasions

Results suggest that the spacing of measurement occasions has implications for the efficiency and accuracy of parameter estimation. For the linear function, results were generally consistent with hypotheses in that extreme spacing led to greater efficiency than the other conditions for the slope term and that left and extreme spacing performed better than the other conditions for the intercept term. These results are consistent with other research (McClelland & Judd, 1993) indicating that extreme spacing provides greater efficiency for detecting interaction effects. As expected, results were more complicated for the quadratic and cubic functions. However, it did appear that the extreme and left conditions performed better than the other conditions for estimating intercepts.

For the intrinsically nonlinear functions, concentrating occasions of measurement near the areas of greatest curvature tended to result in better accuracy and efficiency. However for parameter a_6, which controls the upper asymptote, concentrating the measurement occasions on the right side, nearest to where the asymptote approaches 0, resulted in the lowest SE. More research should be conducted to examine the effect of temporal design on parameters associated with asymptotes. The results of the simulation suggest that temporal design is an important component of designing longitudinal studies and has implications for model estimation in applied research.

Guidelines

From these findings, five preliminary guidelines were formulated. More research should be conducted to support or refute these guidelines:

(1) Adding measurement occasions may increase accuracy for the intrinsically nonlinear functions.
(2) Adding more measurement occasions generally increases efficiency.
(3) When adding measurement occasions, there is a diminishing returns effect, so researchers should think carefully before adding measurement occasions. Using the code provided in the online appendices may aid researchers in making this decision.
(4) For the slope of the linear function, extreme spacing is ideal. For intercepts of the polynomial functions, left (toward the origin) and extreme spacing tend to perform well relative to other types of spacing.
(5) For intrinsically nonlinear functions, it is beneficial to concentrate measurement occasions where curvature is the greatest, although this may not be the case for all parameters.

Implications

Researchers typically pay careful attention to research design. However, the temporal aspects of design often are overlooked. Past research on this topic has demonstrated that suboptimal measurement intervals can decrease accuracy and mislead researchers (Cohen, 1991; Collins, 1996a; Collins & Graham, 1991; Nesselroade & Jones, 1991; Willett, 1989; Windle & Davies, 1999). This study is consistent with past research in that it demonstrates that relatively larger measurement intervals are associated with relatively less accuracy and efficiency. In addition, this study builds on past work by demonstrating that the spacing of measurement intervals, and not simply the number of occasions or the duration of the study, has implications for the accuracy and efficiency of parameter estimates in models applied to longitudinal data.

Several methodologists have suggested that researchers justify their choice of measurement interval (e.g., Collins & Graham, 2002; Nesselroade & Jones, 1991). However, making these justifications is difficult because few empirically based guidelines exist. The present study represents an attempt to generate such guidelines for investigators to use in their own research. Although researchers are sometimes advised to sample most often where change is occurring, a literature search revealed no empirical basis for this claim. The current study gives preliminary evidence that this advice was correct for most parameters tested in our study. In addition, this effect was observed for the cubic function. Although this function is intrinsically linear, the location of greatest curvature did coincide with the most efficient spacing condition, suggesting this advice may also be correct for intrinsically linear functions in which curvature is not constant across the range of the function, though this question should be examined more directly in future work.

It is important to note that the location of greatest curvature differs according to the specific functional form and parameter values used. For this reason, customizable R code is provided in the online appendices. By using this code, researchers can input their own information to determine the number of measurement occasions to include in their studies, as well as how to optimally space those measurement occasions. To use these guidelines, researchers must first make a few assumptions. For example, even if researchers use simulation code to determine the optimal measurement intervals, they must still determine to the best of their ability (1) what type of function is most appropriate for modeling their data, (2) what reasonable parameter values are, and (3) the span over which the change will occur. First, note that these steps are no different from those used in a power analysis. In fact, the procedure can be thought of as a kind of power analysis in which the question is not how many cases are required to achieve a given level of power, but rather how many measurement occasions are necessary—and how they should be timed—to minimize bias and maximize efficiency for theoretically relevant parameters. This technique is no different than current methods, except that it goes one step further by providing quantitatively based guidelines to improve other aspects of the design process. Second, the present technique should not be used in isolation. The importance of theory and previous research is paramount. Without theory, determining how to time measurement occasions is difficult, if not impossible. For areas without a strong theoretical background, the microgenetic method of sampling could be used instead, at least as an initial step. Then, once research areas become more established, more efficient approaches could be employed.

Limitations and Future Directions

Despite the strengths of the study, several limitations must be addressed. As previously discussed, this method should be used only in conjunction with theory and past research. Also, this study focused on linear and nonlinear mixed effects models because such models are commonly used to model change over time. Other methods designed for modeling longitudinal data (e.g., time series models and dynamical systems models) could benefit from investigations into the relative merits of different temporal designs. Next, a limited number of functions were examined. Although it is likely that the results generalize across different types of nonlinear functions, this cannot be known with certainty without investigating more functions. In addition, changing the spacing of measurement occasions may invalidate effect size measures, though it leaves unstandardized parameter estimates unbiased. Effect size measures assume that random sampling was used; yet in longitudinal research, one rarely samples occasions at random, but rather according to a predefined regimen. It is also important to note that these guidelines may not be useful in situations where oversampling might create fatigue or attrition; similarly, the difficulty and complexity of planning longitudinal studies may limit the feasibility of applying these guidelines in particular settings.

It is customary to present an applied example to illustrate new methodological procedures, both to increase accessibility to a general audience and to provide a template for others to follow. However, we do not provide an empirical illustration. In a simulation study, the functional form of the correct trajectory model is known; given this provisional correct functional form, the consequences of spacing for parameter bias and efficiency may be determined while ignoring other factors that may be relevant in practice. In real empirical settings, however, the parametric model chosen by the researcher is never the actual data-generating model. Small departures of the observed data from the hypothesized trajectory can exert large effects on standard errors, making it difficult to separate the effects of model error and spacing of measurements on bias and efficiency, even for close-fitting models. However, for purposes of planning a study, we believe our proposed method is justified. Just as the researcher must posit a provisional fully parameterized model in power analysis—even though the researcher knows this model is not literally correct—in planning temporal design the researcher must condition the procedure on a given parametric trajectory model being true. Regarding the second usual motivation for including an empirical example—providing a template for other researchers—our R code may be adapted to researchers' own measurement schedules and hypothesized functional forms to determine the optimal timing of measurement occasions conditioned on those choices.

Our findings are consistent with the literature on *optimal design* (e.g., Fedorov & Leonov, 2014; Pronzato & Pázman, 2013). Specifically, *c*-optimal designs yield the most efficient (minimum variance) estimate of a linear function of model parameters. Minimization of the SE of a single parameter is a special case of *c*-optimality wherein the linear combination of parameters is simply one chosen parameter. Future work may pursue algorithmic or analytic methods for determining optimal temporal designs for a provisional model given a set number of occasions, a chosen set of parameter values, and distributional assumptions.

Finally, because the parameter values were chosen *a priori*, some conditions had uniformly small SE and PRB. Although the study is informative in that it depicts a general pattern of results, it is important to remember that lack of definite results does not necessarily mean that there was not an underlying effect. Thus, the findings should be interpreted with this in mind. To minimize inaccurate interpretations, general trends were focused upon in the discussion section. However, to gain confidence in the results, future work should ensure that all conditions have adequate variability. The scope and sheer variety of nonlinear functions, temporal designs, and research contexts in which researchers may be interested in applying nonlinear models are far too large to capture in a single simulation study. The R code we provide in the online appendix should help researchers make informed decisions in specific contexts.

CONCLUSIONS

Previous research has demonstrated the importance of temporal design, and this study supports these findings. Furthermore, this study goes beyond past research by demonstrating how the spacing of measurements can affect accuracy and efficiency. Longitudinal research is an especially valuable tool in psychology. It is necessary for determining causality and allows researchers to pose more complex questions about variables under study than they otherwise would be able to ask. However, longitudinal research is expensive and time-consuming. In academic environments where the duration of a study is often limited by external factors, conducting longitudinal research is not always practical. The current study confronts this problem by exploring methods for improving the efficiency and utility of longitudinal research. Finally, in addition to providing guidelines for choosing measurement intervals, we provide R code that allows researchers to conduct customized simulations for their own research projects. Although the present approach is valuable, it should always be used in conjunction with other techniques. Empirical guidelines, past research, and theory should be used in conjunction to create a cyclical, mutually reinforcing system of research design.

FUNDING

This article is based on the first author's senior honors thesis at the University of Kansas and was partially supported by a NSF GRFP Grant No. DGE-0937362 (Timmons, PI) and a University of Kansas Undergraduate Research Award.

REFERENCES

Adolph, K. E., Robinson, S. R., Young, J. W., & Gill-Alvarez, F. (2008). What is the shape of developmental change? *Psychological Review*, *115*, 527–543.

Burchinal, M., & Appelbaum, M. I. (1991). Estimating individual developmental functions: Methods and their assumptions. *Child Development*, *62*, 23–43.

Chaiton, M., Cohen, J., O'Loughlin, J., & Rehm, J. (2010). Use of cigarettes to improve affect and depressive symptoms in a longitudinal study of adolescents. *Addictive Behaviors*, *35*, 1054–1060.

Cheadle, J., & Goosby, B. (2010). Birth weight, cognitive development, and life chance: A comparison of siblings from childhood into early adulthood. *Social Science Research*, *39*, 570–584.

Cohen, P. (1991). A source of bias in longitudinal investigations of change. In L. M. Collins & J. L. Horn (Eds.), *Best methods for analysis of change* (pp. 18–25). Washington DC: American Psychological Association.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediation models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, *112*, 558–577.

Collins, L. M. (1996a). Measurement of change in research on aging: Old and new issues from an individual growth perspective. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (4th ed., pp. 38–56). San Diego: Academic Press.

Collins, L. M. (1996b). Is reliability obsolete? A commentary on "Are simple gain scores obsolete?" [Peer commentary on the paper "Are simple gain scores obsolete?" by R. H. Williams and D. W. Zimmerman]. *Applied Psychological Measurement*, *20*, 289–294.

Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. *Annual Review of Psychology*, *57*, 505–528.

Collins, L. M., & Graham, J. W. (1991). Comment on "A source of bias in longitudinal investigations of change" [Peer commentary on the paper "A source of bias in longitudinal investigation of change" by P. Cohen]. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 26–30). Washington DC: American Psychological Association.

Collins, L. M., & Graham, J. W. (2002). The effect of timing and spacing of observations in longitudinal studies of tobacco and other drug use: Temporal design considerations. *Drug and Alcohol Dependence*, *68*, 85–96.

Collins, L. M., Graham, J. W., & Flaherty, B. P. (1998). An alternative framework for defining mediation. *Multivariate Behavioral Research*, *33*, 295–312.

Fedorov, V. V., & Leonov, S. L. (2014). *Optimal design for nonlinear response models*. Boca Raton, FL: CRC Press.

Gollob, H. F., & Reichardt, C. S. (1991). Interpreting and estimating indirect affects assuming time lags really matter. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 243–259). Washington DC: American Psychological Association.

Graham, J. W., Taylor, B. J., & Cumsille, P. E. (2001). *Planning missing-data designs in analysis of change*. Washington, DC: American Psychological Association.

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, *11*, 323–343.

Hertzog, C., & Nesselroade, J. R. (2003). Assessing psychological change in adulthood: A review of methodological issues. *Psychology and Aging*, *18*, 639–657.

Kyle, F., & Harris, M. (2010). Predictors of reading development in deaf children: A 3-year longitudinal study. *Journal of Experimental Child Psychology*, *107*, 229–243.

McArdle, J. J., & Woodcock, R. W. (1997). Expanding test–retest designs to include developmental time-lag components. *Psychological Methods*, *2*, 403–435.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties in detecting interactions and moderation effects. *Psychological Bulletin*, *114*(2), 376–390.

Nesselroade, J. R. (1991). Interindividual differences in intraindividual change. In L. M. Collins, & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 92–105). Washington, DC: American Psychological Association.

Nesselroade, J. R., & Boker, S. M. (1994). Assessing constancy and change. In T. F. Heatherton, & J. L. Weinberger (Eds.), *Can personality change?* (pp. 121–147). Washington, DC: American Psychological Association.

Nesselroade, J. R., & Jones, C. J. (1991). Multi-model selection effects in the study of adult development: A perspective on multivariate, replicated, single-subject, repeated measures designs. *Experimental Aging Research*, *17*, 21–27.

Nyquist, H. (1928). Certain topics in telegraphic transmission theory. *Proceedings of the IEEE*, *47*, 617–644.

Ployhart, R. E., & Vandenberg, R. J. (2010). Longitudinal research: The theory, design, and analysis of change. *Journal of Management*, *36*, 94–120.

Pronzato, L., & Pázman, A. (2013). *Design of experiments in nonlinear models: Asymptotic normality, optimality criteria and small-sample properties*. New York: Springer.

Rast, R., & Hofer, S. M. (2014). Longitudinal design considerations to optimize power and detect variances and covariances among rates of change: Simulation results based on actual longitudinal studies. *Psychological Methods*, *19*, 133–154.

Seber, G. A. F., & Wild, C. J. (1989). *Nonlinear regression*. New York: Wiley.

Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IEEE*, *37*, 10–21.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental nonexperimental studies: New procedures and recommendations. *Psychological Methods*, *7*, 422–445.

Siegler, R. S. (2006). Microgenetic analysis of learning. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Cognition, perception, and language* (6th ed., Vol. 2, pp. 464–510). New York: Wiley.

Sit, V., & Poulin-Costello, M. (1994). *Catalog of curves for curve fitting*. Biometrics Information, Handbook 4, B.C. Ministry of Forests, Victoria, BC.

Thelen, E., & Ulrich, B. D. (1991). Hidden skills: A dynamic systems analysis of treadmill stepping during the first year. *Monographs of the Society for Research in Child Development*, *56* (1, Serial 223).

Willett, J. B. (1989). Some results on reliability for the longitudinal measurement of change: Implications for the design of studies of individual growth. *Educational and Psychological Measurement*, *49*, 587–602.

Windle, M., & Davies, P. T. (1999). Developmental research and theory. In K. E. Leonard & H. T. Blane (Eds.), *Psychological theories of drinking and alcoholism* (2nd ed., pp. 164–202). New York: Guilford.

Zhang, Z., & Wang, L. (2009). Statistical power analysis for growth curve models using SAS. *Behavior Research Methods*, *41*, 1083–1094.