J Bus Psychol (2011) 26:413–422
DOI 10.1007/s10869-011-9209-6

The “Quick Start Guide” for Conducting and Publishing Longitudinal Research

Robert E. Ployhart · Anna-Katherine Ward

Published online: 15 February 2011
© Springer Science+Business Media, LLC 2011

Abstract Consideration of temporal issues adds precision and insight to our theories, yet most organizational and applied psychological research is based on cross-sectional designs. Calls for longitudinal research have become common in leading journals, but the existing literature provides little prescriptive guidance to overcome the many challenges of this type of research. This article provides a concise summary of challenges to address when theorizing, designing, conducting, and publishing longitudinal research. We structure the article around 12 judgment calls that typically confront researchers when conducting longitudinal studies. We respond to these judgment calls using theory and findings from the relevant literatures, as well as our own experience in designing and conducting longitudinal research across many scholarly domains. Included in these judgment calls is an emphasis on presenting and framing one’s study for publication. We challenge readers to develop theory that addresses the *when*, *why*, and *duration* of change, and to test the theory with the appropriate longitudinal methods. This “quick start guide” is intended to serve as a useful reference for authors and reviewers at any level of methodological expertise.

Keywords Longitudinal research · Study design · Research methods · Time · Modeling change

The last 20 years has seen a dramatic rise in the attention paid to longitudinal theory, methodology, and research. Calls for longitudinal research are now common in leading journals, and funding agencies are increasingly favoring studies that have a longitudinal flavor. And why shouldn’t they? Simply put, we can provide much more specific, actionable advice when we know when, and for how long, an effect will last. For example, when we take a prescription for an ailment, we are told specifically how many doses to take within a certain time frame (and when to seek further medical assistance). Studies that lack insight about the duration or timing of effects and relationships offer little prescriptive advice for practitioners and the general public.

In turn, the precision of our theories is enhanced by considering temporal issues, if for no other reason than most of our theories do not currently do so. At best, most theories treat time in an implicit, casual manner; at worst, time is wholly ignored (Mitchell and James 2001). Further, there is growing evidence that cross-sectional findings may be different from longitudinal findings. Consider some examples. Cross-sectional tests of mediation may be severely biased, to the point where reversals in sign among relationships can occur (Maxwell and Cole 2007). Relationships among two dynamic constructs can be different (and frequently stronger) than examining those same constructs cross sectionally (Ployhart et al. 2009). Mean levels of a predictor (whether static or an average over time) do not predict change in outcomes as strongly as change in the predictor (Chen et al. 2011).

Thus, organizational scholars are presented with a difficult quandary. On the one hand, the field is calling for more longitudinal research, but on the other hand, most theories and empirical research are based on cross-sectional thinking and designs. Researchers trying to pursue longitudinal research will face many challenges. These challenges include how to best theorize change and time, how to best design the longitudinal study, and how to analyze it

R. E. Ployhart (✉) · A.-K. Ward
Darla Moore School of Business, University of South Carolina,
Columbia, SC 29208, USA
e-mail: ployhart@moore.sc.edu

appropriately. Then, authors must face the hurdles of a review process where reviewers and journal editors may have limited experience with longitudinal research and little prior research to consider as precedent.

The purpose of this article is to provide some guidance for authors, and possibly even reviewers, struggling to make sense of longitudinal research. We provide a concise, non-technical primer into longitudinal research because there are many more comprehensive treatments of the topic; most of them technical (e.g., Bollen and Curran 2006; Singer and Willett 2003), some theoretical (e.g., George and Jones 2000), and a few that cover both theory and methods (Mitchell and James 2001; Ployhart and Vandenberg 2010). We will cite the key sources as appropriate, but not get into the details. Further, rather than a purely descriptive review, we have structured this article around the typical judgment calls researchers and reviewers usually face. We have identified 12 judgment calls based on many years of experience with longitudinal research as an author and reviewer, juxtaposed against gaps in the existing literature. That is, we have identified those issues where there is little prescriptive guidance in the existing literature.

To provide a context to the judgment calls, we will consider the challenges faced by a researcher trying to understand how job satisfaction changes over time among new hires. The potential timeframe of this study is 1 year, and s/he has some flexibility in the number and spacing of measurement occasions. S/he would also like to know if the personality trait emotional stability predicts change in job satisfaction.

Judgment Calls About Conceptualization and Design

Do I Have a Longitudinal Study?

The term "longitudinal" gets tossed about so much that it is confusing to know what is a longitudinal study. Do you think you know? Let's start with a quiz. Considering our example above, which of the following is a longitudinal study?

a. The researcher measures job satisfaction several months after measuring emotional stability.
b. The researcher measures job satisfaction twice (several months apart).
c. The researcher measures job satisfaction thrice (several weeks apart).
d. The researcher measures job satisfaction twice (several months apart), and after measuring emotional stability.

The correct answer is "c." Two waves of data do not make a longitudinal study, nor does separating the timing of the predictor variable and outcome variable measurements (Rogosa 1995). With only two waves of data, it is difficult to disentangle true change from measurement error, and it is impossible to model more sophisticated or nonlinear forms of change because of constraints on degrees of freedom (in our experience, these models often fail to converge). Of course, one might conduct a pretest–posttest design and have some form of manipulation intervening between the measurement waves for one subgroup (e.g., training vs. control groups). Such a design is better than a cross-sectional design, but is focused primarily on determining whether the manipulation has an effect. It is not intended to provide detailed information about change and is limited with respect to modeling change and informing questions about duration (e.g., maintenance of training). A two-wave study may be better than a cross-sectional study, but it is not as good as a three-wave or multiple-wave study.

Thus, to study change in a variable or construct, one should collect at least three waves of data, and more waves are better (see Ployhart and Vandenberg 2010). Studies that separate the timing of measures without an explicit focus on understanding change are frequently doing so only to reduce method bias concerns (Conway and Lance 2010). Such studies were typically not designed to provide deep insights into change and temporal processes. Thus, when conceptualizing a longitudinal study, realize that it will require at least three repeated measurements on at least one variable.

It is important to note that the change models we discuss in this article are different from time-series (econometric) models. The former are dominant in the "micro" disciplines (e.g., psychology), while the latter are dominant in the "macro" disciplines (e.g., economics). Time-series models are most frequently used when there are numerous waves of repeated measures data (e.g., 15 or 20), whereas most applications of longitudinal modeling of the type we discuss in this article are usually based on less than 10 waves. For purposes of this article, we do not consider econometric models simply because they are usually less applicable for the kinds of data psychologists and micro HR/OB scholars have available.

How Do I Use Cross-Sectional Theory to Develop Interesting Longitudinal Hypotheses?

Mitchell and James (2001) provide a thoughtful review about how organizational theory and research often fail to incorporate temporal issues. Most theory doesn't consider *when* an effect is likely to occur or for what *duration*. If you pick most any theory in human resources, organizational behavior, and industrial/organizational psychology, you will be hard-pressed to find an answer to these basic questions. In turn, most researchers will present hypotheses that are simple longitudinal variants of cross-sectional hypotheses. They answer the questions of *when* and *duration* based on the data available, rather than being guided by theory. For example, a hypothesis "emotional stability is positively related to job satisfaction" becomes "emotional stability is positively related to job satisfaction over time." These are nearly identical hypotheses, and, consequently, the longitudinal hypothesis is not very compelling.

So how does one use theory that is silent on temporal issues to develop interesting longitudinal hypotheses? The answer lies in focusing on what is unique about change in a construct, versus its static representation. For example, why would job satisfaction change in the first place, and why is variance in change substantively different from variance of a static variable? Chen et al. (2011) offer a good illustration. They argued (and showed) that changes in job satisfaction influence turnover intentions more strongly than static levels of job satisfaction. Turnover intentions were lower when a person's job satisfaction was increasing over time, relative to a person whose satisfaction was decreasing over time, even when the latter person had higher mean satisfaction scores. If things are getting worse, you're more likely to leave, even if you have relatively high levels of satisfaction. As a different illustration, consider the scholar trying to understand how emotional stability relates to change in job satisfaction. If this researcher simply poses the question of whether emotional stability relates to job satisfaction over time, then it is not much more than a simple extension of prior cross-sectional research. However, suppose this researcher instead argues that job satisfaction will decline with time on the job due to unmet expectations and newcomer stress. Further, those with greater emotional stability should be able to better handle unmet expectations and stress, so the decline in job satisfaction is lessened for those with greater emotional stability. Finally, the researcher might propose that declining job satisfaction reaches some point where it levels off, and those more emotionally stable reach that "turning point" sooner. In each hypothesis, one must do more than simply reapply cross-sectional findings, one must explain why change is occurring and how the nature of change differs across levels of emotional stability.

Thus, focusing on change in a construct, and particularly (a) why it changes, (b) what causes the change, and (c) what results from the change, can help create novel and interesting hypotheses. In answering these questions, one is led naturally to confront issues relating to *when* and *duration*.

What Sample Size Do I Need?

The answer to this is easy—as large as you can get! But keep in mind that longitudinal designs can be very economical in terms of subjects versus statistical power. First, consider that in longitudinal research, the total sample size is the number of subjects times the number of repeated measurement occasions. For example, 100 subjects sampled five times produces 500 observations, but so does 20 subjects sampled 25 times. Second, within-subject designs and longitudinal research tend to have smaller residual or error terms relative to cross-sectional designs (Keppel 1991). In experimental designs, the smaller residual term reduces the amount of unexplained variance and hence increases the F test, all else equal. It is also worth emphasizing that more repeated measurements increases reliability (Willett 1989). Thus, the correct way to frame this question is how many *observations* are necessary, where observations are defined as subjects X repeated measurements. Sampling more time periods allows one to increase the number of observations and hence increase power and reliability. However, estimating statistical power is challenging because it is a function of the effect size, the number of subjects, the number of repeated measurements, the spacing between the measurements, reliability, and the type of curve being hypothesized (Raudenbush and Liu 2001). These factors are also interrelated, making it difficult to isolate which one will most affect power. For example, it is unlikely that adding subjects produces the same consequences as adding measurement occasions. We do not know of any research that speaks directly to this point, but we suspect that the relationships between subjects, repeated measurements, and statistical power are complex and nonlinear. Our suspicion is that, assuming minimum thresholds are met, adding more subjects will be more important for statistical tests of between-person effects, while adding more repeated measurements will be more important for statistical tests of within-person effects.

The complexity of this issue is the reason why there are not "simple" power tables available like those found for regression or ANOVA. Lacking such specific guidance, we offer a few suggestions. First, as expected, power is likely to increase with more repeated observations, but observations may be increased through either (a) sample size or (b) repeated observations. If one must choose between adding subjects versus measurement occasions, our recommendation is to first identify the minimum number of repeated measurements required to *adequately* test the hypothesized form of change and then maximize the number of subjects. When it is not possible to obtain more subjects, then fight for more repeated measurement occasions (and vice versa). Second, power increases with effect size, which means power increases as change over time is more rapid, and/or there is more variability in change over time. This is simply a longitudinal variation of range restriction; one needs variance to detect effects, and the larger the effect, the greater the statistical power. Third, the power to detect linear terms tends to be greater than the power to detect nonlinear terms. A linear term will frequently explain the most variance, and to ensure parsimony, most modeling approaches will test the significance of higher-order (nonlinear) terms *after* controlling for the lower-order (linear) term. Hence, a nonlinear term will usually have less variance to explain and will subsequently have a smaller effect size and lower statistical power for the significance test of the effect size. Fourth, power for detecting trends increases as the measurement periods capture the change (see next point). When a construct changes in only modest ways, one should study it for longer periods of time. See Raudenbush and Liu (2001) for more detail about this issue.

When Should I Administer Measures (and How Many)?

Many researchers believe that it is best when measurement occasions are evenly spaced (e.g., all participants are administered job satisfaction surveys at the same time, such as weekly or quarterly). Equal spacing of measurement occasions is something that can be controlled in laboratory experiments, but is usually beyond the control of field researchers. However, there is neither reason nor necessity for ensuring equal measurement occasions. Rather, it is much more important that the measurement occasions occur with enough *frequency* to be capable of detecting the kind of change hypothesized and that the occasions cover a reasonable duration of time (Mitchell and James 2001; Raudenbush and Liu 2001). For example, Fig. 1 shows hypothesized change in job satisfaction among new hires over a 1-year period. Notice that job satisfaction is initially high but then decreases and starts to level out. If one measured job satisfaction at months 1 and 12, a correct inference about declining job satisfaction could be made, but notice it would be a straight linear line and fail to capture the curvilinear trend. Adding a measurement occasion at month 6 would detect (but not perfectly model) the nonlinearity (which is why longitudinal research requires at least three time periods). However, notice the steepest decline in satisfaction occurs during the first 3 months. Should a researcher be limited to three time periods, the spacing of those periods is critical. Measuring job satisfaction for the first 3 months would overestimate the negative slope, relative to measuring job satisfaction in months 1, 6, and 12. Alternatively, measuring job satisfaction for the last 3 months would underestimate the negative slope. Thus, the measurement occasions should be placed at the times most theoretically interesting, with the caveat that the spacing cover a reasonable duration to detect the hypothesized form of change. See Mitchell and James (2001) and Ployhart and Vandenberg (2010) for detailed treatment of this issue.

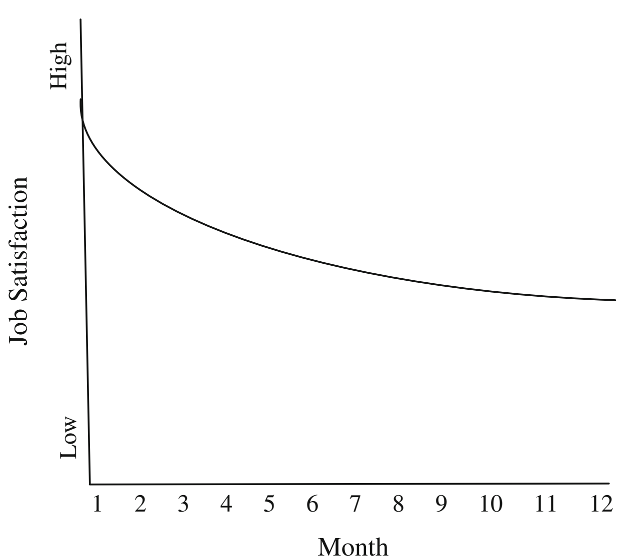

Fig. 1 Hypothetical change in job satisfaction for new hires

Of course, most of the time there is no theoretical guidance on when the measurement occasions should occur. We can offer a few suggestions, based on our experience, to address this situation. First, consider the nature of the phenomenon being investigated and determine whether there are "natural" measurement occasions. For example, financial data are usually collected monthly, quarterly, or yearly, and so studies that use these types of criteria may want to measure psychological predictors (e.g., job attitudes) on the same cycle. Second, conduct interviews or behavioral observations with relevant subject matter experts to determine a measurement schedule. In the job satisfaction example, it would be helpful to meet with employees of varying degrees of experience, to gauge how their job satisfaction seems to be changing. Indeed, such basic descriptive work could lead to new theoretical insights not proposed by existing theory and could potentially lead to separate publications. Third, carefully review related literatures that studied similar phenomena and pay careful attention to the limitations sections of those papers. Finally, talk with your colleagues before collecting data. We constantly run ideas past our friends, colleagues, students, and even spouses and kids, because they frequently offer alternative points of view we had not considered. Ultimately, you need a good justification for the timing and frequency of your measurement occasions, and the more evidence you can provide to show you considered these issues a priori, the stronger your study.

What Should I Do About Missing Data?

Expect and plan for it. Missing data are synonymous with longitudinal research. Like death and taxes, missing data is going to happen despite herculean efforts to avoid it. There are many excellent introductions (e.g., Anseel et al. 2010; Newman 2003; Palmer and Royall 2010) and detailed overviews (e.g., Allison 2001) of missing data. Here we review the topic a bit more from the perspective of the applied researcher: (a) what can one do in advance to reduce missing data and (b) what can one do to handle the missing data after the data are collected?

With respect to reducing missing data, all of the best practices of research design are still relevant (see Anseel et al. 2010, for several excellent recommendations). It is important to provide reasonable instructions and incentives for participation. Response rates will tend to be higher if participants complete the measures during working hours, so that they are being paid, but in a location where they can focus on the survey with confidentiality protected. It is important not to burden participants with too many repeated measurements, because they will either stop completing them or not give them much thought. One should always provide instructions so that participants know why they are being asked similar questions. Asking subject matter experts what they feel is a reasonable amount of time for the survey, and how many repeated measurements will be feasible, goes a long way toward enhancing response rates.

In practice, it is often impossible to survey participants as much as one might like. In this situation, one can spread the measurement occasions across different people. This is called *planned missingness*, because there will be missing data by design. In the job satisfaction example, suppose the researcher does not want to burden participants, so s/he might use only three repeated observations for *each* subject. However, different groups of subjects will get the survey at different times (groups may be determined randomly or based on substantive factors). For example, perhaps everyone gets the survey on months 1 and 12, but when they get the survey in between will differ by group, where one group gets the survey at month 3, another at month 6, and another at month 9. A *cohort-sequential design* (sometimes also called an *accelerated longitudinal design*) is a popular type of planned missing design when the interest is to study change over long periods of time but with a short study duration. For example, if a researcher wanted to study change in satisfaction over one's career, but doesn't want to wait 40 years to collect the data, s/he could approximate the change by sampling different cohorts (e.g., new hires, early career, late career) at the same time periods (e.g., months 1, 6, 12, etc.). Then, the researcher could piece together the trends within each cohort, so that career changes in satisfaction could be observed (of course, an assumption is that the cohorts are similar, which is questionable given generational differences in job satisfaction). Excellent illustrations of these methods are found in Graham et al. (1996, 2001) and Duncan et al. (1996). In general, planned missingness offers an efficient means to increase available data, yet cover the entire duration of interest.

With respect to handling missing data, the most important issue is to determine whether the data are missing at random. Simple approaches involve testing for differences in variables included in the study (usually at time 1). One can compare differences on demographic variables or substantive variables, across groups that have complete data versus those that are missing data at different times (e.g., a group that is missing time 1 scores, a group that is missing time 2 scores, etc.). In the job satisfaction example, one might compare the emotional stability scores or the time 1 job satisfaction scores between a group with complete data versus those who dropped out after the first wave of data collection. Sophisticated approaches involve testing for the randomness of the missing data (e.g., Little and Rubin 2002). If the data are not missing at random, biased parameter estimates can occur. If the data are missing at random, then it is usually possible to include the missing data without introducing bias to the parameter estimates. In practice, we have found the simpler approach to work reasonably well (see also Ployhart and Vandenberg 2010). If the data are missing at random and one is using full information maximum likelihood, then use of structural equation modeling (SEM) or random coefficient modeling (RCM) allows one to simply run the models without much concern because these models will provide unbiased estimates of the missing data. Thus, first provide tests and estimates to show that the data do not differ systematically by the variables within the study and, if they do not, then use SEM or RCM to model the data (if the estimates differ, then it is time to decide whether a lesser scope of generalizability is necessary, or employ more sophisticated data analytic approaches such as those noted in Little ans Rubin 2002).

Analytical Judgment Calls

Which Analytical Method Should I Use?

This is probably the most common question in longitudinal research. The good news is that using a suboptimal method will usually not result in a rejection from a journal, but the bad news is that using a suboptimal method will often require an enormous amount of time in re-analysis and responding to reviewer concerns. Therefore, it is better to give serious thought to this issue in advance. Comprehensive treatments of this question are presented in Bollen and Curran (2006) and Singer and Willett (2003). Ployhart and Vandenberg (2010) provide an overview of the repeated measures

Table 1 Overview of different longitudinal analytical methods

Use the following method…	…when these conditions are present
Repeated measures general linear model	• Focus on group mean change
	• Identify categorical predictors of change (e.g., training vs. control groups)
	• Assumptions with residuals are reasonably met
	• Two waves of repeated data
	• Variables are highly reliable
	• Little to no missing data
Random coefficient modeling	• Focus on individual differences in change over time
	• Identify continuous or categorical predictors of change
	• Residuals are correlated, heterogeneous, etc.
	• Three or more waves of data
	• Variables are highly reliable
	• Model simple mediated or dynamic models
	• Missing data are random
Structural equation modeling	• Focus on individual differences in change over time
	• Identify continuous or categorical predictors of change
	• Residuals are correlated, heterogeneous, etc.
	• Three or more waves of data
	• Want to remove unreliability
	• Model complex mediated or dynamic models

General Linear Model (GLM), RCM (frequently also called Hierarchical Linear Models; HLM), and SEM. Here, and in Table 1, we will summarize the key distinctions.

First, if you want to focus on mean changes over time and missing data is not too problematic, use repeated measures GLM. Second, if you want to study between-person differences in change over time (e.g., growth curves), then use RCM or SEM. With simple growth models, RCM is somewhat easier to use when there is missing data or when the predictors of change are static. With more complex growth models, such as when there are time-varying predictors, dynamic relationships, or mediated change, then SEM is the preferred choice (full information maximum likelihood can be used to model any missing data). For the typical growth modeling question, where the focus is on modeling change over time and using static predictors of change, either RCM or SEM will give you similar, if not identical, answers. For example, studying change in job satisfaction over time, and trying to predict such change with emotional stability, could be equally done in RCM or SEM because emotional stability is static (i.e., it doesn't change over time). If change in job satisfaction was to be predicted by change in mood, either RCM or SEM can be used. However, if change in job satisfaction were to be predicted by change in mood, which is in turn predicted by emotional stability, then SEM would be the preferred choice because of its ability to model mediated relationships.

We offer a few additional suggestions based on our experience as authors and reviewers. First, given that multiple methods can be used with equal appropriateness, researchers should explain why they are using a particular approach. It is often the case that reviewers are familiar with one approach but not another, and so it is helpful to demonstrate to them that there is a good choice for using one particular approach. Second, report and convey the findings in substantive terms. Too often we see results reported based on discussion of the parameter, rather than what the parameter means substantively. For example, rather than noting "the slope is 0.25," one should instead note that "job satisfaction improves 0.25 points per month." Third, graphing and illustrating the data are perhaps the most impactful ways to convey the study's findings. It is difficult to visualize longitudinal data from the typical RCM table or SEM figure, so simple descriptive plots of the trend are powerful means of communication. Finally, don't default to automatically using the most sophisticated or complicated analysis for testing a hypothesis. Rather, go with the simplest approach. For many questions you don't have to use RCM or SEM. There are still many questions that can be answered with the GLM.

How Should I Code Time?

In growth curve models, time is used as a metric for change. In either RCM or SEM, repeated measurements on an outcome are regressed upon a coding for Time, such that change becomes scalable. In general, $\Delta Y = b_0 + b_1(\text{TIME}) + e$. The parameter b_0 is an intercept whose interpretation depends on the coding for TIME. The parameter b_1 is the rate of change over time. TIME is often coded 0, 1, 2, etc. For example, with job satisfaction measured monthly for five months, TIME would be coded 0, 1, 2, 3, 4. Coding the first month as zero allows the intercept to refer to job satisfaction at the first month. For this reason, the intercept is also frequently interpreted substantively as "initial status" or "baseline."

There are a variety of ways in which one may code time, and different types of coding structures will influence interpretation of the intercept, the covariance between the intercept and slope, and possibly the slope parameter and its variability (Biesanz et al. 2004; Mehta and West 2000). For example, one may code time so that the intercept refers to the last time period or the middle period. Theory should determine the best code structure, but lacking a compelling reason to code time differently, the simplest and safest thing to do is code time so that the first time period is zero. Then, one should scale time to reflect when the measurement occasions occur. For example, a satisfaction survey administered in months 1, 3, 7, and 12 might result in TIME being coded 0, 2, 6, and 11. Singer and Willett (2003) and Biesanz et al. (2004) provide excellent guidance on how to most appropriately code time for different substantive questions.

How Do I Know If I am Correctly Modeling Change?

The modeling of trends and curves is challenging not only because there is a paucity of theory to provide guidance, but also because many different types of growth models may produce highly similar fit. In a sense, scholars studying change are faced with a variation of the "equivalent models" problem, where different statistical models provide similar or identical fit to the same data. When equivalent models exist, parsimony and theory should break the tie, but the lack of theory results in a default victory for parsimony. It is very easy to keep adding terms (e.g., linear, quadratic, cubic) to a growth model, so parsimony if nothing else ensures we don't over fit the data.

Lacking clear theory, the best way to build parsimonious longitudinal models is to use a model comparison approach (Bliese and Ployhart 2002). In this approach, simpler models are compared to more complex models, and complexity is only warranted to the extent it improves model fit significantly. A model comparison approach works effectively in both SEM and RCM. First, one should start with what is called an *unconditional growth model*, where repeated measurements on the dependent variable are regressed upon TIME. Second, one should examine the variability that might exist around the intercept and/or slope. Note that these variance components are in many ways more important than the common slope, because the variability suggests between-person (or unit) differences in change over time. Even if the common slope is not different from zero, significant variability around the slope suggests there may be individual differences in change that should be explored. Third, after determining whether there is significant variability in intercepts and/or slopes, it makes sense to try to explain such variability using time-varying or time-invariant predictors. See Bliese and Ployhart (2002) for an introduction and overview of this model testing approach illustrated using RCM. A model comparison approach also ensures that more sophisticated models and methods are only adopted to the extent they provide unique insights.

For publication, it is usually sufficient to just interpret the final model. Most readers, reviewers, and editors don't want to be bogged down in minute details. However, they do want evidence that the author followed those details and interpreted the models correctly. Therefore, it is important to explain that a model comparison approach was used to determine the final or best-fitting model. You should also frequently report the *major* alternative models to show the model comparisons (e.g., model fit using different trends; model fit with different predictors). Report any unusual or unexpected findings that might have occurred while conducting the model comparisons.

Judgment Calls When Publishing Longitudinal Research

How Do I More Strongly Convey a Theoretical Contribution?

If your longitudinal study makes the same predictions and leads to the same conclusions as a cross-sectional study, then there is not a unique theoretical contribution. Indeed, such a study is a generalizability study. This kind of research can still be useful and important, because there is so little longitudinal research on so many phenomena and constructs that any such data is informative. However, to better convey the importance of one's longitudinal study, we suggest the following be considered.

First, go on the offensive against cross-sectional research. Is such research to be trusted? What threats to internal or external validity, or causality, are implicitly assumed in the existing literature? How are these implicit assumptions "laid bare" by the longitudinal study? Report the base-rate of longitudinal studies compared to cross-sectional, to show how serious the problem could be. The most constructive way to critique cross-sectional research is to identify the limits of existing theory—when viewed from a longitudinal perspective. For example, take a hypothesis that is of unquestionable importance and which has considerable empirical support. Now, for this hypothesis, ask questions such as *when does the effect occur*, *for how long*, or *why might it change over time*? For example, cross-sectional research finds that emotional stability influences job satisfaction. Questions that come to mind include, "Does emotional stability always influence job satisfaction?;" "When might emotional stability most/least strongly relate to job satisfaction?;" "When might the effects of emotional stability on job satisfaction wane?" and so on. Asking such basic questions in this manner maintains the integrity of the broader theory, but questions assumptions that may have gone unexamined in prior cross-sectional research.

Second, identify the limits to causal inference in cross-sectional research. We are all taught that correlation is not causation and that causes must precede consequences. Most cross-sectional research will usually conclude with cautions about inferring causality and a need for longitudinal research. Use, and perhaps even quote, these self-reported limitations as the basis for your contribution. However, *proving* causality is difficult and not going to occur in one study, so it is better to emphasize that a well-conducted longitudinal study will *provide stronger inferences* of causality (see Antonakis et al. 2010; Ployhart and Kim, in press; Singer and Willett 2003, for a discussion of this issue).

Finally, focus on how conclusions change between cross-sectional and longitudinal designs. Whenever possible, directly compare and contrast the longitudinal effect sizes to the cross-sectional effect sizes. This can be done within a study or in comparison to prior studies. For example, do the longitudinal effect sizes contradict meta-analytic estimates based on cross-sectional findings and how do these differences weaken prior conclusions? How much of an over/underestimate in cross-sectional effect sizes is likely? The point is to demonstrate that your longitudinal study provided new insights and that those insights could not be garnered through cross-sectional research.

Thus, if the hypotheses, effects, and implications are the same as what is found cross-sectionally, then use of a longitudinal design does not present a very strong theoretical contribution. Of course, if the theory specifies temporal dynamics, and most research has been cross-sectional, then the longitudinal study may provide a theoretical contribution only because no prior research has tested the theory appropriately. Researchers should be able to convey what is unique about their longitudinal study…but simply conducting a longitudinal study does not, by itself, make a strong contribution.

Theory Should Inform the Method, Not the Other Way Around

As obvious as this sounds, it is frequently violated in practice, and especially within the context of longitudinal research. As researchers become exposed to growth models, they begin to understand the power in these models and search for opportunities to apply them. This, coupled with the availability of archival data, often leads to papers being submitted that employ powerful methodologies on rather pedestrian data. Such papers are easy to spot because they are light on theory and heavy on method. Given that longitudinal methods like RCM and SEM are becoming common, authors should emphasize the theory and how the statistic provides the appropriate test of the theory.

The Value of Modeling Nonlinear Trends

With sufficient waves of data, it becomes possible to model nonlinear trends or curves (such as that illustrated in Fig. 1). We noted that more repeated measurements are better and with them often comes the possibility of detecting nonlinear forms of change. But we also noted one should strive to build parsimonious models using a model comparison approach. Yet even with model comparisons, it is often the case that a nonlinear trend may be suggested. The researcher is then faced with the following questions: "Is the nonlinear trend substantively interesting?;" "Does the nonlinear trend provide more insight than a simple linear trend?;" "Why is the trend nonlinear?;" and, lacking clear answers to these questions, "Should I present the nonlinear or linear trend for publication?" The easy answer to this question is to let theory decide, but usually there is not sufficient theory for answering it. The scholar must then either risk developing such theory or fall back into parsimony's comfortable embrace. In this situation, an otherwise straightforward empirical study now adds the challenge of new theory development. In our opinion, organizational science loses in the long term if we model linear trends because of convenience (see also George and Jones 2000). Yet we don't advocate chasing nonlinear trends that have little generalizability or theoretical value (or, searching for nonlinear trends when there is not sufficient statistical power to detect them, as noted above). Therefore, unless there is strong theory to propose a specific trend (and the power to detect it), we suggest scholars limit their examination to quadratic or (at most) cubic models, representing one or two bends in the trend line. And, there should be a very compelling reason to move beyond a quadratic trend. We believe this suggestion is important because it splits the difference between recognizing nonlinearity, versus over fitting models to data.

Table 2 A checklist for conducting and publishing longitudinal research

As you prepare your study for submission, have you:

1. Conveyed the unique theoretical contribution of your longitudinal design?
2. Developed hypotheses that differ from cross-sectional theory?
3. Explained *why* change should occur in your constructs and variables?
4. Described why you measured the variables at the various times, and how this provides a good test of the theory and hypotheses?
5. Considered threats to internal validity (e.g., carryover effects)?
6. Explained how you took steps to reduce missing data?
7. Examined the consequences of missing data, and took appropriate actions?
8. Explained why you chose the analytic method (emphasizing its appropriateness for your situation)?
9. Presented the relevant models so readers can evaluate differences in model fit?
10. Recognized the limitations of your longitudinal design (and not oversold causality)?

When Cross-Sectional Designs Are Preferred

We have a strong belief that the future and evolution of organizational science will depend, in part, on the adoption of longitudinal thinking and research. That said, there may be instances where a good cross-sectional study is better than a poorly-done longitudinal study. This situation occurs when the measurement occasions are spaced inappropriately or at the wrong times; there is no change on the construct or variable; the reasons for attrition are not random and we don't know why; there are threats to internal validity caused by the design (e.g., memory, carryover); the analytical models won't run or converge; or the findings are trivial. In these situations, it may be more informative to analyze and present the data using cross-sectional methods (so long as the findings are consistent between the two approaches).

Further, given the long timeframe for conducting longitudinal field studies, there is no guarantee that the research can be completed. For example, we have had longitudinal research projects terminated because our contact left the organization, the economy tanked, or the firm was bought by a competitor! So how can one save some hope of publishable research? One strategy we have used is to, whenever possible, design publishable "cross-sectional" projects embedded within the longitudinal study. Try to hypothesize interesting questions if you could only collect one wave of data (usually the first), just two waves of data, and so on. In the job satisfaction example, after designing the longitudinal study, the researcher may additionally include a few (short) measures of constructs for a different study (e.g., socialization experiences). In this manner, if only one wave of data is collected, it may still be possible to publish a different cross-sectional study.

Conclusion

In this article, we have presented a fairly concise introduction and summary into longitudinal research. There are many technical treatments of longitudinal statistical models, a few introductions to longitudinal theory and methods, but (to our knowledge) no "quick start guides" for the scholar trying to decide whether to design or pursue a longitudinal study. We have purposefully avoided many analytical details, for example, how to conduct growth modeling in RCM or SEM, because such references already exist, and so we can focus on the "big picture." Table 2 summarizes the major points in this article by giving readers a "checklist" to consider as they prepare their longitudinal research for publication. We now leave readers with the following challenge to their scholarship and practice:

> Pick your favorite theory, relationship, or effect, and critique it using the longitudinal theory and methods introduced in this article. Develop the theory that speaks to the *when*, *why*, and *duration* of change, and test this theory using methods capable of modeling change.
>
> Imagine where the field would be if we all did this…

Acknowledgments We would like to thank Steven Rogelberg and Scott Tonidandel for their helpful and constructive comments on this article.

References

Allison, P. D. (2001). *Missing data*. Thousand Oaks, CA: Sage.

Anseel, F., Lievens, F., Schollaert, E., & Choragwicka, B. (2010). Response rates in organizational science, 1995–2008. A meta-analytic review and guidelines for survey researchers. *Journal of Business and Psychology, 25*, 335–349.

Antonakis, J., Bendahan, S., Jacquart, H., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly, 21*, 1086–1120.

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods, 9*, 30–52.

Bliese, P. D., & Ployhart, R. E. (2002). Growth modeling using random coefficient models: Model building, testing, and illustrations. *Organizational Research Methods, 5*, 362–387.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: John Wiley & Sons.

Chen, G., Ployhart, R. E., Cooper-Thomas, H. D., Anderson, N., & Bliese, P. D. (2011). The power of momentum: A new model of dynamic relationships between job satisfaction change and turnover intentions. *Academy of Management Journal*, in press.

Conway, J. M., & Lance, C. E. (2010). What reviewers should expect from authors regarding common method bias in organizational research. *Journal of Business and Psychology, 25*, 325–334.

Duncan, S. C., Duncan, T. E., & Hops, H. (1996). Analysis of longitudinal data within accelerated longitudinal designs. *Psychological Methods, 1*, 236–248.

George, J. M., & Jones, G. R. (2000). The role of time in theory and theory building. *Journal of Management, 26*, 657–684.

Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research, 31*, 197–218.

Graham, J. W., Taylor, B. J., & Cumsille, P. E. (2001). Planned missing data designs in analysis of change. In L. Collins & M. Sayers (Eds.), *New methods for the analysis of change* (pp. 335–353). Washington, DC: American Psychological Association.

Keppel, G. (1991). *Design and analysis: A researcher's handbook*. Upper Saddle River, NJ: Prentice Hall.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York, NY: Wiley.

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods, 12*, 23–44.

Mehta, P. D., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods, 5*, 23–43.

Mitchell, T. R., & James, L. R. (2001). Building better theory: Time and the specification of when things happen. *Academy of Management Review, 26*, 530–547.

Newman, D. A. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods, 6*, 328–362.

Palmer, R. F., & Royall, D. R. (2010). Missing data? Plan on it!. *Journal of the American Geriatrics Society, 58*, 343–348.

Ployhart, R. E., & Kim, Y. (in press). Dynamic longitudinal growth modeling. Dynamic longitudinal growth models. In J. Cortina & R. Landis (Eds.), *Frontiers of methodology in organizational research*. New York, NY: Routledge.

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

Ployhart, R. E., Weekley, J. A., & Ramsey, J. (2009). The consequences of human resource stocks and flows: A longitudinal examination of unit service orientation and unit effectiveness. *Academy of Management Journal, 52*, 996–1015.

Raudenbush, S. W., & Liu, X. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods, 6*, 387–401.

Rogosa, D. R. (1995). Myths and methods: "myths about longitudinal research" plus supplemental questions. In J. M. Gottman (Ed.), *The analysis of change* (pp. 3–66). Mahwah, NJ: Lawrence Erlbaum Associates.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. New York, NY: Oxford University Press.

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.

Copyright of Journal of Business & Psychology is the property of Springer Science & Business Media B.V. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.