# Longitudinal Methods in the Health Sciences: Four Recommendations

**Gertraud Stadler**
*Columbia University*

**Megan L. Robbins**
*University of California*

**Jean-Philippe Laurenceau**
*University of Delaware*

**Niall Bolger**
*Columbia University*

Longitudinal research with multiple time points has become more popular in health psychology, fueled by the rise of eHealth/mHealth studies. This article will address common challenges using longitudinal designs in the health sciences, including sources of variance and reliability of change, the difference between within-person effects and between-person effects, within-person mediation, and power. We make four recommendations: (1) to select change-sensitive measures and calculate variance components and reliability of change routinely as a starting point of data analysis, (2) to distinguish within-person process from between-person effects in data analysis, (3) to consider within-person mediation processes, and (4) to think of the different sources influencing power in longitudinal designs and to conduct power analyses. We will discuss how the use of advanced longitudinal designs could shape theory and research in the health sciences.

## The Value of Longitudinal Designs in the Health Sciences

In the past decades, health scientists have begun more and more to study health and its correlates and determinants as they fluctuate and change over time. Intentions, self-efficacy, mood, behavior, health - all can fluctuate from day to day, week to week, while growing up from child to adolescent to adult, when acquiring healthy habits and shedding unhealthy ones, becoming sick and getting healthy again.

Longitudinal designs have a number of strengths. In longitudinal studies, researchers can minimize retrospective bias with appropriate assessment instruments, focus on within-person change versus between-person differences, get a better close-up picture of processes as they unfold, and examine how varying contexts influence affect, behavior, and health. However, longitudinal studies also present unique challenges. Therefore, we would like to present four recommendations for longitudinal research in the health sciences.

## Integrating Theoretical Model, Temporal Design, and Statistical Model of Change

When health scientists study change over time it is helpful to consider how to best achieve "integration of theoretical model, temporal design, and statistical model" (Collins, 2006, p. 509). For coming up with a theoretical model of change, researchers need to know quite a bit about the phenomena of interest. How much evidence is there already about the speed of the process you want to study? How quickly do outcome and predictors fluctuate - across minutes, hours, days, weeks, years? What is the meaningful part of that variation in relation to random noise? What are the most important predictors of an outcome over time? In many

cases, there is not much prior longitudinal evidence to answer these questions, particularly for a specific population of interest. In this case, pilot studies can help to make more informed guesses.

When researchers have come up with an - ideally evidence-based - theoretical model of change, they should match the temporal design and data analysis of their study with the hypothesized change in predictors and outcomes as closely as possible. Based on the theory of change, they will decide at what time to begin and end the study, how often to assess, and at what intervals. In a world of limited resources, design decisions often require tough compromises. If there are critical periods where most of the change occurs then most assessments should occur in that critical period and assessments before and after can be more spaced out. For example, researchers would measure more frequently right after the diagnosis of chronic illness, and more rarely later when patients have adapted and developed stable routines. But even with an ideal temporal design researchers still need to select measures and statistical models that fit their theory of
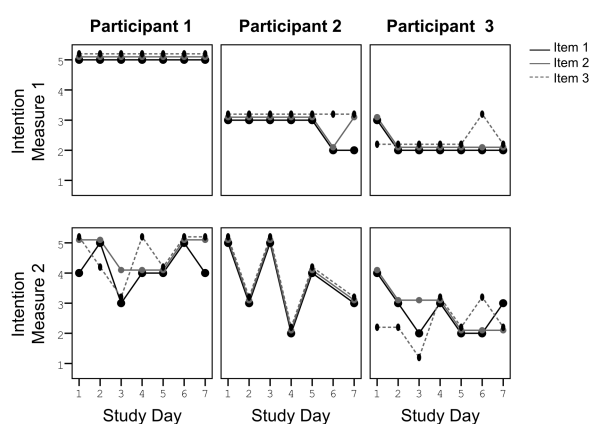


Figure 1: Panel plot of three participants' intentions per study day for 2 intention measures with 3 items each. Intention Measure 1 shows between-person variability, but little within-person variability, while Intention Measure 2 shows both between- and within-person variability.

change (see Recommendation 1), allow to distinguish within- and between-person variation (see Recommendation 2), and get at the processes of interest, including mediation (see Recommendation 3), all with adequate power (see Recommendation 4). We will visit each of these issues with four recommendations.

## Four Recommendations

*Recommendation 1: Select Change-Sensitive Measures and Calculate Reliability of Change*

Because most measures have been optimized for cross-sectional research rather than longitudinal research, finding appropriate measures with good psychometric properties for longitudinal studies can be quite challenging. A good starting point for integrating theoretical model, temporal design, and data exploration is to understand sources of variance and the reliability of each construct of interest in a longitudinal study. Ideally, for building a theory of change, researchers would be able to look up variance components and reliability of change in prior longitudinal studies and have conducted a pilot study in the population of interest.

To illustrate Recommendation 1, we will follow a research team interested in investigating changes in intention and physical activity in patients diagnosed with a chronic illness in a longitudinal intervention study. The researchers may have found two brief intention measures with three items each used in previous studies and tried them out in a brief pilot study in their population of interest. Looking at the data from their pilot study, the research team could start data exploration by drawing panel plots of individual participants' intentions (measured with three items each and the two intention measures) across study days. As Figure 1 shows for three exemplary participants, the

$$M_{pti} = \mu + P_p + T_t + I_i + (PT)_{pt} + (PI)_{pi} + (TI)_{ti} + [(PTI)_{pti} + \varepsilon_{pti}] \qquad (\text{Eq. 1})$$

two intention measures give different information. Intention Measure 1 (Figure 1, upper panels) captures differences in intention level between participants (between-person variability) while Intention Measure 2 (Figure 1, lower panels) captures intention fluctuations within person (within-person variability) in addition to differences in intention level between persons (between-person variability).

Shrout and colleagues have suggested using a generalizability theory framework (Cronbach, Gleser, Nanda, & Rajaratnam, 1972) to analyzing reliability in longitudinal data (Cranford, Shrout, Iida, Rafaeli, Yip, & Bolger, 2006; Shrout & Lane, 2012). Following this approach, researchers divide the available total variance for a certain measure into variance components linked to person, time, and item, and their combinations (Step 1) and then use these variance components to calculate reliabilities (Step 2).

In Step 1, the total variance is divided into variance components for person, time, and item,

based on a three-way, crossed, analysis of variance model (person by time by item). The response of person $p$ at time $t$ to a certain item $i$, $M_{pti}$, can be understood as a combination of the nine components shown in Equation 1. The first component, $\mu$, represents the population mean of the measure. The second component, $P_p$, captures that each person $p$ can have higher or lower responses, regardless of items and time points; this effect reflects between-person differences in how persons respond to the measure of interest. The third component, $T_t$, captures that responses can be higher or lower at time point $t$ compared to other time points, for all persons and all items. The fourth component, $I_i$, captures that item $i$ can receive higher or lower responses than other items, for all persons and time points. The fifth component, $(PT)_{pt}$, captures that person $p$ can give higher or lower responses at time point $t$, on all items. This component is particularly interesting for longitudinal research because it

Table 1: Sources of Variance and Reliabilities for Intention Measure 1 and Intention Measure 2 with three items each measured across 7 days.

| Source of Variance | | Intention 1 (3 items, 7 days) | % | Intention 2 (3 items, 7 days) | % |
|---|---|---|---|---|---|
| Variability across persons | $\sigma^2_{PERSON}$ | 0.45 | 33% | 0.38 | 23% |
| Variability across days | $\sigma^2_{TIME}$ | 0.02 | 1% | 0.01 | 1% |
| Variability across items | $\sigma^2_{ITEM}$ | 0.08 | 6% | 0.07 | 4% |
| Person-by-time variability | $\sigma^2_{PERSON*TIME}$ | 0.03 | 2% | 0.44 | 27% |
| Person-by-item variability | $\sigma^2_{PERSON*ITEM}$ | 0.17 | 12% | 0.22 | 13% |
| Time-by-item variability | $\sigma^2_{TIME*ITEM}$ | 0.01 | 1% | 0.01 | 1% |
| Residual variability | $\sigma^2_{ERROR}$ | 0.6 | 44% | 0.49 | 30% |
| | | | | | |
| Total | $\sigma^2_{TOTAL}$ | 1.37 | 100% | 1.62 | 100% |
| Reliabilities | | | | | |
| Between-person reliability | $R_{KF}$ | 0.95 | | 0.95 | |
| Within-person reliability | $R_C$ | 0.13 | | 0.73 | |

$$R_{KF} = \frac{\sigma^2_{\text{person}} + \frac{\sigma^2_{\text{person}*\text{item}}}{i}}{\sigma^2_{\text{person}} + \frac{\sigma^2_{\text{person}*\text{item}}}{i} + \frac{\sigma^2_{\text{error}}}{t*i}} \qquad \text{(Eq. 2)}$$

$$R_C = \frac{\sigma^2_{\text{person}*\text{time}}}{\sigma^2_{\text{person}*\text{time}} + \frac{\sigma^2_{\text{error}}}{i}} \qquad \text{(Eq. 3)}$$

indicates systematic change over time: Some persons respond higher or lower at a certain time, regardless of the items used for the response. The sixth component, $(PI)_{pi}$, captures that person $p$ can give higher or lower responses to item $i$ than other items, at all time points. The seventh component, $(TI)_{ti}$, captures that item $i$ can get higher or lower responses at time point $t$ by all persons. The eight and ninth components, $(PTI)_{pti}$ and $e_{pti}$, capture that some persons have higher or lower responses on some items at specific time points. We would need repeated assessments of each item at a specific time point, to distinguish the systematic effect, $(TI)_{pti}$, from random error, $e_{pti}$. For most designs where each person provides only one response to each item at each time point, we cannot distinguish this error term from the three-way interaction effect of item, person, and time point, and therefore estimate them together with only one term, as indicated by the brackets around the two terms.

Following the generalizability theory approach for our example, the research team would conduct an analysis of the variance components for the two intention measures. Shrout and Lane (2012) provide code for conducting these analyses in SPSS and SAS. Figure 2 and Table 1 show sources of variance for two intention measures with three items each. Four variance components are of particular interest (see Table 1 and Figure 2 for the example data set): variability across persons, person-by-time variability, person-by-item variability, and residual variability. In the example, variability between persons accounted for about a third of the variance in Intention Measure 1 (33%) while it accounted for a quarter of the variance in Intention Measure 2 (23%). Person-by-time variability accounted for hardly any variance in Intention Measure 1 (2%) while it accounted for another quarter of the variance for Intention Measure 2 (27%). Because systematic change over time is often the main reason for conducting longitudinal research, the research team for our example should be excited to see that Intention Measure 2 seems to capture a good amount of this variance. Person-by-item variability was comparable between the two intention measures (Intention Measure 1: 12%, Intention Measure 2: 13%). Residual variability was larger for Intention Measure 1 (44%) than for Intention Measure 2 (30%)

In Step 2, we then use these variance components to calculate between-person reliability and reliability of change. Assuming fixed time points and items as in the example study, Cranford and colleagues (2006) calculate between-person reliability as shown in Equation
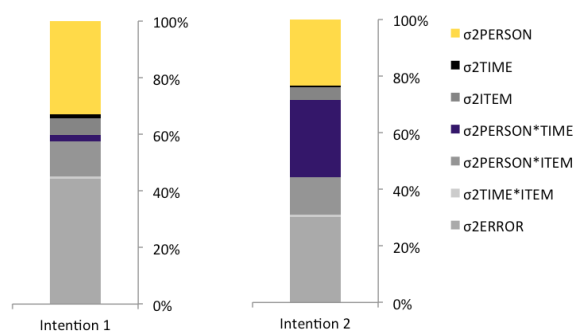


Figure 2: Sources of variance for Intention Measure 1 and Intention Measure 2 with three items each measured across 7 days.

$$\text{Predictor}_p t = \text{Predictor\_W}_{pt} + \text{Predictor\_B}_p \qquad \text{(Eq. 4)}$$

2. Between-person reliability is a ratio, with the numerator being the sum of variability across persons and person-by-item variability, divided by the number of available items, and the denominator this same sum plus residual variability, divided by the product of number of time points $t$ by number of items $i$. Table 1 shows that both intention measures in our example show excellent between-person reliability.

Cranford and colleagues (2006) calculate reliability of change as shown in Equation 3. Reliability of change is a ratio, with the numerator being person-by-time variability, and the denominator being the sum of person-by-time variability plus residual variability, divided by the number of items $i$. Table 1 shows that Intention Measure 1 in our example shows unacceptably low reliability of change (0.13), while Intention Measure 2 shows acceptable reliability of change (0.73) and would therefore be the measure of choice for further studies. An example write-up for Intention Measure 2 in a methods section would be: The measure showed outstanding between-person reliability ($R_{KF}$ > .90) and acceptable reliability of change ($R_C$ > .70).

For more details and syntax for calculating variance components and reliability for longitudinal designs, see Shrout and Lane (2012). Shrout and Lane (2012) give different examples for calculating appropriate reliabilities, depending on the design of the study. The generalizability theory approach presented by Cranford and colleagues (2006) assumes that items and assessment times can be distinguished and are thus fixed within person. That makes sense for items because we usually have a specific set of items and are not selecting

randomly from a pool of items. For assessment times around a critical event, such as diagnosis of a chronic illness or an online intervention where all participants start at the same time, all participants have the same assessment time points and the assumption of fixed time points makes sense as well. However, in other designs, for example, studies with event-contingent assessment or experience sampling studies with random beeps, it makes sense to assume that assessment times are random and thus nested within person. For another approach within a multilevel framework, see also Wilhelm & Schoebi (2007).

*Recommendation 2: Distinguish Within-Person Change From Between-Person Effects*

Our second recommendation is to distinguish within-person processes from between-person effects. Longitudinal data, compared to cross-sectional data, provide the opportunity to observe and analyze changes over time within a person, facilitating the study of health and behavior in daily life (Mehl & Conner, 2012). Time-varying constructs, such as intention and behavior, contain two sources of variation, (a) within-person fluctuations around (b) each person's mean level that varies between persons. Figure 3 illustrates this distinction for the intention data used as an example. Person 1 has on average high intentions, Person 2 has moderate intentions, and Person 3 has low intentions. But all persons show at times higher and lower intentions than their typical level on the change-sensitive Intention Measure 2. Multilevel models for analyzing longitudinal data differentiate within- and between-person variability for outcomes, but not by default for predictors. Therefore, within- and between-person effects have been confounded in many

$$Y_{it} = \gamma_{00} + \gamma_{01}\text{Time}_{pt} + \gamma_{02}\text{Predictor\_W}_{pt} + \gamma_{03}\text{Predictor\_B}_p + \varepsilon_{pt} \qquad \text{(Eq. 5)}$$

prior analyses of longitudinal data but they need to be carefully distinguished to avoid biased conclusions (e.g., Allison, 2009; Bolger & Laurenceau, 2013; Curran & Bauer, 2011; Hamaker, 2012; Raudenbush & Bryk, 2002). The distinction of between- and within-person variability in the predictor variables by using person-level means and within-person deviation scores is one important contribution that many manuscripts have neglected to make so far.

In the theoretical model of change, it is helpful to distinguish between-person effects from within-person processes. Between-person effects reflect stable associations between predictor and outcome, and are prone to all alternative explanations that we are familiar with from cross-sectional research. For example, persons with higher intentions may be more physically active, but the causal mechanism behind this association could be in any stable third variable that is related to both intention and activity. Should the research team find that on days when participants show higher intentions they also show higher physical activity the number of alternative explanations shrinks to constructs that covary with intentions and activity from day to day. A last important theoretical question regarding within-person processes is if increases in a predictor have the same effects as decreases. Most theoretical models assume causal symmetry by default but increases and decreases could have differential effects. For example, increases in intention could have different effects on activity than decreases in intention. Stadler and colleagues have shown an approach to separate effects of increases and decreases in a predictor and found differential effects (Stadler, Snyder, Horn, Shrout, & Bolger, 2012). For the temporal

design, it is important to select time frame and assessment frequency carefully and measure all variables as time-varying constructs with change sensitive measures, to allow for within-person fluctuations in the predictor as well as in the outcome, as delineated in the specific theory of change.

It is relatively straightforward to implement the distinction of within-person process and between-person effects in the statistical model. Each raw predictor score for person $p$ at time $t$, $Predictor_{pt}$, can be split up into the time-varying within-person deviation, $Predictor\_W_{pt}$, from each individual $i$'s average predictor level across all available time points, $Predictor\_B_p$ (see Equation 4).

Thus, we first calculate person means in the predictor variable across all available time points, $Predictor\_B_p$. Then we subtract each person's mean from the raw predictor score, $Predictor_{pt}$, to arrive at deviation scores for the predictor for each time point, $Predictor\_W_{pt}$. To facilitate interpretation of the intercept, we calculate the mean of the person means and center the person means $Predictor\_B_p$ at the
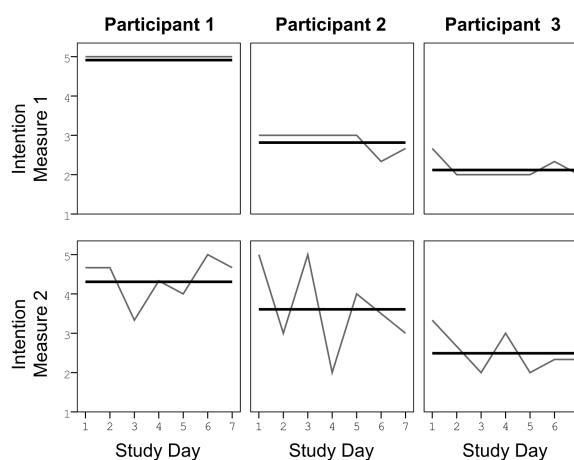


Figure 3: Three participants' intention fluctuations over 7 days around each person's mean for Intention Measure 1 and Intention Measure 2.

grand mean by subtracting the sample mean from the person means. Finally, we enter predictor deviations and person averages, the latter centered at the grand mean, into the multilevel model (Equation 5).

Equation 5 illustrates the data analysis approach for a continuous outcome $Y_{it}$ predicted by time and the within- and between-person predictor. The coefficient $\gamma_{02}$ tests whether at times when a participant is higher on the predictor than usual he or she is higher or lower on the outcome (within-person association); the coefficient $\gamma_{03}$ tests whether persons who are higher in average predictor levels are also higher in the outcome (between-person association).

Interestingly, within- and between-person effects can differ considerably in size and even direction, and can differ in their causal processes. Neglecting these differences can obscure theory building (see Wilson, Stadler, Boone, & Bolger, under review). When researchers keep trying to find an effect on the between-person level that exists in the population on the within-person level and vice versa they will find mixed results (for more in-depth discussion, see Mehl & Conner, 2012; and Bolger & Laurenceau, 2013). Combined with a sound theory of change, the data analytic approach described above can facilitate health scientists' distinction between within- and
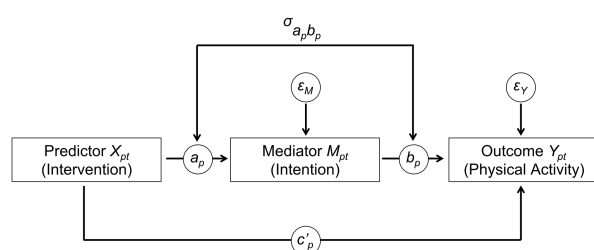


Figure 4: Example of within-person mediation: How much of the effect of the within-person predictor (randomly delivered daily intervention) on the outcome (daily physical activity) is mediated by the within-person mediator (daily fluctuation in intention around the person mean)?

between-persons effects, enhancing our understanding of temporal health processes.

*Recommendation 3: Consider Within-Person Mediation Processes*

Our next recommendation - to consider within-person mediation during theory building, design, and analysis – relies on the two prior recommendations to choose change-sensitive reliable measures and distinguish within-person processes from between-person effects. Given the great interest in developing and testing theories in health psychology, longitudinal researchers who want to understand causal influences on the within-person level can do so by using within-person mediation. This type of mediation is especially suited for intensive longitudinal data (Bolger & Laurenceau, 2013). Because participants are assessed repeatedly in an intensive longitudinal study, each participant can have his/her own mediation effect. Based on each person's mediation effect, researchers can then estimate an average within-person mediated effect as well as between-person heterogeneity around that average. For example, a research team conducting an intervention study aimed at increasing intentions to be physically active would want to see if the intervention actually increased intentions and if the increase in intentions explained the intervention's effect on physical activity. They could pursue these questions with a classic between-person mediation analysis (Baron & Kenny, 1986). But if cause, mediator, and outcome were measured repeatedly over time, they could test whether mediation occurs within each person and to what degree the causal chain explains the intervention effect for different persons. For example, if the intervention was delivered randomly on certain days (and assuming no carry-over effects), the research team would want to know if on intervention

days - compared to control days – intentions were higher in most participants and if, in turn, this led to higher physical activity. Figure 4 provides an example of within-person mediation for a randomly delivered daily intervention aimed at increasing intentions and physical activity. Note that the three mediation coefficients $c'_p$ as well as $a_p$ and $b_p$ are estimated for each person (indicated by a person-specific subscript $p$) allowing estimation of each person's mediation effect in addition to the average mediation effect. Finally, within-person mediation includes a new term, $\sigma_{ab}$, indicating how much the predictor-mediator link (i.e., $a_p$) covaries with the mediator-outcome link (i.e., $b_p$) and which must be included in the calculation of mediated effects if there are substantial $a$ and $b$ random effects (see Kenny, Korchmaros, & Bolger, 2003).

Within-person mediation provides health scientists with another tool for exploring causal mechanism. For hands-on guidance on how to conduct a within-person mediation analysis, see Bolger and Laurenceau (2013) who provide a detailed introduction including syntax and example data to run these analyses.

### Recommendation 4: Pay Attention to Factors Influencing Power

Our last recommendation is to pay attention to factors influencing power in longitudinal studies. Power indicates the probability to detect a hypothesized effect with a given sample, if the effect actually exists in the population. A common threshold for acceptable power is .80, indicating that a study will detect the population effect with a probability of 80%. Many researchers are familiar with the five determinants of power in the cross-sectional context: effect size, sample size, variability of the predictor, unexplained variance in the outcome, and Type I error probability. Studies

have more power if they investigate large effects, with large samples, maximize the variability of the predictor variable, minimize unexplained variance in the outcome, and choose more lenient probability levels (although there is little leeway to stray from the accepted .05 standard level).

Power to detect within-person effects in longitudinal studies has three additional determinants of power: the number of repeated time points, the amount of autocorrelation between time points, and how much the effect varies from person to person - in addition to the size of the within-person effect, sample size, variability of the within-person predictor, unexplained variance in the outcome, and Type I error probability. Studies with more time points per person, lower autocorrelation between time points, and relatively similar effects of the predictor on the outcome across participants have higher power.

In addition to the traditional ways for increasing power discussed above, the three additional determinants can also be addressed through research design. First, one straightforward way to increase power is to add additional time points to the temporal design. However, adding persons is a better way to increase power (see Figure 5). Second, choosing a temporal design that spaces time points not
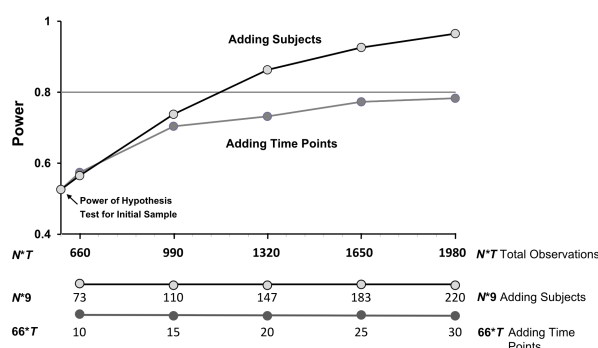


Figure 5: Power curves for the within-person fixed effect of physical activity on depression: What is the benefit of adding persons versus time points to the sample? (reprinted from Bolger, Stadler, & Laurenceau, 2012, p. 299)

too closely together diminishes autocorrelation between time points. Last, researchers can reduce the variability in the predictor-outcome link between persons by using interventions tailored to participants' needs and with standardized implementation or choosing time intervals where the predictor occurs relatively uniformly across participants.

Considering power as part of theory building, study design, and data analysis is a practice that pays off especially in the longitudinal context to optimize the allocation of resources. Bolger, Stadler, and Laurenceau (2012) give more details on conducting power analyses for within-person effects, and provide syntax and an example data set. For conducting power analyses across a wide range of research designs, see Bolger & Laurenceau (2013).

## Discussion

The future of longitudinal research in the health sciences is very promising. With a growing evidence base, researchers can achieve better fit between theory, study design, and data analysis. So far, we have only a vague picture of how health and its determinants change over time for many populations, and we need to rely on pilot studies that are necessarily giving limited information to inform larger studies. Without knowing the dynamics of change, it is hard to know how to best allocate resources. We may need to rely on a more fine-grained temporal resolution, keeping in mind that we can always aggregate measures if change is slower than we thought while we cannot retrieve more details that we have not collected. With more evidence, future research will increasingly become more efficient and sophisticated, relying on measures geared towards capturing change and within-person effects, and even allowing us to look at

mediating and moderating processes with enough power. Longitudinal research, and particularly intensive longitudinal studies and data-burst designs, give us a chance to gain a more complete understanding of stability and change in health across the life course and its causes.

## References

Allison, P. D. (2009). *Fixed effects regression models*. Thousand Oaks, CA: Sage.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.

Bolger, N., & Laurenceau, J-P. (2013). *Intensive longitudinal methods*. New York: Guilford Press.

Bolger, N., Stadler, G., & Laurenceau, J.-P. (2012). Power analysis for intensive longitudinal studies. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 285–301). New York: Guilford Press.

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.

Cranford, J. A., Shrout, P. E., Iida, M., Rafaeli, E., Yip, T., & Bolger, N. (2006). A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably? *Personality and Social Psychology Bulletin*, 32, 917–929.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.

Curran, P. J., & Bauer, D. J. (2011). The

disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, 62, 582-619.

Hamaker, E. L. (2012). Why researchers should think "within-person": A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). New York: Guilford Press.

Kenny, D. A., Korchmaros, J. D., & Bolger, N. (2003). Lower level mediation in multilevel models. *Psychological methods*, 8, 115-128.

Mehl, M. R., & Conner, T. S. (Eds.). (2012). Handbook of research methods for studying daily life. New York: Guilford Press.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: Sage.

Shrout, P. E., & Lane, S. P. (2012). Psychometrics. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 302–320). New York: Guilford Press.

Stadler, G., Snyder, K. A., Horn, A. B., Shrout, P. E. & Bolger, N. (2012). Close relationships and health in daily life: A review and empirical data on intimacy and somatic symptoms. *Psychosomatic Medicine*, 74, 398-409.

Wilhelm, P., & Schoebi, D. (2007). Assessing mood in daily life: Structural validity, sensitivity to change, and reliability of a short-scale to measure three basic dimensions of mood. *European Journal of Psychological Assessment*, 23, 256-267.

Wilson, P. A., Stadler, G., Boone, M. & Bolger, N. (under review). Fluctuations in well-being and sexual risk episodes among HIV-positive gay men.

**Gertraud Stadler**
is Associate Research Scientist at the Department of Sociomedical Sciences and Department of Psychology, Columbia University, United States
**stadler@psych.columbia.edu**



**Megan L. Robbins**
is Assistant Professor at the Department of Psychology, University of California, Riverside, United States
**mrobbins@email.arizona.edu**



**Jean-Philippe Laurenceau**
is Professor of Psychology at Department of Psychology, University of Delaware, United States
**jlaurenceau@psych.udel.edu**



**Niall Bolger**
is Professor of Psychology at Department of Psychology, Columbia University, United States
**bolger@psych.columbia.edu**