

SoulPulse Report 5: Analytic Strategies

The data structure of the SoulPulse data are unique, being unlike most longitudinal data sets studied by sociologists, psychologists, and other social scientists. These data span only two weeks; in contrast, most longitudinal data sets span months if not years. These data contain twice-daily surveys over 14 days, making for 28 surveys collected; most longitudinal studies have only a few data collections spread out over time. These data have broader pool of questions from which the questions on a specific survey are randomly selected; most longitudinal studies ask many of the same questions in each wave. These unique features of SoulPulse have some advantages and some disadvantages. This report evaluates analytic strategies that work well with these data as well as some that do not.

Promising Analytic Strategies

1. Situational analyses. Some of the daily survey questions regard the current situation such as what the participant is doing and who they are with. This makes it possible to examine how an outcome of interest varies by across types of situations. For example, the paper by Kucinskias et al (forthcoming in the Journal for the Scientific Study of Religion) examines how being aware of God varies by participants' activities as well as being with other people. This would take the general form of:

$X \mid \text{Situation}$

2. Time analyses. The daily survey questions are measured at all times of day, which allows for various time analyses, including by hour, day, day of week, month, and age. This allows the analysis of how experiences and characteristics vary across time. For example, the Kucinskias et al article examines how spiritual awareness varies by hour of the day and day of the week.

$X \mid \text{Time}$

3. Simple causation. Because the daily surveys are time ordered, we can use a measure of X at time T to predict an outcome Y at time T+1. The section below presents options for temporal ordering. For example, the article by Gutierrez, Park and Wright examines how financial distress affects later measures of depression.

$X \rightarrow Y$

4. Reciprocal causation. The repeated measures in SoulPulse allow to test the relative impact of X and Y. For example, X at T predicts Y at T+1 and Y at T predicts X at T+1. This type of reciprocal causation could work between two personal states or a personal state and a situation. For example, the article by Park et al. examines the reciprocal interplay stressful life events and ego depletion.

$X \leftrightarrow Y$

5. Interaction effect. The stable characteristics measured in the intake survey work well as moderating variables for processes measured by the daily variables. For example, the effect of X at T on Y at T+1 might vary by Trait Z. For example, the article by Gutierrez, Park and Wright examines if the effect of financial distress on depression varies by participant's religiosity.

$$\begin{array}{c} X \rightarrow Y \\ \uparrow \\ Z \end{array}$$

6. Variance analysis. The measures in the daily surveys vary across the study, and this allows for analysis of both their mean level and variation in them.

$$X \rightarrow \text{Variation}(Y)$$

7. Between and within person. The measures collected on a daily basis can be averaged over the two week period to produce a trait measure. Then, each individual daily observation can be subtracted from this trait measure to produce a state change measure. The two can then be included in the same equation to predict an outcome, allowing for the separate testing of within- and between-person effects.

$$X_{\text{within-person}} + X_{\text{between-person}} \rightarrow Y$$

Temporal ordering

The temporal ordering of the SoulPulse data is complex, for the surveys are administered at all times of the day. Thus, these data require careful attention and nuanced temporal analysis. Advantageously, the complex temporal ordering allows for a variety of analytic strategies for incorporating time into the statistical models. The possibilities for temporal ordering include:

1. Same time. X and Y are measured in the same survey
2. Same day. X is measured in the morning, Y is measured in the evening.
3. By hour. Since X and Y are measured at varying intervals, it's possible to test how the impact of X on Y varies by short periods of time. This could use a plot of how the causal impact of X decays by the hour.
4. Overnight. X is measured in the evening, Y is measured the next morning.
5. Day-to-day. X is measured one day and Y is measured the next. Note: when X and Y are measured twice on a day they would be averaged.
6. Week-to-week. X is measured one week, Y the next. All measures of X for the first week would be averaged as would measures of Y the second week.
7. Sequential. X is measured at one point. Y is measured at some time after X, regardless of when. (Y would have a possible range of half-day to 13.5 days).

There's another layer of complexity to incorporate. Some of the questions include temporal references.

- Most of the questions refer to the moment that the participant received the survey. E.g., Are you happy now?
- Some refer to events that have happened in the past several hours. E.g., How many sweets have you eaten in the past two hours.
- A few refer to the time since the last survey. E.g., since you last took a survey, have you had conflict with someone?
- A few refer to the last 24 hours. E.g., how much have you exercised in the last 24 hours.

The intake variables refer to mostly stable constructs, and they are temporally prior to the daily survey questions.

What doesn't work as well?

Two types of analyses tend not to work well.

Development. The study covers only two weeks in a participant's life, and that is usually too short a period for their personal characteristics to mature or develop in a systematic fashion. Rather, the study works best for analyzing naturally varying states.

Mediation models. As described below, the random sampling of daily survey questions leads to sample size constraints when too many daily variables are included in one model. While it varies by how often a question is asked, the data usually can't support having four or more daily survey variables in a single multivariate model. Note: There is the possibility of complex mediational analysis, however, such as used in structural equation analysis. And this can be done by reading the data in as a correlational matrix rather than reading in raw data with listwise deletion. For example, suppose a researcher wanted to specify a structural equation model with 7 variables—an independent variable, a dependent variable, and 5 mediating variables. Reading in raw data with listwise deletion would leave maybe a dozen surveys that happened to include all 7 variables. However, it would be appropriate to calculate bivariate correlations between all possible pairs of the 7 variables and then enter the data into the structural equation model as half of a 7x7 correlation matrix (i.e., all possible correlations).

Citations

Kucinkas, Jaime, Bradley R.E. Wright, D. Matthew Ray, and John Ortberg. Forthcoming. "States of Spiritual Awareness by Time, Activity, and Social Interaction." *Journal for the Scientific Study of Religion*.

Park, Crystal L., Bradley R.E. Wright, Jeremy Pais, and D. Matthew Ray. In press. "Reciprocal Relations between Daily Stressful Events and Ego Depletion: A Smartphone-Based Experience Sampling Study." *Journal of Social and Clinical Psychology*.

Gutierrez, Ian A., Crystal L. Park, and Bradley R.E. Wright. "When the Divine Defaults: Religious Struggle Mediates the Impact of Financial Stressors on Psychological Distress." Under review at *Psychology of Religion and Spirituality*.