SoulPulse Report 3: Missing Data and Attrition

There are a lot of missing data in SoulPulse; some of it planned, some of it due to attrition. This report reviews the various sources of missing data and estimates their magnitude.

1. Dropping out before the daily surveys. The first source of missing data is participants signing up for the study, completing some or all of the intake survey, but then not transitioning to the daily surveys. This happened more than we anticipated, and anecdotally, we heard of participants signing up without appreciating the two-week commitment. Maybe they thought it would be a one-time survey, maybe they changed their mind about participating. As of October 7, 2016, a total of 7,423 people signed up for SoulPulse; and 5,120 (69%) filled out at least one daily survey. Over the course of the study, we took several correction actions to lessen this frequency of dropping out, but it still remained high.

2. Dropping out amidst the daily surveys. Some participants made it to the daily surveys. They answered some daily surveys but then dropped out of the study altogether. It's difficult to estimate exactly how many participants did this—dropping out amidst the daily surveys—but we think it's rather low. In the first year of the study, participants would sometimes have to contact us to drop out, once they had started, and it wasn't that many. As a rough guess, we estimate that 8-10% of the participants dropped out of the study midway through the daily surveys.

3. Not answering a daily survey. Even when participants remained in the study, they would sometimes choose not to answer a particular survey. They would do this for various reasons including not seeing that it had arrived, being too busy to answer it, not feeling like answering it, and being unable to answer it (e.g., it arrived while they were driving). In previous analyses, we estimated that on average, participants who made it to the end of the completed about 18-20 (about 65%) of the 28 surveys sent to them.

4. Not answering some questions on a survey. When participants received a daily survey and started filling it out, they could answer some questions but not answer others. This is the traditional form of missing data as found in conventional surveys. Somewhat to our surprise, this rarely happened. Perhaps because the surveys were so short, only 20-30 questions long, that once participants started them, it was easy enough to just finish. In the data, this type of missing is designated with .u, and we estimate that about 1-2% of the data points are missing in this way.

5. Questions not asked. The great majority of missing data in this data set is, in fact, not due to attrition. Rather, the daily surveys are designed to include only some of a larger pool of possible questions. Thus, if 30 questions are asked, that still leaves 90 questions that are unasked. This type of missing data is randomized, and so it poses fewer analytic problems than the previous four types missing data that are rooted in attrition. The main consequence of this type of missing data is reduced statistical power, a topic that will be taken up in a later report.

The problem with attrition, of course, is that it can be nonrandom. To better understand the pattern of attribution, here is the demographic patterning of how often participants filled out the daily surveys that were sent to them (i.e., the inverse of #3 above). The mean for the whole sample was 62%.

Men	.61
Women	.63
Age 18-29	.60
Age 30-49	.61
Age 50+	.64
Native American	.58
Asian	.49
Black	.59
White	.64
Mixed Race	.62
Hispanic	.60
Not Hispanic	.62
Protestant	.62
Catholic	.64
Other religion	.59
No religion	.68
Other	.59
Some high school	.46
High school graduate	.61
Some college	.64
College graduate	.62
Graduate work	.63
Have child	.62
No child	.61
Single	.59
Widowed	.60
Divorced	.66
Separated	.66

Cohabitating	.59
Married	.63
Poor health	.58
Fair health	.61
Good health	.63
Very good health	.63
Excellent health	.60
Own residence	.63
Renting	.60
Other	.63