Online Supplement of the Paper:

Well-being, Smartphone Sensors, and Data from Open-access Databases: A Mobile Experience Sampling Study

Stefan Stieger

Department of Psychology

University of Konstanz

Universitätsstraße 10

78464 Konstanz, Germany

Tel. #: 0043 2732 72090 300

E-mail: stefan.stieger@kl.ac.at

Ulf-Dietrich Reips

Department of Psychology

University of Konstanz

Universitätsstraße 10

78464 Konstanz, Germany

Tel. #: +49 7531 88 2564

reips@uni-konstanz.de

Statistical Analyses

Multilevel models with random intercepts and coefficients were calculated using R (Package 'nlme'). Daily diary observations (level 1) were nested within participants (level 2). Continuous level-1 predictors were person-mean centered before analyses. The focus of our research question was on the within-subject processes, therefore we excluded the between-person mean of predictors from the models (Bolger and Laurenceau 2013). A random intercept model was significantly better than the null model (logLikelihood: -15,466.81, $p < .001$). Furthermore, a random intercept—random slope model was significantly better than the random intercept model (logLikelihood: -15,440.90, $p = .003$). Therefore, we present the results of the random intercept—random slope model.

For research question 1, we excluded the night hours (midnight to 7 a.m.) because of the low frequency of well-being ratings in this time frame. Then, we included the following predictor variables into our multilevel model—see variables #1 to #7 in Table 1.

The time predictors #1 (hour of the day) and #2 (day of the week) were assessed with the smartphone app from the present project. The environmental predictors #3 to #5 (average wind speed, rainfall, maximum temperature) were determined via an external database from World Weather Online (https://developer.worldweatheronline.com/). We programmed a script, which automatically retrieved the information through the World Weather Online application program interface (API). The script handed over the longitude and latitude information to the API, which returned the variables #3 to #5. We classified 58 temperature values as outliers and excluded them from further analysis. These data points were produced by three participants who were traveling abroad for several days. For the predictor #6 (altitude), we retrieved the altitude through the Google elevation service API. Again, a script handed over the longitude and latitude to the elevation API from Google, which returned the altitude of each specific geolocation together with a measure of accuracy. We used the Google

Maps altitude instead of the smartphone altitude, because the smartphone altitude could only be determined in 947 cases (vs. 3,599 altitude measures via Google Maps). We will compare the altitude retrieved by the smartphone and Google Maps in detail below in the section on research question 2.

Accuracy of Reminders

One important point when doing field research using the ESM procedure is to determine the reminders' efficiency (i.e., how quickly do participants respond to the reminders that are sent via text messages or WhatsApp). If the response time is very large, it could be a threat to the design (for a discussion, see Stone and Shiffman 2002). For example, if we are interested in participants' well-being in the morning and then send a reminder, but participants were to respond several hours later, we would measure during a different time of the day than intended.

All in all, participants were quite swift in responding to the reminders (Md = 49 minutes) considering that participants were reminded without any advance information about the timing of the reminders (i.e. participants were engaged in any kind of everyday activity) (e.g., driving a car, having a meeting, being in a lecture, etc.). There was no difference in response time depending on whether participants were reminded via text messages or WhatsApp ($Md_{\text{SMS}} = 45$, $Md_{\text{WhatsApp}} = 53$; Mann-Whitney-$U$ = 3,248, $p = .852$, $r = .02$).

Furthermore, we asked participants in a final online questionnaire whether they had had any difficulties with the reminders. Nine participants stated that they sometimes had a bad or no Internet access and could therefore not respond to the reminders when they wanted to and only two participants stated having problems with WhatsApp. No problems were stated regarding text messages. To sum up, using WhatsApp and text messages worked well in reminding participants.

Nevertheless, the reminder procedure has limitations. First, reminders are not always instantly delivered to the recipient. It is well known that text messages can be delayed when there are network problems or long text message are queued on the system. Delays can also emerge on the recipient's side. For example, when the phone is switched off, a connection to the network is not possible due to GSM dead zone areas or if the phone's storage is full. These problems can also apply to WhatsApp messages. Future research might test in-app reminder functions. Although in-app reminders would have several advantages (e.g., no need for Internet access), it would also have a disadvantage in that the experimenter would not know whether the in-app reminder works as it should. Also, the times when participants should be reminded would need to be known in advance.

Intervention Effect

One might argue that giving participants' feedback about their well-being during the assessment of the study could influence their subsequent well-being ratings. To analyze this, we recorded how often and what kind of visual statistics participants requested through the app. Participants could retrieve the following graphics at any time during the field phase of the study (see Figure S2, left panel):

1. A weekly statistic showing the mean well-being score of the participant as well as of the whole sample separated for each day of the week (see Figure S2, middle panel).
2. A daily statistic, again seeing participants mean well-being score over the time of the day as well as the mean well-being score of the whole sample (see Figure S2, right panel).
3. An overall well-being score of the whole sample depicted in the form of a round speedometer.
4. Participants' locations on a map where they filled in the smartphone app's questionnaire.

5. A world map showing the countries participants came from.
6. A calendar where participants could see how often they filled in the smartphone's app questionnaire each day during the field phase of the study.

If the feedback of one's own past well-being scores as well as the mean well-being of the other participants influenced one's well-being ratings, then we should find a correlation between the frequency of requested personalized well-being statistics through the course of the two weeks and the state well-being assessments. However, for all well-being graphics ("overall well-being score," "well-being during the day," and "well-being during the week"), we did not find any significant and substantial correlations with the mean well-being score or with the standard deviation of all well-being scores (all $r_{sp} < .14$, all ps $> .10$).

References

Bolger, N., and J.-P. Laurenceau. 2013. *Intensive longitudinal methods: An introduction to diary and experience sampling research*. New York: Guilford.

Stone, A. A., and S. Shiffman. 2002. Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioural Medicine* 24:236–43.

Figure S1. Geographical dispersion of participants' rating locations.

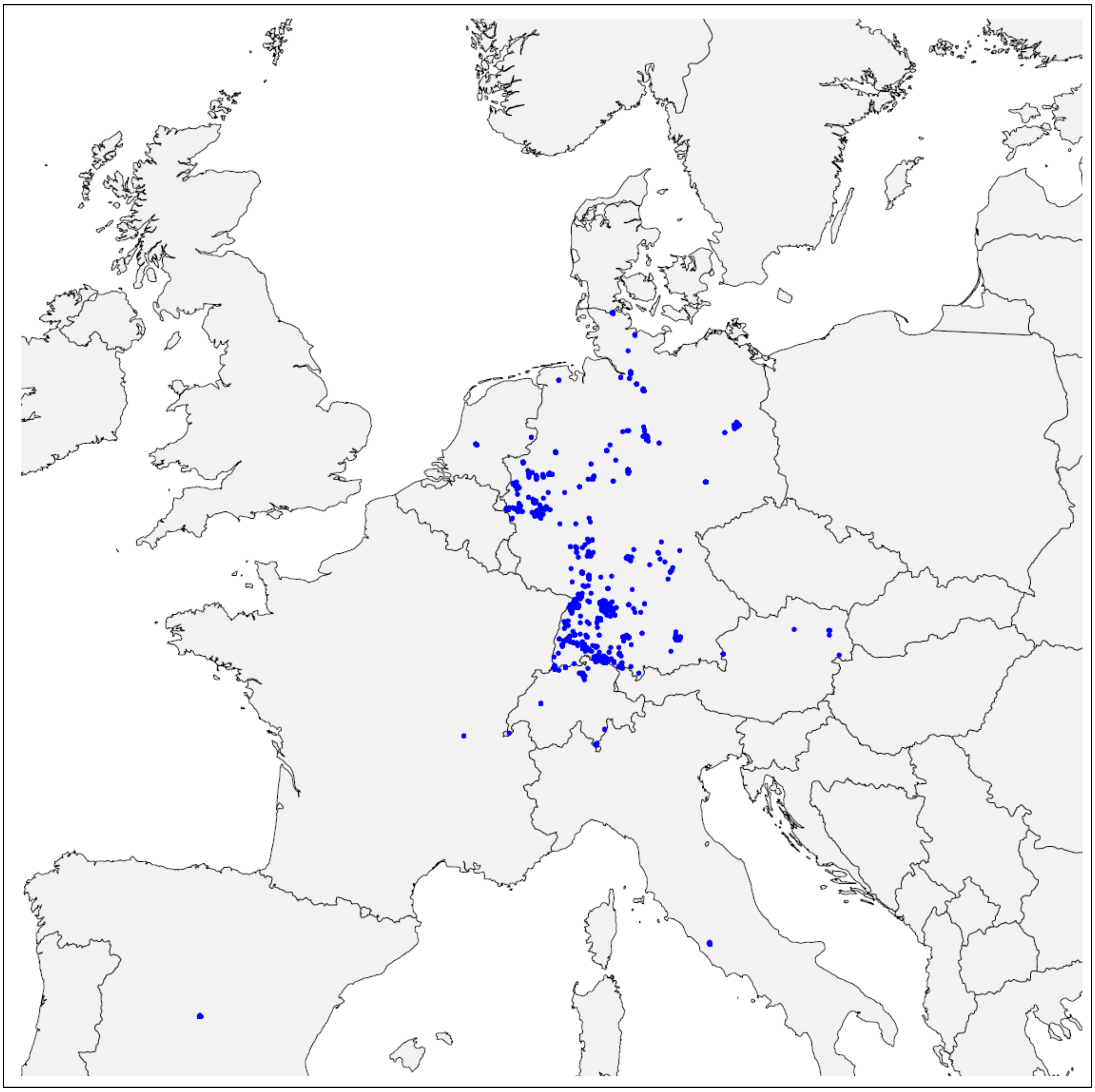

Figure S2. Screenshots from the well-being science app (left to right: Main screen with well-being rating; weekly statistic of mean well-being scores of oneself and others; daily statistic of mean well-being score trajectory of oneself and others).