Compliance in Ambulatory Assessment Studies: Investigating Study and Sample Characteristics as Predictors

Charlotte Ottenstein and Linda Werner

Abstract
Ambulatory assessment (AA) studies are becoming more and more popular. However, it can be challenging to motivate participants to comply with study protocols. The aim of the present study was to investigate possible predictors of compliance in AA studies with diverse samples and study designs. To do so, we extracted compliance information, study characteristics, and sample characteristics from 488 previously published studies. The average compliance across the studies was rather high. The total number of measurement occasions and the number of study days were negatively related to the compliance rate. Moreover, a higher percentage of healthy controls in clinical studies was associated with a higher compliance rate. By contrast, other study characteristics (e.g., the amount of financial compensation) and sample characteristics (clinical vs. healthy sample) were not related to compliance. The findings have implications for the design of future AA studies.

Keywords
compliance, ambulatory assessment, experience sampling method, daily diary study, ecological momentary assessment, adherence

Parallel to the technical development of mobile devices and smartphones, interest in conducting ambulatory assessment (AA) studies has grown rapidly in recent decades (e.g., Iida et al., 2012). The people who participate in these studies are usually asked to fill out short questionnaires multiple times a day. However, there are huge differences between such studies with respect to the characteristics of the study (e.g., study design, incentives) and the characteristics of the underlying sample (e.g., age, clinical vs. healthy participants). Participants’ compliance also differs to a large extent between AA studies. A low overall level of compliance, and hence a reduced sample size, can result in lower statistical power, estimation problems, and issues with the quality of the data (e.g., Button et al., 2013; Maxwell et al., 2008; Schönbrodt & Perugini, 2013). Knowing which study characteristic or sample characteristic are related to compliance might help increase compliance in future AA studies.

To date, only a few studies have investigated predictors of compliance in AA studies. Morren et al. (2009) systematically reviewed 62 studies on pain. The length of the diaries was negatively associated with the compliance, and the compliance was higher when the participants were older. Jones et al. (2018) presented a quantitative meta-analysis on predictors of compliance in AA studies. They included 126 studies on substance use and found that individuals with a substance use disorder showed lower compliance than individuals from community-based samples. In this meta-analysis, the frequency of prompting or the length of the AA period were unrelated to compliance. Vachon et al. (2019) conducted a meta-analysis including 79 studies with samples with severe mental disorders. Compliance was higher in studies with a fixed sampling scheme, fewer measurement occasions per day, and larger time intervals between two consecutive measurement occasions. Rintala et al. (2019) used data from 10 AA studies and found that compliance was higher in older participants, in female participants, and in healthy participants compared with participants with psychosis. They found no difference in the compliance between participants with depression and healthy participants. Stone et al. (2003) experimentally manipulated the sampling density (i.e., the number of prompts per day) in a sample of patients with chronic pain. They found high compliance rates that were not affected by the sampling density.

1University of Koblenz–Landau, Landau, Germany

Corresponding Author:
Charlotte Ottenstein, Center for Methodologies, Diagnostics and Evaluation, University of Koblenz–Landau, Fortstr. 7, Landau 76829, Germany.
Email: ottenstein@uni-landau.de
These results might serve as a basis for deriving recommendations about AA study designs, but the meta-analyses focused on rather special samples (Jones et al., 2018; Vachon et al., 2019), and other results relied on only one study (Stone et al., 2003) or on a small number of studies (e.g., Rintala et al., 2019). Because most of the analyzed studies had rather homogeneous designs, Rintala et al. (2019) recommended that the predictors of compliance should be investigated further in AA studies with more variability in design characteristics. To date, only a little research has incorporated a large number of published studies and investigated predictors of compliance between studies. There are three meta-analyses or systematic reviews that have focused on specific populations (i.e., substance users, Jones et al., 2018; individuals with pain, Morren et al., 2009; individuals with severe mental disorders, Vachon et al., 2019). The present study was aimed at identifying relevant predictors of compliance in AA studies with diverse designs and diverse (clinical and healthy) populations. To do so, we reviewed published AA studies and extracted the compliance and a variety of study and sample characteristics.

**Possible Predictors of Compliance**

**Incentives**

When planning AA studies, researchers must find an optimal design that will allow them to assess the constructs of interest with high validity and reliability. At the same time, researchers want to ensure that participants are motivated to comply with the AA protocol. In practice, researchers apply different strategies to increase participants’ compliance. One strategy is to provide incentives such as financial compensation, course credit, prize rewards, or tailored feedback. Incentives can be interpreted as a symbol of appreciation, which in turn increases the probability of cooperative study participation because participants want to return the appreciation (Pforr, 2015). Therefore, we expected that the amount of incentives would be positively related to compliance in AA studies. Moreover, incentives are often contingent on a certain degree of compliance. Therefore, we conducted an exploratory investigation of whether the mode of incentive was related to compliance.

To date, only a few studies have investigated whether incentives are effective in increasing compliance. In their meta-analysis, Vachon et al. (2019) found that the amount of incentives was positively related to compliance. Harari et al. (2017) compared three studies with student samples and varied the kind of incentive. Compliance was lower in the study with course credit and feedback, and it was higher in studies with compensation and feedback or with prize rewards and feedback. Newcomb et al. (2018) investigated whether weekly tailored feedback would increase compliance in a study with young men who have sex with men. Unexpectedly, they found no effect on compliance. We conducted an exploratory investigation of whether the kind of incentive was related to compliance.

**Contact With Participants**

Another strategy that should help maintain participants’ motivation to continue with the study is to contact them during the study (e.g., Ansell et al., 2015; Gloster et al., 2017). This should reduce the anonymity of the researchers and, thus, should enhance participants’ personal commitment. Moreover, technical issues or other problems can be resolved, which might improve participants’ motivation to continue to participate. Hence, we hypothesized that participants’ compliance would be higher in studies in which the researchers contacted the participants during the study.

**Participant Burden**

Besides numerous advantages of AA studies (e.g., real-time assessment to circumvent recall biases, investigation of within-person processes, real-life assessment to enhance generalizability; e.g., Ebner-Priemer & Trull, 2009), participant burden due to the repeated assessments might reduce the quality of the data and compliance (e.g., Morren et al., 2009; Rintala et al., 2019; Stone et al., 2003). There are several study characteristics that might increase participant burden and, hence, reduce compliance: (a) the length of the AA period, (b) the length of the repeated questionnaires, and (c) the number of daily questionnaires.

The length of AA studies depends on the research question that should be answered. For example, changes that occur over a longer period (e.g., a reduction in depression during therapy) require AA studies that are longer than changes that occur over a shorter period (e.g., fluctuations in momentary mood). The length can range from 1 day (e.g., Krkovic et al., 2018) to 1 year (e.g., Burg et al., 2017). Compliance typically decreases over the course of an AA study (Comulada et al., 2015; Fuller-Tyszkiewicz et al., 2013). We therefore expected that compliance would be lower for longer AA periods.

The longer the repeated questionnaires are, and thus, the more time participants must spend with the questionnaires, the higher the participant burden should be. In a systematic review of diary studies investigating pain, Morren et al. (2009) found that the shorter the questionnaires were, the higher compliance was. By contrast, Rintala et al. (2019) found no association between the number of questions and compliance. However, the authors mentioned that the low level of variability in the lengths of the questionnaires included in the study might explain this lack of association. Thus, a study with more diverse questionnaire lengths is needed. We hypothesized
that compliance would be lower when single repeated assessments took longer to complete.

AA studies differ greatly in the number of questionnaires that must be completed each day (i.e., the number of prompts per day). Wen et al. (2017) reviewed AA studies with minors and reported that in some studies, participants were prompted more than 25 times each day. Characteristics that have high change rates (e.g., mood) should be assessed more often than characteristics that change more slowly (e.g., attitude toward one’s employer) or that occur only at a specific time of day (e.g., sleep quality). A study by Burke et al. (2017) revealed that the participants were not accepting of more than five measurement occasions per day. We expected that compliance would be lower for larger numbers of measurement occasions per day.

**Intervention Studies**

Participants can benefit from AA intervention studies more than from AA studies that focus on only the assessment of everyday characteristics. The intrinsic motivation to be compliant with the study protocol might be higher because participants can use their study participation to increase desirable outcomes or to decrease undesirable outcomes. Hence, we hypothesized that compliance would be higher in intervention studies than in studies without any intervention.

**Clinical Populations**

Conducting AA studies with clinical patients can be challenging. Ramsey et al. (2016) stated that “common mental health problems such as depression or anxiety disorders may diminish motivation and reduce confidence in using technology” (p. 211). There are only a few empirical results on this question. Participants with psychosis were less compliant than healthy participants (Rintala et al., 2019; Vachon et al., 2019). However, there was no difference in the compliance between participants with depression and healthy participants (Rintala et al., 2019). A study by Gershon et al. (2019) revealed that youth with bipolar disorder showed lower compliance than healthy controls. We therefore expected that compliance would be lower in clinical populations than in healthy control populations.

**Overview of the Present Research**

The aim of the present study was to identify relevant predictors of compliance in AA studies. We reviewed published AA studies and extracted compliance and a variety of study and sample characteristics. This approach enabled us to study predictors of compliance with sufficient variability. We did not investigate individual differences as most psychological studies do; rather, we investigated differences between studies.

Regarding compliance, we investigated two different indicators: the compliance rate (i.e., the percentage of beeps that were answered) and the percentage of non-completers (i.e., the percentage of individuals who completed the study). We hypothesized that the amount of incentives would be positively related to compliance and that the length of the AA period, the length of the single repeated assessments, and the number of measurement occasions per day would be negatively related to compliance. Moreover, we hypothesized that compliance would be higher in studies with contact between the researchers and the participants and in intervention studies. Compliance was hypothesized to be lower in clinical populations than in healthy control populations. These hypotheses were pre-registered at the OSF (https://doi.org/10.17605/OSF.IO/ DT49F). Furthermore, we explored whether additional characteristics (individual feedback, age, sex, and mode of incentive) were associated with compliance. We report how we determined our sample size, all data exclusions, and all measures in the study.

**Method**

**Study Selection**

When we computed the necessary sample size, we could not find any comparable research that had investigated similar hypotheses or used a similar design. We expected low to medium effect sizes. To be conservative, our anticipated sample size comprised 400 studies to achieve a power of .90 for detecting a correlation of .16. To select studies that reported results from AA studies, we used the quarterly literature screenings from the Society for Ambulatory Assessment (http://www.saa2009.org/?page_id=62). We started with the most recent quarter (the first quarter of 2018) and went backward until we collected data from 481 studies (until the third quarter of 2014).

Because of the wide range of AA topics, we focused on studies that addressed the topics of affect and well-being. This is a broad research topic and covers research interests from different fields (Diener et al., 2018). Articles that were not considered included topics like physical diseases, physical activity, or substance use (without any connection to affect and well-being). Figure 1 presents the entire process by which studies were excluded. Next, we removed articles that were listed multiple times. We additionally checked for whether the reported analyses were based on the same study. In this case, we included only the article with the largest sample size in our analyses. Some articles were not available through an online search. In this case, we tried to contact the corresponding author. However, a small number of articles could not be procured. Some studies did not include AA data (e.g., meta-analyses, reviews,
or theoretical articles) and were therefore excluded. Finally, we excluded studies that reported no information about the compliance rate. In some cases, this was because the listed AA studies had been solely event-based, which means that there was no planned maximum number of measurement times, and in turn, there was no way to measure the compliance rate. Because of this, all of the studies that we included in the analyses refer to signal- and/or interval-based AAs.

**Study Description**

A total of 99.8% of the 481 studies we analyzed were first published between 2014 and 2018. The average sample size was 121.85 (SD = 217.56, Mdn = 81, Min = 5, Max = 3,867). A total of 436 studies reported the mean age of the underlying sample (studies that reported only an age range were not included in the calculation of the average age). The mean age was 31.78 years (SD = 13.07, Min = 9.53, Max = 81.18). A total of 461 studies reported information about the sex distribution. On average, 63% were female (Min = 0%, Max = 100%). A total of 265 studies provided information about the professional status of the participants. The average percentage of students was 68% (n = 223, SD = 42%, Mdn = 100%, Min = 0%, Max = 100%), the average percentage of employed participants was 33% (n = 235, SD = 41%, Mdn = 0%, Min = 0%, Max = 100%), and the average percentage of unemployed participants was 6% (n = 224, SD = 15%, Mdn = 0%, Min = 0%, Max = 82%).

A total of 468 out of 481 studies reported information about the data collection medium (402 studies used one data collection medium, 54 studies used two data collection media, 12 studies used three or more data collection media): 185 used an electronic device (e.g., a handheld
device), 138 used smartphones (95 used smartphones that the study lent to the participants, 33 used the participants’ own smartphones, 20 gave no information about the smartphone owner), 59 used email/online questionnaires, 30 collected data with paper-and-pencil questionnaires. A total of 69 studies reported that they used another data collection method in addition to the questionnaires: 16 studies used accelerometers, 11 studies used actigraph devices, 10 studies used saliva samples, 7 used GPS information, and 35 studies used other methods that were not identified any further.

**Extraction and Coding**

**Compliance Rate.** We defined the compliance rate as the percentage of completed measurement occasions for participants who did not drop out across the entire survey period. This value was reported in different ways across different studies (e.g., the number of completed measurement occasions, percentage of completed measurement occasions, average number of completed measurement occasions per person or per day), and some studies reported contradictory values. To decide which of the reported values to use in our analyses, we applied the following rules: If the number of planned and completed measurement occasions was reported, we calculated the exact compliance rate. If one of these values was not available, we used the reported percentage. If none of these values were available, we used any other reported value (e.g., the average number of measurement occasions completed per person or per day) and calculated the compliance rate out of this information (combined with the sample size or the number of days in the study).

**Nondropout.** As another compliance measure, we used the percentage of participants who completed the study (who did not drop out during the study). Thus, a high percentage of nondropouts indicates a high compliance.

**Incentives.** The compensation for participating in AA studies was categorized into four categories (money, voucher, course credit, or material gift). Moreover, we documented the maximum possible amount of the financial compensation (in US$) if it was given in the article. If the currency was other than US$, we converted it to US$ (we used the exchange rates from December 19, 2018; that was the time of data preparation).

**Contact.** This indicated whether the research staff had contact with the participants, if they called them to ask about any technical problems, or if the participants visited the research staff during the study period. If nothing was reported about contact, it was interpreted as no contact.

**Length of AA Period.** The length of the AA period was recorded in two ways: (a) the total number of scheduled measurement occasions and (b) the number of study days on which the repeated measurement occasions took place. Days without scheduled measurement occasions within the study period were not included.

**Duration of Measurement Occasions.** Some studies reported the average amount of time participants needed to finish one measurement occasion (in seconds). For the studies that reported no time but instead reported the number of items, we converted the number of items into a measure of time by assuming that participants took 5 seconds to answer each item. The conversion factor of 5 seconds was chosen because of previous research. In three different AA studies, Arndt et al. (2018) found that the average time participants needed to respond to simple emotion intensity items was 1.37 to 4.17 seconds. Because these previous items were rather short, we rounded up to 5 seconds per item.

**Number of Measurement Occasions Per Day.** This measure included the number of scheduled signal- and interval-based measurement occasions per day. If the number of measurement occasions differed between days, the average number of measurement occasions was taken. Studies that reported only a range of daily measurement occasions were not included in the analyses.

**Intervention Study.** Intervention studies were studies that described any intervention that occurred within the AA survey period. If there was no mention of an intervention, we assumed that there had been no intervention.

**Population.** We categorized whether the sample was a clinical sample or not. As clinical studies often include healthy control participants, we additionally included the relative percentage of healthy controls in our analyses.

**Additional Characteristics.** We extracted additional study and sample characteristics that were not part of the presented analyses (e.g., topic of the study, kind of device, ethnic background). The entire data set can be retrieved from https://doi.org/10.17605/OSF.IO/DT49F.

**Data Analytic Models**

Relations between metric variables were analyzed with correlations. Mean comparisons were computed with $t$ tests for independent samples (in case of variance heterogeneity, the Welch correction of the degrees of freedom was applied). All hypotheses were tested separately with both of the dependent variables that were used to measure compliance (compliance rate, percentage of nondropouts). All data were
Table 1. Descriptive Statistics and Bivariate Correlations for Metric Variables.

|                  | 1            | 2            | 3            | 4            | 5            | 6            | 7            | 8            |
|------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| 1. Compliance rate (%) | —            | .23***       | —            | —            | —            | —            | —            | —            |
| 2. Percentage of nondropouts (%) | [.11, .34]   | —            | —            | —            | —            | —            | —            | —            |
| (n = 254)        |              |              |              |              |              |              |              |              |
| 3. Maximum possible amount of financial compensation (US$) | .01          | .03          | —            | —            | —            | —            | —            | —            |
| (n = 211)        |              |              |              |              |              |              |              |              |
| 4. Number of scheduled measurement occasions | .39***       | —            | —            | —            | —            | —            | —            | —            |
| (n = 425)        |              |              |              |              |              |              |              |              |
| 5. Number of study days | [.20, -.02]  | [-.20, .04]  | [.05, .31]   | [.71, .79]   | —            | —            | —            | —            |
| (n = 463)        |              |              |              |              |              |              |              |              |
| 6. Duration of measurement occasions (s) | -.11*        | -.09         | -.03         | -.01         | .03          | —            | —            | —            |
| (n = 130)        |              |              |              |              |              |              |              |              |
| 7. Number of measurement occasions per day | .00          | .02          | -.05         | .07          | -.24***      | -.06         | —            | —            |
| (n = 438)        |              |              |              |              |              |              |              |              |
| 8. Percentage of healthy controls (%) | [.02, .29]   | [-.22, .15]  | [-.34, .10]  | [-.26, .04]  | [.31, .38]   | [.11, .39]   | —            | —            |
| (n = 200)        |              |              |              |              |              |              |              |              |

Note. 95% confidence intervals are displayed in square brackets.
*p < .05. **p < .01. ***p < .001.

analyzed with R (R Core Team, 2018). The data and R script can be retrieved from the OSF (https://doi.org/10.17605/OSF.IO/DT49F). The number of studies that were used for the different analyses differed because some characteristics were not included in every study.5

Results

Descriptive Results

Descriptive statistics and correlations between metric variables can be found in Table 1. Overall, the compliance rate (n = 481, M = 77%, SD = 14%, Mdnn = 79%, Min = 18%, Max = 100%) and the percentage of nondropouts (n = 254, M = 89%, SD = 11%, Mdnn = 93%, Min = 28%, Max = 100%) were rather high.

Preregistered Hypotheses

Incentives. We first hypothesized that the amount of incentives would be positively related to compliance. A total of 274 (57%) of the 481 studies reported that they provided incentives for study participation. The maximum possible amount of compensation was unrelated to the compliance rate and was also unrelated to the percentage of nondropouts (see Table 1). Controlling for the length of the AA period (number of study days and number of scheduled measurement occasions) did not change these results.

A total of 271 studies provided information about the kind of incentive. Only 11 studies provided material gifts. Thus, we refrained from analyzing the compliance of those studies separately. Most studies (65%) provided money as the only incentive. Figure 2 shows the compliance rates of studies with different incentive structures. The highest compliance rate was achieved when money and course credit were given, followed by studies that gave money only. We did not analyze whether mean differences were significantly different from zero due to some small group sizes. A total of 147 studies provided information about the percentage of nondropouts and about the kind of incentive. The highest percentage of nondropouts (i.e., the highest compliance) was achieved
in studies that used course credit, followed by studies that used course credit and money and studies that used course credit and vouchers (Figure 3).

**Contact.** To investigate the second hypothesis, we compared studies that explicitly reported that they had contact with the participants during the study with studies that

---

**Figure 2.** Separate illustration of the compliance rate depending on the incentive structure.

*Note. n = 271; the five studies from the first group provided material gifts only; boxes are drawn with widths proportional to the square roots of the number of observations in the groups.*

---

**Figure 3.** Separate illustration of the percentage of non-dropouts depending on the incentive structure.

*Note. n = 147; the three studies from the first group provided material gifts only; boxes are drawn with widths proportional to the square roots of the number of observations in the groups.*
reported nothing about contact. In 80 studies, the researchers stayed in contact with the participants during the study, but the compliance rate did not differ between studies with contact versus those without contact, $t(479) = 0.59, p = .554$, 95% CI $[-2.37, 4.42], d = 0.07$. Moreover, the percentage of nondropouts did not differ between studies with contact versus without contact, $t(252) = 0.07, p = .942$, 95% CI $[-3.47, 3.74], d = 0.01$.

Length of AA Period. For the third hypothesis, we tested whether (a) the number of scheduled measurement occasions and (b) the number of study days was related to compliance. The number of measurement occasions was negatively related to the compliance rate, but it was unrelated to the percentage of nondropouts (see Table 1). Similarly, the number of study days was also negatively related to the compliance rate, but it was unrelated to the percentage of nondropouts (see Table 1). However, the effect sizes of the associations between the two indicators of the length of the AA period and the compliance indicators were similar (the $r$ ranged from $-0.08$ to $-0.11$).

Duration of Measurement Occasions. Fourth, we hypothesized that the longer the single repeated assessment lasted, the lower the compliance would be. Contrary to this hypothesis, the average amount of time the participants needed to fill out a single assessment was not related to the compliance rate or to the percentage of nondropouts (see Table 1).

Number of Measurement Occasions Per Day. The fifth hypothesis stated that the higher the number of measurement occasions per day, the lower the compliance would be. The results showed that the number of daily measurement occasions was not related to the compliance rate or to the percentage of nondropouts (see Table 1).

Intervention Study. In the sixth hypothesis, we assumed that compliance would be higher in intervention studies than in studies without any intervention. In contrast to our hypothesis, the compliance rate was higher in studies without an intervention ($n = 428, M = 77\%, SD = 14\%$) than in intervention studies ($n = 54, M = 73\%, SD = 15\%), $t(479) = 2.29, p = .022, 95\% CI [0.67, 8.71], d = 0.33$. Similarly, the percentage of nondropouts was higher in studies without an intervention ($n = 227, M = 90\%, SD = 11\%$) than in intervention studies ($n = 27, M = 84\%, SD = 13\%$), $t(252) = 2.89, p = .004, 95\% CI [2.02, 10.73], d = 0.59$.

Clinical Versus Healthy Population. Last, we investigated the hypothesis that compliance would be lower in clinical populations than in healthy populations. To do so, we compared studies that included only healthy participants with clinical studies. We did not find any difference between these samples in the compliance rate, $t(470.16) = 0.51, p = .613$, 95% CI $[-1.85, 3.12], d = 0.05$, or in the percentage of nondropouts, $t(251) = 0.06, p = .951, 95\% CI [-2.68, 2.85], d = 0.01$. The clinical samples often included not only individuals with a mental disorder but also healthy control participants. We, therefore, additionally tested for whether the relative percentage of healthy controls in clinical studies was related to compliance (see Table 1). The results showed that the percentage of healthy controls was positively related to the compliance rate. By contrast, it was unrelated to the percentage of nondropouts.

Exploratory Analyses

Feedback. A total of 63 studies (13%) reported that they promised individual feedback to the participants. We explored whether compliance was higher in studies in which participants could receive individual feedback. The comparison revealed no difference in the compliance rate between studies with feedback compared with studies without feedback, $t(479) = 0.64, p = .523, 95\% CI [-2.53, 4.97], d = 0.09$. The percentage of nondropouts also did not differ between these study groups, $t(252) = -0.85, p = .397, 95\% CI [-5.59, 2.22], d = 0.15$.

Age. Age was not related to the compliance rate ($r = .04, p = .387$) or to the percentage of nondropouts ($r = -1.0, p = .121$).

Sex. The percentage of female participants was not related to the compliance rate ($r = .04, p = .410$) or to the percentage of nondropouts ($r = .00, p = .958$).

Mode of Incentive. In some studies, the incentives (e.g., money, vouchers, course credit) were given irrespective of the level of compliance, whereas in other studies, a certain level of compliance must have been achieved for the incentive to be given. There was no difference regarding the compliance rate between those studies, $t(112.05) = -1.37, p = .174, 95\% CI [-7.05, 1.29], d = 0.23$. In contrast, the percentage of nondropouts was higher when the incentive was linked to a certain degree of compliance, $t(32.51) = -2.30, p = .028, 95\% CI [-15.97, -0.97], d = 0.60$. We therefore checked whether the mode of incentive moderated the group differences between studies that promised a certain incentive versus studies that did not promise this incentive. There was no interaction between the mode of incentive and the kind of incentive or the amount of financial compensation in predicting the percentage of nondropouts (all $|r|s \leq 1.62$, all $p$s $\geq .110$).

Discussion

The aim of the present study was to investigate possible predictors of compliance in AA studies with diverse samples
and study designs. The average level of compliance was rather high and similar to previous meta-analyses and systematic reviews (de Vries et al., 2020; Jones et al., 2018; Morren et al., 2009; Vachon et al., 2019). Some of the pre-registered hypotheses were supported: The total number of measurement occasions and the number of study days were negatively related to the compliance rate. Moreover, clinical studies with higher percentages of healthy controls had higher compliance rates. All other hypotheses (regarding the amount of financial compensation, contact between researchers and participants during the study, length of the repeated assessments, number of measurement occasions per day, intervention study vs. nonintervention study, and clinical vs. healthy sample) were not supported by the results. One comparison even went in the opposite direction: The compliance rate and the percentage of nondropouts (i.e., the compliance) were lower in intervention studies compared with studies without any intervention.

Initially, we assumed that participants in intervention studies would be more motivated to comply with the study protocol because they benefited from their participation. This effect (if present) may have been covered up by other characteristics that differed between the intervention studies and the “pure assessment studies.” Participating in an intervention study might come along with increased time expenditure and a personally demanding engagement with oneself. This might have increased the participant burden and in turn reduced the compliance. Previous studies have shown that a higher participant burden due to a longer AA period was associated with diminished compliance (Comulada et al., 2015; de Vries et al., 2020; Fuller-Tyszkiewicz et al., 2013). Our findings were in line with those results: We found that a larger total number of scheduled measurement occasions and a larger number of study days were associated with a reduction in the compliance rate. The size of those associations was rather small. However, the length of the AA period should be carefully determined to optimize the compliance, especially when it is anticipated that the compliance level would be low due to other factors. The other indicators of participant burden (duration of measurement occasions, number of measurement occasions per day) were not associated with the compliance. Consistent with those results, experimental AA studies (Eisele et al., 2020; Hasselhorn et al., 2021) have shown that the sampling frequency (i.e., the number of measurement occasions per day) had no effect on the compliance. Regarding the length of the AA questionnaires, previous results were mixed (Eisele et al., 2020; Hasselhorn et al., 2021), suggesting that the processes under which possible effects occur, should be further investigated in future research. It might be that participants of AA studies cope differently with the diverse design characteristics that should enhance or reduce participant burden (Hasselhorn et al., 2021). Hence, future studies should carefully determine the length of the AA period that is needed to answer the research questions. To ensure high compliance, the burden from participating in a study should not exceed a critical level. Moreover, a high participant burden might be associated with other undesirable outcomes, such as reactivity or careless responding (Eisele et al., 2020; Hasselhorn et al., 2021). Thus, it remains a challenging task to find an optimal balance between the necessary amount of data and a low participant burden.

In our study, a higher percentage of healthy controls was associated with a higher compliance rate in clinical studies. Even if this association was small, it points out that a higher effort is needed to ensure a sufficient compliance in clinical samples. Our finding is consistent with previous findings on differences in compliance between healthy participants and participants with a mental disorder (Gershon et al., 2019; Rintala et al., 2019; Vachon et al., 2019). However, this difference did not occur when comparing participants suffering from depression with healthy controls (Rintala et al., 2019). Moreover, Schiepek et al. (2016) found lower compliance rates in participants diagnosed with personality disorders compared with affective disorders and neurotic, stress-related, and somatoform disorders. Hence, the difference in complying with the study protocol seems to depend on the kind of mental disorder. More research is required to investigate the compliance-relevant differences between these disorders. This could help researchers who are interested in specific mental disorders find solutions to optimally motivate their clinical participants and, thus, to reach the best possible level of compliance.

We conducted an exploratory test of whether age and sex were associated with compliance. Age was unrelated to compliance. This finding is in line with the meta-analysis by Vachon et al. (2019), but there are also studies that found higher compliance in older participants (Morren et al., 2009; Rintala et al., 2019). Vachon et al. (2019) assumed that these different findings might be due to differences in the inclusion criteria (e.g., specific sample, specific presentation form of AA questionnaires). Sex was also unrelated to compliance in the present study. This is inconsistent with previous findings that female participants were more compliant than male participants (Rintala et al., 2019; Vachon et al., 2019). However, there are also findings from single studies that showed no difference in compliance between female and male participants (e.g., Courvoisier et al., 2012; Hartley et al., 2014). It seems that whether there is a sex difference in compliance or not depends on other study or sample characteristics. More research is needed to identify the conditions under which sex differences occur.

We expected that a larger amount of incentives and contact between researchers and participants would be two motivation strategies that could help improve compliance. However, our results did not support these assumptions. Contacting participants during the study might have other
positive effects, such as resolving technical issues, but it does not seem to increase compliance. The amount of financial compensation was also unrelated to compliance. This is in line with the findings from de Vries et al. (2020) who found no difference in the compliance levels between studies with or without incentives. However, exploratory analyses have shown that the common strategy of linking the incentive to a certain degree of compliance helps increase the percentage of nondropouts (but not the compliance rate). Furthermore, it might be the case that some study characteristics that are known before starting the study (e.g., the amount of incentives) unfold their positive effect before the study rather than during study participation. Thus, the positive effect might not emerge in a higher compliance but in a greater willingness to participate in the first place (e.g., Sharp et al., 2006).

**Strengths and Limitations**

The present analyses included up to 481 studies. This enabled us to compare a wide range of different study characteristics and sample characteristics. Thus, we followed the demand for an investigation of predictors of compliance in AA studies with more variability in design characteristics (Rintala et al., 2019).

Before we began collecting the data (i.e., the study selection and information coding), we computed the required sample size to achieve a sufficient level of statistical power. However, most analyses were based on a subsample of studies because of missing information. Thus, the statistical power was in fact somewhat lower for some analyses. These analyses should be interpreted with caution. Regarding the missing information, several studies did not report any information about the number of measurement occasions. Hence, they had to be excluded from the analyses because no compliance rate could be computed (see Figure 1). Stone and Shiffman (2002) already highlighted the importance of reporting compliance information. We want to emphasize this point and recommend that researchers report compliance whenever they report the results of an AA study.

The present study focused on studies on the topics of affect and well-being. Hence, the generalizability of the current results to other topics is unknown. However, it seems implausible that the results were strongly affected by the topics of the studies. Nevertheless, this assumption should be investigated in future studies.

A lot of studies reported a percentage of nondropouts of (nearly) hundred. This led to a negatively skewed distribution that might have reduced the associations with other variables. Hence, the true associations between the independent variables and the percentage of nondropouts were likely underestimated.

Our study design allowed only correlational conclusions. Thus, it is not possible to interpret the associations between study and sample characteristics and compliance in a causal manner. There might be confounding variables that were not considered (e.g., educational or socioeconomic background of study participants). This limits the explanatory power of some results, for example, the findings regarding the kind of incentive because some incentives are used only in studies with students. A valuable supplement might be experimental studies that vary certain design characteristics (e.g., the number of daily measurement occasions or the length of the repeated questionnaires) and then compare the compliance levels of these groups. Such studies could circumvent another drawback of our study: The information that was entered into our analyses was aggregated in nature (e.g., the mean compliance across participants and across time). In experimental studies, intraindividual change in compliance can be modeled (including time-of-day effects or day-of-the-week effects; Rintala et al., 2019) and can be predicted by the experimental factor or by individual characteristics.

**Conclusion**

Most of the preregistered hypotheses were not supported by the results. Thus, it seems that compliance is not that prone to some researchers’ decisions regarding study characteristics or sample characteristics that have been made in the hope that data quantity and quality will be high. Nevertheless, the results suggest that the length of the AA period (number of scheduled measurement occasions and number of study days) should be chosen carefully to optimize compliance. Moreover, money as an incentive seems to work better than course credit or vouchers in optimizing compliance. Another practical recommendation is that linking the incentives to a certain degree of compliance might reduce the dropout during the AA period. However, such recommendations are based on correlational data and should be complemented by experimental studies.

The present research focused solely on predicting the compliance of study participants. It would be interesting for future studies to complement these results by predicting individuals’ reasons for participating in an AA study. This might also be affected by study characteristics (e.g., amount of incentives, length of AA period) and sample characteristics (e.g., clinical vs. healthy sample, age, sex).

**Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.
References

Ansell, E. B., Laws, H. B., Roche, M. J., & Sinha, R. (2015). Effects of marijuana use on impulsivity and hostility in daily life. *Drug and Alcohol Dependence, 148*(March), 136-142. https://doi.org/10.1016/j.drugalcdep.2014.12.029

Arndt, C., Lischetzke, T., Crayen, C., & Eid, M. (2018). The assessment of emotional clarity via response times to emotion items: Shedding light on the response process and its relation to emotion regulation strategies. *Cognition and Emotion, 32*(3), 530-548. https://doi.org/10.1080/02699931.2017.1322039

Burg, M. M., Schwartz, J. E., Kronish, I. M., Diaz, K. M., Alcantara, C., Duer-Hefele, J., & Davidson, K. W. (2017). Does stress result in you exercising less? Or does exercising result in you being less stressed? Or is it both? Testing the bi-directional Stress-exercise association at the group and person (N of 1) level. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine, 51*(6), 799-809. https://doi.org/10.1007/s12160-017-9902-4

Burke, L. E., Shiffman, S., Music, E., Styn, M. A., Kriska, A., Smialowicz, A., Siewiorek, D., Ewing, L. J., Chasens, E., French, B., Mancino, J., Mendez, D., Strollo, P., & Rathbun, S. L. (2017). Ecological momentary assessment in behavioral research: Addressing technological and human participant challenges. *Journal of Medical Internet Research, 19*(3), e77. https://doi.org/10.2196/jmir.7138

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews. Neuroscience, 14*(5), 365-376. https://doi.org/10.1038/nrn3475

Comulada, W. S., Lightfoot, M., Swendeman, D., Grella, C., & Wu, N. (2015). Compliance to cell phone-based EMA among Latino youth in outpatient treatment. *Journal of Ethnicity in Substance Abuse, 14*(3), 232-250. https://doi.org/10.1080/153264014.986354

Courvoisier, D. S., Eid, M., & Lischetzke, T. (2012). Compliance to a cell phone-based ecological momentary assessment study: The effect of time and personality characteristics. *Psychological Assessment, 24*(3), 713-720. https://doi.org/10.1037/a0026733

de Vries, L. P., Baselmans, B. M. L., & Bartels, M. (2020). Smartphone-based ecological momentary assessment of well-being: A systematic review and recommendations for future studies. *Journal of Happiness Studies, 21*(6), 510. https://doi.org/10.1007/s10902-020-00324-7

Diener, E., Oishi, S., & Tay, L. (2018). Advances in subjective well-being research. *Nature Human Behaviour, 2*(4), 253-260. https://doi.org/10.1038/s41562-018-0307-6

Ebner-Priemer, U. W., & Trull, T. J. (2009). Ambulatory assessment. *European Psychologist, 14*(2), 109-119. https://doi.org/10.1027/1016-9040.14.2.109

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myingermey, I., & Viechtbauer, W. (2020). The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment. Advance online publication*. https://doi.org/10.1177/1073191120957102

Fuller-Tyszkiewicz, M., Skouteris, H., Richardson, B., Blore, J., Holmes, M., & Mills, J. (2013). Does the burden of the experience sampling method undermine data quality in state body image research? *Body Image, 10*(4), 607-613. https://doi.org/10.1016/j.bodyim.2013.06.003

Gershon, A., Kaufmann, C. N., Torous, J., Depp, C., & Ketter, T. A. (2019). Electronic Ecological Momentary Assessment (EMA) in youth with bipolar disorder: Demographic and clinical predictors of electronic EMA adherence. *Journal of Psychiatric Research, 116*(September), 14-18. https://doi.org/10.1016/j.jpsychires.2019.05.026

Gloster, A. T., Miché, M., Wersonbe, H., Mikoteit, T., Hoyer, J., Imboden, C., Bader, K., Meyer, A. H., Hatzinger, M., & Lieb, R. (2017). Daily fluctuation of emotions and memories thereof: Design and methods of an experience sampling study of major depression, social phobia, and controls. *International Journal of Methods in Psychiatric Research, 26*(3), e1578. https://doi.org/10.1002/mpr.1578

Harari, G. M., Müller, S. R., Mishra, V., Wang, R., Campbell, A. T., Rentfrow, P. J., & Gosling, S. D. (2017). An evaluation of...
students’ interest in and compliance with self-tracking methods: Recommendations for incentives based on three smartphone sensing studies. *Social Psychological and Personality Science*, 8(5), 479-492. https://doi.org/10.1177/1948550617712033

Hartley, S., Varese, F., Vasconcelos, S., Udachina, A., Barrowelough, C., Bentall, R. P., Lewis, S. W., Dunn, G., Haddock, G., & Palmier-Claus, J. (2014). Compliance in experience sampling methodology: the role of demographic and clinical characteristics. *Psychosis*, 6(1), 70-73. https://doi.org/10.1080/17522439.2012.752520

Hasselhorn, K., Ottenstein, C., & Lischetzke, T. (2021). The effects of assessment intensity on participant burden, compliance, within-person variance, and within-person relationships in ambulatory assessment [Manuscript submitted for publication]. Department of Psychology, University of Koblenz-Landau.

Iida, M., Shrout, P. E., Laurenceau, J.-P., & Bolger, N. (2012). Emotional stress and paranoia. *European Psychiatry: The Journal of the Association of European Psychiatrists*, 27(5), 537-563. https://doi.org/10.1016/S0304-3959(03)00040-X

Jones, A., Remmerswaal, D., Verveer, I., Robinson, E., Franken, I. H. A., Wen, C. K. F., & Field, M. (2018). Compliance with ecological momentary assessment protocols in substance users: A meta-analysis. *Addiction (Abingdon, England)*, 114(4), 609-619. https://doi.org/10.1111/add.14503

Krkovic, K., Krink, S., & Lincoln, T. M. (2018). Emotion regulation as a moderator of the interplay between self-reported and physiological stress and paranoia. *European Psychiatry: The Journal of the Association of European Psychiatrists*, 49(March), 43-49. https://doi.org/10.1016/j.eurpsy.2017.12.002

Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology*, 59, 537-563. https://doi.org/10.1146/annurev.psych.59.103006.093735

Morren, M., van Dulmen, S., Ouwerkkerk, J., & Bensing, J. (2009). Compliance with momentary pain measurement using electronic diaries: A systematic review. *European Journal of Pain (London, England)*, 13(4), 354-365. https://doi.org/10.1016/j.ejpain.2008.05.010

Newcomb, M. E., Swann, G., Estabrook, R., Corden, M., Begale, M., Ashbeck, A., Mohr, D., & Mustanski, B. (2018). Patterns and predictors of compliance in a prospective diary study of substance use and sexual behavior in a sample of young men who have sex with men. *Assessment*, 25(4), 403-414. https://doi.org/10.1177/1073191116667584

Pförr, K. (2015). *Incentives*. GESIS. https://doi.org/10.15465/GESIS-SG_001

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ramsey, A. T., Wetherell, J. L., Depp, C., Dixon, D., & Lenze, E. (2016). Feasibility and acceptability of smartphone assessment in older adults with cognitive and emotional difficulties. *Journal of Technology in Human Services*, 34(2), 209-223. https://doi.org/10.1080/15228835.2016.1170649

Rintala, A., Wampers, M., Myin-Germeys, I., & Viechtbauer, W. (2019). Response compliance and predictors thereof in studies using the experience sampling method. *Psychological Assessment*, 31(2), 226-235. https://doi.org/10.1037/pas0000662

Schiepek, G., Aichhorn, M., Gruber, M., Strunk, G., Bachler, E., & Aas, B. (2016). Real-time monitoring of psychotherapeutic processes: Concept and compliance. *Frontiers in Psychology*, 7, 604. https://doi.org/10.3389/fpsyg.2016.00604

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, 47(5), 609-612. https://doi.org/10.1016/j.jrp.2013.05.009

Sharp, E. C., Pelletier, L. G., & Lévesque, C. (2006). The double-edged sword of rewards for participation in psychology experiments. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement*, 38(3), 269-277. https://doi.org/10.1037/cjbs2006014

Stone, A. A., Broderick, J. E., Schwartz, J. E., Shiffman, S., Lifcher-Kelly, L., & Calvanese, P. (2003). Intensive momentary reporting of pain with an electronic diary: Reactivity, compliance, and patient satisfaction. *Pain*, 104(1), 343-351. https://doi.org/10.1016/S0304-3959(03)00040-X

Stone, A. A., & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine*, 24(3), 236-243. https://doi.org/10.1207/S15324796abm2403_09

Vachon, H., Viechtbauer, W., Rintala, A., & Myin-Germeys, I. (2019). Compliance and retention with the experience sampling method over the continuum of severe mental disorders: Meta-analysis and recommendations. *Journal of Medical Internet Research*, 21(12), e14475. https://doi.org/10.2196/jmir.6641

Wen, C. K. F., Schneider, S., Stone, A. A., & Spriuijt-Metz, D. (2017). Compliance with mobile ecological momentary assessment protocols in children and adolescents: A systematic review and meta-analysis. *Journal of Medical Internet Research*, 19(4), e132. https://doi.org/10.2196/jmir.6641