

Daily Diary Designs in Lifespan Developmental Psychology

Shevaun D. Neupert & Jennifer A. Bellingtier

This material has been published after copy-editing in *The Oxford Research Encyclopedia of Psychology* edited by O. Braddick
<https://doi.org/10.1093/acrefore/9780190236557.013.347>. This version is free to view and download for private research and study only. Not for re-distribution or re-use. ©Oxford University Press.

Summary

Daily diary designs allow researchers to examine processes that change together on a daily basis, often in a naturalistic setting. By studying within-person covariation between daily processes, one can more precisely establish the short-term effects and temporal ordering of concrete daily experiences. Additionally, the daily diary design reduces retrospective recall bias because participants are asked to recall events that occurred over the previous 24-hour period as opposed to a week or even a year. Therefore, a more accurate picture of individuals' daily lives can be captured with this design. When conclusions are drawn between people about the relationship between the predictors and outcomes, the covariation that occurs within people through time is lost. In a within-person design, conclusions can be made about the simultaneous effects of within-person covariation as well as between-person differences. This is especially important when many interindividual differences (e.g., traits) may exist in within-person relationships (e.g., states).

Daily diary research can take many forms. Diary research can be conducted with printed paper questionnaires, divided into daily booklets where participants mail back each daily booklet at the end of the day or entire study period. Previous studies have called participants on the telephone to respond to interview questions each day for a series of consecutive days, allowing for quantitative as well as qualitative data collection. Online surveys that can be completed on a computer or mobile device allow the researcher to know the specific day and time that the survey was completed while minimizing direct involvement with the collection of each daily survey. There are many opportunities for lifespan developmental researchers to adopt daily diary designs across a variety of implementation platforms to address questions of important daily processes.

The benefits and drawbacks of each method along with suggestions for future work are discussed, noting issues of particular importance for aging and lifespan development.

Keywords

Daily Diary Designs; Paper and Pencil; Telephone; Online; mTurk

Author Copy: Not for Distribution

Daily Diary Designs in Lifespan Developmental Psychology

Introduction

Daily diary methods are used to obtain repeated measurements from individuals during their daily lives (Almeida, 2005). Typically, participants report on events, experiences, behaviors, and emotional states for a set number of consecutive days; although the questions do not change from day to day, the expectation is that the participants' responses can change, capturing within-person (intraindividual) variability. By obtaining information about individuals' actual events, behaviors, etc. over short-term intervals, daily diaries circumvent concerns about ecological validity (applicability to real life) that constrain findings from laboratory research (Almeida, 2005; Bolger & Laurenceau, 2013). As Allport noted in 1942, "Acquaintance with particulars is the beginning of all knowledge – scientific or otherwise. In psychology the font and origin of our curiosity in, and knowledge of, human nature lies in our acquaintance with concrete individuals. To know them in their natural complexity is an essential first step. ... psychology needs to concern itself with life as it is lived, with significant total-processes of the sort revealed in consecutive and complete life documents" (p. 56).

Daily diary designs allow researchers to examine processes that change together on a daily basis, often in a naturalistic setting (Almeida, Wethington, & Kessler, 2002; Shiffman & Stone, 1998; Tennen, Suls, & Affleck, 1991). Perhaps the most valuable feature of diary methods is the ability to assess within-person processes (Almeida, 2005). This represents a shift from assessing mean levels of events and well-being between individuals to charting the day-to-day fluctuations in events and well-being within an individual as well as to identify their predictors, correlates and sequelae (Reis & Gable, 2000). For example, instead of asking whether individuals with high levels of work stress experience more distress than individuals with less

stressful jobs, a researcher can ask whether a worker experiences more distress on days when he or she has too many deadlines (or is reprimanded) compared to days when work has been stress free. This within-person approach allows the researcher to rule out temporally stable personality and environmental variables as third variable explanations for the relationship between events and well-being (Almeida, 2005). By studying within-person covariation between daily processes, one can more precisely establish the short-term effects and temporal ordering of concrete daily experiences (Almeida & Kessler, 1998; Bolger, DeLongis, Kessler & Schilling, 1989; Larson & Almeida, 1999; Lewinsohn & Talkington, 1979; Stone, Reed, & Neale, 1987). Additionally, the daily diary design reduces retrospective recall bias because participants are asked to recall events that occurred over the previous 24-hour period as opposed to a week or even a year (Kessler, Mroczek, & Belli, 1999). Therefore, a more accurate picture of individuals' daily lives can be captured with this design, which is critically important for increasing the external validity of lifespan developmental work (Freund, 2015). When conclusions are drawn between people about the relationship between the predictors and outcomes, the covariation that occurs within people through time is lost. In a within-person design, conclusions can be made about the simultaneous effects of within-person covariation as well as between-person differences. This is especially important when many interindividual differences (e.g., traits) may exist in within-person relationships (e.g., states). A focus on the individual as the unit of analysis across time, conditions, and situations can be used in important ways to examine behavior and development (Diehl, Hooker, & Sliwinski, 2015).

Daily Diary Designs and Lifespan Development

Birren and Bengtson (1988) criticized the field of aging for being data rich and theory poor, which may still be a relevant criticism (Freund, 2015). Daily diary designs tend to be

naturally data rich, but are also poised to make tremendous contributions to theories of aging. For example, Strength and Vulnerability Integration (SAVI; Charles, 2010) explicitly describes the importance of the passage of time for understanding changes in emotion regulation across adulthood. For example, the process of approaching and reacting to daily stressors may be experienced differently across the lifespan. SAVI predicts that older adults have both strengths and vulnerabilities that impact their emotional reactions to stressors. With advancing age, individuals may display emotion regulatory strengths in the form of strategies to avoid or limit the impact of negative experiences. These skills may translate to preventing the occurrence of a stressor or reframing the meaning of stressful events (Charles, 2010). SAVI also posits, however, that advancing age is associated with vulnerabilities in the form of physiological inflexibility (Charles, 2010) or fewer social supports (Schilling & Diehl, 2014) which may result in greater difficulty in responding to stressors that produce large and sustained responses. Importantly, SAVI suggests that there are limits to the age-related strengths, such that time functions as a moderator to increase or reduce age-related benefits in emotional functioning. Specifically, age-related improvements in emotion regulation abilities should be minimized immediately prior to or following a stressor, but reappear as time passes, and that situations of prolonged stress will reduce age-related emotion-response advantages. Thus, daily diary designs appropriately acknowledge the importance of time with respect to potential strengths and vulnerabilities by applying within-person models to examine antecedents, correlates, and consequences of daily events and states. In the remainder of this entry the various modes of data collection and recruitment, methodological considerations, and future directions of daily diary designs are described.

Modes of Data Collection and Recruitment

Before launching into a full-scale daily diary study, researchers new to the method may wish to consider conducting a small-scale pilot study with a few participants. Unanticipated issues are common regardless of the mode of data collection or recruitment, so a pilot study will likely save tremendous time and effort when the full study is launched.

Paper and Pencil

VA Normative Aging Study

A common mode of collecting daily diary data is with paper-and-pencil reports. The VA Normative Aging Study (NAS) is a longitudinal study of normal aging processes in men that began in the 1960s (see Spiro & Bossé, 2001, for additional information). Starting in August 2002, recruitment began for a paper-and-pencil 8-day daily diary study regarding stressors, physical symptoms, positive and negative affect, memory failures, pain, and social support (Neupert, Almeida, Mroczek, & Spiro, 2006a). Between August 2002 and April 2003, 529 NAS participants and their wives were contacted and invited to participate. Of these, 374 agreed, and 333 (181 men and 152 women, mean age = 72) returned usable surveys. Most participants completed all 8 days of the study, yielding a compliance rate of 99% and resulting in 2,649 days available for analysis (Neupert, Almeida, Mroczek, & Spiro, 2006b). Participants who completed the diary did not differ significantly from those who refused or from NAS participants who were not invited to participate. Instructions indicating when to complete the diary (approximately 30 minutes before going to bed) and when to return the surveys (when all eight were completed) were sent to each participant. Husbands and wives were instructed to complete their surveys independently. For eight consecutive evenings, participants completed short semistructured questionnaires about their daily experiences, and at the conclusion of the 8-day period, returned the diaries via postal mail. If they completed 5 or more of the 8 study days, they received \$30; if

they completed 4 or fewer days, they received \$15. Since this original paper-and-pencil daily diary study within the NAS, there have been three additional waves (or bursts) of daily diary data, each using the paper-and-pencil method but adding biomarkers, which are discussed in more detail in the Future Direction section.

The ACED study.

The authors have also conducted several paper-and-pencil daily diary studies with older adults recruited from local communities. Participants from the ACED (Anticipatory Coping Every Day) study (Neupert, Ennis, Ramsey, & Gall, 2016) were recruited through presentations at community activity groups targeted for older adults in central North Carolina. Potential participants were informed about the purpose of the study and were screened for cognitive impairment with the Short Blessed Test (Katzman et al., 1983). Individuals who scored ≤ 8 were included in the study. Participants read and signed informed consent forms and provided contact information for mailing of compensation. They were then given the packets of daily diary questionnaires, along with pre-paid envelopes to return to the primary investigator. Of 51 initial participants, 43 returned diary packets. Participants were aged 60-96 ($M = 74.65$, $SD = 8.19$), and included 39 women (90.7%), and 4 men (9.3%). Twenty-two (51.2%) were African American, 20 (46.5%) were White, and 1 (2.9%) was Asian. Participants completed diaries over nine consecutive days at home. The first packet collected baseline and demographic information (e.g., personality and SES). The following packets - to be opened on each of the eight days - contained items assessing daily stressors, anticipatory coping, memory failures, affect, and physical health symptoms. Participants mailed back completed packets and were subsequently debriefed over the phone. A \$20 gift card was sent via mail for completing five or more study days, and a \$10 card for four or fewer days. The compliance rate was 98.2%, with 380 out of a

possible 387 days completed. The NAS and ACED studies both had extremely high rates of daily compliance, but were recruited in different ways and represent different populations. The NAS diary data are an extension of an existing longitudinal dataset that was already selective regarding initial health status. The ACED study recruited new participants who were representative of the geographic region in terms of race, but over-represented women. Selection effects are important to keep in mind when conducting any study, but may become especially salient for a more burdensome design like a daily diary design.

Considerations.

One drawback of the paper and pencil mode of diary data collection is the extent to which participants comply with researchers' instructions, particularly with respect to the timing of diary reports (Bolger, Davis, & Rafaeli, 2003; Reis & Gable, 2000; Stone, Shiffman, Schwartz, Broderick, & Hufford, 2002). When participants complete diaries later than required by the protocol, they may rely on retrospection, and such retrospection could reintroduce the cognitive biases in self-reporting that diaries were initially designed to avoid (Green, Rafaeli, Bolger, Shrout, & Reis, 2006). One pattern is when participants forget to complete entries or miss some items within an entry. Often, participants attempt to make up for these by completing the forgotten or missed entries when they are completing the next diary (Green et al., 2006). This phenomenon is sometimes referred to as *hoarding* or *backfilling*, and can occur, for example, when participants in a daily diary study complete three days' entries in a row after missing two of the preceding days (Green et al., 2006). The frequency of hoarding or backfilling is not widely known, but in one report (Gable, Reis, & Elliot, 2000), hoarding was quite common in that 2/3 of the participants completed at least two diaries simultaneously.

Green et al. (2006) tested whether and to what extent paper diary data and electronic diary data were different when both are collected using comparable procedures. Across three studies using paper and electronic diaries, they concluded that compliance is more strongly related to issues of study design and participant motivation than it is an issue of whether a diary is administered in paper-and-pencil form or electronically. Flood, Zazzali, and Devlen (2013) also found evidence for measurement equivalence between paper and electronic daily diaries. Depending on the design of a study and the variables of interest, researchers can choose the mode of data collection that best suits both their own needs and the needs of their participants (Green et al., 2006).

Telephone

Considerations.

Stone, Kessler, and Haythornthwaite (1991) recommended telephone interviews as the most feasible way to collect nationally representative daily diary data. Telephone interviews make it possible to interview large numbers of people at a reasonable cost (Almeida et al., 2002). They also make it easier to obtain more detailed and accurate information about events such as daily stressors (e.g., timing and duration) through the use of question probes and complex skip patterns. In addition, the gain in greater control over data recording in telephone interviews can also lead to higher response rates and less missing data (Almeida et al., 2002). Telephone interviews typically have higher response rates than self-administered questionnaires in general population samples (Dillman, 1991). The researcher may also have more control over the quality of the interviews (e.g., whether the participant is paying full attention to the task, whether diaries are completed every day) (Almeida et al., 2002). Data are recorded more completely in telephone interviews than self-administered diaries because the interviewer can ensure that no questions are

skipped. Telephone interviews can also enhance the quality of data through probing incomplete or unclear responses (Almeida et al., 2002). Telephone administration also permits rapid feedback about nonresponse such as missed phone appointments, making it possible to implement special efforts to complete the interview (Almeida et al., 2002). When considering potential age differences in mode preference for participation in a daily diary study, Sacco, Smith, Harrington, Svoboda, and Resnick (2016) noted that older adults preferred telephone diaries to other electronic methods, and preferred to report in the morning regarding the previous day rather than in the evening about the current day.

National Study of Daily Experiences.

The largest telephone diary study to date is the nationally representative National Study of Daily Experiences (NSDE), part of the Midlife in the United States (MIDUS) project (Almeida et al., 2002). Participants were obtained through random digit dialing of telephone numbers and data were collected for the first wave in 1996-1997. Participants in the NSDE received \$20 for their participation in the project, and over the course of 8 consecutive evenings, participants completed short (approximately 10-15 minutes) interviews about their daily experiences. On the final evening of interviewing, participants also answered several questions about the previous week. Data collection spanned an entire year (March 1996 to April 1997) and consisted of 40 separate “flights” of interviews, with each flight representing the 8-day sequence of interviews from approximately 38 participants. The initiation of interview flights was staggered across the day of the week to control for the possible confounding between day of study and day of week. In all, 1,031 participants agreed to participate and completed an average of seven of the eight interviews, resulting in a total of 7,221 daily interviews (Almeida et al., 2002). Notably, an interviewer was involved in a one-on-one telephone call for each of those

7,221 daily interviews, representing a high level of effort and research staff burden. Although the advent of personal cell phones has changed the way in which many people use their phones and screen their calls, the second wave of the NSDE was conducted between 2004 and 2009 and was able to increase the total sample size to 2,022. Telephone survey centers can be hired to deploy telephone diary studies, but some researchers may not have the resources to hire one of these firms.

Interactive Voice Response (IVR) with a justice-involved sample.

IVR technology collects daily diary information over the phone, but without the need for a one-on-one interview. Neupert et al. (2017) recently completed a daily diary study of 117 offenders (100% men) participating in a community-based alcohol and drug abuse treatment program (76% probationers, 24% parolees). Potential study participants were recruited at the time of their referral to treatment, which was made by his substance abuse counselor, in consultation with the offender's probation or parole officer.

After a brief IVR training, participants called a toll-free phone number to access the IVR system and completed their first call-in IVR survey with the help of the study staff. The IVR system used recorded voice prompts to ask questions and participants answered the questions by pressing numbers on the telephone keypad. Participants were provided with a pocket reference card that included the toll-free number to call each day and several of the response stems to assist when making the daily call. On the card, the research staff wrote the participant's identification number (used to ensure confidentiality of responses when they called the IVR system) and the dates to begin and end calls to the IVR survey line. If the participant had a personal cell phone, the research staff assisted him in programming the number into his phone's contact list. Each IVR call lasted approximately 5 minutes. Participants were instructed to answer the daily IVR

questions using the timeframe of “since this time yesterday” and to complete one call each day for 14 consecutive days. Participants received \$10 gift cards for each IVR week (\$20 total for the IVR portion of the study).

A two-stage data validity check was used for the IVR dataset. At the first step, each case was required to have a valid study code. Based on this criterion 103 calls were dropped. Many of these calls were more than 95% incomplete and assumed to be either wrong numbers or they were known to be tests/demonstrations of the system by the research staff. At the second step, incomplete cases were checked to see if the survey was restarted by calling back (e.g., cases of a cell phone problem such as low battery or an accidental end to the call). Based on this criterion, 34 calls were dropped because the participant called back and completed a subsequent call (within minutes of the incomplete). This resulted in a final sample of 117 participants who collectively made 860 (out of 1638 possible) calls to the IVR survey line. Analyses were conducted to see if the compliance rates were associated with any study variables. Those who were more compliant reported lower craving for illegal drugs and lower use of illegal drugs, and older participants were also more compliant than younger participants.

Considerations.

The IVR method has a number of both strengths and limitations. IVR does not require a consistent phone number or location for the participant, which is important for resource-poor settings (Wiseman, Conteh, & Matovu, 2005), such as when participants may be illiterate, homeless, or at-risk for homelessness. The level of information regarding timing of assessment is the same in telephone-based or IVR-based diary studies, but IVR data are not as rich as telephone interviews. Specifically, all of the data collected through IVR are captured with the ten digits on a telephone keypad; there is no option for a narrative explanation of each response, and

there is no interviewer to monitor the quality of the responses. For researchers with a limited budget or for those interested in resource-poor populations, IVR may be an ideal mode of daily diary data collection.

Online

Daily diaries can also be conducted online, but age differences may be important to consider. Conner and Lehman (2012) noted that older adults may not use the internet at the same rate as younger adults and may not be as familiar with electronic diary equipment as younger adults. Online surveys that can be completed on a computer or mobile device allow the researcher to know the specific day and time that the survey was completed while minimizing direct involvement with the collection of each daily survey. One of the first decisions to be made when collecting data online is the decision of which survey management tool to use. Neupert and Bellingtier (2017) recently completed an online daily diary study with older adults using Amazon's Mechanical Turk (mTurk).

mTurk.

MTurk is online marketplace where “requesters” can post Human Intelligence Tasks (HITs), i.e., jobs, for “workers” to complete. It has become popular inside academia as a method for collecting survey data, especially for cross-sectional studies. Neupert and Bellingtier (2017) used mTurk to recruit older adult participants to participate in an online daily diary study, providing researchers with another way to recruit a nationally (or internationally) representative sample.

MTurk allows for qualifications to screen participants. There are a few “system” qualifications which have been premade by mTurk. These include “location” which screens individuals by the location they reported when they signed up for mTurk. Location can be

double-checked by asking individuals to self-report what country they lived in. Less than 1% reported countries other than the USA when the qualification of living in the USA was used (Neupert & Bellingtier, 2017).

HIT approval rates (i.e., the percentage of HITs the worker has completed that have been approved) and number of HITs approved can also be used to screen participants. These two qualifications could be used to weed out poor workers or inexperienced workers, but may require special considerations in an IRB application.

Until recently, all other qualifications had to be created by the researcher and individually assigned to participants. This would generally be done by creating a qualification HIT or using a Day 1 HIT for this purpose. However, subsequent to Neupert and Bellingtier's (2017) study, mTurk has introduced premium qualifications which allow researchers to pay (prices range from \$.05 to \$.65 per participant requested) to have workers pre-screened for additional traits (e.g., age, income, education). Currently, the premium qualifications allow researchers to pay to pre-screen workers in different age brackets at a cost of \$.50 per worker requested. However, the oldest age bracket available is "age 55 or older." Given the cost (e.g., for 200 50+ participants researchers would pay \$100) and the limited age brackets, creating a qualifying HIT and eschewing the premium qualification for a target sample of older adults may be optimal. A short survey that screens for birthdate and any other required characteristics could be created. One could pay a few cents for this type of qualification HIT and gain information beyond age.

Much has been written elsewhere about how much one should pay mTurk participants. Neupert and Bellingtier (2017) compensated participants \$1/day. One thing to be aware of when making this decision is Turkopticon.ucsd.edu. This is a website mTurk workers use to rate mTurk

requesters (individuals posting HITs). It is similar to ratemyprofessor.com with workers reporting on a requester's fairness, fastness, pay, and communication. Other workers look at these rating before deciding whether or not to do a HIT. If very low compensation is offered, the researchers are nonresponsive to messages from participants, or fail to pay them in a timely and fair manner, bad ratings or comments could result. This could then result in future participants being unwilling to "work" for the research team. Researchers can create a [turkopticon](http://turkopticon.com) account which allows them to view and/or respond to their ratings or comments. Although this site suffers from all the sampling problems of a voluntary internet survey, it is probably best for researchers to be aware of what is being posted about their lab or study. Appendix A provides a tutorial for researchers interested in recruiting participants for a daily diary study from mTurk with data collected via [Qualtrics](http://Qualtrics.com).

TurkPrime.

TurkPrime is designed for Amazon Mechanical Turk Requesters with an MTurk account, but also provides a service so that researchers can use the TurkPrime.com MTurk account from where studies can be conducted. TurkPrime offers Prime Panels with access to over 10 million participants. Data can be collected outside of Mechanical Turk with a larger and more diverse population. Researchers can also selectively recruit participants based on age, gender, race, education, income, and many other variables. An additional benefit of TurkPrime is that Prime Panel participants have had much less exposure to psychological studies and are more naïve to standard manipulations than participants from Mechanical Turk. TurkPrime is more expensive than Mechanical Turk, charging an additional 2 cents + 5% per complete HIT beyond Mechanical Turk, which may be a strong consideration for some researchers.

CrowdFlower.

CrowdFlower (CF; www.crowdflower.com) combines artificial intelligence with crowdsourced labor and allows users access to an online workforce. Given the integration with artificial intelligence, answers that are subjective in nature (as in psychological surveys) would bypass all of CF's quality control mechanisms. CF explicitly acknowledges the difficulty of demographic and psychological surveys because workers can easily share codes with other workers. CF participants provided a better response rate and were more diverse than mTurk participants, but CF participants failed more attention-check questions and did not reproduce known effects replicated on mTurk and other platforms (Peer, Brandimarte, Samat, & Acquisti, 2017).

Prolific Academic.

Prolific Academic (<https://www.prolific.ac/researchers/>) provides demographic screening and ensures high data quality so researchers can rapidly recruit target participants. Prolific Academic offers reliable, on-demand participants anytime for a survey or task and researchers can check new data before approving participant rewards. However, the stated current participant pool is over 25,000, which is substantially smaller than mTurk and TurkPrime. The service integrates with any software (e.g., Qualtrics, SurveyMonkey) and endorses the principle of ethical rewards, meaning that they ask researchers to reward participants with at least \$6.50/hour and then a 30% service charge is added to each participant payment. Thus, Prolific Academic is likely the most expensive option, but it may produce higher quality data than other platforms (Peer et al., 2017).

Regardless of the recruitment or survey management tool chosen, the ability to have a timestamp for the exact date and time of starting and completing a diary entry may be especially important for lifespan developmental questions. For example, older adults tend to prefer to respond in the morning (Sacco et al., 2016). A circadian preference for morningness is also associated with conscientiousness (Lipnevich, Credè, Hahn, Spinath, & Roberts, 2017). Because online diaries are well-suited for deploying timed cognitive tasks, age differences in time of day of completing diaries may be important for understanding potential age differences in daily cognitive performance (Cavallera, Boari, Giudici, & Ortolano, 2011). Time of day could be a valuable within-person covariate in daily diary studies of lifespan development.

In-person

Daily diary designs can also be conducted in person. Röcke, Li, and Smith (2009) had 37 participants (18 younger adults, 19 older adults) come to the lab for individual 1-hour test sessions between 9AM and 8PM for 45 consecutive weekdays (Monday-Friday). Time of day was self-selected by the participants and kept constant across the study period. Of the possible 1,665 daily occasions (37 participants X 45 days), 1,649 (99.04%) were obtained. Participants received 10 Euros per hour (the study was conducted in Germany) and a bonus of 200 Euros if they completed most of the daily sessions. A clear benefit of conducting an in-person daily diary study is the ability to standardize a laboratory-based data collection, ensuring a high degree of control over the timing and collection of daily assessments. Having 45 daily measurement occasions also provides reliable indicators of intraindividual variability along with within-person relationships.

Conducting 45 daily assessments in person has some drawbacks. Although Röcke et al. had many daily observations and sufficient power to detect within-person effects, they had a

relatively small sample of individuals which limited the possibility for between-person analyses. It is possible that the time demand of the study created a select sample who was willing to commit to the intense protocol (Iida, Shrout, Laurenceau, & Bolger, 2012) and therefore not representative of the larger population. An additional drawback would be participant and research staff burden. Each participant spent up to 45 hours in the lab, but also had to travel to and from the lab. Because daily testing was done individually, the research staff spent 1,649 hours (total number of daily sessions) in a one-on-one setting with participants.

Sliwinski, Smyth, Hofer, and Stawski (2006) conducted an in-person microlongitudinal study to examine changes in stress, health, and cognition over time. In contrast to Röcke et al. (2009) who had 37 participants come to the lab each weekday for 45 occasions, Sliwinski et al. conducted six testing sessions over a period of 8 to 14 days with 108 participants. The focus of within-person variability in cognitive performance on laboratory-based tasks made the in-person assessment a reasonable mode of data collection.

Although the in-person mode is perhaps the most burdensome for participants and research staff, it can be a good option for researchers with questions that require direct interactions or observations of participants. For example, those interested in day-to-day fluctuations in performance on a physical task such as balance, walking speed, etc. could find an in-person assessment extremely valuable. When choosing a diary format, researchers should take into account a range of considerations (Green et al., 2006). For example, studies requiring equally spaced reports are likely to benefit from features of telephone, IVR, online, and in-person methods that verify the time of completion. On the other hand, studies of special populations with members who are not familiar with electronic data collection devices or have

mobility issues that may prevent in-person assessments may find that paper-and-pencil methods produce better data (Green et al., 2006).

Methodological Considerations

During the planning stages for a new daily diary collection, researchers need to consider a variety of issues. Here we describe some of the most common issues involving how to determine 1) the number of days (duration); 2) the number of participants; and 3) power to detect effects. These issues are important because they are closely related to participant burden, and therefore selection effects (Iida et al., 2012), as well as the ability of the data to adequately address the research questions.

Duration of Study

The duration of a daily diary study should ideally be guided by a theory of how the phenomenon of interest changes (Bolger & Laurenceau, 2013; Collins, 2006). Phenomena that are slow moving or have little variability should be assessed less frequently and less densely than those that are faster moving or have high variability (Bolger & Laurenceau, 2013). As noted by Bolger and Laurenceau (2013), theories of social and behavioral phenomenon rarely specify the temporal course of associations between predictors and outcomes or the shape of an outcome's trajectory over a certain period of time. Gunthert and Wenzel (2012) noted that daily diary studies typically last 7 to 30 days (with some exceptions lasting longer, up to 15 months) but did not suggest an optimal length of assessment. The number of days and duration of a daily diary should be closely linked to the research questions and balanced with participant burden. It is important that the study be long enough to capture within-person variability in the constructs of interest, yet not so long that the participants become highly selected because they do not agree to participate in the first place or drop out midway through the protocol. The National Study of

Daily Experiences uses an 8-day paradigm and has found evidence for significant within-person variability in key constructs of interest such as affect, physical health, daily stressors, substance use, etc. The 8-day paradigm has also yielded significant within-person variation in anticipatory coping (Neupert et al., 2016), control beliefs, memory failures (Neupert, Almeida, Mroczek, & Spiro, 2006), and medication adherence (Neupert, Patterson, Davis, & Allaire, 2011). As others (e.g., Martin & Hofer, 2004) have noted, sampling time and study duration will influence analysis and interpretation of intraindividual variability and short-term change.

Number of Participants

Daily diary studies have a wide range of sample sizes, with some containing a few dozen participants (e.g., Röcke et al., 2009; Neupert et al., 2016) to some containing thousands (e.g., NSDE; Almeida et al., 2002). As the sample size increases the power to detect effects related to person-level differences also increases, but so can the burden on the resources of the researchers. To the extent that research staff are heavily involved in recruiting, collecting, and coding, teams will need to consider how many participants are feasible. Importantly, researchers will need to ascertain whether the research questions and constructs of interest have strong individual differences (necessitating more participants) or if the main focus is instead on within-person change processes (necessitating perhaps fewer participants but maybe more days). For example, when researchers are interested in studying changing process within a narrow age range of older adults, the potential influence of individual differences in age is less important than the daily fluctuations. This within-group focus acknowledges the heterogeneity of daily experiences of older adults, rather than using something like an extreme age groups design that focuses on between-group differences. Studies where the focus is simultaneously situated in between-person

differences and within-person processes will need to balance power and selection effect considerations.

Power

As Bolger, Stadler, and Laurenceau (2012) noted, until relatively recently researchers had very few resources to draw upon when making estimates of power for intensive longitudinal studies such as daily diaries. Conducting a power analysis for a daily diary study is considerably more challenging than for simpler designs because diary studies involve multiple sources of random variation (i.e., days and persons), requiring researchers to make assumptions about each in order to do the required calculations (Bolger et al.). When possible, it is very helpful to have some prior data available for making the assumptions.

Bolger et al. (2012) presented three options for conducting power analyses for intensive longitudinal (e.g., daily diary) designs: 1) working with the power formulae available in books of multilevel modeling and longitudinal designs (Fitzmaurice, Laird, & Ware, 2004; Gelman & Hill, 2007; Hox, 2010; Moerbeek, Van Breukelen, & Berger, 2008; Snijders & Bosker, 1999); 2) using specialized software designed for power analyses for multilevel and longitudinal models, for example, the freely available PinT (Bosker, Snijders, & Guldemon, 2007), Optimal Design (Raudenbush, Spybrook, Liu, & Congdon, 2006), and RMASS2 (Hedeker, Gibbons, & Waternaux, 1999); and 3) using simulation methods in general purpose programming software, such as Mplus (Muthén & Muthén, 1998-2007), SAS (Littell, Milliken, Stroup, Wolfinger, & Schabenberger, 2006; SAS Institute, Inc., 2010), R (R Development Core Team, 2011), or MATLAB (MathWorks, Inc., 2011).

Bolger et al. (2012) advocated for simulation methods because they are able to determine whether it is better to increase the number of time points per person or the number of persons per

time point in order to detect a within-person effect. Results from power simulations involve combinations of increased participants and time points. Although increasing upper-level units (i.e., persons) can often result in more power than increasing the number of lower-level units (i.e., days), the cost of increasing participants is substantially greater than the cost of increasing time points (Bolger et al.; Snijders & Bosker, 1999).

Selection Effects

Given the increased burden in participating in a daily diary study relative to a cross-sectional study, it is important to consider who is likely to agree to participate and persist in a daily diary study. Ram et al. (2014) combined age heterogeneity, longitudinal panel, daily diary, and experience sampling protocols and collected data with smartphone and web-based technologies to obtain intensive longitudinal data from 150 persons age 18 to 89 years as they completed three 21-day measurement bursts, spanning 8,557 days and 64,112 social interactions. These data are clearly rich in their comprehensive assessments of various timescales, but the richness can come at a cost in terms of sample selection.

There could be special considerations when recruiting older adults for a daily diary study. Older adults tend to prefer telephone diaries to electronic diaries and also prefer to respond in the morning about the previous day (as opposed to before bedtime) (Sacco et al., 2016). One potential consequence of this preference for morning reporting is that older adults may have a longer recollection period than younger adults, which may introduce a systematic retrospective recall bias that disadvantages older adults. Gathering the time of day of completion and then testing for age differences or differential effects by age is an important consideration for daily diary researchers. Harber, Zimbardo, and Boyd (2003) looked at compliance in a daily diary study of undergraduates. Those with a future time perspective were more compliant throughout a

daily diary study, but not more likely to sign up or to dropout. This may be due to the nature of students' requirements to do research. In older adults, those who sign up for daily diary research and who stick with it may also be more future oriented.

It is important to balance the intensity of the study with the potential burden on participants. Reducing barriers to initial participation agreement, such as the ability to collect data in naturalistic environments with relatively unobtrusive tools should decrease selection at baseline. Once participants have agreed to participate, attempts to reduce attrition become important ways to reduce selection effects. Keeping the duration of the daily assessment brief and the number of total days required to a minimum are important to consider. To state this idea another way, it is likely a very different participant who agrees to travel to the lab for repeated assessments compared to someone who agrees to participate in online assessments at home (or from a mobile device). Further, the kind of person who agrees to a study over 45 assessments is likely to be meaningfully different from someone who agrees to a study over 8 assessments. These differences have important implications for external validity. There will be research questions that will necessitate intense, long-term daily diary studies to capture processes that may unfold over relatively long periods. Although these studies may benefit from an increased precision to identify change processes, they will also likely have decreased person-level generalizability. Daily diary researchers should strongly consider both and make design decisions that best match their research questions while minimizing participant selection effects.

Future Directions

Biomarkers

Daily diary designs are poised to be at the forefront of many exciting scientific discoveries that cross many disciplines and impact individuals' lives. One way in which daily

diaries are being used in interdisciplinary research is through combining participant self-report with biomarkers. Self-report daily diaries have been integrated with biomarkers such as blood glucose (Berg et al., 2013) and cortisol. Cortisol, a product of the HPA axis, is the best biomarker for understanding the effects of negative psychosocial stressors on health and disease (Kemeny, 2003) and age-related health declines (Dmitrieva, Almeida, Dmitrieva, Loken, & Pieper, 2013). Cortisol follows the same diurnal pattern across age groups (Van Cauter, Leproult, & Kupfer, 1996), sharply increasing to a peak about 30 min after waking (the cortisol awakening response [CAR]; Kirschbaum & Hellhammer, 2000) and steadily declining until reaching a nadir in the late evening (the diurnal cortisol slope [DCS]; Lovallo & Thomas, 2000). An important innovation of the second wave of the NSDE was the collection of psychophysiological data from a nationally-representative sample (<http://www.icpsr.umich.edu/icpsrweb/NACDA/studies/26841>). 1,736 participants collected saliva via salivettes four times per day across 4 days, allowing for the calculation of the cortisol awakening response, diurnal cortisol slope, and total cortisol output. The research team created detailed instructions with pictures and color-coded tubes and record sheets so that participants could properly collect, record, and store their own saliva. Saliva collection boxes were mailed to each participant, and a subsample of participants received a “Smart Box” to store their salivettes. These boxes contained a computer chip to record the time participants opened and closed the box. When all 16 tubes were ready to be returned to the research team, participants used a pre-addressed, paid courier package for the return mailing. The enclosed salivettes were shipped to the MIDUS Biological Core at the University of Wisconsin, where there were stored in an ultracold freezer at -60°C before being assayed.

Moving beyond negative events where cortisol is typically the best biomarker, Sin, Graham-Engeland, and Almeida (2015) examined the role of daily positive events with three inflammatory markers (interleukin-6 [IL-6], C-reactive protein [CRP], and fibrinogen) in the second wave of the NSDE. 969 adults aged 35-86 reported on positive experiences that occurred over the past 24 hours and then provided blood samples that were obtained at a separate clinic visit. On average, participants experienced positive events on 73% of the days, and more daily positive events were associated with lower IL-6 and CRP. The effects were especially pronounced for participants in the lowest quartile of positive event frequency, suggesting that lack of positivity in daily life may be particularly consequential for inflammation. In addition, people who experienced a greater loss in positive affect on days when they encountered stressors had elevated IL-6, compared to those who were better able to maintain positive affect when stressors occurred (Sin, Graham-Engeland, Ong, & Almeida, 2015). Sin and Almeida (in press) proposed a theoretical model that positive affect and positive events in everyday life can promote physical health through favorable physiological functioning, better health behaviors, and by mitigating the effects of stress on health.

In addition to positive events, daily positive affect can mitigate the within-person association between daily negative affect and systolic blood pressure (Ong & Allaire, 2005) and buffer against the effects of daily stress on depressive symptoms in recently bereaved widows (Ong, Bergeman, & Bisconti, 2004). Daily positive emotions also attenuate negative affective reactivity to stressors and predict accelerated emotional recovery from prior-day stressors (Ong, Bergeman, Bisconti, & Wallace, 2006). Daily diary researchers are also using reports of daily negative and positive affect to examine individual differences in emotional complexity as the co-

occurrence of negative and positive affect (e.g., Hay & Diehl, 2011; Ong & Bergeman, 2004; Ramsey, Neupert, Mroczek, & Spiro, 2016; Scott, Sliwinski, Mogle, & Almeida, 2014).

Ambulatory Assessments

An additional way that daily diaries can contribute to future interdisciplinary and transdisciplinary work is with the integration of ambulatory assessments. Capturing within-person variability of developmental processes as they occur within context-specific influences of individuals' natural environments is a powerful research tool (Hoppman & Riediger, 2009). As Brose and Ebner-Priemer (2015) note, ambulatory assessments can be utilized to detect person-specific critical periods and for designing immediate person-specific interventions. Hoppman and Riediger (2009) suggest that ambulatory assessments could help balance the internal and external validity of findings because of the combination of subjective and objective indicators being collected in naturalistic settings. One example of ambulatory assessment within the context of a daily design is the Effects of Stress on Cognitive Aging, Physiology and Emotion (ESCAPE; Scott et al., 2015) project. Ambulatory cognition was assessed repeatedly via smartphones in naturalistic settings as participants went about their daily activities. Scott et al. (2015) suggested that ambulatory assessments provided a more ecologically valid characterization of participants' cognitive functioning that will complement and extend traditional laboratory-based assessments. Interested readers should consult Hoppman and Riediger's (2009) in-depth review of four research themes in developmentally relevant ambulatory assessment studies where electronic devices are used as ambulatory assessment instruments: a) affective-motivational development; b) social contexts of development; c) age-related challenges and everyday functioning; and d) cognitive development.

Sampling the Future

The temporal space of daily diary research is expanding. Daily diaries have primarily been used to report on past events and experiences, but they can also ask about future expectations and intentions. For example, in the ACED study (Neupert et al., 2016) 43 older adult participants reported for eight consecutive days on the likelihood that they would experience a stressor in one of seven domains and then to report the coping that they were doing at that time to possibly prepare for the upcoming stressor. This coping *before* the stressor, anticipatory coping, involves efforts to prepare for the stressful consequence of an upcoming event that is likely to happen (Folkman & Lazarus, 1985). Although anticipatory coping is posited to be situation-specific and associated with reduced response (or reactivity) to a stressor (Aspinwall & Taylor, 1997; Schwarzer & Knoll, 2003), Neupert et al. were the first to use a daily diary design to examine anticipatory coping from a within-person perspective within changing contexts (i.e., various stressor domains).

The results suggested that daily anticipatory coping is dynamic; people are not using the same coping tool in all circumstances or at all times. Daily anticipatory coping was also linked to well-being outcomes within the context of interpersonal stressors, but the directions of the effects differed once again highlighting the contextual and dynamic process. In this sample of older adults, a coping form previously considered maladaptive when examined as a between-person individual difference (Stagnant Deliberation) was adaptive; on days when people reported increases in thought patterns that they believed were not making any progress, they reported fewer memory failures in response to arguments the following day.

Stagnant Deliberation has been positively associated with an active cognitive style (Feldman & Hayes, 2005), which Fresco, Frankel, Mennin, Turk, and Heimberg (2002) suggested may help individuals to make sense of and cope with their experiences more

effectively. It is possible that the specific items for Stagnant Deliberation (e.g., Even though I really concentrate on it, I don't seem to get any answers; I think about how to solve the problem, but the thoughts just spin around in my head) represent the cognitive benefits of deliberating on a possible solution, even if the participant appraises that the deliberation is not helpful at the time. This is especially important when one considers that anticipatory coping was captured in real time; that is, anticipatory coping was reported on one day, and the argument in question did not occur until the following day (if it happened at all). Although speculative, it is possible that the participant did not think the deliberation was helpful at the time, but in fact the deliberation could have been helpful if the argument happened the following day.

The use of daily diaries to collect data about future experiences, expectations, and intentions is an exciting future prospect. One of the strengths of daily diaries is to connect temporal associations between antecedents, correlates, and consequences, but this has primarily been done with retrospective accounts. The consideration of future-oriented questions within daily diaries to capture the thoughts, behaviors, and actions that may occur earlier within the chain of events and outcomes is encouraged

Publicly Available Datasets

Because daily diary datasets take a significant amount of expertise and resources to execute, it can be valuable to have access to existing daily diary data when one is interested in working with these kinds of data. ICPSR is an international consortium of more than 750 academic institutions and research organizations that maintains a data archive of more than 250,000 files of research in the social and behavioral sciences. ICPSR collaborates with a number of funders, including U.S. statistical agencies and foundations, to create thematic data collections. The vast majority of ICPSR data holdings are public-use files with no access

restrictions. Researchers interested in accessing archived daily diary data should look at the list of [datasets currently available](#).

Analyzing Daily Diary Data

Daily diary data are frequently analyzed using multilevel models (Raudenbush & Bryk, 2002) where the daily data (Level 1) are nested within the person-level data (Level 2). To examine temporal effects, lagged models can be employed to track the effects of a given independent variable on one day to changes in a given dependent variable the next day (Neupert et al., 2017). These models have typically been analyzed with software packages such as SAS, SPSS, or HLM (see Bolger & Laurenceau, 2013 for step-by-step examples with SAS and SPSS). A future direction that we see emerging within the analytic realm of daily diary research is the use of R, an open source software package that is freely available. As noted on the website, “R is a language and environment for statistical computing and graphics. R provides a wide variety of statistical (e.g., [multilevel models]) and graphical techniques, and is highly extensible. One of R’s strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed. Great care has been taken over the defaults for the minor design choices in graphics, but the user retains full control.”

There are several benefits in a shift toward R for analyzing daily diary data. First, because it is free, anyone could replicate the models for a given study which is an important consideration within the context of current concerns regarding a “replication crisis” (Maxwell, Lau, & Howard, 2015). Specifically, the cost of many software programs restricts the field of potential researchers to those with access the expensive tools. For example, a researcher working with limited resources in a developing country would not be able to access SPSS to conduct a replication. Second, R is known for its extensive graphing capabilities that would give

researchers new ways to visualize complex daily diary data. Heatmaps are tables of numbers, where the numbers are substituted with colored cells. It gives a visual impact to the matrix, which makes it easier to understand large amounts of data. Last, R has the capability to generate visuals that show data moving dynamically, allowing researchers to move beyond static, 2-dimension figures.

Diaries as Interventions

Participating in a daily diary study may act as a sort of intervention that influences ratings of experiences, often called reactivity (Bolger & Laurenceau, 2013). Barta, Tennen, and Litt (2012) noted that participating in a daily diary could cause reactivity effects, whereby participants experience changes as a function of being part of the study. For example, participants who are repeatedly asked to report on daily stressors may develop a heightened sensitivity to stressful events in their environment, which would result in reporting an increase in stressors over the course of the study. Conversely, it is also possible that by repeatedly asking participants to report on stressors in their environment, participants' potential increased awareness may be associated with more effective avoidance strategies that would result in decreases in stressor exposure reports.

There are a few studies reporting positive effects on participants' lives as a function of daily diary study participation. In an explicit effort to increase gratitude and well-being in non-clinically depressed older adults, Killen and Macaskill (2015) used a "Three good things in life" gratitude intervention on hedonic and eudemonic well-being and perceived stress levels over the course of two weeks. Every evening for 14 consecutive days, participants recorded briefly in their diaries three events occurring that day that seemed positive to them, and why they viewed them positively. The intervention produced significant differences in eudemonic well-being as

measured by flourishing from baseline to Day 15 that was maintained at Day 45. There was also a significant decrease in perceived stress from Day 1 to Day 15. Notably, there was no difference between online and paper delivery of the intervention, but the older adult participants in this study reported that they preferred the online delivery. Gratitude diaries seem to be a promising and cost-effective method of producing beneficial improvements in well-being for older adults.

Zautra et al. (2012) conducted a randomized controlled trial of a brief, daily intervention targeting either personal control or mindful awareness in a community sample with symptoms of depression. The interventions were delivered in prerecorded automated messages via phone each morning. Each evening, participants completed an online daily diary that included the outcome measures. Results revealed significantly greater improvements in emotional health and self-reported physical health for the treatment conditions across the 31-day trial in comparison to controls. We agree with Zautra et al. (2012) that these findings should encourage further development and testing of innovative and accessible intervention methods to address mental health problems of older adults in the community.

Neupert has also seen daily diaries function as incidental interventions in her lab. For example, participants in the MACE project reported significantly fewer daily mindfulness errors (e.g., “I found it difficult to stay focused on what was happening in the present”) as the study progressed ($\gamma_{10} = -0.14$, $t = -5.78$, $p < .0001$), and study day explained 8% of the within-person variance in daily mindfulness errors. Across each of Neupert’s 8-day daily diary studies, there is consistently a significant decrease in negative affect over the course of the study. Participants have also acknowledged the potential role of daily diaries as a sort of intervention. One older adult participant from the MACE project emailed after the study was completed and shared, “Thank you for allowing my participation in this survey. A side affect [sic] of the survey for me

was that I did become more attentive and aware of how much less stress I have in my life since I retired I realized that I have become more accepting of the foibles of my fellow family, neighbors, and friends.... Thank you for allowing me to learn this.”

Conclusion

Examining “life as it is lived” (Allport, 1942) through daily diary designs continues to be a promising avenue to expand theoretical and empirical understanding of lifespan development. Researchers can conduct studies for varying lengths of time via paper and pencil, telephone, IVR, online, or in person, matched to appropriate research questions and samples. Promising future directions include expanding the temporal space of daily diaries to include thoughts and behaviors before, during, and after events and states and considering ways in which daily diaries could be used as purposeful interventions to improve the daily lives of adults across the lifespan.

References

- Adam, E.K., & Kumari M. (2009). Assessing salivary cortisol in large-scale, epidemiological research. *Psychoneuroendocrinology*, *34*, 1423–1436.
doi:10.1016/j.psyneuen.2009.06.011
- Allport, G.W. (1942). *The Use of Personal Documents in Psychological Science*. Social Science Research Council. New York: NY.
- Almeida, D. M. (2005). Resilience and vulnerability to daily stressors assessed via diary methods. *Current Directions in Psychological Science*, *14*(2), 62-68. doi: 10.1111/j.0963-7214.2005.00336.x
- Almeida, D.M., & Kessler, R.C. (1998). Everyday stressors and gender differences in daily distress. *Journal of Personality and Social Psychology*, *75*, 670-680.
- Almeida, D. M., Wethington, E., & Kessler, R. C. (2002). The Daily Inventory of Stressful Events (DISE): An interview based approach for measuring daily stressors. *Assessment*, *9*, 41–55. doi: 10.1177/1073191102091006
- Aspinwall L. G., Taylor S. E. (1997). A stitch in time: self-regulation and proactive coping. *Psychological Bulletin*, *121*, 417–436. doi: 10.1037/0033-2909.121.3.417
- Barta, W. D., Tennen, H., & Litt, M. D. (2012). Measurement reactivity in diary research. In M. R. Mehl, T. S. Conner, M. R. Mehl, T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 108-123). New York, NY, US: Guilford Press.
- Berg, C.A., Butner, J.E., Butler, J.M., King, P.S., Hughes, A.E., & Wiebe, D.J. (2013). Parental persuasive strategies in the face of daily problems in adolescent Type 1 diabetes management. *Health Psychology*, *32*, 719-728. doi: 10.1037/a0029427
- Birren, J.E., & Bengtson, V.L. (1988). *Emergent theories of aging*. New York, NY: Springer.

- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, *54*, 579-616.
- Bolger, N., DeLongis, A., Kessler, R.C., & Schilling, E. (1989). Effects of daily stress on negative mood. *Journal of Personality and Social Psychology*, *57*, 808-818.
- Bolger, N., & Laurenceau, J.-P. (2013). *Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research*. New York, NY: 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, NY: Guilford Press.
- Bosker, R.J., Snijders, T.A.B., & Guldemon, H. (2007). *PinT (Power in two-level designs): Estimating standard errors of regression coefficients in hierarchical linear models for power calculations* (Version 2.12). Groningen, The Netherlands: Author.
- Brose, A., & Ebner-Priemer, U. (2015). Ambulatory assessment in the research on aging: Contemporary and future applications. *Gerontology*, *61*(4), 372-380.
doi:<http://dx.doi.org/prox.lib.ncsu.edu/10.1159/000371707>
- Cavallera, G.M., Boari, G., Giudici, S., & Ortolano, A. (2011). Cognitive parameters and morning and evening types: Two decades of research (1990-2009). *Perceptual and Motor Skills*, *112*, 649-665. doi: 10.2466/02.09.17.PMS.112.2.649-665.
- Charles, S. T. (2010). Strength and vulnerability integration: A model of emotional well-being across adulthood. *Psychological Bulletin*, *136*, 1068-1091. doi:10.1037/a0021232
- Chida, Y., & Steptoe, A. (2009). Cortisol awakening response and psychosocial factors: a systematic review and meta-analysis. *Biological Psychology*, *80*(3), 265-278.
<https://doi.org/10.1016/j.biopsycho.2008.10.004>

- Conner, T. S., & Lehman, B. J. (2012). Getting started: Launching a study in daily life. In M. R. Mehl, T. S. Conner, M. R. Mehl, T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 89-107). New York, NY, US: Guilford Press.
- DeSantis, A. S., DiezRoux, A. V., Hajat, A., Aiello, A. E., Golden, S. H., Jenny, N. S., ... Shea, S. (2012). Associations of salivary cortisol levels with inflammatory markers: The Multi-Ethnic Study of Atherosclerosis. *Psychoneuroendocrinology*, *37*(7), 1009–1018.
<https://doi.org/10.1016/j.psyneuen.2011.11.009>
- Diehl, M., Hooker, K., & Sliwinski, M. J. (2015). A brief historical overview of intraindividual variability research across the life span. In M. Diehl, K. Hooker, M. J. Sliwinski, M. Diehl, K. Hooker, M. J. Sliwinski (Eds.), *Handbook of intraindividual variability across the life span* (pp. 3-15). New York, NY, US: Routledge/Taylor & Francis Group.
- Dillman, D.A. (1991). The design and administration of mail surveys. *Annual Review of Sociology*, *17*, 225-249.
- Dmitrieva, N. O., Almeida, D. M., Dmitrieva, J., Loken, E., & Pieper, C. F. (2013). A day-centered approach to modeling cortisol: Diurnal cortisol profiles and their associations among U.S. Adults. *Psychoneuroendocrinology*, *38*(10), 2354-2365.
[doi:10.1016/j.psyneuen.2013.05.003](https://doi.org/10.1016/j.psyneuen.2013.05.003)
- Feldman, G. C. & Hayes, A.M. (2005). Preparing for problems: A measure of mental anticipatory processes. *Journal of Research in Personality*, *39*, 487-516. doi:
[10.1016/j.jrp.2004.05.005](https://doi.org/10.1016/j.jrp.2004.05.005)
- Fitzmaurice, G.M., Laird, N.M., & Ware, J.H. (2004). *Applied longitudinal analysis*. Hoboken, NJ: Wiley.

- Flood, E. M., Zazzali, J. L., & Devlen, J. (2013). Demonstrating measurement equivalence of the electronic and paper formats of the urticaria patient daily diary in patients with chronic idiopathic urticaria. *The Patient: Patient-Centered Outcomes Research*, 6(3), 225-231. doi:10.1007/s40271-013-0021-4
- Folkman, S., & Lazarus, R. S. (1985). If it changes it must be a process: A study of emotion and coping during three stages of a college examination. *Journal of Personality and Social Psychology*, 48, 150-170. doi: 10.1037/002f2-3514.48.1.150
- Fresco, D. M., Frankel, A., Mennin, D. S., Turk, C. L., & Heimberg, R. G. (2002). Distinct and overlapping features of rumination and worry: The relationship of cognitive production to negative affective states. *Cognitive Therapy and Research*, 26, 179-188. doi: 10.1023/A:1014517718949
- Freund, A.M. (2015). Getting at developmental processes through experiments. *Research in Human Development*, 12, 261-267. doi: 10.1080/15427609.2015.1068033
- Gable, S.L., Reis, H.T., & Elliot, A.J. (2000). Behavioral activation and inhibition in everyday life. *Journal of Personality and Social Psychology*, 78, 1135-1149.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.
- Green, A.S., Rafaeli, E., Bolger, N., Shrout, P.E., & Reis, H.T. (2006). Paper or plastic? Data equivalence in paper and electronic diaries. *Psychological Methods*, 11, 87-105. doi: 10.1037/1082-989X.11.1.87
- Gunthert, K. C., & Wenze, S. J. (2012). Daily diary methods. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 144-159). New York, NY: Guilford Press.

- Harber, K. D., Zimbardo, P. G., & Boyd, J. N. (2003). Participant self-selection biases as a function of individual differences in time perspective. *Basic and Applied Social Psychology, 25*(3), 255-264. doi:10.1207/S15324834BASP2503_08
- Hay, E.L., & Diehl, M. (2011). Emotion complexity and emotion regulation across adulthood. *European Journal of Ageing, 8*, 157-168. doi: 101007/s10433-011-0191-7
- Hedeker, D., Gibbons, R.D., & Waternaux, C. (1999). Sample size estimation for longitudinal designs with attrition: Comparing time-related contrasts between two groups. *Journal of Educational and Behavioral Statistics, 24*, 70-93.
- Heim, C., Ehler, U., & Hellhammer, D. H. (2000). The potential role of hypocortisolism in the pathophysiology of stress-related bodily disorders. *Psychoneuroendocrinology, 25*(1), 1–35. [https://doi.org/10.1016/S0306-4530\(99\)00035-9](https://doi.org/10.1016/S0306-4530(99)00035-9)
- Hoppmann, C. A., & Riediger, M. (2009). Ambulatory assessment in lifespan psychology: An overview of current status and new trends. *European Psychologist, 14*(2), 98-108. doi:10.1027/1016-9040.14.2.98
- Hox, J.J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York, NY: Routledge.
- Iida, M., Shrout, P. E., Laurenceau, J., & Bolger, N. (2012). Using diary methods in psychological research. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol 1: Foundations, planning, measures, and psychometrics* (pp. 277-305). Washington, DC, US: American Psychological Association. doi:10.1037/13619-016
- Karlamangla, A. S., Friedman, E. M., Seeman, T. E., Stawski, R. S., & Almeida, D. M. (2013). Daytime trajectories of cortisol: Demographic and socioeconomic differences—Findings

- from the national study of daily experiences. *Psychoneuroendocrinology*, 38(11), 2585-2597. doi:10.1016/j.psyneuen.2013.06.010
- Katzman, R., Brown, T., Fuld, P., Peck, A., Schechter, R., & Schimmel, H. (1983). Validation of a Short Orientation-Memory-Concentration Test of Cognitive Impairment. *American Journal of Psychiatry*, 140, 734–739.
- Kemeny, M. E. (2003). The psychobiology of stress. *Current Directions in Psychological Science*, 12(4), 124-129. doi:10.1111/1467-8721.01246
- Kessler, R.C., Mroczek, D.K., & Belli, R.F. (1999). Retrospective adult assessment of childhood psychopathology. In D. Shaffer, C.P. Lucas, & J.E. Richters (Eds.), *Diagnostic assessment in child and adolescent psychopathology* (pp. 256-284). New York, NY: Guilford Press.
- Killen, A., & Macaskill, A. (2015). Using a gratitude intervention to enhance well-being in older adults. *Journal of Happiness Studies*, 16(4), 947-964. doi:10.1007/s10902-014-9542-3
- Kirschbaum, C., & Hellhammer, D. H. (2000). Salivary cortisol. In G. Fink (Ed.), *Encyclopedia of stress* (Vol. 3, pp. 379–383). New York: Academic Press.
- Larson, R.W., & Almeida, D.M. (1999). Emotional transmission in the daily lives of families: A new paradigm for studying family process. *Journal of Marriage and the Family*, 61, 5-20.
- Lewinsohn, P.M., & Talkington, J. (1979). Studies on the measurement of unpleasant events and relations with depression. *Applied Psychological Measurement*, 31, 83-101.
- Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., & Schabenberger, O. (2006). *SAS for mixed models* (2nd ed.). Cary, NC: SAS Institute.
- Lipnevich, A.A., Credè, M., Hahn, E., Spinath, F.M., & Roberts, R.D. (2017). How distinctive are morningness and eveningness from the big five factors of personality? A meta-

- analytic investigation. *Journal of Personality and Social Psychology*, *112*, 491-509. doi: 10.1037/pspp0000099
- Lovallo, W. R., & Thomas, T. L. (2000). Stress hormones in psychophysiological research: Emotional, behavioral, and cognitive implications. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 342–367). Cambridge, England: Cambridge University Press.
- Martin, M., & Hofer, S.M. (2004). Intraindividual variability, change, and aging: Conceptual and analytical issues. *Gerontology*, *50*, 7-11.
- MathWorks, Inc. (2008). *Matlab* (Version 2008b). Sherborn, MA: Author.
- Maxwell, S.E., Lau, M.Y., & Howard, G.S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *American Psychologist*, *70*, 487-498. doi: 10.1037/a0039400
- Miller, G. E., Chen, E., & Zhou, E. S. (2007). If it goes up, must it come down? Chronic stress and the hypothalamic-pituitary-adrenocortical axis in humans. *Psychological Bulletin*, *133*(1), 25–45. <https://doi.org/10.1037/0033-2909.133.1.25>
- Moerbeek, M., Van Breukelen, G.J.P., & Berger, M.P.F. (2008). Optimal designs for multilevel studies. In J. de Leeuw & E. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 177-205). New York: Springer.
- Muthén, L.K., & Muthén, B.O. (1998-2007). *Mplus user's guide* (5th ed.). Los Angeles, Author.
- Neupert, S. D., Almeida, D. M., Mroczek, D. K., & Spiro, A., III. (2006a). Daily stressors and memory failures: Findings from the VA Normative Aging Study. *Psychology and Aging*, *21*, 424 – 429. doi: 10.1037/0882-7974.21.2.424

- Neupert, S. D., Almeida, D. M., Mroczek, D. K., & Spiro, A. III (2006b). The effects of the Columbia shuttle disaster on the daily lives of older adults: Findings from the VA Normative Aging Study. *Aging & Mental Health, 10*, 272-281. doi: 10.1080/13607860500409682
- Neupert, S.D., & Bellingtier, J.A. (2017). Aging attitudes and daily awareness of age-related change interact to predict negative affect. *The Gerontologist*. Advance online publication. doi: 10.1093/geront/gnx055
- Neupert, S.D., Desmarais, S.L., Gray, J.S., Cohn, A., Doherty, S., & Knight, K. (2017). Daily stressors as antecedents, correlates, and consequences of alcohol and drug use and cravings in community-based offenders. *Psychology of Addictive Behaviors*. Advance online publication. doi: 10.1037/adb0000276
- Neupert, S. D., Ennis, G. E., Ramsey, J. L., & Gall, A. A. (2016). Solving tomorrow's problems today? Daily anticipatory coping and reactivity to daily stressors. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 71*(4), 650–660. <https://doi.org/10.1093/geronb/gbv003>
- Ong, A. D., & Allaire, J. C. (2005). Cardiovascular Intraindividual Variability in Later Life: The Influence of Social Connectedness and Positive Emotions. *Psychology and Aging, 20*(3), 476–485. <https://doi.org/10.1037/0882-7974.20.3.476>
- Ong, A.D., & Bergeman, C.S. (2004). The complexity of emotions in later life. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 59*, P117-P122. doi: 10.1093/geronb/59.3.P117

- Ong, A.D., Bergeman, C.S., & Bisconti, T.L. (2005). Unique effects of daily perceived control on anxiety symptomatology during conjugal bereavement. *Personality and Individual Differences, 38*, 1057-1067.
- Ong, A. D., Bergeman, C. S., Bisconti, T. L., & Wallace, K. A. (2006). Psychological resilience, positive emotions, and successful adaptation to stress in later life. *Journal of Personality and Social Psychology, 91*(4), 730–749. <https://doi.org/10.1037/0022-3514.91.4.730>
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*. Online Advance Access. doi: 10.1016/j.jesp.2017.01.006
- Piazza, J. R., Almeida, D. M., Dmitrieva, N. O., & Klein, L. C. (2010). Frontiers in the use of biomarkers of health in research on stress and aging. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 65B*(5), 513-525. doi:10.1037/a0022262
- R Development Core Team. (2011). *R: A language and environment for statistical computing* (Version 2.7.2). Vienna: R Foundation for Statistical Computing.
- Ram, N., Conroy, D. E., Pincus, A. L., Lorek, A., Rebar, A., Roche, M. J., & ... Gerstorf, D. (2014). Examining the interplay of processes across multiple time-scales: Illustration with the Intraindividual Study of Affect, Health, and Interpersonal Behavior (iSAHIB). *Research in Human Development, 11*(2), 142-160.
doi:10.1080/15427609.2014.906739
- Ramsey, J.L., Neupert, S.D., Mroczek, D.K., & Spiro, A. III. (2016). The effects of daily co-occurrence of affect on older adults' reactivity to health stressors. *Psychology and Health, 31*, 364-378. doi: 10.1080/08870446.2015.1111368

- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.
- Raudenbush, S.W., Spybrook, J., Liu, X.-F., & Congdon, R. (2006). *Optimal Design* (Version 1.77). Ann Arbor, MI: HLM Software. Retrieved from sitemaker.umich.edu/group-based/files/od-manual-20080312-v176.pdf.
- Reis, H.T., & Gable, S.L. (2000). Event sampling and other methods for studying everyday experience. In H.T. Reis & C.M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (pp. 190-222). New York: Cambridge University Press.
- Röcke, C., Li, S.-C., & Smith, J. (2009). Intraindividual variability in positive and negative affect over 45 days: Do older adults fluctuate less than young adults? *Psychology and Aging, 24*, 863-878. doi: 10.1037/a0016276
- Sacco, P., Smith, C. A., Harrington, D., Svoboda, D. V., & Resnick, B. (2016). Feasibility and utility of experience sampling to assess alcohol consumption among older adults. *Journal of Applied Gerontology, 35*(1), 106-120. doi:10.1177/0733464813519009
- SAS Institute, Inc. (2010). SAS 9.1.3. Cary, NC: Author.
- Schilling, O.K. & Diehl, M. (2014). Reactivity to stressor pile-up in adulthood: Effects on daily negative and positive affect. *Psychology and Aging, 29*, 72-83. doi: 10.1037/a0035500
- Schwarzer, R., & Knoll, N. (2003). Positive coping: Mastering demands and searching for meaning. In S. J. Lopez & C. R. Snyder, *Positive Psychological Assessment: A Handbook of Models and Measures* (pp. 393-409). Washington, DC: American Psychological Association. doi: 10.1037/10612-025
- Scott, S.B., Graham-Engeland, J.E., Engeland, C.G., Smyth, J.M., Almeida, D.M., Katz, M.J., Lipton, R.B., Mogle, J.A., Munoz, E., Ram, N., & Sliwinski, M.J. (2015). The Effects of

- Stress on Cognitive Aging, Physiology and Emotion (ESCAPE) Project. *BMC Psychiatry*, *15*, 146. doi: 10.1186/s12888-015-0497-7
- Scott, S.B., Sliwinski, M.J., Mogle, J.A., & Almeida, D.M. (2014). Age, stress, and emotional complexity: Results from two studies of daily experiences. *Psychology and Aging*, *29*, 577-587. doi: 10.1037/a0037282
- Shiffman, S., & Stone, A.A. (1998). Introduction to the special section: Ecological momentary assessment in health psychology. *Health Psychology*, *17*, 3-5.
- Sin, N. L., & Almeida, D. M. (in press). Daily positive experiences and health: Biobehavioral pathways and resilience to daily stress. In C. D. Ryff & R. F. Krueger (Eds.), *Oxford Handbook of Integrative Health Science*. New York: Oxford University Press.
- Sin, N. L., Graham-Engeland, J. E., & Almeida, D. M. (2015). Daily positive events and inflammation: Findings from the National Study of Daily Experiences. *Brain, Behavior, and Immunity*, *43*, 130–138. <https://doi.org/10.1016/j.bbi.2014.07.015>
- Sin, N. L., Graham-Engeland, J. E., Ong, A. D., & Almeida, D. M. (2015). Affective reactivity to daily stressors is associated with elevated inflammation. *Health Psychology*, *34*(12), 1154–1165. <https://doi.org/10.1037/hea0000240>
- Sliwinski, M.J., Smyth, J.M., Hofer, S.M., & Stawski, R.S. (2006). Intraindividual coupling of daily stress and cognition. *Psychology and Aging*, *21*, 545-557. doi: 10.1037/0882-7974.21.3.545
- Snijders, T.A.B., & Bosker, R.J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.
- Spiro, A. III, & Bossé, R. (2001). The Normative Aging Study. In G. Maddox (Ed.), *Encyclopedia of aging* (3rd ed.). New York: Springer Press.

- Stawski, R.S., Cichy, K.E., Piazza, J.R., & Almeida, D.M. (2013). Associations among daily stressors and salivary cortisol: Findings from the National Study of Daily Experiences. *Psychoneuroendocrinology*, *38*, 2654-2665. doi:10.1016/j.psyneuen.2013.06.023
- Stone, A.A., Reed, B.R., & Neale, J.M. (1987). Changes in daily event frequency precede episodes of physical symptoms. *Journal of Human Stress*, *13*, 70-74.
- Stone, A.A., Kessler, R.C., & Haythornthwaite, J.A. (1991). Measuring daily events and experiences: Methodological considerations. *Journal of Personality*, *59*, 575-607.
- Stone, A.A., Shiffman, S., Schwartz, J.E., Broderick, J.E., & Hufford, M.R. (2002). Patient non-compliance with paper diaries. *British Medical Journal*, *324*, 1193-1194.
- Tennen, H., Suls, J., & Affleck, G. (1991). Personality and daily experience: The promise and the challenge. *Journal of Personality*, *59*, 313-337. doi: 10.1111/j.1467-6494.1991.tb02387.x
- Van Cauter, E., Leproult, R., & Kupfer, D. J. (1996). Effects of gender and age on the levels and circadian rhythmicity of plasma cortisol. *Journal of Clinical Endocrinology & Metabolism*, *81*(7), 2468-2473.
- Wiseman, V., Conteh, L., & Matovu, F. (2005). Using diaries to collect data in resource-poor settings: questions on design and implementation. *Health Policy Plan* *20*, 394-404. doi: 10.1093/heapol/czi042
- Zautra, A. J., Davis, M. C., Reich, J. W., Sturgeon, J. A., Arewasikporn, A., & Tennen, H. (2012). Phone-based interventions with automated mindfulness and mastery messages improve the daily functioning for depressed middle-aged community residents. *Journal of Psychotherapy Integration*, *22*(3), 206-228. doi:10.1037/a0029573

Appendix A

Survey data can be collected directly inside mTurk, however for researchers already familiar with other popular survey management tools (e.g., Qualtrics and Survey Monkey) it will likely be easier to simply link mTurk users to these surveys. MTurk provides easy to follow [tutorials](#) to help facilitate this process. Regardless of your survey management choice two important points to consider before starting data collection are the activation/deactivation of surveys and survey auto-submission.

Activation/deactivation of surveys.

Activation simply requires the researchers to “turn-on” the survey before the start of data collection. Should the researchers forget to activate the survey, participants will be unable to participate. An equally important step is deactivating the surveys when data collection has finished. Failure to do so could result in survey responses being recorded after the mTurk participants had finished completing their surveys. This data is often junk or completed too quickly to be accurate responding.

Survey auto-submission.

Researchers should consider when they want responses to auto-submit (i.e., when should the survey be submitted if the participant does not submit the survey themselves). Individuals may leave the survey open on their computer without actively working on it. This can result in response times of nearly a day, which will positively skew mean response times. If all responses should be recorded in one sitting then the researcher may want to consider having the survey auto-submit after a reasonable completion time (perhaps twice your anticipated completion time). On the other hand, if the researcher wants individuals to be able to return to the survey

should the participants be interrupted then the researcher may choose to have the survey auto-submit after 24 hours.

IRB Concerns.

It is important to consult with your Internal Review Board (IRB) to ensure a smooth review of an mTurk study. The researcher may need to plan for a longer than normal review process as many IRBs have not dealt with mTurk yet.

The authors' IRB initially required them to constrain their participants to the USA only. Although they eventually gained approval to open the HITs to other countries that spoke English, however they were told they would need to undergo a cultural review before opening to any other countries. Consider the population you want to sample before applying for IRB approval.

Additionally, the authors' IRB said that there have been reports of individuals being forced to work on mTurk in some developing countries. They had ethical concerns about approving studies wishing to sample from these areas.

Preparing for mTurk.

Researchers should be prepared before launching an mTurk study. Once a HIT is posted on mTurk, the researchers should expect to receive messages about the HIT that workers will want answered quickly. If the design includes qualifying individuals (e.g., verifying their age before they continue with the remainder of your survey questions, see below) the researchers will need to be ready to review the incoming responses and then assign appropriate qualifications on mTurk. If new to mTurk, the researcher may want to consider rolling their study out in "batches" (e.g., if the goal is to collect data on 200 people the researchers might not want to start by requesting 200 HITs. Instead start with 20 and see what types of questions and responses occur. Then the researcher can adjust instructions and questions before proceeding further. This also

allows the researcher to get a feel for how quickly people will complete the work and how many people can be managed at the same time).

Prepare for mTurk workers to freely share their opinions about the study. The authors' workers sent many messages. At the beginning some people messaged to report they did not like the HIT or that they thought it should be designed differently. On the other hand, many individuals messaged to report that they enjoyed doing the surveys and that they wished the authors well in their research efforts. The authors were not prepared for the emotional rollercoaster of participant messages. This will likely vary depending on what is being asked of participants, but be prepared to hear from the participants in a way that one might not when working in more traditional methods.

In the authors' experience it worked best to post HITs in the morning (e.g., around 9am EST). This ensured that all the HITs would be finished before the end of the work day. In the few cases where HITs needed to be posted after 3pm EST, it became necessary to respond to questions and qualifying individuals into the wee hours of the night. If posting early on the east coast, the researchers may have a slower response time than if they waited to post around 12pm EST. This is because the rest of the country is likely not up yet. Noon EST was the best time to post for getting quick responses during normal business hours.

How many HITs for a daily diary?

The authors set up their daily diary survey as 2 separate mTurk HITS: 1 for Day 1 and 1 for Days 2-9. This would be the minimum number of HITS to create for daily diary research. The Day 1 HIT serves as a qualifier to make sure the participant should be invited to continue on with the study. The Day 1 HIT can also be used to collect baseline data. Depending on the budget and what the researchers intend to pay for Day 1, the researchers might want to set up a pre-Day 1

screening HIT. This would be a HIT where participants are paid a very small amount, but only asked the questions needed to determine if an individual was eligible to continue with the study. These individuals could then be assigned a “qualification” that would be needed to complete all future HITs (i.e., Days 1-9).

Researchers could also consider creating a HIT for each study day. In the authors’ study not everyone completed all 9 study days. In order to pay these individuals for their work it became necessary to provide a bonus to the participants’ Day 1 HIT (i.e., adding \$1 bonus for each day completed). MTurk will only allow requesters to approve or not approve a worker’s HIT. Thus on the Days 2-9 HIT the authors could not approve an individual who had not completed all 9 days. Furthermore, mTurk requires participants to leave the window for the HIT open while they complete the survey and then return to the window to enter the code that verifies they completed the study (this is something that can be set up in Qualtrics). Leaving a window open for 9 days will not work for most people. Researchers can advise individuals to not accept the Days 2-9 HIT until they finished on Day 9 to avoid this problem. Creating a HIT for each day would also avoid this technical problem, and not require the researcher to verify that all 9 surveys have been completed before approving the Days 2-9 HIT. However, using daily HITs would require the researcher to be constantly monitoring and approving each daily HIT.

One of the most common qualifications researchers may want to create is a qualification that denotes “previous participants.” This could be used in two ways. First, for any future Day 1 HITs the researcher could set a qualification that individuals who have been assigned the “previous participant” qualification are not eligible, thus ensuring that only new participants are enrolled. Second, the researcher could use the “previous participant” qualification to ensure that

only participants who have previously completed the Day 1 HIT are allowed to continue on to the Day 2 HIT.

Researchers can also create qualifications related to any other participant characteristics (e.g., age, income, etc.). Researchers do not have to pay for qualification that they create. However, individuals will need to respond to these questions first and then the researcher can assign the qualification based on the participant's response. One way to ensure accurate responses is to create a qualification HIT that gives no hint of the desired response. This can help avoid desirable responding by workers.

MTurk Participants.

The authors asked for participants who were at least 60 years old. Despite stating this in the title of the HIT and in the HIT description, about 10% of those who completed the Day 1 HIT were under 60. On the plus side, this indicates that mTurk workers will answer honestly about their age and birthday even when it means they do not qualify for the study.

The authors had no difficulty recruiting participants aged 60+. All of the HIT requests were completed within hours of posting them. However, it is possible that the difficulty of recruiting would increase as the age cutoff was raised. The mode age for the authors' sample was 60. Eighty percent of the participants were in their 60s (10% were younger, 9% were in their 70s, and 1% were in their 80s, with 1 90 year old). Researchers may be interested in surveying individuals 65+; in the authors' research they made up 30% of the sample. However, it is important to note that the authors were asking individuals to commit to 9 study days. It is possible that it may be easier to recruit older participants for less demanding studies.

In the aforementioned sample participants were screened for dementia/MCI by asking if a doctor had ever told them they had either. For this sample, 10% self-reported these diagnoses.

Given the three screening questions (age, location, and dementia/MCI) the authors disqualified 15% of individuals who took the Day 1 HIT. It was not uncommon for someone to meet more than one disqualification criteria (e.g., to be too young and been diagnosed with an MCI).

Daily Diary Concerns.

The authors asked individuals to self-generate a code to enter when taking each survey so that they could link the surveys together. The same instructions were provided each day to help ensure that each participant used the same code each day (i.e., *Please enter the code you generated on day 1. You created your code by typing the first 2 letters of your first name, the first 2 letters of your last name, the first 2 numbers of your zip code, and the first 2 numbers of your phone number. For example if your name is Bill Smith, your zip code is 98765 and your phone number is 123-456-7890, you would type: bism9812.*) This seemed like a good plan but many people had trouble following these instructions. First, three individuals listed bism9812 as their code thus requiring additional work to tell them apart. Second, many individuals did not generate the same code for each study day. One individual generated four different codes across the nine study days. Most errors appeared to be inversions of two letters or numbers (e.g., instead of bism9812 they wrote bims9812 or bism8912) or common typos (i.e. entering an 8 instead of a 5 on a 10-key pad). The authors requested individuals report their birthday on each study day as well. This practice is highly recommended as both a check on accurate age reporting and as a way to verify participants' identity across study days.

Another option would be to generate individualized Qualtrics links for each person on each study day and individually send those links. The trouble with this approach is that mTurk rules do not allow for the collection of personally identifying information including email

addresses. Thus researchers cannot simply enter emails into a survey management tool like Qualtrics. One way to communicate with participants is by sending them a message attached to a “bonus.” If researchers used this approach it would require them to individually send a message to each participant with their own link. This would require considerably more time and effort on the part of the researcher.

When conducting daily diary research, investigators will be working with participants for an extended period of time. Even though the Day 1 HIT might have been set up to occur during hours that work well for the researcher, there is really no telling when the participants will message during the next 9 days (note: researchers are free to provide their email to mTurk workers and should do so in order to address any questions). If participants are in different time zones, researchers may be getting messages at 2AM. Have a plan for how to handle messages. The most common message the authors received was from participants who needed a password to access a survey. The authors wanted to respond to them as quickly as possible so that the participants would not give up, leading to a loss of data. If lab members are involved who sleep odd hours this could be a solution. Researchers could also set up “vacation” messages that answer common questions as an auto-respond during sleeping hours.

Paying participants on mTurk.

mTurk charges fees as a percentage of what participants are paid. If requesting 10 or more participants the fee is 40%. If requesting 9 or fewer participants the fee is 20%. To avoid paying the higher 40% fee post the survey in “batches” with each batch requesting 9 or less participants. The downside to this option is that it will require more work to ensure that participants do not repeat the survey. Researchers can use the mTurk setting to state that each mTurk worker only complete 1 HIT inside any given batch. Thus if posting a batch with 90 HITs

the researcher can ensure each of those 90 surveys is completed by a different mTurk worker. However, if posting 10 batches of 9 HITs the researcher could in theory end up with 9 people taking the survey 10 times each. This can be avoided by posting the first 9 HIT batch and waiting for it to finish. Once it has finished the researcher can assign each person who took that survey a “qualification.” When posting the 2nd batch the researcher can specify that only individuals without that qualification are eligible to complete a HIT. The process would need to be repeated after the 2nd batch was finished and so on. This will require more time and effort on the part of the researcher and will result in a longer completion time than posting 1 batch with 90 HITs, however if paying \$1 a HIT you would pay \$18 in fees vs. \$36.

Another way to get around this situation is to post 1 batch with 90 HITs but only offer to pay \$.01 per HIT. The researcher would need to post in the HIT description that participants will be given a bonus reflecting the remaining \$.99 (or whatever amount is chosen). The reason researchers may want to do this is that bonuses are assessed the 20% fee regardless of the original HIT size. In this scenario you would pay \$21.42 in fees and avoid the need to constantly update your participant’s qualification. However, researchers would need to individually apply bonuses to each participant’s HIT. Furthermore, mTurk allows workers to search for HITs based on how much they pay. In this scenario the HIT would be listed as paying \$.01. Individuals may see this and choose not to complete the HIT or not to read the description that states the actual take-home pay will add up to \$1 (or the amount of the original HIT plus bonus).

Author Copy: Not for Distribution