Measuring Diurnal Cortisol Change in a Population-Based Field Study:

Don't Try This at Home (Alone)

Carolyn Tucker Halpern,[1,4] Kathleen Mullan Harris,[3,4] Eric A. Whitsel,[2] Brandon Wagner[3,4]

[1] Department of Maternal and Child Health, University of North Carolina at Chapel Hill (UNC-CH)

[2] Department of Epidemiology, UNC-CH

[3] Department of Sociology, UNC-CH

[4] Carolina Population Center, UNC-CH

Paper to be presented at the Population Association of America Meeting

Dallas TX April 15-17, 2010

Biomarkers are increasingly used in population-based[1] field research (National Research Council, 2008). One biomarker of interest is cortisol, an endogenous corticosteroid that affects multiple physiological systems and has been implicated in a wide range of physical and psychological illnesses. Cortisol has a strong diurnal profile that emerges early in infancy (Gunnar & Vazquez, 2006). Circulating concentrations of cortisol are influenced by the effects of stress on the hypothalamic-pituitary-adrenal (HPA) axis. In response to major and minor stressors, the hypothalamus secretes corticotropin releasing hormone (CRH), which causes the pituitary to secrete corticotropin, which in turn prompts the adrenal cortex to secrete cortisol. The HPA axis has an important role in supporting normal physiological functions and in regulating other systems. For example, cortisol affects gluconeogenesis (stimulating the conversion of amino acids and other substrates to glucose in the liver), lipolysis (breaking down fats for energy), vascular reactivity, as well as function of the inflammatory, immune, and central nervous systems.

Although the negative biological and health effects of prolonged cortisol activation (e.g., produced by chronic exposure to stressors) have been documented (McEwen & Stellar, 1993), more work is needed, especially longitudinal studies examining relationships between daily stress, cortisol concentrations, and physiological parameters over time (Smyth et al., 1998). There is interest, for example, in determining whether change over time is typical or atypical, as atypical patterns of change may reflect dysregulation of the HPA axis. Thus linking context and life experiences with more naturalistic cortisol measurement outside of laboratory settings is of increasing significance.

[1] By population-based, we are referring to studies of large probability samples that combine demographic, social, and behavioral data and yield representative findings that can be generalized to a defined population.

Technical advances have allowed for assay of cortisol in saliva. Saliva collection is low-risk and non-invasive, making cortisol measurement outside of laboratory and clinical settings both practical and affordable. Sampling ease makes salivary cortisol measurement especially useful for large scale studies. Moreover, salivary cortisol concentrations are highly correlated with serum and plasma-based measures of non protein-bound cortisol concentrations, allowing for inference about its most physiologically active fraction (Kirschbaum & Hellhammer, 1994).

Of particular interest in the context of field collection is the cortisol response to awakening (CRA), which is thought to be a reliable indicator of the "acute reagibility of the HPA axis" (Fries, Dettenborn, & Kirschbaum, 2009).The CRA is a rapid increase of cortisol within a 20- to 30-minute period after awakening (approximately 38 to 75% of awakening levels). It is typically quantified as the difference between an awakening sample and a sample taken 30 minutes after awakening; a CRA slope may also be calculated.[2] The CRA is thought to be linked to activation of memory and the anticipation of upcoming life demands, which can stimulate HPA activity (Fries et al, 2009). The CRA is also positively correlated with reactivity to laboratory stressors and ACTH administration. Changes in CRA have been reported in association with experience of chronic stress. The CRA can be observed in about 75% of adults, and has been reported to have high intra-individual stability (Fries et al., 2009; Hucklebredge et al., 2005). However CRA measurement appears to require exact timing (Kunz-Ebrecht et al., 2004; Pruessner et al. 1997), increasing the importance of protocol adherence.

There are different types of non-adherence. A respondent (R) may refuse to provide samples, or even if they agree, they may not actually produce the sample(s). This yields missing data, which if patterns of missingness are correlated with other relevant respondent characteristics, may bias analyses. Alternatively, respondents may produce the sample(s) at the

[2] The slope may be calculated by taking the difference between the two log-transformed cortisol values and dividing by the amount of time between samples, yielding an index of change in natural log units per hour (Almeida, Piazza, & Stawski, 2009b).

“wrong” time (i.e., not according to protocol). If the R accurately reports actual time of collection, these issues can be addressed analytically. However, if they provide inaccurate times that appear to follow the protocol, but actually do not, additional error is introduced and, depending on the levels of non-adherence, findings could be biased, possibly leading to spurious associations.

A number of methodological studies have been conducted to evaluate optimal saliva sampling strategies for cortisol assessment, different “instructional sets,” and their implications for protocol adherence. These studies typically rely on small samples and employ electronic (time-stamp) monitors in caps (hereafter referred to as TrackCaps) covering containers used to store tubes or Salivette cotton rolls to be used for saliva sample collection. Among widely cited studies using such technology is one conducted by Kudielka, Broderick, and Kirschbaum (2003). They used a sample of 42 “community-dwelling subjects” who were 15 to 75 years old to evaluate adherence to a one-day protocol requiring the collection of six samples (directly after awakening, 30 minutes later, and at 11 AM, 3 PM, 8 PM, and 10 PM) in participants’ usual environment. Half the participants were randomly assigned to a condition in which they were told their protocol adherence was being monitored. Because cortisol levels change rapidly after awakening and then decrease at a slower rate over the rest of the day, adherence was defined as: within 10 minutes of awakening (sample 1), 30 +/- 7 minutes after awakening (sample 2), and +/- one hour of the remaining targeted times (samples 3-6).

Kudielka et al. (2003) found a quarter of participants to be non-adherent (failing to obtain one or more of the six samples within the designate time windows); about 10% of samples were out of specified time ranges. More than half of non-adherent samples were those obtained during the early morning rise, a critical sample for characterizing an individual’s circadian profile. Analyses indicated that diurnal profiles were different for respondents who did and did not adhere to protocol, and that the response to awakening was especially sensitive to non-

adherence. Non-adherent participants showed a much smaller cortisol increase, and without information on adherence, their CRA pattern would have been erroneously labeled as “blunted” (Kudielka et al., 2003). Investigators speculated that protocols entailing more days of saliva collection were likely to yield even poorer adherence. However, one promising study finding was that telling respondents that their sample collection was being electronically monitored resulted in greater protocol adherence and increased accuracy of respondents’ self-reports of collection time.

Broderick, Arnold, Kudielka, and Krischbaum (2004) obtained similar results in a sample of 65 female fibromyalgia patients and matched controls using a 7-day protocol requiring collection of five samples per day (including samples to measure CRA), with collection monitoring again accomplished via TrackCaps. As in the earlier study, self-reported adherence was inflated; verified adherence was 71% among Rs who did not know they were monitored. Differences in self-reported versus verified adherence were especially large in healthy controls, whose overall verified compliance was 62% compared with their self-report of 92%. Because of the number of sampling days, Broderick et al. (2004) could evaluate, within individuals, the performance of adherent and non-adherent samples for about half of participants. They found that for both the early morning rise and across-the-day slope, the change in cortisol was significantly greater for the adherent samples compared with the non-adherent. As in the Kudielka et al. (2003) study, knowledge of monitoring appeared to significantly improve adherence and accuracy of self-report. Broderick et al. also concluded that a 7-day collection protocol did not appear to exacerbate adherence problems.

In a later 2007 study, Kudielka, Hawkely, Adam, and Cacioppo replicated the 2003 and 2004 findings using an older sample of 83 participants who were 50 to 67 years old, and a 3-day sampling protocol requiring an awakening and +30 minute post-awakening sample each day. A TrackCap was again used to obtain a proxy for time of sample collection. In this analysis they found that 60% of participants were non-adherent (using time windows described above) on one

or more days, that non-adherence was inconsistent across days, and as demonstrated before, that non-adherence strongly affected the accuracy of cortisol measurement in the field setting. Further, adherence was positively correlated with participants' reports of social support, suggesting the possibility that if adherence is associated with psychological, behavioral, or contextual factors under investigation, spurious relations could result or "real" associations might be obscured (Kudeilka et al., 2007). Taken together, findings from these small scale studies based largely on convenience samples of older adults indicate a high likelihood of non-adherence that may or may not be affected by the number of sampling days, but that appears to be improved by the perception that protocol adherence is being monitored.

Several relatively large studies have incorporated non-laboratory saliva collection protocols. The Coronary Artery Risk Development in Young Adults Study (CARDIA) collected saliva for cortisol measurement from a subset of 33 to 45 year-old participants who lived within 50 miles of two study sites at the year-15 examination (6^{th} interview). The CARDIA protocol entailed collection of six cortisol samples over the course of one weekday (awakening, 45 minutes post-awakening, 2.5 hours post-awakening, 8 hours post-awakening, 12 hours post-awakening, and at bedtime). Participants also received an alarm watch to remind them to collect samples. Samples were collected by investigators the following day (Matthews, Schwartz, Cohen, & Seeman, 2006; Cohen et al., 2006). Of the 1336 eligible participants, 63% agreed to collect saliva and 60% of eligibles actually returned samples. With elimination of samples showing extremely atypical cortisol patterns, 58% of eligibles (about 775 respondents) produced samples for analysis (Cohen et al., 2006). Reported cortisol values demonstrated the CRA, although the amount of average change reported (Cohen et al., 2006) appears to be less than those reported from other smaller scale studies. To our knowledge, assessment of protocol adherence was not conducted.

Perhaps the most ambitious effort to date in field collection of saliva for cortisol measurement in the U.S. is that conducted in the National Study of Daily Experiences (NSDE). The NSDE is a telephone diary study (Daily Diary portion of MIDUS II - Midlife in the United States Survey) in which participants ages 35 to 84 years completed short telephone interviews at the end of eight consecutive days, and in the second wave of data collection, also collected saliva samples four times a day (awakening, 30 minutes after awakening, before lunch, at bedtime) on days 2 through 5 of participation, using a "home saliva collection kit" received about a week before the initial phone interview (Almeida, McGonagle, & Kim, 2009a). The kit included numbered and color-coded salivettes, and a detailed instruction sheet. Telephone interviewers also reviewed collection procedures with respondents. Self-reported times of sample collection were obtained each evening, and participants also recorded times in a paper log. About a quarter of participants (N=430) had their collection (salivette) materials stored in boxes with a time stamp lid. After all days of collection were completed, participants returned samples using a pre-addressed, paid courier package.

Of 2,022 respondents who participated in the Wave II NSDE, 86% (1736) provided usable saliva samples. Correlations between self-reported times and time stamps ranged from .75 (evening) to .95 (morning). Cortisol values for the awakening and +30 minute samples were similar to those generated in smaller studies conducted under more controlled conditions (Almeida et al., 2009a), as were slopes across the course of days. These promising findings suggested the utility of this protocol, at least for samples composed of middle aged adults and for study sizes where daily telephone contact is feasible.

The National Longitudinal Study of Adolescent Health (Add Health) is a nationally representative survey of U.S. adolescents enrolled in grades 7 through 12 in the 1994–1995 school year. Over 90,000 adolescents in 132 schools participated in the Wave I in-school survey, with 20,745 also completing subsequent in-depth Wave I in-home interviews (1994–

1995 school year). Follow up in-home interviews were completed in 1996 (Wave II), 2001-2002 (Wave III), and 2008 (Wave IV). Understanding the interplay between environment, behavior, and biology in their contributions to health trajectories over time was a focus of the Wave IV in-home assessment. Examination of the longitudinal relationships between stress, cortisol, and health is a central component, so saliva collection for cortisol assay was planned as one indicator of stress.

This paper describes the evaluation of a three-sample, one-day, post-interview protocol for collecting saliva and assaying cortisol concentrations in the Add Health Wave IV pretest, which was conducted in 2007. In this evaluation of the unsupervised, participant-conducted biospecimen collection protocol, we ask the following questions:

1. What percentages of respondents consent to collect samples, consent to sample archive[3], and actually return (any) samples?
    a. Do consent and return vary by sociodemographic characteristics (age, biological sex, race/ethnicity, attained education, employment status, presence of children under age 12 in the household)?
    b. Do incentive amount and/or reminder phone calls improve consent and/or sample return?
2. What is the degree of adherence to the collection protocol? We examine this in terms of respondents' self-reports of collection times, and verified collection times as validated by TrackCap time stamps.

[3] Add Health survey and biomarker data are used by researchers around the world. In Wave IV, one goal was to preserve an archive of all biospecimens collected to allow for subsequent analyses of other relevant analytes.

3. What is the intra-individual reliability of cortisol measures, as based on samples collected approximately 2 weeks apart according to the same protocol?

Methods

Pretest sample. Three hundred Add Health respondents residing in three states (NC, OH, TX) were selected for the pretest, of whom 193 were located and interviewed within the two-month window allotted for pretest field work.

Saliva Collection Protocol. We used a protocol that incorporated elements thought to improve consent and adherence. Due to time constraints, Rs were asked to self-collect saliva samples for cortisol assay on the day after their interviewer-administered in-home interview. Because multiple samples are needed to capture diurnal change in cortisol levels, Rs were instructed to collect 3 samples on a single day: upon awakening, 30 minutes post awakening, and at bedtime. We defined "awakening" as "**before you get out of bed** (awakening is when your eyes are open and you are ready to get up for the day." Respondents were encouraged to keep the saliva kit next to their bed and to set (their own) alarm to prompt them to make the second sample. Respondents were also asked to complete a brief checklist after each collection, noting time of sample collection, any stressful events that occurred on the collection day, food/beverage consumption, drug use, and physical activity that could affect cortisol concentrations. Interviewers gave Rs detailed oral and written instructions about the collection protocol, after which Rs practiced an interviewer-supervised sample collection (expectorating into a small tube).

The collection protocol was designed based on the expectations that: (1) it would not be unduly burdensome for Rs, (2) it would maximize participant consent, (3) it would be associated with high adherence to the fixed time-of-day collection protocol (Kudielka et al. 2003), and

assuming good adherence, (4) mean slopes of the cortisol concentration-time association would be similar across single- and multiple-day collection protocols.

Saliva Collection Materials. Respondents received three small, color-coded and pre-labeled collection tubes (#1, 2, 3) stored in a plastic bottle closed by a MEMS TrackCap that recorded the dates and times when the cap was removed from the bottle. Respondents were instructed to only open the TrackCapped bottle when they were about to collect saliva, and to close the bottle after removing each tube. Respondents used a pre-addressed, postage-paid envelope to mail collected saliva samples to the lab.

Salivary Assay. Saliva samples were assayed for cortisol in duplicate by Salimetrics Laboratories. The Salimetrics HSCortisol kit is a competitive enzyme immunoassay specifically designed for the quantitative measurement of salivary cortisol. The assay has a range of sensitivity from .007 to 1.8 µg/dl, and average intra- and inter-assay coefficients of variation less than 5% and 9%, respectively.

Measures and embedded experiments. To assess ways of maximizing consent, adherence to protocol, and reliability, we embedded several experiments in the pretest. Regarding consent, Rs were divided into two subgroups to test two incentive amounts. Group 1 was paid $40.00 (by a mailed check) for their participation in the cortisol sample collection and Group 2 was paid $20.00 (by a mailed check). Regarding protocol adherence, Rs were divided into 3 sub-groups: one third were told that saliva collection times would be monitored for all Rs, one third were told that saliva collection times would be monitored for a randomly selected subset, and one third were not told about the possibility of collection time monitoring. In actuality, all Pretest Rs were monitored via the TrackCap, which allowed comparison of self-reported and TrackCap-recorded collection times. Knowing that it would not be possible to monitor protocol adherence for the entire Add Health Wave IV study population (which later achieved completed interviews with

15,701 respondents) given the prohibitive cost of TrackCaps ($100 each), demonstration of the accuracy of self-report and adherence to the morning protocol was critical.

Analyses. To assess sociodemographic differences in consent and sample collection, we examined sample information and protocol adherence according to a number of respondent characteristics. These were age at the time of the Wave IV pretest interview, biological sex, race/ethnicity, highest completed education, whether the respondent was employed at the time of the interview (more than 10 hours/week), and whether there were children under age 12 living in the respondent's household. To test whether different response levels of variables in the analysis were associated with significantly different success in our response categories (consent, consent to archive, sample return, and adherence), we calculated t-tests for differences in group means using the first group in the table listings as the reference category.

Intra-Individual Variation (IIV) Study. Wave IV also incorporated an intra-individual variation study of 100 Add Health Rs, 43 of whom were interviewed during the pretest. Of the 43 Rs, 58% were female, 65% White, 16% Black, 12% Latino, and 7% other. The primary goal of the IIV study was to estimate the short-term reliability of the study's biomarkers, including salivary cortisol concentrations. To this end, IIV Rs were seen twice, 1-2 weeks apart (mean = 8.2 days). At Visit 1, Rs completed the full 2-hour interview and post-interview saliva collection protocol. At Visit 2, an abbreviated interview and post-interview saliva collection were repeated. Salimetrics laboratory staff responsible for processing the saliva were masked to the common origin of saliva collected repeatedly from the same respondent.

Results

Sample Characteristics

Table 1 includes information about the sociodemographic characteristics of the 193 respondents participating in the pretest. The pretest sample was 52% female, with a mean age

27.75 years (standard error = 0.14; range: 24-31). Race/ethnicity composition was 69% white, 19% black, 8% Latino, and 4% other. Nine of 10 pretest Rs had completed high school (diploma/GED) and 31% had a college degree or some education beyond college. Sixty one percent were employed at the time of the interview. Of interviewed respondents, 46% had children under age 12 living in the household at the time of the interview. The $40 versus $20 incentive amounts were distributed approximately equally.

Consent and Sample Return

Table 2 displays the percentages of sociodemographic and incentive groups who consented to provide saliva samples for cortisol assay, consented to archive samples, and who actually returned any samples to the lab (even if collection were incomplete). Overall, 188 of 193 pretest Rs agreed to provide saliva samples, yielding a 97% consent rate. One hundred fifty-six Rs (or 83% of those consenting, 81% of interviewed Rs) agreed to have their samples archived after cortisol measurement. However, only 146 Rs actually returned any samples to the lab (78% of consenting and 76% of interviewed pretest Rs).

Likelihood of consent, both for immediate cortisol assay or archive, was unrelated to all sociodemographic characteristics examined, and unrelated to the incentive amount offered. Actually collecting/returning any saliva samples was related only to completed education. Respondents who are high school graduates or have higher education were more likely to return samples than were those Rs who did not graduate high school. (Note, however, that there are only eight Rs in the pretest who did not complete high school.) Respondents with education beyond college were the most likely to return samples. Amount of incentive was marginally associated with returning samples ($p = .06$), with the higher incentive group being more likely to follow through.

Reminder calls to maximize Respondents' sample return. To increase receipt of saliva from Rs who had agreed to provide samples, reminder calls were attempted for Rs whose samples had not been received at the lab within three days post-interview (47% of those consenting to saliva collection). Of 29 Rs successfully contacted before sample receipt who confirmed their intent to collect and mail samples, 62% never returned samples.

Protocol adherence

Of saliva packages received, 66% included all checklists, TrackCaps (which Rs had been instructed to return), and three samples. Of samples received, 25% were missing self-reported collection times.

Table 3 displays percentages of sociodemographic and incentive groups with varying levels of protocol adherence among the 136 respondents who returned at least one sample and for whom adherence could be verified.[4] Seventy nine percent of respondents completed at least one sample according to protocol (57% of those who consented). Sixty nine percent completed the first sample on time, 63% completed the second on time, and 60% completed the third on time. This translates to 56%, 45%, and 43%, respectively, of Rs who had consented to collect samples. Only 46% of the 136 Rs for whom adherence, as defined by comparison with the TrackCap stamped time, is calculable fully adhered to the collection protocol (or one in three of respondents who had consented to collect saliva). None of the sociodemographic characteristics examined or incentive level was associated with adhering to protocol for any of the samples at the .05 level of significance. Being employed was marginally associated with some type of protocol adherence.

[4] Following the work of Kudielka et al., 2003, adherence was defined according to accuracy windows of +/- 10 minutes for sample 1; +/- 7 minutes for sample 2; +/- 60 minutes for sample 3. Broadening accuracy windows did not change findings.

Cortisol Response to Awakening. As noted earlier, adherence to the 30-minute collection protocol for the first and second samples was crucial because of interest in the cortisol response to awakening (CRA). For Rs who were not missing self-reported times for samples 1 and 2, and who correctly reported collection times for those samples (74), 73% (54 Rs) adhered to the 30 minute lag (+/- 5 minutes). This translates to 29% of consenting respondents who appear to have adhered to the 30” awakening protocol.

Intra-Individual Reliability

Table 4 includes information about the average concentrations of cortisol for each of the three samples for the 136 Rs who returned samples and for whom timeliness of collection was calculable. Mean cortisol concentrations provide some evidence of the CRA, with an average 63% increase over mean wakening levels. However, the short-term reliability of samples for the 27 pretest respondents who participated in the IIV study (and returned samples) was poor, especially for the wakening sample, where the intraclass correlation was essentially zero.

Effects of varying information provided about monitoring

Results of separately conducted multiple regression models indicated that varying information told to respondents about sample collection monitoring (all, random, none) had no effect on sample receipt, self-reported adherence to protocol, or verified adherence to protocol (results not shown). There was also no interaction between the incentive amounts and information received about monitoring.

Discussion

Based on protocols tested in smaller extant studies (e.g., Kudielka et al., 2003; Broderick et al., 2004), we had expected that Rs would be more likely to adhere to protocol—or more likely to self-report the actual collection time had they deviated from it—if they believed there

was a good chance they were being monitored. Protocol adherence among study participants who are aware they are being monitored has been reported as greater than 90% (Kudielka et al., 2003). We also expected that given respondents' past participation in Add Health and the relatively low burden of our collection protocol, we would see not only high levels of consent, but also good adherence to protocol and accuracy in reported sample collection times. None of these expectations was supported in our pretest data. Rather, our analyses suggest that although virtually everyone consents to saliva collection for cortisol measurement, only about 8 in 10 actually return samples. Despite our liberal testing strategy, only education, of the sociodemographic characteristics we examined, was significantly related to sample return. Highly educated participants were the most likely to collect and return samples (but of note, not different in terms of protocol adherence).

We did see a trend suggesting the potential for a higher monetary incentive to improve levels of sample return. However the advisability of implementing an incentive on the order of $40 should be considered in the context of study size, total cost, and expected levels of protocol adherence and data quality. For example, given our final sample size of 15,701 interviewed respondents at Wave IV, applying the 97% consent and 81% return rates yields a cost close to $494,000 in incentives alone when a $40 incentive is used. Consideration of adherence levels is also discouraging. About a quarter of self-reported collection times were missing for the 76% of the sample who actually returned samples. In the main field work, we would have to rely exclusively on self-reported collection times; pretest data suggested we could expect large amounts of collection time data to be missing. Of respondents for whom protocol adherence could be calculated in the pretest, less than half adhered to the full protocol. Although we saw some indication of a CRA based on mean values, ICCs for the awakening and +30 minute samples were extremely poor, calling the interpretation of these cortisol concentration values

into question[5]. The estimate of morning rise, key for the Add Health project purposes, is very dependent on the wake up sample being collected according to protocol. Of consenting respondents who returned samples, less than a third verifiably adhered to the morning protocol required for the calculation of the critical CRA measure. Because of cost anticipated in the main field work, we were unable to include protocol elements that smaller scale studies have used (e.g., timers that would alarm when it was time to collect sample #2) to enhance protocol adherence. Finally, use of labor intensive reminder calls – which would be prohibitively expensive for the full Add Health sample, yielded little return in terms of samples received. These considerations all contributed to our ultimate decision not to include saliva collection for cortisol assay in the protocol for main field work in Wave IV.

It is not clear why our return and adherence rates were poorer than those obtained in some other larger scale studies, nor why knowledge of monitoring appeared to make no difference in adherence. The Add Health sample at Wave IV is younger (24-32 years) than participants in CARDIA (33 to 45) and NSDE (35 to 84); differential time demands across age groups may have contributed to differences across studies. However, our consent and sample return rates were actually higher than those achieved in CARDIA. To our knowledge external protocol adherence (i.e., via TrackCaps or other technology) was not assessed in CARDIA. Adherence was examined in the NSDE sample, and the correlation between self-reported times and time stamps was .95 for morning samples. We speculate that the combination of an older sample and feasibility of making nightly phone contact with participants significantly enhanced protocol adherence and accuracy of self-report.

Given the general absence of associations between sociodemographic factors and indicators of sample return and protocol adherence in our pretest sample, it is possible that poor

[5] We want to emphasize that assay performance from our lab (Salimetrics) was excellent and was not a factor in poor intra-individual reliability.

adherence is more a reflection of a “day-level” rather than “person-level” problem (Thorn, Hucklebridge, Evans, & Clow, 2006). That is, circumstances of a given day may have contributed to poor adherence. If budgets permit, given study sample size, implementing a protocol that includes multiple days of data collection might facilitate superior adherence on at least some days and provide more usable values. However, the characteristics of events on collection days used in analyses would have to be examined, as those days may differ in systematic ways from others. For studies where diurnal slope can be used and CRA omitted, random time sampling (beeper strategies) might be a better choice than a fixed-time sampling protocol (Jacobs et al., 2005).

Limitations

There are a number of limitations in our evaluation. First, our sample size is small, especially for the pretest component of the IIV study, where we only had 27 pairs of samples for analysis. Findings might be different with larger numbers of persons and/or more days of sample collection. Second, our respondents, now ages 24-32, are at an extremely busy point in their lives. This was quite evident in our attempts to contact respondents and schedule interviews. We suspect this was a factor in the level of non-adherence. Based on the experiences of other studies, it appears that samples with different characteristics (e.g., older participants) may demonstrate better protocol adherence.

The initial awakening sample is crucial to quantify the CRA. Use of track-cap technology, although expensive, is somewhat crude and incomplete. We cannot definitely determine which opening times were or were not actually done in association with sample collection. Further, as in other studies, we cannot objectively determine whether sample collection occurred at awakening, as we defined it in our collection protocol.

Conclusions

The Add Health Wave IV pretest experience indicates that large population-based field studies should carefully evaluate the feasibility of, adherence to, and reliability of biomarkers assayed in unsupervised, participant-conducted biospecimen collection protocols. Multiple factors related to the protocol itself and the characteristics of the sample may enhance or detract from collection and adherence. The implications of our results should also be considered in light of a number of unresolved issues surrounding identification of influences on and interpretation of the CRA. In a recent review Fries et al. (2009) suggest that age, gender, and menstrual cycle phase do not appear to influence CRA, although results across studies are not completely consistent. (They also note that the relevance of age appears to vary depending on sample size – this could be related to issues of protocol adherence and the ability to monitor adherence as sample sizes increase.) There is more consistent evidence of the effects of stress related factors on CRA, but here too there are multiple unanswered questions and some inconsistencies in findings. For example, effects of stress-related factors may vary, depending on the duration of stress. A number of studies suggest the possibility that the CRA on any given day is greatly influenced by situational factors, and may vary based on anticipation of activities and demands *in the coming day.* Disentangling these nuanced influences will require minimization of measurement error via protocol adherence, and detailed information about experienced stressors both long term and in detail about yesterday, today, and tomorrow.

Additional experimental methodological research, necessarily based on smaller samples that can be more closely monitored, is needed to inform the protocols of large studies of geographically dispersed individuals. Incorporating additional technology to objectively asses the temporal sequence of sampling is needed. As others have noted (e.g., Almeida et al., 2009a), the more complete monitoring design is to use both actigraphy to monitor wake-up time (movement monitoring), and TrackCaps to (theoretically) monitor when a sample has been collected. Some data are available suggesting the improvements of these methods over self

report (e.g., Eissa, Poffenbarger, & Portman, 2001). However, short of time-stamped videotaping – certainly not an option for large population-based studies, a certain amount of faith about protocol adherence may always be needed.

References

Almeida, D.M., McGonagle, K., King, H. 2009a. Assessing daily stress processes in social surveys by combining stressor exposure and cortisol. *Biodemography and Social Biology* 55(2):219-237.

Almeida, D.M., Piazza, J.R., Stawski, R.S. 2009b. Interindividual differences and intraindividual variability in the cortisol awakening response: An examination of age and gender. *Psychology and Aging* 24(4):819-827.

Broderick, J. E., Arnold, D., Kudielka, B. M., & Kirschbaum, C. 2004. Salivary Cortisol Sampling Compliance: Comparison of Patients and Healthy Volunteers. *Psychoneuroendocrinology* 29:636-650.

Cohen, S., Schwartz, J.E., Epel, E., Kirschbaum, C., Sidney, S., & Seeman, T. 2006. Socioeconomic status, race, and diurnal cortisol decline in the Coronary Artery Risk Development in Young Adults (CARDIA) study. *Psychosomatic Medicine 68:*41-50.

Eissa, M.A., Poffenbarger, T., & Portman, R.J., 2001. Comparison of the actigraph versus patients' diary information in defining circadian time periods for analyzing ambulatory blood pressure monitoring data. *Blood Pressure Monitoring* 6(1):21–25.

Fries, E., Dettenborn, L., & Kirschbaum, C. 2009. The cortisol awakening response (CAR): Facts and future directions. *International Journal of Psychophysiology* 72:67-73.

Gunnar, M.R., & Vazquez, D. 2006. Stress neurobiology and developmental psychopathology. In: Cicchetti, D., Cohen, D. (Eds.), *Developmental Psychopathology: Developmental Neuroscience, vol. 2.* Wiley, New York, pp.533-577.

Hucklebridge, F., Hussain, T., Evans, P., Clow, A. 2005. The diurnal patterns of the adrenal steroids cortisol and dehydroepiandrosterone (DHEA) in relation to awakening. *Psychoneuroendocrinology* 30:51-57.

Jacobs, N., Nicolson, N.A., Derom, C., Delespaul, P., van Os, J., & Myin-Germeys, I. 2005. Electronic monitoring of salivary cortisol sampling compliance in daily life. *Life Sciences* 76:2431-2443.

Kirschbaum, C., & D. H. Hellhammer. 1994. Salivary Cortisol in Psychoneuroendocrine Research: Recent Developments and Applications. *Psychoneuroendocrinology* 19:313-333.

Kudielka, B. M., J. E. Broderick, & C. Kirschbaum. 2003. Compliance with Saliva Sampling Protocols: Electronic Monitoring Reveals Invalid Cortisol Daytime Profiles in Non-Compliant Subjects. *Psychosomatic Medicine* 65:313-319.

Kudielka, B. M., Hawkley, L.C., Adam, E.K., Cacioppo, J.T. 2007. Compliance with ambulatory saliva sampling in the Chicago Health, Aging, and Social Relations Study and associations with social support. *Annals of Behavioral Medicine* 34(2):209-216.

Kunz-Ebrecht, S.R., Kirschbaum, C., Marmot, M., Steptoe, A. 2004. Differences in cortisol awakening response on work days and weekends in women and men from the Whitehall II cohort. *Psychoneuroendocrinology* 29(4):516-528.

Matthews, K., Schwartz, J., Cohen, S., Seeman, T. 2006. Diurnal cortisol decline is related to coronary calcification: CARDIA study. *Psychosomatic Medicine* 68:657-661.

McEwen, B. S. & E. Stellar. 1993. Stress and the Individual: Mechanisms Leading to Disease. *Archives of Internal Medicine* 153:2093-2101.

National Research Council. 2008. *Biosocial Surveys* Committee on Advances in Collecting and Utilizing Biological Indicators and Genetic Information in Social Science Surveys. M. Weinstein, J.W. Vaupel, and K.W. Wachter, eds. Committee on Population, Division of Behavioral & Social Sciences and Education. Washington, DC: National Academies Press.

Pruessner, J.C., Wolf, O.T., Hellhamer, D.H., Buske-Kirschbaum, A., von Auer, K., Jobst, S., Kaspers, F., Kirschbaum, C. 1997. Free cortisol levels after awakening: A reliable biomarker for the assessment of adrenocortical activity. *Life Sciences* 61:2539-2549.

Smyth, J., M. C. Ockenfels, L. Porter, C. Kirschbaum, D. H. Hellhammer, & A. A. Stone. 1998. Stressors and Mood Measured on a Momentary Basis are Associated with Salivary Cortisol Secretion. *Psychoneuroendocrinology* 23:353-370.

Thorn, L., Hucklebridge, F., Evans, P., & Clow, A. 2006. Suspected non-adherence and week-end versus week day differences in the awakening cortisol response. *Psychoneuroendocrinology* 31:1009–1018.

Table 1. Add Health Wave IV Pretest Sample Composition	
	Percentage of Sample (N)
	100% (193)
Age at Wave IV (years)	
Mean (SE)	27.75 yrs (0.14)
Biological Sex	
Male	47 (92)
Female	52 (101)
Race/Ethnicity	
White	69 (135)
Black	19 (36)
Hispanic	8 (17)
Other	4 (4)
Completed Education	
Less Than High School	10 (20)
High School Graduate or GED	19 (37)
Some College or Vocational Training	38 (74)
College Graduate	25 (50)
Graduate Education beyond College	6 (12)
Employed 10 or More Hours/Week	
No	38 (75)
Yes	61 (118)
Children Under Age 12 in Household	
No	53 (104)
Yes	46 (89)
Incentive amount	
$40.00	52 (101)
$20.00	47 (92)

Table 2. Percentages of Sociodemographic Groups Consenting to Sample Collection, Consenting to Archive, and Returning Samples (Pretest N = 193)

	Consent	Consent to Archive	Any Samples Returned
Respondent Characteristics	Percentage (n)	Percentage (n)	Percentage (n)
	97% (188)	81% (156)	76% (146)
Age			
Mean (SE)	27.73 (0.15)	27.62 (0.16)	27.67 (0.18)
Biological Sex			
Male	96 (88)	78 (72)	71 (66)
Female	99 (100)	83 (84)	79 (80)
Race/Ethnicity			
White	97 (132)	81 (110)	78 (106)
Black	100 (36)	83 (30)	69 (25)
Hispanic	100 (17)	76 (13)	76 (13)
Other Race#	---	---	---
Completed Education			
Less Than High School (ref)	100 (20)	79 (14)	40 (8)
High School Graduate or GED	97 (36)	86 (32)	81 (30) *
Some College or Vocational Training	97 (72)	82 (61)	74 (55) *
College Graduate	96 (48)	80 (40)	84 (42) *
Graduate Education beyond College	100 (12)	75 (9)	91 (11) *
Employed 10 or More Hours/Week			
No	97 (73)	59 (62)	78 (59)
Yes	97 (115)	79 (94)	73 (87)
Children Under Age 12 in Household			
No	97 (101)	81 (85)	77 (81)
Yes	98 (87)	79 (71)	73 (65)
Incentive amount			
$40.00	97 (98)	76 (77)	81 (82)
$20.00	98 (90)	85 (79)	69 (64) +

Denominator for all groups is their total number in the sample

Cell sizes too small to report

* $p < .004$

+ $p = .06$

Table 3. Percentages of sociodemographic groups with varying levels of saliva collection protocol adherence (of 136 respondents who returned samples & for whom timeliness was calculable)

Respondent Characteristics	Any protocol adherence	Time 1 (Wake) verified correct	Time 2 (+30") verified correct	Time 3 (Bedtime) verified correct	Complete protocol adherence
	% (n)	% (n)	% (n)	% (n)	% (n)
Characteristic (number available samples)	78.68% (107)	69.12% (94)	63.23% (86)	59.56% (81)	45.59% (62)
Age (years)					
Mean (SE)	27.65 (0.20)	27.66 (0.22)	27.51 (0.24)	27.75 (0.24)	27.61 (0.28)
Biological Sex					
Male (60)	75.00 (45)	63.33 (38)	61.67 (37)	53.33 (32)	40.00 (24)
Female (76)	81.58 (62)	73.68 (56)	64.47 (49)	64.47 (49)	50.00 (38)
Race/Ethnicity					
White (97)	78.35 (76)	69.07 (67)	64.95 (63)	57.73 (56)	47.42 (46)
Black (24)	75.00 (18)	66.67 (16)	62.50 (15)	54.17 (13)	41.67(10)
Hispanic (13)	84.62 (11)	69.23 (9)	53.85 (7)	76.92 (10)	38.46 (5)
Other #	---	---	---	---	---
Completed Education					
Less Than High School (8)	87.50 (7)	75.00 (6)	75.00 (6)	75.00 (6)	62.50 (5)
High School Graduate or GED (28)	82.14 (23)	67.86 (19)	75.00 (21)	67.86 (19)	53.57 (15)
Some College or Vocational Training (48)	77.08 (37)	72.92 (35)	58.33 (28)	64.58 (31)	50.00 (24)
College Graduate (41)	78.05 (32)	63.41 (26)	58.54 (24)	48.78 (20)	34.15 (14)
Graduate Education beyond College (11)	72.73 (8)	72.73 (8)	63.64 (7)	45.45 (5)	36.36 (4)
Employed 10 or More Hours/Week					
No (55)	70.91 (39)	61.82 (34)	56.36 (31)	56.36 (31)	41.82 (23)
Yes (81)	83.95 (68)+	74.07 (60)	67.90 (55)	61.73 (50)	48.15 (39)

Children Under Age 12 in Household					
No (76)	76.32 (58)	67.11 (51)	65.79 (50)	55.26 (42)	44.74 (34)
Yes (60)	81.67 (49)	71.67 (43)	76.67 (46)	65.00 (39)	46.67 (28)
Incentive amount					
$40.00 (78)	79.49 (62)	67.95 (53)	64.10 (50)	62.82 (49)	46.15 (36)
$20.00 (58)	77.59 (45)	70.69 (41)	62.07 (36)	55.17 (32)	44.83 (26)

Note: Denominator for all groups is their total number in the sample
NOTE: number of samples for which timeliness is calculable (denominator) varies by row

Cell sizes too small to report
+ p = .07

Table 4. Cortisol concentrations of received saliva samples from Add Health pretest (n=136) and reliability based on IIV Study (n=27)

Cortisol Sample	Mean (ug/dl)	Std	Min	Max	ICC	CV
Wake	.43	.39	.03	3.00	.06	73.0
+30 min	.70	2.55	0.04	30.00	.25	97.2
Bedtime	.15	.67	.01	7.74	.43	29.6

Acknowledgments:

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). Work by Halpern, Harris, and Whitsel on these analyses was supported by grant P01-HD31921.