

Published in final edited form as:

Psychoneuroendocrinology. 2012 April ; 37(4): 499–508. doi:10.1016/j.psyneuen.2011.07.019.

Challenges of Measuring Diurnal Cortisol Concentrations in a Large Population-Based Field Study

Carolyn Tucker Halpern, PhD^{*,1,4}, Eric A. Whitsel, MD, PhD², Brandon Wagner, MA^{3,4}, and Kathleen Mullan Harris, PhD^{3,4}

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

²Departments of Epidemiology and Medicine, UNC-CH

³Department of Sociology, UNC-CH

⁴Carolina Population Center, UNC-CH

Abstract

Objectives—Longitudinal examinations of associations between daily stress, diurnal cortisol concentrations, and physiological parameters in population-based studies are needed. This study evaluates issues related to consent, collection, and protocol adherence for a low-burden saliva collection protocol.

Methods—In the 2007 pretest (N = 193) for Wave IV of the National Longitudinal Study of Adolescent Health (Add Health) a three-sample, one-day, unsupervised saliva collection protocol was pilot tested. Embedded experiments allowed for examination of adherence and effects of monetary incentives.

Results—Although most (97%) study participants consented to collection, only about 80% actually mailed back samples. Use of a time-stamping TrackCap allowed comparison of self-reported and stamp-recorded collection times. Of returned samples, self-report of collection time was missing for about a quarter, and only about one in three respondents (of those for whom adherence was calculable) fully adhered to the collection protocol, indicating significant potential for bias. Consent, return, and protocol adherence were unrelated to key sociodemographic characteristics, and did not improve with higher monetary incentives or knowledge of being monitored.

Conclusions—Despite the relatively low-burden collection protocol and use of multiple strategies thought to improve collection and protocol adherence, response and adherence were poor, leading to a decision to drop cortisol measurement from the Wave IV Add Health protocol. Large field studies should carefully evaluate the feasibility of collection and protocol adherence

© 2011 Elsevier Ltd. All rights reserved

*Address correspondence to: Carolyn Tucker Halpern Carolina Population Center, CB #8120 University of North Carolina at Chapel Hill Chapel Hill, NC 27599-8120 Telephone: 919-9664462 Fax: 919-966-6638 carolyn_halpern@unc.edu.

Conflict of Interest: None declared

Contributors: Carolyn Tucker Halpern conceptualized the paper, planned analyses, and wrote most text. Eric A. Whitsel collaborated in the interpretation of results and edited sections of the paper. Brandon Wagner collaborated in planning analyses, conducted analyses, and wrote sections of text. Kathleen Mullan Harris is the PI of the parent project, collaborated in the interpretation of results, and edited sections of the paper. All authors collaborated in the original protocol development.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

for unsupervised collection protocols before implementing costly, and potentially unusable, biological measurements.

Keywords

consent; protocol adherence; population-based; USA

Biomarkers are increasingly used in population-based¹ 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 initiates a pathway that ultimately results in cortisol release from the adrenal cortex. 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.

Salivary Cortisol

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 reactivity of the HPA axis” (Fries et al., 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) that can be observed in about 75% of adults. It is typically quantified as the difference between an awakening sample and a sample taken 30 minutes after awakening. 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 positively correlated with reactivity to laboratory stressors and ACTH administration, and changes in CRA have been reported in association with experience of chronic stress. Although the CRA is reported to have high intra-individual stability (Fries et al., 2009; Hucklebridge et al., 2005), CRA measurement

¹By 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.

appears to require exact timing in sample collection (Kunz-Ebrecht et al., 2004; Pruessner et al., 1997), increasing the importance of protocol adherence.

Protocol Non-adherence

There are different types of non-adherence. Respondents may refuse to provide samples, or even if they agree, may not actually produce and return the samples. This yields missing data, which if patterns of missingness are correlated with other relevant respondent characteristics, may bias analyses. Alternatively, respondents may produce the samples at the “wrong” time (i.e., not according to protocol). If respondents accurately report 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.

Non-adherence is problematic because diurnal profiles are different for respondents who do and do not adhere to protocol, and the response to awakening appears to be especially sensitive to non-adherence (Kudielka et al., 2007; Kudielka et al., 2003). Further, one study has demonstrated that adherence is positively correlated with participants' reports of social support, raising 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 (Kudielka et al., 2007).

Saliva Sampling Strategies

A number of methodological studies have been conducted to evaluate optimal saliva sampling strategies for cortisol assessment, different “instructional sets,” and the implications of each for protocol adherence. These studies typically rely on small samples of older adults and employ electronic (time-stamp) monitors in caps (hereafter referred to as TrackCaps) covering containers used to store small tubes or cotton rolls to be used for saliva sample collection. Methodological studies have examined single and multiple day collection protocols with varying numbers of samples collected per day. In general, this body of work has revealed high levels of non-adherence to protocol, ranging across studies from a quarter to 60% of respondents not following protocol on one or more days (Kudielka et al., 2007; Broderick et al., 2004; Kudielka et al., 2003). Work to date also suggests that respondents' knowledge of being monitored (via TrackCaps) improves protocol adherence and the accuracy of self-report. The implications of multi-day collection protocols for protocol adherence are not yet clear. Some studies (e.g., Broderick et al., 2004; 7 day collection protocol) find no statistically significant deterioration across multiple days, whereas other investigators suggest that, given poor compliance in less burdensome (e.g., 3 day) protocols, requesting sample collection across several days would exacerbate adherence problems (Kudielka et al. 2007). **Work by** Hellhammer and colleagues (2007) suggests that single-day collection protocols yield CRAs that largely reflect situational (state) factors and that two or more consecutive weekday collection protocols yield more reliable trait assessments. This study, however, did not appear to monitor protocol adherence.

Several relatively large studies have incorporated non-laboratory saliva collection protocols with some success, including the Coronary Artery Risk Development in Young Adults Study (CARDIA; six saliva samples over the course of one weekday) and the National Study of Daily Experiences (NSDE; the Daily Diary portion of MIDUS II - Midlife in the United States Survey; four saliva samples per day for four days). In CARDIA 58% of eligibles (about 775 respondents) produced samples for analysis (Cohen et al., 2006). In NSDE, 86% of eligibles (1736) provided usable saliva samples. Correlations between self-reported times and time stamps ranged from .75 (evening) to .95 (morning), suggesting excellent protocol adherence in this latter sample of middle-aged and older adults (Almeida

et al., 2009). (See Adam & Kumari, 2009, for an in-depth review of cortisol assessment issues in large-scale studies.)

The present paper describes the evaluation of a three-sample, one-day, post-interview protocol for collecting saliva and assaying cortisol concentrations in the National Longitudinal Study of Adolescent Health (Add Health) Wave IV pretest, which was conducted in 2007. Add Health survey and biomarker data are used by researchers around the world. Add Health uses tiered informed consent for biospecimens. Respondents can: 1) decline to provide specimen, 2) consent to provide specimen only for those assays specified in the consent form (in this case, cortisol), after which the specimen is destroyed, or 3) consent to provide specimen for specified assays and archival of remaining specimen for future assays that fall within the scientific goals of Add Health, but are unspecified in the consent form. A focus of the Wave IV Add Health program project is understanding the interplay between environment, behavior, and biology in their contributions to health trajectories over time. Longitudinal relationships between stress, cortisol, and health are a central component; therefore saliva collection for cortisol assay was planned as one indicator of stress in Wave IV.

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

Methods

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). For more details on the Add Health design, see Harris et al. (2009).

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 is relatively non-burdensome and that incorporated elements thought to improve consent and adherence, based on findings from earlier studies. Due to time constraints, respondents 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, respondents were instructed to collect three samples on a single day: upon awakening, 30 minutes post awakening, and at bedtime. We defined “awakenin” 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 for 30 minutes post-awakening 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 respondents detailed oral and written instructions about the collection protocol, after which respondents practiced an interviewer-supervised sample collection (spitting into a small tube).²

The collection protocol was designed based on the expectations that: (a) it would not be unduly burdensome for respondents, (b) it would maximize participant consent, (c) it would be associated with high adherence to the fixed time-of-day collection protocol (10), and assuming good adherence, (d) 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, respondents were divided into two subgroups to test two incentive amounts. Group 1 was paid \$40.00 for their participation in the cortisol sample collection and Group 2 was paid \$20.00. Both groups were paid by a check mailed after sample receipt. Regarding protocol adherence, respondents were divided into three sub-groups: one third were told that saliva collection times would be monitored for all respondents, one third were told that saliva collection times would be monitored for a randomly selected subset of respondents, and one third were not told about the possibility of collection time monitoring. In actuality, all pretest respondents were monitored via the

²Cortisol pretest data will not be released. However, written instructions given to respondents, as well as checklists respondents completed, are available on the Add Health web page (<http://www.cpc.unc.edu/projects/addhealth/data/guides>).

TrackCap, which allowed comparison of self-reported (on the checklist) and TrackCap-recorded collection times. Knowing that it would not be possible to actually 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), demonstrations of the accuracy of self-report and adherence to the morning protocol were critical.

Analyses

To assess sociodemographic differences in consent and sample collection, we examined saliva sample information and protocol adherence according to a number of respondent characteristics. These are age at the time of the Wave IV pretest interview, biological sex, race/ethnicity, respondent's 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. Chi-Square tests of proportion were calculated to assess whether different sociodemographic characteristics were associated with different success in our response categories (consent, consent to archive, sample return and adherence). As multiple tests were conducted in each response category, a Bon Ferroni correction was applied to all critical values. Significance levels reported in tables represent this correction.

Intra-Individual Variation (IIV) Study

Wave IV also incorporated an intra-individual variation study of 100 Add Health respondents, 43 of whom were interviewed during the pretest. Of the 43 IIV respondents, 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 respondents were seen twice, 1–2 weeks apart (mean = 8.2 days). At Visit 1, respondents 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

Of the 193 respondents participating in the pretest, 52% were female, with a mean age 27.8 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 respondents 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 1 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 respondents agreed to provide saliva samples, yielding a 97% consent rate. One hundred fifty-six respondents (or 83% of those consenting, 81% of interviewed respondents) agreed to have their samples archived after cortisol measurement. However, only 146

respondents actually returned any samples to the lab (78% of consenting and 76% of interviewed pretest respondents).

Likelihood of consent, both for immediate cortisol assay or archiving, 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 had not completed high school were less likely to return samples than those with higher levels of education.

Reminder calls to maximize respondents' sample return—To increase receipt of saliva from respondents who had agreed to provide samples, reminder calls were attempted for respondents whose samples had not been received at the lab within three days post-interview (47% of those consenting to saliva collection). Of 29 respondents 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 three checklists, TrackCaps (which respondents had been instructed to return), and three samples. Of samples received, 25% were missing self-reported collection times.

Table 2 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. Following the work of Kudielka et al. (2003), adherence was defined as collecting saliva according to accuracy windows of ± 10 minutes for sample 1; ± 7 minutes for sample 2; ± 60 minutes for sample 3. Seventy nine percent of respondents completed at least one sample according to protocol (57% of those who originally consented to collection). 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 respondents who had consented to collect samples. Only 46% of the 136 respondents 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). Broadening accuracy windows did not change findings. None of the sociodemographic characteristics examined nor incentive level was associated with adhering to protocol for any of the samples.

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. For respondents who were not missing self-reported times for samples 1 and 2, and who correctly reported collection times for those samples ($n=74$), 73% (54 respondents) 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 3 includes information about the average concentrations of cortisol for each of the three samples for the 122 respondents who returned all three samples and for whom timeliness of collection was calculable. There is little evidence of the CRA, with an average 16% increase over mean waking levels. Reliability (**right-hand panel of Table 3**) was computed **on log-transformed data** as the ratio of the between-respondent to total variance, that is, an intraclass correlation coefficient with 95% confidence intervals computed using the delta method under the assumption of normality (ICC, 95% CI) (Oehlert, 1992). The ICC represents the proportion of variance in cortisol level that is not due to measurement

variance and can be interpreted as the correlation between repeated measurements on the same individual. Short-term reliability of samples for the 27 pretest respondents who participated in the IIV study (and returned samples) was poor, especially for the key waking sample, where the intraclass correlation was essentially zero.

Effects of varying information provided about monitoring

Results of separately conducted bivariate logistic 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., Broderick et al., 2004; Kudielka et al., 2003), we had expected that respondents 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 compared with that used in other field studies, we would see not only high levels of consent, but also good adherence to collection protocol and accuracy in reported sample collection times. Respondents knew they would not be penalized (i.e., receive less incentive money than had been offered to them) for accurately reporting times that did not follow protocol. 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 can be expected to actually return samples. Only education, of the sociodemographic characteristics we examined, was significantly related to sample return. Respondents with low levels of educational attainment were the least likely to collect and return samples (but of note, were not different in terms of protocol adherence). Even increasing the monetary incentive, which would have been prohibitively expensive to implement for the full sample, did not appear to significantly improve levels of sample return.

Consideration of adherence levels is also discouraging. About a quarter of self-reported collection times were missing for the 76% of respondents 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 **slight** 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.³ The estimate of morning cortisol rise, key for Add Health project purposes, is dependent on the wake up sample being collected according to protocol. Although self-reported wake-time has been found to be reasonably accurate, it appears that inaccuracies are predictive of protocol non-adherence (DeSantis et al., 2010) and delays of 15 minutes or more yield misleading CRA results (Dockray et al., 2008). Of consenting respondents **in the Add Health pretest** 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.,

³Assay performance from our lab (Salimetrics) was excellent and was not a factor in poor intra-individual reliability.

supplying timers that would alarm when it was time to collect sample #2) to enhance protocol adherence. However, we did follow other successful strategies to the extent possible, such as using numbered and color-coded collection tubes, a detailed instruction sheet, and the collection of a practice sample with the field interviewer present. For reasons of cost and logistics we were not able to have interviewers check in with respondents by phone on a daily basis, as was done in the NSDE. However, our use of labor intensive reminder calls after a period of non-receipt, which would be prohibitively expensive for the full Add Health sample, yielded relatively little return.

Work in earlier studies has demonstrated that diurnal profiles are different for respondents who do and do not adhere to protocol, and that the response to awakening is especially sensitive to non-adherence. Non-adherent participants show a much smaller cortisol increase, and without information on adherence, their CRA pattern would likely be erroneously labeled as “blunted” (Broderick et al., 2004; Kudielka et al., 2003). Similarly, change in cortisol in the across-the-day slope is significantly greater for adherent samples compared with non-adherent samples (Broderick et al., 2004). Given the high levels of non-adherence in our pretest sample, and their implications for data quality and interpretation, we ultimately decided not to include saliva collection for cortisol assay in the protocol for Wave IV main field work.

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 directly 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 in the NSDE.

To our knowledge, only one other study (Hall et al., 2011) has examined characteristics of participants who do and do not adhere to collection protocols. In contrast to our findings of little association between sociodemographic factors and protocol adherence, Hall et al., found, in a select sample of 262 breast cancer survivors, that African-American and fatigued respondents were less likely to adhere to protocols, though almost half the sample had “high adherence” (80% or more of samples within 15 to 45 minutes of protocol). Differences in findings across studies may be due to design differences. In addition to sample selectivity, the Hall et al. study is limited by exclusive reliance on self-reported times (TrackCap data were collected but not used in analyses because many respondents failed to use the container as directed) and failure to collect awakening time (to compare with time of collection of first sample).

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 adherence is more a reflection of a “day-level” rather than “person-level” problem (Thorn et al., 2006). That is, circumstances of a given day may have contributed to poor adherence. There is some empirical support for this; Broderick et al. (2004) found that more than half of their sample had both adherent and non-adherent days. **Although not directly assessed**, Hellhammer et al.'s (2007) findings of greater reliability of “trait-based” data derived from multiple sampling days also suggest that multiple days of collection increase the likelihood

of seeing “adherent” days. Thus, 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 a greater number of 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 had only 27 pairs of samples for analysis. Thus our estimates may be less reliable than those based on larger samples; findings regarding adherence, however, may not differ according to sample size. 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. It is possible that collecting samples across multiple days could have yielded one or more days with better adherence, and therefore a larger number of usable samples. There is currently no consensus on the optimal number of days to collect samples in unsupervised field contexts. We speculate that the optimal protocol to achieve reliable diurnal profiles may vary, depending on sample characteristics (e.g., age, disease status, etc.) and study resources relative to sample size (e.g., ability to follow up with reminders and/or to provide external collection prompts such as programmed timers).

Finally, the initial awakening sample is crucial to quantify the CRA. Use of track-cap technology, although helpful, is expensive as well as somewhat crude and incomplete. We cannot definitely determine which opening times did or did not correspond to actual 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. 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 respondent diversity and protocol variation (e.g., number of samples and sampling days) is needed to determine which protocols are best suited to individuals in different contexts and in different stages of the life course. Technology to objectively assess the temporal sequence of sampling is also needed to evaluate experimental findings. As others have noted (e.g., Almeida et al., 2009), 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 et al., 2001). However, even under ideal circumstances substantial effort will be required to reduce potential for protocol non-adherence.

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 (NICHD), 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 NICHD grant P01-HD31921.

Roles of Funding Source: NICHD had no role in study design, data collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

References

- Adam EK, Kumari M. Assessing salivary cortisol in large-scale, epidemiological research. *Psychoneuroendocrinology*. 2009; 34:1423–1436. [PubMed: 19647372]
- Almeida DM, McGonagle K. Assessing daily stress processes in social surveys by combining stressor exposure and cortisol. *Biodemography Soc Biol*. 2009; 55(2):219–237. [PubMed: 20183906]
- Broderick JE, Arnold D, Kudielka BM, Kirschbaum C. Salivary cortisol sampling compliance: Comparison of patients and healthy volunteers. *Psychoneuroendocrinology*. 2004; 29:636–650. [PubMed: 15041086]
- Cohen S, Schwartz JE, Epel E, Kirschbaum C, Sidney S, Seeman T. Socioeconomic status, race, and diurnal cortisol decline in the Coronary Artery Risk Development in Young Adults (CARDIA) study. *Psychosom Med*. 2006; 6:41–50. [PubMed: 16449410]
- DeSantis AS, Adam EK, Mendelsohn KA, Doane LD. Concordance between Self-Reported and Objective Wakeup Times in Ambulatory Salivary Cortisol Research. *Int. J. Behav. Med*. 2010; 17:74–78. [PubMed: 19597997]
- Dockray S, Bhattacharyya MR, Molloy GJ, Steptoe A. The cortisol awakening response in relation to objective and subjective measures of waking in the morning. *Psychoneuroendocrinology*. 2008; 33:77–82. [PubMed: 17996380]
- Eissa MA, Poffenbarger T, Portman RJ. Comparison of the actigraph versus patients' diary information in defining circadian time periods for analyzing ambulatory blood pressure monitoring data. *Blood Press Monit*. 2001; 6(1):21–25. [PubMed: 11248757]
- Fries E, Dettenborn L, Kirschbaum C. The cortisol awakening response (CAR): Facts and future directions. *Int J Psychophysiol*. 2009; 72:67–73. [PubMed: 18854200]
- Gunnar, MR.; Vazquez, D. Stress neurobiology and developmental psychopathology. In: Cicchetti, D.; Cohen, D., editors. *Developmental psychopathology: Developmental neuroscience*. Vol. vol. 2. Wiley; New York: 2006. p. 533-577.
- Hall DL, Blyler D, Allen D, Mishel MH, Creandell J, Germino BB, Porter LS. Predictors and patterns of participant adherence to a cortisol collection protocol. *Psychoneuroendocrinology*. 2011; 36:540–546. [PubMed: 20863625]

- Harris, KM.; Halpern, CT.; Whitsel, E.; Hussey, J.; Tabor, J.; Entzel, P.; Udry, JR. The National Longitudinal Study of Adolescent Health: Research Design. 2009. [WWW document]. URL: <http://www.cpc.unc.edu/projects/addhealth/design>
- Hellhammer J, Fries E, Schweisthal OW, Schlotz W, Stone AA, Hagemann D. Several daily measurements are necessary to reliably assess the cortisol rise after awakening: state- and trait components. *Psychoneuroendocrinology*. 2007; 32(1):80–86. [PubMed: 17127010]
- Hucklebridge F, Hussain T, Evans P, Clow A. The diurnal patterns of the adrenal steroids cortisol and dehydroepiandrosterone (DHEA) in relation to awakening. *Psychoneuroendocrinology*. 2005; 30:51–57. [PubMed: 15358442]
- Jacobs N, Nicolson NA, Derom C, Delespaul P, van Os J, Myin-Germeys I. Electronic monitoring of salivary cortisol sampling compliance in daily life. *Life Sci*. 2005; 76:2431–2443. [PubMed: 15763075]
- Kirschbaum C, Hellhammer DH. Salivary cortisol in psychoneuroendocrine research: Recent developments and applications. *Psychoneuroendocrinology*. 1994; 19:313–333. [PubMed: 8047637]
- Kudielka BM, Broderick JE, Kirschbaum C. Compliance with saliva sampling protocols: Electronic monitoring reveals invalid cortisol daytime profiles in non-compliant subjects. *Psychosom Med*. 2003; 65:313–319. [PubMed: 12652000]
- Kudielka BM, Hawkley LC, Adam EK, Cacioppo JT. Compliance with ambulatory saliva sampling in the Chicago Health, Aging, and Social Relations Study and associations with social support. *Ann Behav Med*. 2007; 34(2):209–216. [PubMed: 17927559]
- Kunz-Ebrecht SR, Kirschbaum C, Marmot M, Steptoe A. Differences in cortisol awakening response on work days and weekends in women and men from the Whitehall II cohort. *Psychoneuroendocrinology*. 2004; 29(4):516–528. [PubMed: 14749096]
- McEwen BS, Stellar E. Stress and the individual: Mechanisms leading to disease. *Arch Intern Med*. 1993; 153:2093–2101. [PubMed: 8379800]
- National Research Council. Committee on Advances in Collecting and Utilizing Biological Indicators and Genetic Information in Social Science Surveys. In: Weinstein, M.; Vaupel, JW.; Wachter, KW., editors. *Biosocial Surveys*. National Academies Press; Washington, DC: 2008.
- Oehlert GW. A note on the delta method. *Am Stat*. 1992; 46(1):27–29.
- Pruessner JC, Wolf OT, Hellhammer DH, Buske-Kirschbaum A, von Auer K, Jobst S, Kaspers F, Kirschbaum C. Free cortisol levels after awakening: A reliable biomarker for the assessment of adrenocortical activity. *Life Sci*. 1997; 61:2539–2549. [PubMed: 9416776]
- Smyth J, Ockenfels MC, Porter L, Kirschbaum C, Hellhammer DH, Stone AA. Stressors and mood measured on a momentary basis are associated with salivary cortisol secretion. *Psychoneuroendocrinology*. 1998; 23:353–370. [PubMed: 9695136]
- Thorn L, Hucklebridge F, Evans P, Clow A. Suspected non-adherence and weekend versus week day differences in the awakening cortisol response. *Psychoneuroendocrinology*. 2006; 31:1009–1018. [PubMed: 16876958]

Table 1

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

Respondent Characteristics	Consent	Consent to Archive	Any Samples Returned
	Percentage (n)	Percentage (n)	Percentage (n)
	97% (188)	81% (156)	76% (146)
Age			
Mean (SE)	27.7 (0.15)	27.6 (0.16)	27.7 (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 ^a	---	---	---
Completed Education			
Less Than High School (ref)	100 (20)	79 (14)	40 (8) ^b
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.

^a Cell sizes too small to report

^b $p < 0.05$

Table 2

Percentages of sociodemographic groups with varying levels of saliva collection protocol adherence (of 136 respondents who returned samples and 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)	(n)	% (n)	(n)	% (n)	(n)
<u>Characteristic (number available samples)</u>										
Age (years)	78.7%	(107)	69.1%	(94)	63.2%	(86)	59.6%	(81)	45.6%	(62)
Mean (SE)	27.7	(0.20)	27.7	(0.22)	27.5	(0.24)	27.8	(0.24)	27.6	(0.28)
Biological Sex										
Male (60)	75.0	(45)	63.3	(38)	61.7	(37)	53.3	(32)	40.0	(24)
Female (76)	81.6	(62)	73.7	(56)	64.5	(49)	64.5	(49)	50.0	(38)
Race/Ethnicity										
White (97)	78.4	(76)	69.1	(67)	65.0	(63)	57.7	(56)	47.4	(46)
Black (24)	75.0	(18)	66.7	(16)	62.5	(15)	54.2	(13)	41.7	(10)
Hispanic (13)	84.6	(11)	69.2	(9)	53.9	(7)	76.9	(10)	38.5	(5)
Other ^a	---	---	---	---	---	---	---	---	---	---
Completed Education										
Less Than High School (8)	87.5	(7)	75.0	(6)	75.0	(6)	75.0	(6)	62.5	(5)
High School Graduate or GED (28)	82.1	(23)	67.9	(19)	75.0	(21)	67.9	(19)	53.6	(15)
Some College/Vocational Training (48)	77.1	(37)	72.9	(35)	58.3	(28)	64.6	(31)	50.0	(24)
College Graduate (41)	78.1	(32)	63.4	(26)	58.5	(24)	48.8	(20)	34.2	(14)
Education beyond College (11)	72.7	(8)	72.7	(8)	63.6	(7)	45.5	(5)	36.4	(4)
Employed 10 or More Hours/Week										
No (55)	70.9	(39)	61.8	(34)	56.4	(31)	56.4	(31)	41.8	(23)
Yes (81)	84.0	(68)	74.1	(60)	67.9	(55)	61.7	(50)	48.2	(39)
Children Under Age 12 in Household										
No (76)	76.3	(58)	67.1	(51)	65.8	(50)	55.3	(42)	44.7	(34)
Yes (60)	81.7	(49)	71.7	(43)	76.7	(46)	65.0	(39)	46.7	(28)
Incentive amount										
\$40.00 (78)	79.5	(62)	68.0	(53)	64.1	(50)	62.8	(49)	46.2	(36)

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)	
\$20.00 (58)	77.6	(45)	70.7	(41)	62.1	(36)	55.2	(32)	44.8	(26)

Denominator for all groups is their total number in the sample.

Number of samples for which timeliness is calculable (denominator) varies by row.

^a Cell sizes too small to report

Table 3

Cortisol concentrations of received saliva samples from Add Health pretest (n=122)^J and reliability based on IIV Study (n=27)

Cortisol Sample	Mean (ug/dl)	Std	Min	P25	P75	Max	IIV	
							ICC	CV
Wake	.43	.32	.11	.27	.49	3.00	.06	73.0
+30 min	.50	.25	.12	.32	.63	1.42	.25	97.2
Bedtime	.09	.13	.01	.03	.10	.76	.43	29.6

^J Calculations are based on data from respondents who returned all three requested samples, and for whom timeliness of collection was calculable. Values $\leq .1$ or ≥ 5.0 were excluded from analyses. Calculations based on more observations (i.e., data not subject to these exclusion criteria) yielded similar patterns.