Published in final edited form as:
Res Gerontol Nurs. 2014 March ; 7(2): 56–65. doi:10.3928/19404921-20131018-01.

Biobehavioral Measures as Outcomes: A Cautionary Tale

Christine R. Kovach, PhD, RN, FAAN [Professor],
University of Wisconsin-Milwaukee, Milwaukee, Wisconsin

Diana Lynn Woods, PhD, RN, APRN-BC, FGSA [Associate Professor],
Azusa Pacific University School of Nursing, Azusa, CA 91702-7000

Elizabeth C. Devine, PhD, RN [Professor Emeritus],
University of Wisconsin-Milwaukee, Milwaukee, Wisconsin

Brent R. Logan, PhD [Professor], and
Medical College of Wisconsin, Milwaukee, Wisconsin

Hershel Raff, PhD [Professor of Medicine, Surgery, and Physiology]
Medical College of Wisconsin Endocrine Research Laboratory, Aurora St. Luke's Medical Center, Aurora Research Foundation Milwaukee, Wisconsin

Diana Lynn Woods: dwoods@apu.edu

Abstract

This article discusses the use of biobehavioral measures as outcomes for healthcare intervention studies. Effect size (ES) values for salivary cortisol, and observation-based measures of pain and agitation are examined. Effects pre to post treatment were assessed separately for nursing home (NH) residents with and without acute psychotic symptoms. This study revealed large positive effects on both pain and agitation measures in the group with acute psychotic symptoms and small-to-medium positive effects on these same measures in the group without acute psychotic symptoms. In both of these groups the ES values were not consistently positive on the cortisol measures. Prior to determining if a measure can be used to estimate minimum clinically important differences, it is essential to consider if the biomarker will be responsive to therapy in the populations and contexts being studied.

Keywords

aging; dementia; stress; endocrine system

In recent years there has been increased emphasis on including biological measures in health care research and in evaluating patient outcomes. The National Institute for Nursing Research (NINR) has emphasized the need for more biobehavioral research (https://www.ninr.nih.gov/sites/www.ninr.nih.gov/files/ninr-strategic-plan-2011.pdf) and the Institute of Medicine (IOM) of the National Academies (IOM, 2010) released a report discussing the issues associated with using biological measures as surrogate endpoints for patient outcomes. Biobehavioral research is grounded in the premise that understanding and influencing health outcomes is enhanced through examining the links between biological, psychosocial, and behavioral factors and health status/outcomes (Pellmar, Brandt, & Baird, 2002). This is a worthy goal; however, the success of biobehavioral science in making new

Correspondence concerning this article should be addressed to: Christine R. Kovach, College of Nursing, University of Wisconsin Milwaukee, 1921 E Hartford Ave, Milwaukee, WI 53201-0413. ckovach@uwm.edu.

The authors declare no conflict of interest.

inroads that influence health outcomes is dependent on understanding and attending to the complexities of a host of specialized conceptual and methodological challenges associated with using biological measures in research.

As noted in the IOM report cited above, consistent scientific processes and frameworks are needed in order to ensure the rigorous and transparent use of biological measures. In addition to possibly being invasive and costly, given the state of the science, the selection and interpretation of biological measures must be done with caution. For some biological measures we are lacking strong evidence on the natural history of variability over time and on predictors of variability in the measure. There is a need to build a strong research base on the ability of biological measures to capture meaningful change and to determine minimum clinically important differences (MCID). The purpose of this article is to explore some issues regarding the use of biological measures as outcomes in nursing studies. Data from a study testing the effects of a nurse assessment and treatment protocol used to treat nursing home (NH) residents with dementia are presented (Kovach, Logan, et al., 2006). The purpose of these results is not to look for treatment effects; rather it is to highlight the specific methodological issue of measurement responsiveness to therapy. Effect sizes (ES) for salivary cortisol, observer-rated pain and agitation are presented for two groups of NH residents who were being assessed and treated for new problems over six weeks. For this secondary analysis we compared NH residents who developed new acute psychotic symptoms to NH residents who developed other new problems (e.g. arthrtitic pain, constipation, infection). Because acute psychosis is associated with high stress, these two groups provided an opportunity to examine outcomes following treatment in two different populations.

MCID AND RESPONSIVENESS TO CHANGE

The determination of MCID is needed for evaluating the clinical meaningfulness of statistically significant research findings. MCID is defined as an estimation of the smallest difference in a measurable clinical parameter that indicates a meaningful change in a health care outcome (Kiley, Sri Ram, Croxton, & Weinmann, 2005). The development and use of standardized MCIDs will increase the evidence-base for nursing practice, assist in making decisions regarding resource allocation, and allow clearer interpretation of research findings.

While MCIDs have value, there are notable challenges to developing metrics for MCIDs (Gilbert, Brown, Cappelleri, Carlsson, & McKenna, 2009; Jordon, Dunn, Lewis & Croft, 2006). Even with fairly well-accepted biological measures, it is possible that as science develops we will learn that a measure may be able to detect change under some conditions, but not under other conditions. Therefore it is important to consider theory, previous research, and the expected change from treatment when deciding if it is reasonable to select a particular biological measure for a specific study.

An essential quality of the measure used to develop the MCID, is the ability to accurately detect health change in response to a specific treatment or stimuli. The measure should validly represent the intended construct. While determining this is not trivial, it can be particularly challenging for biological measures which may have well-established measurement validity for the biological phenomenon being measured (e.g., it validly measures the serum level of a hormone) but which may, or may not have, well-established measurement validity for the intended outcome construct of the study (e.g., stress). In addition, the outcome measure used should remain relatively unchanged if the treatment is not administered (i.e., it must have test-retest reliability), there must be room for change on the measure (e.g., it should not have a floor or ceiling effect), and the timing of measurement must be appropriate for the onset and duration of the intervention’s effect.

Once a measure is deemed reasonable, pilot work can estimate the direction and magnitude of changes on the measure following the proposed treatment or stimuli condition. This estimate can be used in ad hoc power analysis and to determine if the estimated amount of change is clinically significant.

A statistically significant change in group scores following treatment on a biomarker does not necessarily indicate an important change in health or quality of life. In order for a biomarker to qualify as a MCID, the change in score needs to relate to other measures (Jordon et al., 2006). For example, in people with dementia, these might include family report of improvement or changes in psychotropic drug use. The change in the biomarker should also not have a negative consequence on the overall health status of the individual. For example ezetimibe significantly lowers cholesterol but is related to severe hepatic side effects (Stolk, Becx, Kuypers & Seldenrijk, 2006).

BACKGROUND TO STUDY OF PEOPLE WITH ADVANCED DEMENTIA

The Serial Trial Intervention (STI) is a decision support tool to address the problem of under assessment and treatment of newly emerging problems of nursing home residents with advanced dementia that are unable to clearly or consistently verbally report symptoms. The estimates of treatment effect used in this study are based on data from a parent study that compared the effects of two versions of a nursing assessment and treatment protocol (5-step and 9-step Serial Trial Intervention) on stress, pain, and agitation among NH residents with advanced dementia (Kovach et al., 2012).

The 5-step version of the STI was tested for the outcomes of pain and agitation and found to be effective when compared to standard care (Kovach, Kelber, Simpson, & Wells, 2006). However, it was also found that thoughtful follow through was often lacking as evidenced by failure to get effective treatments scheduled for regular use and failure to add adjunctive and preventive therapies to the therapeutic regimen (Kovach, Cashin, et al., 2006). The 9-step version attempted to address these deficits by adding steps that directed nurses to provide more long-term and more comprehensive treatments and to perform more thorough evaluations (Kovach et al., 2012).

The physical condition and functional limitations of people with advanced dementia pose challenges to measurement of variables of interest in nursing research. People with advanced dementia have communication and cognitive deficits that preclude the use of most self-report measures. Proxy reports from family and professional caregivers on measures such as pain are highly biased. For example, multiple studies have found that family and professional caregivers grossly underestimate pain and depression in people with dementia (Horgas & Dunn, 2001; Cohen-Mansfield & Lipson, 2002). Also, invasive measures or those that involve application of sensors to the body may induce a response that confounds the measurement (Aung et al., 2008).

The use of physiological variables to measure indices of stress and depression has been driven, in part, by technological advances that enable the noninvasive measurement of some biological measures. Cortisol can be easily accessed by saliva collection that can be accomplished in everyday contexts with minimal interruption to the natural flow of activities. This is an important consideration for debilitated NH residents. Several studies have demonstrated a strong correlation (r = .805) between time-matched salivary and plasma cortisol (Hellhammer, Wust, & Kudielka, 2009; Kirschbaum & Hellhammer, 1994; Pruessner et al., 1997). Saliva collection has been shown to be well tolerated by NH residents (Woods et al., 2008). The acceptability of saliva collection is important so that cortisol levels are not confounded by an induced stress response from sample collection.

These samples are also stable for up to 2 weeks at room temperature (Malamud, 1992), thus eliminating the need for immediate freezing.

Key conceptual issues in attempting to incorporate cortisol as an outcome measure into intervention studies for intervention studies include justifications for : a) cortisol as a marker of physical and emotional stress; b) the ability of people with dementia to experience stress from a threat or challenge; and c) the relationship of treatment of the threat or challenge to lower stress. While salivary cortisol has been used as a measure in psychophysiological research for the past 20 years (Kirschbaum & Hellhammer, 1989), this measure has begun to be used more recently in nursing and dementia care research (Hanrahan, McCarthy, Kleiber, Lutgendorf, & Tsalikian, 2006; Herrington, Olomu, & Geller, 2004; Williams, Hagerty, & Brooks, 2004; Woods & Dimond, 2002; Woods & Martin, 2007). Salivary cortisol is frequently used as a biomarker of psychological stress (Piazza, Almeida, Dmitrieva, & Klein, 2010) and measurement of cortisol in saliva provides a reliable marker of stress resulting from both physical discomfort and emotional stress (Chrousos, 2011; Weibel, 2003; Young, Abelson, & Lightman, 2004).

Stress is defined as a physiological response of the body to situations or stimulus that are perceived as dangerous or threatening. People with dementia have a decreased threshold for tolerating such stressors (Lawton, 1986; Hall & Buckwalter, 1987). Multiple studies have shown an association between stress, emotional arousal and salivary cortisol levels (Blair et al, 2008; El-Sheikh, Erath, Buckhalt Granger, & Mize, 2008; Fortunato, Dribin, Granger, & Buss, 2008). Age can influence cortsiol values. The most consistent finding with advanced age is an increase in the nighttime nadir which flattens the slope (Ice, 2005; Fiocco, Wan, Weekes, Pim, & Lupien, 2006; Raff et al, 1999; Smyth et al, 1997). Treatment of physical and emotional stress is associated with lowering of cortisol level (Carlson, Speca, Patel & Goodey, 2004; Gaab, Blattler, Menzi, Pabst, Stoyer & Ehlert, 2003; Woods & Dimond, 2002).

METHODS

Design and Sample

Study participants for this secondary analysis were from twelve NHs, all of which had a typical schedule of early awakening (i.e. 5:00 a.m. to 8:00 a.m.) and early bedtime (5:40 p.m. to 8:00 p.m.). The inclusion criteria from the parent study were that the person had a diagnosis of a dementing illness, had no other chronic psychiatric diagnosis, had no known diseases of the hypothalamic-pituitary-adrenal axis, was not taking a corticosteroid, had a pre-study length of stay in the NH of at least 4 weeks, and had no diagnosed acute illness at pretesting. Participants from the two study conditions were combined because the direction and magnitude of effects for the two-week time period examined in the data used for this study were similar across the two versions of the protocol.

The study was approved by the Institutional Review Board. Written consent was obtained from the NH resident’s guardian and verbal assent was obtained from each participant asked to participate. While all participants provided assent, five participants clearly refused or expressed displeasure regarding saliva collection. Samples were not collected when a participant refused or expressed any sign of displeasure regarding the data collection.

It is worth noting that as is typical for NH residents, many patients were being treated for common conditions that included infection (e.g., of the skin, urinary tract, respiratory tract, or gastrointestinal tract), pain (e.g., musculoskeletal or neuropathic), mobility or body alignment problems, skin breakdown, constipation, leg swelling from venous insufficiency, impacted cerumen, poor dentition, and dysphagia. For the current study two samples were

identified: participants being treated for new acute problems without acute psychosis (n = 95) and participants who were being treated for new acute psychotic symptoms (n = 16).

Thirty-one NH residents from the parent study were excluded from this study: 10 were in the dying process, 9 were hospitalized for acute illness (e.g. stroke, bradycardia, heart failure, unresponsive episode), 6 had saliva containing exogenous hydrocortisone documented by liquid chromatograph/tandem mass spectrometry (Raff & Singh, 2012), and 6 had missing saliva samples due to collection refusal (n =5) or quantity not sufficient (n =1).

Measures and Procedures

Salivary cortisol was used to measure HPA axis response to stress (Raff, 2000). The Wisconsin Agitation Intensity (WAI) Scale, used to measure agitated behavior, is an observational scale that uses number, duration and intensity of behaviors as parameters for agitation severity. Possible scores are 0 to 100. Interrater reliability using this tool is .95–.98 and the scale demonstrated responsiveness to change following intervention (Kovach et al., 2004). Pain was assessed using the 9-item Discomfort-DAT scale (Hurley, Volicer, Hanrahan, Houde, & Volicer, 1992). Possible scores are 0 to 75. It is an observational tool that assesses overall level of discomfort rather than pain in response to acute stimuli. Internal consistency alpha coefficients have been reported between .86 and .89 (Hurley et al., 1992) and interrater reliability of .9 (Kovach, Noonan et al., 2006). In the latter study, which was a randomized experiment, the tool demonstrated responsiveness to change following intervention.

Saliva and observation-based measures of agitation and pain were collected on the same weekday, both before the nurses began using the assessment and treatment protocol and two weeks after treatment was initiated. Three nurse research assistants, not employed by the nursing homes, were trained by the principal investigator to screen for eligibility, to conduct unobtrusive observations of agitation and pain, and to collect the saliva samples. Data collectors were blinded to study conditions.

Measures of pain and agitation were collected at least 30 minutes past the time of any potentially discomfort or stress-producing event (e.g., bath, medical exam). Since agitation typically fluctuates over the day, multiple measurements at various time points are recommended to provide a reasonable measure of overall agitation (Kovach et al., 2004). Eight agitation measures were taken per day (2 during breakfast, midmorning, before and after dinner). To capture possible differences in overall level of discomfort in the morning and afternoon, two Discomfort-DAT measures were taken per day. To derive an overall assessment of daily pain and agitation, scores were averaged to yield one pretest and one posttest score on each measure for each study participant. Evidence supports the use of a similar physiological stimulus such as food intake during regular mealtimes to stimulate the HPA axis and control for some intraindividual variation in diurnal rhythm (Gibson et al., 1999; Rosmand, Holm, & Bjorntorp, 2000). It is also recommended that saliva for cortisol assay be obtained close to awakening and bedtime to capture diurnal variation in cortisol levels (Stone et al., 2001). Because residents of many nursing homes retire to bed early in the evening, 45 minutes after dinner was a feasible data collection time that is close to nadir. Four saliva samples were collected (±15 minutes) from under the tongue: 30 minutes after waking (T1-waking), 45 minutes after breakfast (T2-morning), and 45 minutes before (T3-afternoon) and after dinner (T4-evening). Following collection, samples were inserted into a cryogenic vial or storage tube, centrifuged, batched by subject, and frozen at −20°C.

Saliva samples were analyzed using an FDA-cleared cortisol ELISA (Salimetrics, LLC, State College, PA; Raff, Homar, & Skoner, 2003). Samples from each subject were assayed in one batch. The intraassay imprecision, expressed as a coefficient of variation, (CV) was

5.2% at 3.1 (SD, 0.2) nmol/L (n = 10) and 2.6% at 10.4 (0.3) nmol/L (n = 10). Intrassay imprecision is the analytical error for samples analyzed within a single run of the assay. In the current study, samples from each subject were analyzed within one run to minimize interassay imprecision described below. It is typical of immunoassays to have higher CVs at the lowest concentrations of the analyte measured. It is important to note that an intraassay CV of <6% is excellent and provides highly reliable, objective results. Interassay (total) imprecision (CV) was 11% at 2.8 (SD = 0.3) nmol/L (n = 10), 11% at 10.1 (SD = 1.1) nmol/L (n = 10), and 6.9% at 25.0 (SD = 1.7) nmol/L (n = 10) (Raff et al., 2003). The interassay imprecision accounts for the variability from run to run of the assay. Note, in this case, that the interassay imprecision was ≤ 11%, which is also excellent for an immunoassay of this type. The measurement of salivary cortisol was done objectively as the assay technician was blinded to the sample identifier. There are several pre-analytic and analytic factors that can confound the measurement of salivary cortisol (Raff, 2000; Raff, 2004; Raff, 2012; Raff, 2013). They include inter-subject biological variability, incorrect sampling time, contamination of the saliva sample with topical hydrocortisone as described elsewhere in this paper, concomitant acute stress unrelated to the study design, and improper handling of the sample. These can all potentially contribute to variability over and above the inherent analytic variability of the assay itself.

The effect size (ES) value used in this study is based on Cohen's (1969) statistic delta (the standardized mean difference). For statistic delta, Cohen defines ES values as small (0.20), medium (0.50), and large (0.80). To facilitate the discussion of ES values, in this report ES values will be referred to as negligible (≤ 0.19), small (0.20–0.30), small-to-medium (0.31–0.49), medium (0.50–0.79), and large (0.80 or larger).

Analysis

Cortisol measures used for analysis included waking cortisol (T1), morning cortisol (T2), afternoon cortsiol (T3), evening cortisol (T4). Slope was calculated using least squares curve to the four data points for each participant (Adam & Kumari, 2009), and area under the curve (AUC) was calculated with respect to ground using the trapezoidal rule and the four data points for each participant (Fekedulegn et al., 2007). Prior to calculating ES values, log transformations were completed for cortisol levels, T1 – T4, to normalize the data for linear analysis. Pain and agitation scores were normally distributed. ES values were calculated for each variable by dividing the mean difference scores (i.e., the posttest mean minus pretest mean) by the pooled, within-group standard deviation. The order of subtraction was determined in order to yield a positive value when, on average, there was less of the attribute after the intervention (e.g., less stress, anxiety, or pain) and a negative value when the reverse was true. Two sample t-tests were used to compare the mean difference scores between patients with and without acute psychotic symptoms.

Nine subjects began to receive a benzodiazepine medication between the pre- and post-test measure. Benzodiazepines suppress the basal and stress-related activation of the HPA system (Grottoli et al., 2002). Because ESs for the five subjects in the acute psychotic symptoms group and four subjects in the group without acute psychotic symptoms were substantially different than for those that did not receive a benzodiazepine, these nine subjects were dropped from the study. Using mixed model regression analysis (Davis, 2002), the possible confounding effect of other factors (including Mini Mental State Exam score, wake time, the use of narcotics, and the use of 5 or 9-step STI) were examined and no interactions with time were found, indicating that none of these variables significantly affected the pre-post ES.

RESULTS

A total of 102 NH residents were included in the final analysis. Eleven of the 102 were being treated for acute psychotic symptoms. Subjects in the both groups with and without acute psychotic symptoms had typical NH demographics; they were female (78 and 83%), very old (Mean = 86 (SD = 7.7) and 89 (SD = 5.0), and severely demented (median MMSE 5 and 10).

Agitation and pain were the two behavioral measures used in this study. For both of these behavioral measures, the effect of the treatment was large for the group with acute psychotic symptoms (ES = 1.58 for agitation and 1.89 for pain) while the effect on these same measures was small-to-medium for the group without acute psychotic symptoms (ES = 0.33 for agitation and 0.42 for pain) (See Table 1). The difference in the magnitude of the treatment effect between those with and without acute psychotic symptoms is noteworthy (e.g., 1.58 versus .33 on agitation), and the mean differences are statistically significant (for agitation p = .001 and for pain p = .005). In contrast, for the six cortisol measures, the treatment effect was variable with a high ES value of 0.95 and a low ES value of −0.19 for those with acute psychotic symptoms and a high ES value of 0.31 and a low ES value of −0.14 in those without acute psychotic symptoms. Only the mean change in waking cortisol was statistically significantly different between those with and those without acute psychotic symptoms (p = .022).

For area under the curve (AUC), a global measure of the cortisol response (Pruessner, Kirschbaum, Meinlschmid, & Hellhammer, 2003; Stewart & Seeman, 2000), the effect of the treatment was large in those with acute psychotic symptoms (ES = 0.95), but a negligible effect in those without acute psychotic symptoms (ES = 0.07). For the group with acute psychotic symptoms, there is a small-to-medium ES in waking cortisol, which decreased following treatment (ES = 0.46). This dampening of the waking cortisol flattened the slope (ES = −0.19). For those without acute psychotic symptoms, the treatment yielded a negligible increase in waking cortisol level (ES = − 0.14), and a small decrease at T4, the evening cortisol (ES = 0.21). These changes increased the slope, yielding an ES for slope that is small-to-medium (ES = 0.31).

DISCUSSION

This study revealed large positive effects on both observational measures in the group with acute psychotic symptoms and small-to-medium positive effects on these same measures in the group without acute psychotic symptoms. In both of these groups the ES values were not consistently positive on the cortisol measures. The mean differences in the change from pre to post are statistically significantly different between those with and without acute psychosis on agitation, pain, and wakening cortisol. These findings underscore questions about the complexity inherent in using hormonal measures as outcome variables for any intervention. Especially for older adults, between-individual heterogeneity is high resulting in a decreased ability to determine intervention ES with smaller sample sizes (Smyth et al., 1997; Stone et al., 2001). The higher ESs for the acute psychotic group could, in part, be explained by (a) there being greater room for change in a group likely to be more stressed due to their acute psychosis, (b) statistical regression toward the mean, or (c) error as a result of selection bias from a small sample. Statistical regression toward the mean refers to the tendency for higher scores to have a more positive random error that can push the score up at the pre-test, and this random error tends to be less extreme on later measures (Shadish, Cook, & Campbell, 2002). The difference also could be due to a floor effect. Participants without acute psychotic symptoms could have habituated to the stress of their chronic conditions and thus have less room for change on a measure of physiologic stress.

Measurement error could have influenced ESs, though consistent procedures were used for observations and to collect saliva samples in the same manner and from the same location in the mouth. It is noteworthy that measures were taken only during the daytime and nighttime treatment response (e.g. decreased arthritic pain or hallucinations at night) may have been more sensitive to cortisol changes in the daytime. Cortisol levels can also be impacted by multiple non-HPA axis factors. In this study we found and controlled for the confounding effect of benzodiazepine use. We tested and did not find that cognitive exam score, wake time, the use of narcotics, and the use of 5 or 9-step STI were confounding influences. It is possible that other factors in the nursing home or within the individual, such as environmental stressors, caregivers demeanor, or other medications, could have acted as confounding variables and influenced the ESs obtained.

The current study found a large ES using AUC in those with acute psychotic symptoms compared to those without acute psychotic symptoms. Interpreting these results is challenging, given that AUC is a measure of total cortisol secreted throughout the day.

The clinical significance of the ESs for the differences in slopes must be interpreted with caution. While it is noteworthy that the slopes for the groups with and without acute psychotic symptoms differ in direction, it is essential to note that the mean differences in slopes are minimal and the standard deviations are extremely small (Table 1). When standard deviations are very small, relatively modest mean difference can yield large ES values. Cortisol slope has been used most extensively to ascertain the effect of an intervention on stress (Sephton, Sapolsky, Kraemer, & Spiegel, 2000), however the calculation of slope may not provide the best measure when few points are utilized. Ice (2005) examined factors associated with cortisol levels measured 8 times per day starting at 8:00 a.m. and every 2 hours until 10:00 pm. Factors such as positive affect, positive mood state, age, and time of day were all significantly associated with cortisol level. These factors can have a major influence on slope calculations, given that the slope is frequently determined using only 2 or 3 daily cortisol levels. Increased age is frequently associated with a flatter slope and an overall higher mean cortisol, thought to be associated with a higher evening cortisol level (Magri et al., 2006; Wilkinson et al., 2001), although Giubilei and colleagues (2001) demonstrated elevated levels of both morning and evening cortisol in persons with Alzheimer's disease (AD) compared to controls. The small difference in slope exhibited in the current study may result from the advanced age of participants, decreased awakening cortisol levels or increased T4 cortisol levels.

As one considers the similarity and differences in ES values for NH residents with or without symptoms of acute psychosis there is yet another factor worth noting. Contrary to conventional wisdom, estimates of ES values from extremely small studies are likely to yield larger effect than those from studies with a larger sample size. This has been demonstrated by theoretical mathematical analysis (Hedges, 1981) and is described as the effect of sample bias by Hedges and Olkin (1985) who advocate correcting for the effect of sample bias when one tests the homogeneity of ES values across a series of studies.

Currently it is unclear which cortisol variables are best utilized under which circumstances. It is also unclear which specific cortisol variables should be targeted for any given treatment. Lupien, McEwen, Gunnar, and Hiem (2009) suggest that future studies must pay attention to altered cortisol patterns and pattern consistency. More recently measures such as Group-based trajectory methods (GBTM), a form of mixture modeling, have been utilized to identify and describe the distinct trajectories or patterns of change that exist within a population. Each unique trajectory is assumed to belong to a group, with the members of each group following a given response pattern. While originally introduced to examine patterns of change over many years (Lacourse, Nagin, Tremblay, Vitaro, & Claes, 2003),

these procedures can also be applied in cortisol analysis (Van Ryzin, Chatham, Kryzer, Kertes, & Gunnar, 2009) in which change is examined over the course of a day. This method, designed for heterogenous populations, can identify atypical patterns over time and thus may provide insight into characteristics that may be modified by a specific treatment.

In this study cortisol was measured only in individuals with dementia residing in a NH before and after a nursing assessment and treatment protocol. To build on this, other research is needed to examine cortisol variables and attributions of meaning of findings in at least five situations for this and other populations: (1) the stable description of groups or subgroups, (2) the description of response to acute stressors, (3) the description of altered diurnal cortisol rhythms as indicators of HPA axis integrity and chronic stress, (4) response to treatment for an acute stressor, and (5) response to treatment for stressors that are more chronic or habituated. There is also a need need for data supporting the impact of changes in cortisol and other biomarkers on actual health outcomes

CONCLUSION

Incorporating biobehavioral measures into nursing intervention studies offers important opportunities for understanding the complex mechanisms that interact to influence health behavior, health outcomes and the effectiveness of health care intervention. Enthusiasm for the use of these measures must be accompanied by appropriate cautionary effort to maximize the match between the biomarker, the target concept, and the study design, as well as to minimize potential sources of error variance. Given the lack of knowledge at this point about many issues that surround the use of biobehavioral measures in nursing intervention studies, our recommendation is that pilot studies provide needed data on procedural issues, possible confounds, responsiveness to change, and within and between subject variance when the intervention is not occurring. It is also important for future studies to comprehensively record and report measurement procedures, to assess the quality of the measurement tool in capturing responsiveness to change under specific conditions, and conduct analysis of the influence of multiple potential confounding variables. In conclusion, this article adds to an emerging literature that documents an array of issues that must be carefully considered when using biobehavioral measures. The use of biobehavioral measures without attention to if the biomarker will be responsive to therapy in the populations and contexts being studied will compromise the accumulation of knowledge regarding the effectiveness of healthcare interventions on health outcomes.

Acknowledgments

Funding: DHHS PHS NIH NINR 5R01NR07765, DHHS PHS NIH NINR. 1P20NR010674

References

Adam EK, Kumari M. Assessing salivary cortisol in large-scale, epidemiological research. Psychoneuroendocrinology. 2009; 34(10):1423–1436. [PubMed: 19647372]

Aung, APW.; Fook, FS.; Jayachandran, M.; Song, Z.; Biswas, J.; Nugent, C.; Yap, L. Smart wireless continence management system for elderly with dementia. HealthCom; Proceedings IEEE 10th International Conference on e-Health Networking, Application & Services; 2008. p. 33-34.URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4600105&isnumber=4600097

Blair C, Granger DA, Kivlighan KT, Mills-Koonce R, Willoughby M, Greenberg MT. Family Life Project Investigators. Maternal and child contributions to cortisol response to emotional arousal in young children from low-income, rural communities. Developmental Psychology. 2008; 44(4): 1095–1109. [PubMed: 18605837]

Carlson LE, Speca M, Patel KD, Goodey E. Mindfulness-based stress reduction in relation to quality of life, mood, symptoms of stress and levels of cortisol, dehydroepiandrosterone sulfate (DHEAS) and melatonin in breast and prostate cancer outpatients. Psychoneuroendocrinology. 2004; 29:448–474. [PubMed: 14749092]

Cohen, J. Statistical power analysis for the behavioral sciences. New York: Academic Press; 1969.

Cohen-Mansfield J, Lipson S. Pain in cognitively impaired NH residents: How well are physicians diagnosing it? Journal of the American Geriatrics Society. 2002; 50(6):1039–1044. [PubMed: 12110063]

Davis, CS. Statistical Methods for the Analysis of Repeated Measurements. New York: Springer-Verlag; 2002.

El-Sheikh M, Erath SA, Buckhalt JA, Granger DA, Mize J. Cortisol and Children's Adjustment: The Moderating Role of Sympathetic Nervous System Activity. Journal of Abnormal Child Psychology. 2008; 36:601–611. [PubMed: 18197472]

Fekedulegn DB, Andrew ME, Burchfiel CM, Violanti JM, Hartley TA, Charles LE, Miller DB. Area under the curve and other summary indicators of repeated waking cortisol measurements. Psychosomatic Medicine. 2007; 69:651–659. [PubMed: 17766693]

Fiocco AJ, Wan N, Weekes N, Pim H, Lupien SJ. Diurnal cycle of salivary cortisol in older adult men and women with subjective complaints of memory deficits an/or depressive symptoms: Relation to cognitive functioning. Stress. 2006; 9(3):143–152. [PubMed: 17060048]

Fortunato CK, Dribin AE, Granger DA, Buss KA. Salivary alpha-amylase and cortisol in toddlers: Differential relations to affective behavior. Developmental Psychobiology. 2008; 50(8):807–815. [PubMed: 18688807]

Gaab J, Blattler N, Menzi T, Pabst B, Stoyer S, Ehlert U. Randomized controlled evaluation of the effects of cognitive–behavioral stress management on cortisol responses to acute stress in healthy subjects. Psychoneuroendocrinology. 2003; 28:767–779. [PubMed: 12812863]

Gibson EL, Checkley S, Papadopoulos A, Poon L, Daley S, Wardle J. Increased salivary cortisol reliably induced by a protein-rich midday meal. Psychosomatic Medicine. 1999; 61(2):214–224. [PubMed: 10204975]

Gilbert C, Brown MCJ, Cappelleri JC, Carlsson M, McKenna SP. Estimating a minimally important difference in pulmonary arterial hypertension following treatment with sildenafil. Chest. 2009; 135:137–142. [PubMed: 18812447]

Giubilei F, Patacchioli F, Antonini G, Monti MS, Tisei P, Bastianello S, Angelucci L. Altered circadian cortisol secretion in Alzheimer's disease: Clinical and neuroradiological aspects. Journal of Neuroscience Research. 2001; 66(2):262–265. [PubMed: 11592122]

Grottoli S, Giordano R, Maccagno B, Pellegrino M, Ghigo E, Arvat E. The stimulatory effect of canrenoate, a mineralocorticoid antagonist, on the activity of the hypothalamus-pituitary-adrenal axis is abolished by alprazolam, a benzodiazepine, in humans. The Journal of Clinical Endocrinology & Metabolism. 2002; 87(10):4616–4620. [PubMed: 12364444]

Hall GR, Buckwalter KC. Progressively lowered stress threshold: A conceptual model for care of adults with Alzheimer's disease. Archives of Psychiatric Nursing. 1987; 1:399–406. [PubMed: 3426250]

Hanrahan K, McCarthy AM, Kleiber C, Lutgendorf S, Tsalikian E. Strategies for salivary cortisol collection and analysis in research with children. Applied Nursing Research. 2006; 19(2):95–101. [PubMed: 16728293]

Hedges LV. Distribution theory for Glass's estimator of effect size and related estimators. Journal of Educational Statistics. 1981; 6:107–128.

Hedges, LV.; Olkin, I. Statistical methods for meta-analysis. Orlando, FL: Academic Press; 1985.

Hellhammer DH, Wüst S, Kudielka BM. Salivary cortisol as a biomarker in stress research. Psychoneuroendocrinology. 2009; 34(2):163–71. [PubMed: 19095358]

Herrington CJ, Olomu IN, Geller SM. Salivary Cortisol as Indicators of Pain in Preterm Infants. Clinical Nursing Research. 2004; 13(1):53–68. [PubMed: 14768767]

Horgas AL, Dunn K. Pain in nursing home residents. Comparison of residents' self-report and nursing assistants' perceptions. Incongruencies exist in resident and caregiver reports of pain; therefore,

pain management education is needed to prevent suffering. Journal of Gerontological Nursing. 2001; 27(3):44–53. [PubMed: 11915257]

Hurley AC, Volicer BJ, Hanrahan PA, Houde S, Volicer L. Assessment of discomfort in advanced Alzheimer patients. Research in Nursing and Health. 1992; 15(5):369–377. [PubMed: 1529121]

Ice GH. Factors influencing cortisol level and slope among community dwelling older adults in Minnesota. Journal of Cross-Cultural Gerontology. 2005; 20(2):91–108. [PubMed: 16917746]

Micheel, CM.; Ball, JR., editors. Institute of Medicine & Committee on Qualifications of Biomarkers and Surrogate Endpoints in Chronic Disease. Evaluation of biomarkers and surrogate endpoints in chronic disease. Washington, DC: National Academies Press; 2010.

Jordon K, Dunn KM, Lewis M, Croft P. A minimal clinically important difference was derived for the Roland-Morris Disability Questionnaire for low back pain. Journal of Clinical Epidemiology. 2006; 59(1):45–52. [PubMed: 16360560]

Kiley JP, Sri Ram J, Croxton TL, Weinmann GG. Challenges Associated with Estimating Minimal Clinically Important Differences in COPD—The NHLBI Perspective. Journal of Chronic Obstructive Pulmonary Disease. 2005; 2(1):43–46. [PubMed: 17136960]

Kirschbaum C, Hellhammer DH. Salivary cortisol in psychobiological research: an overview. Neuropsychobiology. 1989; 22:150–169. [PubMed: 2485862]

Kirschbaum C, Hellhammer DH. Salivary cortisol in psychoneuroendocrine research: recent developments and applications. Psychoneuroendocrinology. 1994; 19(4):313–333. [PubMed: 8047637]

Kovach CR, Taneli Y, Dohearty P, Schlidt AM, Cashin S, Silva-Smith AL. Effect of the BACE intervention on agitation of people with dementia. Gerontologist. 2004; 44(6):797–806. [PubMed: 15611216]

Kovach CR, Noonan PE, Schlidt AM, Wells T. The Serial Trial Intervention: An Innovative Approach to Meeting Needs of People with Dementia. Journal of Gerontological Nursing. 2006a; 32(4):18–26. [PubMed: 16615709]

Kovach CR, Logan B, Noonan PE, Schlidt AM, Smerz J, Simpson M, Wells T. Effects of the Serial Trial Intervention on Discomfort and Behavior in Demented Nursing Home Residents. American Journal of Alzheimer's Disease & Other Dementias. 2006b; 21(3):147–155.

Kovach CR, Simpson MR, Joosse LL, Logan BR, Noonan PE, Reynolds S, Woods DL, Raff H. Comparison of the Effectiveness of Two Protocols for Treating Nursing Home Residents with Advanced Dementia. Research in Gerontological Nursing. 2012; 5(4):251–263. [PubMed: 22998656]

Lacourse E, Nagin D, Tremblay RE, Vitaro F, Claes M. Developmental trajectories of boys' delinquent group membership and facilitation of violent behaviors during adolescence. Developmental Psychopathology. 2003; 15:183–197.

Lawton, MP. Environment and Aging. Albany: Center for the Study of Aging; 1986.

Lupien SJ, McEwen BS, Gunnar MR, Hiem C. Effects of stress throughout the lifespan on the brain, behaviour and cognition. Nature Reviews Neuroscience. 2009; 10(6):434–445.10.1038/nrn2639

Magri F, Cravello L, Barili L, Sarra S, Cinchetti W, Salmoiraghi F, Ferrari E. Stress and dementia: The role of the hypothalamic-pituitary-adrenal axis. Aging Clinical and Experimental Research. 2006; 18:167–170. [PubMed: 16702789]

Malamud D. Saliva as a diagnostic fluid. British Medical Journal. 1992; 305(6847):207–208. [PubMed: 1290500]

National Institute for Nursing Research. Bringing Science to Life: NINR Strategic Plan. 2011. Retrieved May 5 2013, https://www.ninr.nih.gov/sites/www.ninr.nih.gov/files/ninr-strategic-plan-2011.pdf

Pellmar TC, Brandt EN Jr, Baird MA. Health and behavior: the interplay of biological, behavioral, and social influences: summary of an Institute of Medicine report. American Journal of Health Promotion. 2002; 16(4):206–219. [PubMed: 11913326]

Piazza JR, Almeida DM, Dmitrieva NO, Klein LC. Frontiers in the use of biomarkers of health in research on stress and aging. Journal of Gerontology: Psychological Sciences. 2010; 65B(5):513–525.

Pruessner JC, Kirschbaum C, Meinlschmid G, Hellhammer DH. Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. Psychoneuroendocrinology. 2003; 28:916–931. [PubMed: 12892658]

Pruessner JC, Wolf OT, Hellhammer DH, Buske-Kirschbaum A, Von Auer K, Jobst S, Kirschbaum C. Free cortisol levels after awakening: a reliable biological marker for the assessment of adrenocortical activity. Life Sciences. 1997; 61:2539–2549. [PubMed: 9416776]

Raff H. Salivary cortisol: A useful measurement in the diagnosis of Cushing's syndrome and the evaluation of the hypo-thalamic-pituitary-adrenal axis. The Endocrinologist. 2000; 10:9–17.

Raff H, Homar PJ, Skoner DP. New enzyme immunoassay for salivary cortisol. Clinical Chemistry. 2003; 49(1):203–204. [PubMed: 12507989]

Raff H. Cushing's syndrome: diagnosis and surveillance using salivary cortisol. Pituitary. 2012; 15:64–70. [PubMed: 21833616]

Raff H. Update on late–night salivary cortisol for the diagnosis of Cushing's Syndrome: methodological considerations. Endocrine 2013. 2013 (e-pub). 10.1007/s12020-013-0013-0

Raff H. Salivary cortisol: a useful measurement in the diagnosis of Cushing's syndrome and the evaluation of the hypothalamic-pituitary adrenal axis. The Endocrinologist. 2000; 10:9–17.

Raff H. The role of salivary cortisol determinations in the diagnosis of Cushing's syndrome. Current Opinion in Endocrinology and Diabetes. 2004; 11:271–275.

Raff H, Raff JL, Duthie EH, Wilson CR, Sasse EA, Rudman I, Mattson D. Elevated salivary cortisol in the evening in healthy elderly men and women: Correlation with bone mineral density. The journals of gerontology Series A, Biological sciences and medical sciences. 1999; 54(9):M479–M483.

Raff H, Singh RJ. Measurement of late night salivary cortisol and cortisone by liquid chromatography – tandem mass spectrometry to assess pre-analytical sample contamination with topical hydrocortisone. Clinical Chemistry. 2012; 58:947–948. [PubMed: 22377529]

Rosmand R, Holm G, Bjorntorp P. Food-induced cortisol secretion in relation to anthropometric, metabolic and hemodynamic variables in men. Internation Journal of Obesity. 2000; 24:416–422.

Salimetrics, LLC. High sensitivity salivary cortisol enzyme immunoassay kit. 2010. Available from http://www.salimetrics.com

Sephton SE, Sapolsky RM, Kraemer HC, Spiegel D. Diurnal cortisol rhythm as a predictor of breast cancer survival. Journal of the National Cancer Institute. 2000; 92:994–1000. [PubMed: 10861311]

Shadish, WR.; Cook, TD.; Campbell, DT. Experiemental and Quasi-experimental Designs for Generalized Causal Inference. Boston: Houghton Mifflin Co; 2002.

Smyth JM, Ockenfels MC, Gorin AA, Catley D, Porter LS, Kirschbaum C, Stone AA. Individual differences in the diurnal cycle of cortisol. Psychoneuroendocrinology. 1997; 22:89–105. [PubMed: 9149331]

Stewart, J.; Seeman, TE. Salivary cortisol measurement. MacArthur Research Network on SES & Health; 2000. Retrieved from http://www.macses.ucsf.edu/research/allostatic/salivarycort.php

Stolk MF, Becx MC, Kuypers KC, Seldenrijk CA. Severe hepatic side effects of ezetimibe. Clinical Gastroenterology and Hepatology. 2006; 4(7):908–911. [PubMed: 16797241]

Stone AA, Schwartz JE, Smyth J, Kirschbaum C, Cohen S, Hellhammer D, Grossman S. Individual differences in the diurnal cycle of salivary free cortisol: a replication of flattened cycles for some individuals. Psychoneuroendocrinology. 2001; 26:295–306. [PubMed: 11166492]

Chrousos G. Hypothalamic–pituitary–adrenal axis, neuroendocrine factors and stress. Journal of Psychosomatic Research. 2011; 53(4):865–871. [PubMed: 12377295]

Van Ryzin MJ, Chatham M, Kryzer E, Kertes DA, Gunnar MR. Identifying atypical cortisol patterns in young children: The benefits of group-based trajectory modeling. Psychoneuroendocrinology. 2009; 34:50–61. [PubMed: 18838226]

Weibel L. Methodological guidelines for the use of salivary cortisol as biological marker of stress. Presse Medicale. 2003; 32(18):845–51.

Wilkinson CW, Petrie EC, Murray SR, Colasurdo EA, Raskind MA, Peskind ER. Human glucocorticoid feedback inhibition is reduced in older individuals: Evening study. Journal of Clinical Endocrinology and Metabolism. 2001; 86:545–550. [PubMed: 11158007]

Williams RA, Hagerty BM, Brooks G. Trier social stress test: A method for use in nursing research. Nursing Research. 2004; 53(4):277–280. [PubMed: 15266167]

Woods DL, Dimond M. The Effect of Therapeutic Touch on Agitated Behavior and Cortisol in Persons with Alzheimer's Disease. Biological Research in Nursing. 2002; 4(2):104–114.

Woods DL, Martin J. Cortisol and wake time in nursing home residents with behavioral symptoms of dementia. Biologic Research Nursing. 2007; 9:21–29.

Young EA, Abelson J, Lightman SL. Cortisol pulsatility and its role in stress regulation and health Frontiers. Neuroendocrinology. 2004; 25(2):69–76.

Woods DL, Kovach CR, Raff H, Joosse L, Basmadjian A, Hegadoren K. Using saliva to measure endogenous cortisol in nursing home residents with advanced dementia. Research in Nursing and Health. 2008; 31(3):283–94. [PubMed: 18228609]

Table 1

Mean differences, standard deviations, and effect sizes (ES) for study variables.

Pre to Post variable measured	Acute psychotic symptoms (*n*=11)			No Acute psychotic symptoms (*n*=91)			t	p
Variables	**Mean diff**	**SD**	**ES**	**Mean diff**	**SD**	**ES**		
T1 Waking cortisol*	0.42	0.91	0.46	−0.09	0.61	−0.14	2.33	.022
T2 Morning cortisol*	0.29	0.54	0.54	0.03	0.71	0.04	1.13	.263
T3 Afternoon cortisol*	0.29	0.58	0.51	0.04	0.84	0.05	0.90	.372
T4 Evening cortisol*	0.03	0.73	0.04	0.15	0.70	0.21	0.51	.609
Slope for cortisol*	−0.01	0.07	−0.19	0.02	0.06	0.31	1.49	.141
AUC for cortisol*	3.52	3.70	0.95	0.40	5.92	0.07	1.60	.115
Observed agitation	31.82	20.19	1.58	7.54	22.86	0.33	3.36	.001
Observed pain	23.90	12.66	1.89	7.73	18.34	0.42	2.84	.005

*Cortisol values T1–T4 were log transformed prior to analyses. The slope and AUC were calculated using log-transformed values. Mean differences refer to the pre-post difference, so that positive differences indicate more improvement.