The Utility and Dynamics of Salivary Sex Hormone Measurements in the National Social Life, Health, and Aging Project, Wave 2

Michael J. Kozloski,1,2,3 L. Philip Schumm,4 and Martha K. McClintock2,3,5,6

1Department of Sociology, 2Center on Demography and Economics of Aging, 3Institute for Mind and Biology, 4Department of Health Studies, 5Department of Comparative Human Development, and 6Department of Psychology, University of Chicago, Illinois.

Objectives. Sex hormones affect physical, mental, and social health, yet their role in mediating social effects on aging is understudied. To facilitate such analyses with the National Social Life, Health & Aging Project Wave 2, we summarize the conceptual background, collection protocols, laboratory assays, and data analysis strategies for biologically active (free) levels of testosterone, estradiol, progesterone, and dehydroepiandrosterone (DHEA).

Method. Saliva from passive drool was collected from returning Wave 1 respondents and non-respondents as well as their partners during an in-home interview. Specimens were frozen and sent to Dresden LabService GmbH for duplicate assays of biologically active steroids using identical assay kits from National Social Life, Health, and Aging Project (NSHAP) Wave 1 (SaliCap, Catalog No. RE69995). Overall, 2,772 testosterone, 2,504 estradiol, 2,714 progesterone, and 2,800 DHEA measurements are publically available for Wave 2 analyses. Through a series of weighted linear regressions, all 4 steroids are compared by gender and age and to Wave 1 measurements.

Results. Men had higher levels of both free testosterone and progesterone than women; women and men had the same levels of estradiol and DHEA. Both free testosterone and DHEA decreased with age. We also found significant wave effects for all 4 sex hormones.

Conclusion. NSHAP Waves 1 and 2 are the first U.S. probability sample studies to measure these 4 salivary sex hormones simultaneously, providing individual profiles 5 years apart. Wave 2 data demonstrate differences by gender and trends by age that are similar to those found in other saliva-based and serum-based studies of free steroid levels. The differences between waves arising from the change in assay laboratory need to be adjusted in future longitudinal analyses using NSHAP Wave 1 and Wave 2 steroid data.

Key Words: DHEA—Estradiol—Progesterone—Saliva collection—Sex hormones—Testosterone.
To test both the internal and external validity of our findings, we compare our Wave 2 results to those already obtained in Wave 1 as well as other smaller-scale studies that have employed salivary enzyme immunoassays. Our goal is to provide a comprehensive report, building on Wave 1 (Gavrilova and Lindau, 2009) with the technical details and interpretation guidelines for Wave 2 and longitudinal analyses. We aim to provide a reference to be cited by multidisciplinary analysts using these steroid data to address a variety of questions about their role in the interplay among social life, psychological condition, and health during the aging process.

Testosterone
Testosterone is the principle male reproductive hormone, playing not only the primary role in physical masculine development but is also central for health, energetic, and overall well-being in men (Bassil, Alkaade, & Morley, 2009) and male behavior and personality (Zak et al., 2009). And although there is a natural decrease with age in testosterone and its precursors produced by the testes and adrenals, it can be accelerated by chronic illness, obesity, excessive drinking, and other conditions that affect overall physical health and quality of life (Feldman et al., 2002). And in turn, significant decreases in testosterone can exacerbate these conditions as well as lead to loss of libido and chronic depression (Kraemer et al., 1976; Roney, Mahler, & Maestripieri, 2003). Although testosterone levels in women also decrease with age due to an age-related decrease in secretion from the ovary and precursors from the adrenal, further research is needed to fully elucidate testosterone’s role, as it influences quality-of-life measures such as general well-being and mood as well as the restoration of sexual desire after menopause (Davis & Tran, 2001). Further, although androgen replacement therapy (ART) in later life is becoming increasingly common among men, it remains unclear at what age and stage it becomes most appropriate and carries with it the least amount of cardiovascular risk (American Society of Andrology, 2006).

Estradiol
Estradiol (E2) is the only biologically potent steroid of the three estrogens secreted by the ovary and is also produced by fat cells, particularly visceral fat (Nelson and Bulun, 2001; Yamatani, Takahashi, Yoshida, Takata, & Kurachi, 2013). It is responsible for the development of sex characteristics in women during growth and development and, with progesterone, plays a key role during the luteal phase of the menstrual cycle in preparing the endometrium for implantation (Hall & Anthony, 1993). In women (and with findings in men proving inconclusive), low estradiol in later life has been associated with mental and physical deterioration such as memory loss and decreases in cognition and osteoporosis (Yaffe et al., 2007). Low estradiol levels are also linked to decreases in sexual function, whereas high levels of estradiol have been linked to postmenopausal breast cancer (Key, Appleby, Barnes, & Reeves, 2002).

Although estradiol has been shown to remain relatively stable throughout the life course in men, there is a natural and substantial decrease in estradiol in women after menopause, with male and female levels approximately equal in later life (ZRT Laboratory, LLC, 2007). Hormone therapy of estradiol is not uncommon (in NSHAP Wave 1, approximately 4% of women were documented using estradiol in hormone therapy), but its overall beneficial effects have come into question in recent years, as the cardiovascular and malignancy risks of hormone therapy are believed to outweigh the benefits (Barnabei et al., 2005).

Progesterone
Commonly called the “hormone of pregnancy,” progesterone plays several key roles throughout the human gestation period, including uterine development, inhibition of lactation during pregnancy, increasing milk production after labor, and, possibly, facilitating the onset of labor (Schumacher et al., 2004). Peaking during pregnancy and in the luteal phase of the menstrual cycle, levels in women and men are otherwise approximately equal (Oettel & Mukhopadhyay, 2004). Depending on the specific tissue, progesterone can either counteract estrogen activity, such as thinning the vaginal wall, or act synergistically to regulate pain.

Although progesterone is produced by both men and women, its function and physiological importance in men is not entirely understood. Via research on progesterone and its possible relationship with cortisol, increased progesterone levels are likely to occur during periods of increased stress, suggesting an adrenal activation or a mechanism to counteract the effects of anxiety. Although this has been found in both sexes, the relationship with cortisol may be especially strong in men (Wirth, Meier, Fredrickson, & Schultheiss, 2007). Although further research is needed to understand the neuroendocrine mechanisms of natural progesterone in men, exogenous progesterone has been associated with the modulation or reduction of male sexual behavior in both human and animal research, thereby providing possible insights into future research on male sexual dysfunction (Andersen & Tufik, 2006).

Dehydroepiandrosterone
DHEA is a prohormone secreted by the adrenal glands and gonads in both sexes. It can be generally classified as a sex hormone because of its influence on sexual physiology in both men and women (Labrie et al., 2005) and also binds at a variety of receptors regulating diverse physiological systems potentially involved in the morbidity of aging (Webb, Geoghegan, Prough, & Miller, 2006). DHEA and its biologically inactive excretory product, DHEA-S, steadily decrease with age in both sexes, and low levels have been linked to numerous pathologies including cardiovascular disease, rheumatoid arthritis, depression, and sexual dysfunction.
(Savineau, Marthan, & Dumas de la Roque, 2013; Spark, 2002; Traish, Kang, Saad, & Guay, 2011). Although DHEA is made naturally by the adrenal glands, there have long been claims that DHEA replacement can work to reduce the risks of age-related diseases without the adverse events associated with other hormone replacement therapies. This remains controversial to date particularly because it is so easily converted to testosterone and estradiol, which can promote cancers (Labrie, 2010; Webb et al., 2006).

Despite decades of research on serum hormone levels, there is comparatively very little data on levels of free steroids, the biologically active fraction, particularly on a nationwide scale, either cross-sectionally or longitudinally, and the standardization of measurements remains a work in progress (Stanczyk, Lee, & Santen, 2007). Moreover, high-quality studies on salivary hormone levels among the elderly people are especially rare, and nationally representative studies among the elderly people are all but non-existent. With NSHAP, we are filling in these gaps and we intend to show that this alternative method of collection is an accurate, practical, and informative way of studying the effects of declining sex hormone levels with age.

**METHOD**


**Saliva Collection**

Of the total respondents, 3,230 (96%) agreed to submit a sample of their saliva via a passive drool technique. Details of the saliva collection protocol are available in the Supplementary Material (Saliva & Medication Log in NSHAP Wave 2) and described elsewhere for both Wave 1 (Mendoza, Curran, & Lindau, 2007a, 2007b, 2007c; Nallanathan, Mendoza, Curran, & Lindau, 2007) and Wave 2 (O’Doherty et al., 2014). In short, respondents were encouraged not to eat during the interview, and answered questions about their food consumption on that day, even though eating and dental care are reported to have no effect on the replicability of steroid levels in saliva (Gröschl, Wagner, Rauh, & Dörr, 2001). Approximately an hour and a quarter of their spouses or cohabiting partners, plus 161 individuals who were selected into the Wave 1 sample but had declined to participate and 48 of their spouses or partners, yielding a total sample size of 3,377. Data are publicly available (NSHAP Wave 1: Waite, Linda J., Edward O. Laumann, Wendy Levinson, Stacy Tessler Lindau, and Colm A. O’Muircheartaigh. National Social Life, Health, and Aging Project (NSHAP): Wave 1. ICPSR20541-v6. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2014-04-30. doi:10.3886/ICPSR20541.v6. NSHAP Wave 2: Waite, Linda J., Kathleen Cagney, William Dale, Elbert Huang, Edward O. Laumann, Martha K. McClintock, Colm A. O’Muircheartaigh, L. Phillip Schumm, and Benjamin Cornwell. National Social Life, Health, and Aging Project (NSHAP): Wave 2 and Partner Data Collection. ICPSR34921-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2014-04-29. doi:10.3886/ICPSR34921.v1.).

**Table 1. Valid and Missing Measurements of Salivary Sex Hormones, Unweighted NSHAP Wave 2**

<table>
<thead>
<tr>
<th>Sex hormone</th>
<th>Testosterone</th>
<th>Estradiol</th>
<th>Progesterone</th>
<th>DHEA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Total interviews</td>
<td>3,377</td>
<td>100</td>
<td>3,377</td>
<td>100</td>
</tr>
<tr>
<td>Incomplete interviews</td>
<td>1</td>
<td>&lt;1</td>
<td>1</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Special sample*</td>
<td>4</td>
<td>&lt;1</td>
<td>4</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Refused</td>
<td>142</td>
<td>4</td>
<td>142</td>
<td>4</td>
</tr>
<tr>
<td>Agreed to provide sample</td>
<td>3,230</td>
<td>96</td>
<td>3,230</td>
<td>96</td>
</tr>
<tr>
<td>Tried, unable to do</td>
<td>71</td>
<td>2</td>
<td>71</td>
<td>2</td>
</tr>
<tr>
<td>Equipment problem</td>
<td>4</td>
<td>&lt;1</td>
<td>4</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Adequate specimen</td>
<td>3,155</td>
<td>93</td>
<td>3,155</td>
<td>93</td>
</tr>
<tr>
<td>Lost in transit</td>
<td>5</td>
<td>&lt;1</td>
<td>5</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Received by laboratory*</td>
<td>3,150</td>
<td>93</td>
<td>3,150</td>
<td>93</td>
</tr>
<tr>
<td>Saliva volume too low*</td>
<td>106</td>
<td>3</td>
<td>299</td>
<td>9</td>
</tr>
<tr>
<td>Low sample quality</td>
<td>50</td>
<td>1</td>
<td>54</td>
<td>2</td>
</tr>
<tr>
<td>&lt;Lower limit of sensitivity*</td>
<td>15</td>
<td>&lt;1</td>
<td>116</td>
<td>3</td>
</tr>
<tr>
<td>&gt;Upper limit of sensitivity*</td>
<td>39</td>
<td>1</td>
<td>45</td>
<td>1</td>
</tr>
<tr>
<td>Valid measurements</td>
<td>2,943</td>
<td>87</td>
<td>2,636</td>
<td>78</td>
</tr>
<tr>
<td>Out of age range</td>
<td>171</td>
<td>5</td>
<td>132</td>
<td>4</td>
</tr>
<tr>
<td>Used in analysis</td>
<td>2,772</td>
<td>82</td>
<td>2,504</td>
<td>74</td>
</tr>
</tbody>
</table>

*Notes. DHEA = dehydroepiandrosterone; NSHAP = National Social Life, Health, and Aging Project.

*Respondents with HIV infection were not asked to submit a sample.

*Order of laboratory analysis if insufficient sample: estradiol, testosterone, progesterone, and DHEA.

*Lower limit of sensitivity: testosterone = 1.0 pg/mL, estradiol = 1.0 pg/mL, progesterone = 5.0 pg/mL, DHEA = 5.0 pg/mL.

*Upper limit of sensitivity: testosterone = 500 pg/mL, estradiol = 64 pg/mL, progesterone = 1,000 pg/mL, and DHEA = 3,000 pg/mL.
half into the interview, respondents were asked to submit a sample of approximately 1 mL of unstimulated saliva into a vial (Salicap, Catalog No. RE69995; IBL International GMBH, Hamburg, Germany) via a small household plastic straw. The field interview gave them unpressured time by concurrently recording information from the containers of all of their prescribed and over the counter medications, some of which might affect their saliva production.

Of the 3,230 respondents, only 71 (2.2%) were unable to provide any saliva, and equipment problems hindered four of the collections (0.1%, Table 1). To minimize bacterial growth and sample degradation, the samples were transferred immediately after collection via cold packs to the field interviewers’ home freezers. Once a month, the specimens were shipped overnight on dry ice to the Central Survey Biomasure Laboratory where they were stored at −80 °C before final shipment to Clemens Kirschbaum, PhD Director, Technisch Universität, Dresden LabService GmbH (Dresden, Germany) where they were subsequently analyzed for content. Five samples were lost during transit, yielding a total of 3,150 analyzable saliva specimens.

**Assay Analysis and Quality Control**

Although the Wave 2 specimens were analyzed using identical enzyme immunoassay kits provided by Salimetrics in Wave 1, the assays were conducted in different laboratories: at Salimetrics, LLC in Wave 1 and at Dresden LabService GmbH in Wave 2. On the day of assay, specimens were thawed completely, vortexed, and centrifuged at 3,000 rpm for 15 min, and clear samples were pipetted in duplicate into test wells using a 96-well plate format. Given sufficient sample volume, analyses were scheduled in the priority order of (a) progesterone, (b) testosterone, (c) estradiol, and (d) DHEA to minimize the effects of freeze/thaw cycles on results. In the case of insufficient sample volume, the order of priority was adjusted to (a) estradiol, (b) testosterone, (c) progesterone, and (d) DHEA. Assays were performed in duplicate and retested when the variation exceeded 15% between duplicates. All hormone values were measured in picograms per milliliter, and the average of the duplicates are used in all analyses with two exceptions: (a) a single value was used when the sample volume was inadequate to run duplicates or when assay results yielded only one valid measurement and (b) when a sample was retested, the two values with the smallest variation were averaged and used. The exact intra- and interassay coefficients of variation are given at the top of Table 2 and are similar to those reported by Gavrilova and Lindau (2009) in Wave 1.

For each hormone, measurements that fall outside the reportable range of the standard curve, that is, below the lower limit of sensitivity of the assay or are higher than the upper limit, are flagged for further scrutiny. The lower limits of sensitivity were determined by interpolating the mean minus two standard deviations of the optical densities at

### Table 2. Descriptive Statistics of Salivary Sex Hormones (pg/mL) in NSHAP Waves 2 and 1 Projected to the U.S. Population of Older Adults

<table>
<thead>
<tr>
<th>Descriptive statistics</th>
<th>Testosterone</th>
<th>Estradiol</th>
<th>Progesterone</th>
<th>DHEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wave 2: ages 62–90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intraassay CV</td>
<td>8%</td>
<td>14%</td>
<td>12%</td>
<td>11%</td>
</tr>
<tr>
<td>Interassay CV</td>
<td>9%</td>
<td>15%</td>
<td>18%</td>
<td>17%</td>
</tr>
<tr>
<td>Unweighted counts</td>
<td>1,422</td>
<td>1,350</td>
<td>1,264</td>
<td>1,325</td>
</tr>
<tr>
<td>Measurements (pg/mL)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>1.18</td>
<td>5.13</td>
<td>1.0</td>
<td>5.09</td>
</tr>
<tr>
<td>Maximum</td>
<td>460.49</td>
<td>484.63</td>
<td>33.69</td>
<td>50.17</td>
</tr>
<tr>
<td>Mean</td>
<td>44.57</td>
<td>81.12</td>
<td>4.10</td>
<td>12.72</td>
</tr>
<tr>
<td>Mean ln(y)</td>
<td>3.58</td>
<td>4.24</td>
<td>1.16</td>
<td>1.19</td>
</tr>
<tr>
<td>Standard error of mean</td>
<td>0.81</td>
<td>1.21</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td>Median</td>
<td>40.07</td>
<td>77.26</td>
<td>2.93</td>
<td>2.97</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>30.48</td>
<td>44.44</td>
<td>4.03</td>
<td>4.01</td>
</tr>
<tr>
<td>Wave 1: ages 57–85</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unweighted counts</td>
<td>1,118</td>
<td>1,195</td>
<td>1,171</td>
<td>1,182</td>
</tr>
<tr>
<td>Measurements (pg/mL)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>2.30</td>
<td>3.81</td>
<td>1.0</td>
<td>1.15</td>
</tr>
<tr>
<td>Maximum</td>
<td>217.42</td>
<td>393.42</td>
<td>64.0</td>
<td>64</td>
</tr>
<tr>
<td>Mean</td>
<td>46.42</td>
<td>83.05</td>
<td>9.48</td>
<td>9.52</td>
</tr>
<tr>
<td>Mean ln(y)</td>
<td>3.72</td>
<td>4.32</td>
<td>1.95</td>
<td>1.94</td>
</tr>
<tr>
<td>Standard error of mean</td>
<td>0.71</td>
<td>1.07</td>
<td>0.26</td>
<td>0.26</td>
</tr>
<tr>
<td>Median</td>
<td>42.93</td>
<td>77.5</td>
<td>6.98</td>
<td>6.66</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>23.59</td>
<td>37.06</td>
<td>8.94</td>
<td>8.90</td>
</tr>
</tbody>
</table>

**Notes.** CV = coefficient of variation; DHEA = dehydroepiandrosterone; ln = natural log; NSHAP = National Social Life, Health, and Aging Project; pg/mL = picograms per milliliter.

*Mean of the measurements after a natural log transformation.*
the 0 pg/mL standard (i.e., “zero standard”), using 10 sets of duplicates for testosterone, estradiol, and DHEA and 20 sets for progesterone. For measurements greater than those of the highest calibrator, samples were diluted to bring the optical density readings within accepted range and flagged as the likely products of interference. Measurements outside of the reportable ranges are excluded from our analyses in this article until further investigation into respondent medications and conditions can address these extreme values.

A comprehensive quality control analysis was performed to quantify inter-assay and intraassay variation as well as test for possible effects of field interviewers, shipping duration, and run effects throughout time within the lab environment, and findings were discussed with laboratory director Clemens Kirschbaum, PhD. Details of quality control protocols are described in full elsewhere (O’Doherty et al., 2014), and we provide detail on the types of missing values in our data and how to best account for them in future analyses.

Statistical Analyses

All analyses were conducted using Stata 13 software (StataCorp, 2013). Respondents were limited to those who were age-eligible in Wave 1, and all analyses are weighted to account for non-response rates based on age and urbanicity of residence, thereby representing the U.S. population born between 1920 and 1947 (age 62–91 at the time of 2010–2011 interview). Descriptive statistics of sex hormone levels are given for men and women in both NSHAP Waves 1 and 2 and the distributions of each measure are shown graphically for both waves using boxplots, split by gender.

To provide further insight into differences by age and gender in Wave 2 data, each of the hormone measures are regressed against (1) no demographic variables for a constant-only model; (2) gender, with men assigned “0” and women assigned “1”; (3) gender, age, and age²; and finally, (4) gender, age, age², a gender by age interaction, and a gender by age² interaction. Age is centered at its mean in Models 3 and 4, and each sex hormone measurement is transformed by the natural log to minimize the effects of extreme outliers.

Because the four models are nested, we use Wald tests with α = .05 to determine the most parsimonious model for each sex hormone. Finally, based on Model 4 (the full model), the predicted means of each sex hormone are plotted onto a two-dimensional margins plot by age and gender with pointwise 95% confidence intervals around the means.

To quantify the longitudinal effects of age and gender while controlling for the methodological change in laboratory, we conduct four repeated measures analysis of variance models for each hormone measure with the following covariates: (1) wave of survey with NSHAP Wave 1 assigned “0” and NSHAP Wave 2 assigned “1”; (2) wave of survey and gender; (3) wave of survey, gender, age, and age²; (4a) wave of survey, gender, age, age² plus a wave by gender interaction; and (4b) wave of survey, gender, age, age² plus a wave by age interaction and a wave by age² interaction. Age is once again centered at its mean, hormone measurements are log transformed, and Wald tests conducted for nested models.

Under the assumption of no significant secular changes in the population between 2006 and 2011, the “wave of survey” factor should encompass the effect of laboratory. This assumption comes with a caveat, however, as recent studies have shown that there actually may be an age-independent, period-independent decrease in population-level male testosterone levels in both the United States and Europe (Andersson et al., 2007; Travison, Araujo, O’Donnell, Kupelian, & McKinlay, 2007). However, if this decline is indeed due to cohort effects rather than any unaccountable environmental factors, its effect on our findings should theoretically be null because the men added to the study in Wave 2 are, on average, from the same cohorts as the returning Wave 1 respondents (mean age for both groups = 73 years).

Results

Gender Differences

Graphically, men appeared to have higher overall levels of testosterone and progesterone than women, whereas men and women had nearly identical distributions of estradiol and DHEA (Figure 1, boxplots for Wave 2). Indeed, the gender coefficients for women in the most parsimonious models were negative and significant for both testosterone (Table 3; Model 3: coef = −0.665, p < .001) and progesterone (Model 2: coef = −0.138, p < .001). No significant gender differences were seen in estradiol or DHEA levels.

Age Differences

There was also a significant linear decrease by age in androgen levels and an absence of any observed curvilinearity (testosterone Model 3: coef = −0.008, p < .001; DHEA Model 4: coef = −0.004, p < .01). In contrast, neither estradiol nor progesterone manifested significant age effects. When interactions for gender and age were added to the models, however, we got a clearer picture of how age-related decreases in androgen levels function. For testosterone, the addition of the Gender × Age and Gender × Age² effects were not significant by themselves, but when considered together give a p value for model inclusion of p = .003. Because both interactions were positive, this means that the reduction in testosterone levels with age was significantly larger in men. Similarly, for DHEA, age was actually not significant in the main effects model (Model 3), but when the interactions for gender and age were added in Model 4, the coefficient for linear age was negative and significant (coef = −0.004, p < .01). This also suggests that the significant decrease in DHEA with age was exclusive to men.
KozlosKi et al. showed that testosterone and estradiol levels were significantly lower in Wave 2 than they were in Wave 1 (testosterone Model 3: coef = −0.107, \( p < .001 \); estradiol Model 1: coef = −0.754, \( p < .001 \)), whereas progesterone and DHEA levels were significantly higher (progesterone Model 2: coef = 0.104, \( p < .001 \); DHEA Model 3: coef = 0.397, \( p < .001 \)). The distributions in Figure 1 illustrate these wave differences clearly and most notably for estradiol in both men and women.

**NSHAP Wave Differences**

The most parsimonious main effects models for each hormone in Table 4 showed that testosterone and estradiol levels were significantly lower in Wave 2 than they were in Wave 1 (testosterone Model 3: coef = −0.107, \( p < .001 \); estradiol Model 1: coef = −0.754, \( p < .001 \)).

![Figure 1. The distributions of salivary sex hormone measurements (pg/mL) for women and men (columns) in NSHAP Waves 1 and 2 (rows). Notes. DHEA = dehydroepiandrosterone; NSHAP = National Social Life, Health, and Aging Project. For simplicity and graph visibility, two Wave 1 male respondents with DHEA > 1,000 pg/mL excluded.](image-url)
Although these wave effects for estradiol or progesterone did not vary by age or gender, they did for both androgens. For testosterone, there was a steeper decline across ages in NSHAP Wave 2 (Model 4b: coef = −0.007, p < .001), whereas for DHEA, the age effect was stronger in Wave 1 (Model 4b: coef = −0.004, p < .001). Further, gender differences in DHEA levels were significantly larger in Wave 2 from what they had been in Wave 1 (Model 4a: coef = −0.004, p < .001).

**Discussion**

**Benefits and Costs of Salivary Measures**

Although saliva has been used to scientifically measure biologically active sex steroids in humans for decades, concerns remain about its applicability and accuracy. These were addressed by careful attention to sample collection, storage, and shipping (O’Doherty et al., 2014) as discussed here. For adults, collecting saliva is virtually stress free and provides an accurate picture of the dynamics of a stress-labile hormone such as cortisol (Vining, McGinley, Maksvytis, & Ho, 1983). The NSHAP protocol ensured privacy and reduced time pressure during collection, which could be stressors.

The storage and shipping of saliva requires specimens to be frozen to decrease the likelihood of bacterial contamination and steroid metabolism (Whembolua, Granger, Singer, Kivlighan, & Marguin, 2006; Wood, 2009), and NSHAP protocols were tightly standardized to control for potential effects temperature and freeze–thaw cycles (Gröschl et al., 2001; O’Doherty et al., 2014).

Although the level of invasiveness of salivary testing is a considerable improvement over that of venipuncture, dried blood spot testing is also less invasive, does not require samples to be frozen for shipping, and does not suffer from the contamination risks of salivary testing (McDade, 2014). However, the small amount of material obtained per participant can make measurement accuracy challenging and leave little room for error due to sample requirements for retesting (McDade, Williams, & Snodgrass, 2007), and projections to serum levels have been shown to be more accurate in premenopausal women than men (Shirtcliff, Reavis, Overman, & Granger, 2001), thereby raising questions concerning its accuracy in samples of older adults. As is the case with

<table>
<thead>
<tr>
<th>Model</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>R²</td>
<td>0.205</td>
<td>0.213</td>
<td>0.216</td>
</tr>
<tr>
<td>Wald p value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.003</td>
</tr>
<tr>
<td>Constant</td>
<td>3.906***</td>
<td>4.244***</td>
<td>4.254***</td>
</tr>
<tr>
<td>Women</td>
<td>−0.668***</td>
<td>−0.665***</td>
<td>−0.686***</td>
</tr>
<tr>
<td>Overall age p value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.003</td>
</tr>
<tr>
<td>Age</td>
<td>−0.008***</td>
<td>−0.013***</td>
<td>−0.003</td>
</tr>
<tr>
<td>Age</td>
<td>−2.08e−4</td>
<td>−4.60e−4</td>
<td>−3.79e−4</td>
</tr>
<tr>
<td>Overall interactions p value</td>
<td>.003</td>
<td>.826</td>
<td></td>
</tr>
<tr>
<td>Women × Age</td>
<td>4.45e−4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 4</td>
<td>1.206***</td>
<td>0.003</td>
<td></td>
</tr>
</tbody>
</table>

**Notes.** DHEA = dehydroepiandrosterone; NSHAP = National Social Life, Health, and Aging Project; W1 = Wave 1.

1. Most parsimonious model (main effects only).
2. Most parsimonious model (with significant effect modification).
3. Wald test p value comparing nested models.
4. Men coded as “0”; women coded as “1.”
5. Age is centered around its mean.

*p < .05. **p < .01. ***p < .001.
salivary testing, the convenience factors and lower levels of invasiveness to participants come with the higher costs necessary in assay development and validation (McDade, 2014). Ultimately, NSHAP chose saliva over dried blood spots because the assays for free levels of steroids were well validated, and we wished to use the dried blood spots collected for other measures otherwise unobtainable, that is, HbA1c.

**Missing Values**

Our sample sizes for all four salivary sex hormones are large compared with that in many other studies (Table 1 and Figure 2). Nonetheless, we know that the NSHAP missing values are not Missing Completely At Random, and our data indicate that they may not be Missing At Random either. For example, participants who were unable to provide a saliva specimen or provided a sample of inadequate volume are more likely to be physically compromised and in the oldest age bracket—respondents age 80 and older comprise 24% (testosterone), 25% (estradiol), 27% (progesterone), and 25% (DHEA) of the total share of missing values in the sex hormone measurements compared with 19% in the overall sample. Thus, when conducting statistical analyses with reproductive hormones and other biological measures in NSHAP, analysts should keep in mind the possible biases due to excluding these cases and instead consider handing them through multiple imputation. In doing the latter, analysts should also consider including covariates related to medications that reduce salivary flow in the imputation model.

**Extreme Values**

In any assay, measurements have greater error and variation at the lowest levels of detection (Lewis, 2006). We encounter similar issues in measurements that fall below the reportable range for each hormone, the most problematic being estradiol (N = 115). The difficulty in capturing low estradiol readings via immunoassays is well documented (Handelsman et al., 2014) and can result in attenuated measures of association, suggesting a weaker association with disease (or other outcome) than actually exists (Rosner, Hankinson, Sluss, Vesper, & Wierman, 2013). Because both
low and high hormone levels in older ages can be indicative of deteriorating health, we again highly recommend handling the missing data at both ends of the spectrum through multiple imputation rather than exclusion.

Among respondents with valid measurements (Table 2 and Figure 1), there are some with salivary hormone readings that are more than twice the levels of the average person in their age range and even greater than those of their much younger counterparts (Mendoza, Curran, & Lindau, 2007a, 2007b; Nallananath, Mendoza, Curran, & Lindau, 2007). For the majority of these respondents, this is likely due to blood contamination of the saliva from oral injury or bleeding gums that had gone undetected through quality control protocols (Granger et al., 2007) or the use of prescription hormone replacement therapy (Gavrilova & Lindau, 2009). At the time of this writing, medication data for Wave 2 are currently being coded and linked to the Multum drug database, and the data will be available in future releases. For forthcoming analyses, it will be necessary to control for or remove individuals who are artificially increasing their hormone levels in order to get an accurate picture of how naturally decreasing hormonal levels are related to morbidity and mortality in old age.

**Gender Differences**

For external validity of our measurements, we turn to existing literature that has successfully measured sex hormones via saliva collection. Even in the absence of a gold standard for salivary hormone levels, we can use the literature to assess whether our measurements fall within typical ranges of other studies. Figure 2 shows salivary sex hormone measurements in external sources (including both NSHAP waves), ordered by increasing average age of the sample respondents. Although sample sizes for external sources are small and their measurements show few age-related patterns, statistical differences for salivary hormone levels between men and women in NSHAP Wave 2 (as shown in Change 3 and Figure 3) coincide with gender differences in these sources. ZRT Laboratory, LLC (2007)
and IBL Immuno-Biological Laboratories (2005) both report salivary testosterone measures for older men and women, and all show higher levels for men. Analysts may also wish to consider the established gender differences in the association of testosterone levels between saliva and serum, due in part to gender differences in SHBG. Granger, Shirtcliff, Booth, Kivlighan, and Schwartz (2004) propose a method to control for this difference in statistical models.

We also found higher levels of progesterone in men in Wave 2, and ZRT Laboratory, LLC affirms this, showing higher levels in older men (44 pg/mL) compared with older women (36 pg/mL). Although there are usually no sex differences in levels of progesterone outside of the luteal phase in young adults (Oettel & Mukhopadhyay, 2004), some studies have shown a decrease in progesterone levels after menopause (Dennerstein, Alexander, & Kotz, 2003; Schindler et al., 2003). No significant gender differences were seen in average levels of estradiol, as has been documented in other sources (ZRT Laboratory, LLC, 2007).

Gender differences in DHEA levels are mixed, with no significant differences seen between men and women in Wave 2, but when gender by wave effects are examined (Model 4a in Table 4), the main effects for both gender and wave and the Wave × Gender interaction become significant, suggesting that the gender differences seen in Wave 1 were not present in Wave 2. Although seemingly contradictory, these findings are actually quite representative of findings from existing research, some of which find evidence of higher DHEA levels in men at older ages (Kraemer et al., 2001; Young, Skibinski, Mason, & James, 1999), whereas others show significantly higher levels in women (Šulcová, Hill, Hampl, & Stárka, 1997). In their detailed analysis of DHEA and DHEA-S levels throughout the life span, Šulcová and coworkers (1997) find that there may actually be a slight increase in DHEA levels around age 80. One of their explanations for this increase is the process of natural selection; in other words, people with higher levels of DHEA may tend to live longer lives. If this is indeed the case, it may provide further explanation for the elevated levels of DHEA in Wave 2 beyond the effects of laboratory.

Figure 3. Sex differences in levels of salivary sex hormones by age in NSHAP Wave 2 (weighted to represent the U.S. population aged 62–90). Notes. CI = confidence interval; DHEA = dehydroepiandrosterone; NSHAP = National Social Life, Health, and Aging Project; pg/mL = picograms per milliliter.
Age Differences

Significant and negative linear age effects are seen for both testosterone and DHEA (Figure 3 and Model 3 in Table 3). Both of these hormones also show significant curvilinear effects throughout the life span with testosterone’s decline becoming more rapid with age and DHEA’s decline slowing down. All of these age trends are documented in existing literature and thereby strengthen the validity of our findings. During the aging process, testosterone declines in both sexes: steadily throughout adulthood in men and more gradually between the ages of 20 and 45 in women, likely due to the age-related decrease in adrenal secretion (Davis & Tran, 2001; Ellison et al., 2002; Feldman et al., 2002). Further, with advanced age comes greater risk of chronic illness and a more rapid deterioration of physical health, and in turn, the decrease in testosterone levels is likely to accelerate (Feldman et al., 2002). Johnson, Bebb, and Sirrs (2002) claim that DHEA falls steadily with age in both sexes due to the slowdown in the adrenals, eventually dropping to levels only 1%–20% of normal values among young people. Because adrenal secretions show the largest rates of decline before age 60 and gradually slow down after, declining rates of DHEA production tend to follow the same pattern (Labrie, Bélanger, Cusan, Gomez, & Candas, 1997).

NSHAP Wave Differences

Although all gender and age-related differences parallel the literature on salivary and serum sex hormone levels, the wave effects for DHEA and testosterone cannot be explained via external sources and require further scrutiny. The simplest explanation would be that combining data from the two waves allows us to detect smaller differences with increased precision, even if the differences are not clinically relevant. Given the sizes of the effects, however, we believe that the change in laboratory truly plays a significant role. Because probability of a compromised salivary flow increases with age, it is possible that one of the two chosen laboratories was able to more effectively capture accurate results even with higher rates of compromised salivary flow (Jones, Watkins, Hand, Warren, & Cowen, 2002). Further, salivary flow rates are affected by the sizes of the salivary glands, which are biologically larger in men (Inoue et al., 2006), thereby providing another plausible explanation for the significant gender by wave interaction for DHEA. Regardless of its true meaning, however, NSHAP wave and its interaction with gender and age throughout older ages need to be included as confounders in all statistical models that merge sex hormone data from Waves 1 and 2.

Conclusions

Next Steps

Future analyses that may help explain some of the differences between Wave 1 and Wave 2 measurements include the flagging or exclusion of respondents in either wave undergoing hormone replacement therapy. In Wave 1, 12.3% of women and 2.5% of men were using exogenous sex hormones, and ultimately, these hormones only had a strong effect on the mean levels of estradiol, as most of the women were using estrogen therapy (Gavrilova & Lindau, 2009). There exists strong evidence, however, that 5 years of additional medical research on the hot-button issue of hormone therapy and pharmaceutical promotion of ART may have increased those percentages substantially between 2005 and 2010. Using data from Clinformatics Datamart, one of the nation’s largest and most comprehensive health insurance populations, Baillargeon, Urban, Ottenbacher, Pierson, and Goodwin (2013) show that the percentage of older American men receiving prescriptions for ART has more than doubled since 2006—the year NSHAP Wave 1 was fielded—and has more than tripled since 2001. Moreover, Handelsman (2012) confirms that this huge increase of ART in such a short period of time is not a phenomenon unique to the United States but has occurred in parallel in Australia and perhaps in other highly developed nations worldwide.

Although this rapid increase in hormone therapy rates may explain some of the inconsistencies seen in NSHAP sex hormone data overall, it will also give rise to a wealth of future research endeavors examining its population-level effects on the health and well-being of older Americans. It has already been shown in multiple studies that the current rapidly increasing rates of ART in men have led to increased strength and mobility at older ages (Basaria et al., 2010), but such physical improvements have come with increased rates of adverse cardiovascular events (including myocardial infarctions, heart failure, and stroke), adverse respiratory events, and adverse dermatological events (Basaria et al., 2010; Bhasin et al., 2010; Finkle et al., 2014; Vigen et al., 2013). With the release of NSHAP Wave 2 medication data, we will not only be able to investigate all of the above findings with a nationally representative sample but we will also be able to understand how the availability of hormone therapy is affecting the life decisions of older men, whether they understand the risks they are taking by undergoing hormone therapy and how these risks may be affecting their relationships, both sexual and non-sexual.

Another goal of future research is to identify respondents displaying symptoms of hormone depletion, examine their corresponding measured hormone levels, and attempt to uncover a biologically accurate cause-and-effect relationship. For example, we can identify respondents who are overweight, fatigued, frail, depressed, or have a loss of libido and compare their mean levels of testosterone, estradiol, progesterone, and DHEA in hopes to pinpoint which of these hormone depletions may be most strongly contributing to the telltale “symptoms” of aging. The same can be done for respondents with severe hormone imbalances, such as men with high levels of estradiol compared with testosterone, which is a condition leading to higher risk of
stroke (Abbott et al., 2007). Most studies that have linked certain conditions with hormone depletion and imbalances have relied on serum measurements of hormones (often measuring total and not just the biologically active levels). Therefore, if we are able to make the same ties with salivary measurements, it will further increase the validity of salivary hormone collection as a non-invasive and accurate tool for elucidating health during aging. Most importantly, measuring these hormones will reveal reciprocal links between the social environment, psychological states, and the aging process.

Supplementary Material
Supplementary material can be found at: http://psychsocgerontology.oxfordjournals.org/

Funding
The National Social Life, Health, and Aging Project is supported by the National Institutes of Health, including the National Institute on Aging (R37AG030481; R01AG033903), the Office of Women’s Health Research, the Office of AIDS Research, and the Office of Behavioral and Social Sciences Research (R01AG021487), and by NORC which was responsible for the data collection. M. J. Kozloski was supported by a National Institute on Aging Postdoctoral Traineeship (T32 AG00243).

Acknowledgment
The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Correspondence
Correspondence should be addressed to Michael J. Kozloski, PhD, Department of Sociology, University of Chicago, 1126 East 59th Street, Chicago, IL 60637. E-mail: kozloski@uchicago.edu.

References
Davis, S., & Tran, J. (2001). What are “normal” testosterone levels for women? The Journal of Clinical Endocrinology and Metabolism, 86, 1842–1844. doi:10.1210/jcem.86.4.7436-10


Young, D. G., Skibinski, G., Mason, J. I., & James, K. (1999). The influence of age and gender on serum dehydroepiandrosterone sulphate (DHEA-S), IL-6, IL-6 soluble receptor (IL-6 sR) and transforming growth factor beta 1 (TGF-beta1) levels in normal healthy blood donors. Clinical and Experimental Immunology, 117, 476–481. doi:10.1111/j.1365-2249.1999.01003.x
