Reproducibility Studies and Interlaboratory Concordance for Androgen Assays in Female Plasma

Thomas R. Fears,[1] Regina G. Ziegler, Jennifer L. Donaldson, Roni T. Falk, Robert N. Hoover, Frank Z. Stanczyk, Jimmie B. Vaught, and Mitchell H. Gail

Biostatistics Branch [T. R. F., J. L. D., M. H. G.], Environmental Epidemiology Branch [R. T. F.], Epidemiology and Biostatistics Program [R. N. H.], and Nutritional Epidemiology Branch [R. G. Z.], National Cancer Institute, Bethesda, Maryland 20892; Women's and Children's Hospital, Los Angeles, California 90033 [F. Z. S.]; and MA Bioservices, Rockville, Maryland 20850 [J. B. V.]

Abstract

We conducted studies to determine the magnitude and sources of variability in androgen assay results and to identify laboratories capable of performing such assays for large epidemiological studies. We studied androstanediol (ADIOL), androstanediol glucuronide (ADIOL G), androstenedione (ADION), androsterone glucuronide (ANDRO G), androsterone sulfate (ANDRO S), dehydroepiandrosterone (DHEA), dehydroepiandrosterone sulfate (DHEA S), dihydrotestosterone (DHT), and testosterone (TESTO). A single sample of plasma was obtained from five postmenopausal women, five premenopausal women in the midfollicular phase of the menstrual cycle, and five women in the midluteal phase, divided into aliquots, and stored at −70°. Four sets of two coded aliquots from each woman were then sent to participating labs for analysis at monthly intervals over 4 months.

Using the logarithm of assay measurements, we estimated the components of variance and three measures of reproducibility. The usual coefficient of variation is a function of the components that are under the control of the laboratory. The intraclass correlation between measurements for a given individual is the proportion of the total variability that is associated with individuals. The minimum detectable relative difference is important to evaluate study feasibility. Results suggest that a single sample of ADIOL G, DHEA, DHEA S, and ANDRO G (with two lab replicates per sample) can be used to discriminate reliably among women in a given menstrual phase or menopausal status. The results for DHT, TESTO, ADION, and ANDRO S are more problematic and suggest that the present measurement techniques should be used with care, especially with midluteal phase women. The results for ADIOL suggest that this assay is not yet ready for use in epidemiological studies.

Introduction

Endogenous steroid hormones are believed to play a major role in breast cancer etiology, although a consensus does not yet exist about the precise endocrine patterns that maximize risk (1–3). Estrogens, especially estradiol, stimulate division of breast epithelial cells and have long been linked to the promotion and growth of breast cancer (4). More recently, androgens have also been postulated to be important in breast carcinogenesis, possibly as a source of estrogens, or by other mechanisms (3, 5–7). Epidemiological studies to explore the effects of endogenous androgens require reliable and accurate assays.

To help identify appropriate techniques and laboratories for measuring endogenous hormone in blood and urine samples collected in large epidemiological studies, the reproducibility of several capable laboratories was determined and compared. In earlier reports, Gail *et al.* (8) estimated the sources of variability and reproducibility of assays of estradiol, estrone, estrone sulfate, and progesterone in plasma from pre- and postmenopausal women; and Ziegler *et al.* (9) determined the reproducibility and validity of new measurement techniques for 2-hydroxyestrone and 16α-hydroxyestrone in urine. The present report presents similar results for nine androgens measured in plasma samples from women: ADIOL,[2] ADIOL G, ADION, ANDRO G, ANDRO S, DHEA, DHEA S, DHT, and TESTO. These include the ovarian and adrenal androgens previously analyzed in epidemiological studies of breast cancer, as well as other androgen metabolites of potential importance in breast cancer etiology. We have estimated assay variability over the time required to assay samples from a large epidemiological study by using four measurements spaced over 3 months. Because androgen relationships have been reported to differ for premenopausal and postmenopausal breast cancer, we present variability and reproducibility separately for follicular phase premenopausal women, luteal phase premenopausal women, and postmenopausal women.

Materials and Methods

Experimental Design. Plasma was collected from 15 women who were volunteers from the National Cancer Institute. Five were in the midfollicular phase of the menstrual cycle (4–6 days after the start of menses; mean age, 40 years), five were in the midluteal phase (4–6 days before menses; mean age, 36

Received 6/11/99; revised 1/12/00; accepted 2/10/00.
The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] To whom requests for reprints should be addressed, at National Cancer Institute, Executive Plaza South, Room 8040, MSC 7368, 6120 Executive Boulevard, Bethesda, MD 20892.

[2] The abbreviations used are: ADIOL, androstanediol; ADIOL G, ADIOL glucuronide; ADION, androstenedione; ANDRO G, androsterone glucuronide; ANDRO S, androsterone sulfate; DHEA, dehydroepiandrosterone; DHEA S, DHEA sulfate; DHT, dihydrotestosterone; TESTO, testosterone; CV, coefficient of variation; ICC, intraclass correlation; MDRD, minimum detectable relative difference.

years), and five were postmenopausal (natural menopause; mean age, 56 years). None were taking exogenous hormones. Immediately after separation, plasma was stored at 4°C. Within 4 h of draw, the plasma was mixed, aliquoted, and stored at −70°. Further details on the material acquisition and handling are provided by Gail *et al.* (8).

Each participating lab received four batches of samples, with one batch to be assayed at the beginning of each of 4 consecutive months. Each batch contained two aliquots from each of the 15 subjects. The identifying numbers for the 30 samples within each batch were randomly assigned, separately for each batch. Lab personnel were told only whether a sample was from a premenopausal or postmenopausal woman. Each aliquot was assayed in duplicate. Thus, this study provides information on assay variability among women, among days on which assays were performed, among aliquots, and among lab replicates, but it does not provide information on temporal variations in hormone levels within women.

Laboratory Methods. Four laboratories, two academic and two commercial, recognized for their skill and experience in measuring endogenous hormones, were invited and willing to participate in this study. Each lab was asked to use their standard assay procedures and to perform only those assays with which they had experience.

Laboratory 1. Lab 1 assayed ADIOL G, ADION, DHEA, DHEA S, TESTO, and DHT in plasma. The assay for ADIOL G included organic extraction of the plasma to remove unconjugated 3α ADIOL and other unconjugated steroids, followed by incubation of the aqueous phase with β-glucuronidase. After enzyme hydrolysis and celite chromatography, the product of hydrolysis, ADIOL, was measured by RIA (10, 11). ADION was measured by extracting plasma with ethyl acetate (20%) in hexane, celite column chromatography, and RIA (12–14). DHEA was also determined by extraction with hexane:ethyl acetate (80:20), celite column chromatography, and RIA (15–17). DHEA S was measured by RIA after diluting the specimens 1:2500 with assay buffer (15, 18, 19). TESTO was measured by RIA, preceded by extraction of plasma with ethyl acetate (20%) in hexane and celite column chromatography (20–22). DHT was also measured by RIA involving ethyl acetate:hexane extraction and celite column chromatography (13, 23). Lab 1 reported the sensitivity of the assays to be 3 ng/dl for ADION, 25 ng/dl for ADIOL G, 15 ng/dl for DHEA, 5 μg/dl for DHEA S, 2 ng/dl for TESTO, and 5 ng/dl for DHT.

Laboratory 2. Lab 2 measured ADIOL G, ADION, DHEA, DHEA S, TESTO, and DHT in plasma. ADIOL G was assayed using a method developed at lab 2. Plasma was extracted with a polar solvent. The dried extract was subjected to complete enzymatic hydrolysis, followed by extraction of free ADIOL with hexane:ethyl acetate and purification by high performance liquid chromatography. ADIOL in the purified eluate was quantitated by RIA. ADION was measured by extracting plasma with hexane:ethyl acetate, followed by RIA developed at lab 2. DHEA was also determined by extraction with hexane:ethyl acetate and RIA developed in lab 2. The assay for DHEA S was similar to that for DHEA, except that the initial step was the removal of sulfate by overnight hydrolysis with sulfatase. TESTO was measured by RIA after extraction and column chromatography according to the method of Furuyama *et al.* (24). DHT was measured by an RIA developed at lab 2. Plasma samples were first extracted with hexane:ethyl acetate, followed by treatment with a strong oxidizer to destroy all unsaturated steroids, and purification on alumina columns. Lab 2 reported the sensitivity of the assays to be 14 ng/dl for ADION, 10 ng/dl for ADIOL G, 20 ng/dl for DHEA, 5 μg/dl for DHEA S, 3 ng/dl for TESTO, and 2 ng/dl for DHT.

Laboratory 3. Lab 3 assayed DHEA, DHEAS, ADION, TESTO, DHT, ADIOL, ADIOL G, ANDRO S, and ANDRO G. DHEA S was quantified by direct RIA after a 1000-fold dilution of the plasma sample with assay buffer (25). DHEA, ADION, TESTO, DHT, and ADIOL were measured by RIA after extraction of plasma with diethyl ether and subsequent purification by celite column chromatography (26–29). ADIOL G was quantified directly in plasma using a validated commercial kit (Diagnostic Systems Laboratories, Webster, Texas; Ref. 30). ANDRO S and ANDRO G were measured after unconjugated steroids were removed by extraction with diethyl ether, and the remaining conjugated steroids were hydrolyzed using hydrochloric acid and β-glucuronidase to cleave the sulfate and glucuronide moieties, respectively (31). In both assays, the product of hydrolysis, androsterone, was quantified by RIA after extraction with ethyl acetate and purification by celite column chromatography. Lab 3 reported the sensitivity of the assays to be as follows: 2 μg/dl for DHEA S, 20 ng/dl for DHEA, 10 ng/dl for ADION, 4 ng/dl for TESTO, 2 ng/dl for DHT, 2 ng/dl for ADIOL, 5 ng/dl for ADIOL G, 3 ng/dl for ANDRO G, and 6 ng/dl for ANDRO S.

Laboratory 4. Lab 4 performed measurements of ADION, DHEA, DHEA S, and TESTO in plasma. ADION was measured by carbon tetrachloride extraction of plasma followed by an RIA kit (ICN Biochemicals, Diagnostics Division). DHEA was measured by dichloromethane extraction and an RIA kit from Coat-A-Count. DHEA S was assayed directly in plasma using a double antibody RIA kit from ICN Biochemicals, Diagnostics Division. TESTO was also measured directly in plasma using an ICN RIA kit. The sensitivity of the assays, as reported in the kit documentation, was 0.1 ng/ml for ADION, 0.04 ng/ml for DHEA, 0.5 ng/ml for DHEA S, and 0.1 ng/ml for TESTO.

Statistical Methods. Measurements were analyzed on the natural logarithmic scale. This transformation reduces the dependence of the SD of the response on the mean so that variance can be assumed to be unrelated to subject. For each of the three groups of women (midfollicular, midluteal, and postmenopausal), a nested component of variance analysis was performed. Components were estimated for subjects (σ^2_a), month (σ^2_b), aliquots on the same day (σ^2_c), and replicates from the same aliquot (σ^2). Letting z_{ijkl} denote the hormone measurement for woman i (i = 1,2,3,4,5) on analysis day $j(i)$ (j = 1,2,3,4), using aliquot $k(ij)$ (k = 1,2) and replicate $l(ijk)$ (l = 1,2), the statistical model is written (1):

$$log_e(z_{ijkl}) = \mu + a_i + b_{j(i)} + c_{k(ij)} + \epsilon_{l(ijk)} \qquad (1)$$

where log_e denotes the natural logarithm (base e). In the model, μ is the average level of the hormone, and a_i, $b_{j(i)}$, $c_{k(ij)}$, and $\epsilon_{l(ijk)}$ are normal independent variates with means zero and variances σ^2_a, σ^2_b, σ^2_c, and σ^2, respectively. Restricted maximum likelihood estimates of the variance components were obtained using the SAS procedure PROC VARCOMP (32). The procedure also provides estimates of the SE of the estimated variance components. Restricted maximum likelihood estimates cannot be less than zero, and they agree with the usual ANOVA estimates when all estimates are greater than zero.

Knowledge of the variance components allows a careful quantitative consideration of assay reproducibility. We use three measures of reproducibility derived from these components: the CV, the ICC coefficient, and the MDRD. The three

measures are quite different, but each measure is useful depending on the application.

The common measure of reproducibility used by the labs is the CV, namely the population SD of a measurement divided by its mean. The components associated with day, aliquot, and replicate are the components that are under the control of the lab. An application of the δ method (33) shows that the sum of these components is a good estimate of the square of the CV (8). Because the labs all used two replicates, the CV expressed as a percentage is estimated by $100(\sigma^2_b + \sigma^2_c + \sigma^2/2)^{1/2}$, where hats denote estimates of corresponding parameters. This CV incorporates the variation associated with day and may be much larger than a CV based only on variability due to aliquots and replicates on a single day. The validity of this approximation depends heavily on the model and particularly on the assumption that after logarithmic transformation the variance is unrelated to subject.

The assay would not be useful if the differences in true values between subjects were not large compared to total assay variability. For this to be the case, the variability associated with subjects should be large compared to the variability under the control of a lab. It is appropriate then to compare σ^2_a, the component associated with subjects, with the sum of all components. The ratio is close to unity when the biological component is large relative to the components associated with the lab. In fact, the ICC between measurements on different days from a given individual is exactly this ratio. If two replicates are used for each sample, the estimated ICC between two measurements on different days is ICC = $\sigma^2_a/(\sigma^2_a + \sigma^2_b + \sigma^2_c + \sigma^2/2)$. We express the ICC in percent by multiplying by 100. If σ^2_a is small, the ICC may not be near one, even when the CV is small. The ICC is of importance to the epidemiologist because it indicates the effect of measurement error on study results. Specifically, regression analyses relating the log relative risk of disease to the log hormone assay level will be attenuated by the ICC. If the ICC is 0.90, there will be a downward bias that is slight, only 10%, but an ICC of <0.80 results in bias that may be important.

Assay variability can decrease the power of a study to detect a difference in hormone levels between cases and controls. Knowing the variance components allows one to determine the minimum difference that is reliably detected with a given number of cases and controls. Specifically, for a two-sided $\alpha = 0.05$ level test, the minimum difference in average log assay values ($\delta = \mu_1 - \mu_2$) detectable with power 0.90 is the solution to

$$\delta^2 = (\sigma^2_a + \sigma^2_b + \sigma^2_c + \sigma^2/2)(1/n_1 + 1/n_2)(1.96 + 1.282)^2$$

where 1.96 is the 97.5th percentile of the standard normal distribution, 1.282 is the 90th percentile, and n_1 and n_2 are the case and control sample sizes. From δ, one can calculate the minimum percentage difference detectable with power 0.90 as $100\{\exp(\mu_1) - \exp(\mu_2)\}/\exp(\mu_2) = 100\{\exp(\delta) - 1\}$. We call this quantity the MDRD. Usually an investigator has a sense of what differences exist and what sample sizes can be used so the MDRD is useful. If σ^2_a is small, the value of δ and therefore the MDRD may be small even when the CV is large.

For each hormone and laboratory, we examined graphs of grand means, daily means, and aliquot means and examined these quantities for stability over time. We also examined differences in variability and agreement of results among the lab assays. These graphs are presented for ADIOL G and for DHT. Graphs for other androgens are available upon request. Spearman rank correlations are used to measure concordance of grand means among lab assays. The estimated components of variance and SEs of the estimates are tabulated in the "Appendix." The components of variance are used to obtain estimates of the CVs, ICCs, and MDRDs, which are compared among the labs. To calculate ICCs and CVs in Tables 1 and 2, we assume that the measurement used is the mean of the two logarithmic-transformed replicates. To calculate MDRD, we assume, in addition, that $n_1 = 300$ cases and $n_2 = 600$ controls are used; these numbers approximate the sample sizes available in an ongoing study of Asian-American women that motivated these assay reliability studies.

Results

ADIOL G. Fig. 1, *a*, *b*, and *c* displays the results for Log_e-(ADIOL G), respectively, among the five midfollicular phase women, five midluteal phase women, and five postmenopausal women. Each woman is represented by a different symbol. The leftmost symbols represent the grand means of the 16 measurements for each of the five women. The next symbols represent the daily means of the four measurements for each woman on analysis day 1, and connected to each of these symbols are symbols that represent the aliquot means of the two replicate readings from each of the aliquots on that day. Results for the four measurements for each woman on analysis days 2, 3, and 4 are then presented in a similar manner.

No consistent time trends were evident, although there was a decreasing trend in measurements from lab 2 for some midfollicular phase women (Fig. 1*a*). There was overlap of the women's aliquot means from labs 1 and 2 in Fig. 1, *a*, *b*, and *c*; thus the ranks of the subjects' aliquot means were not completely consistent over time. Results from lab 3 were more consistent in that aliquot means were well separated for women in all phases, except for minimal overlap of two for midfollicular and midluteal phase women.

The geometric means of all ADIOL G measurements were 151 ng/dl at lab 3, 74.4 at lab 1, and 64.1 ng/dl at lab 2. These differences are not surprising because the labs do not correct for molecular weight differences, hydrolysis, and procedural losses in the same way. The correlations of the ranks of the subjects' mean responses were 0.94 between labs 1 and 2, but only 0.88 between labs 1 and 3 and 0.80 between labs 2 and 3.

For midfollicular and midluteal women, the CVs ranged from 13 to 17% at labs 1 and 3 but were somewhat higher, about 30%, at lab 2 (Table 1). The CV was 16% for measurements from postmenopausal women at lab 1 and much higher at labs 2 and 3 (25–35%). The ICCs were all >80% and >90% for labs 1 and 3. The estimated MDRDs were 14–18% using measurements from lab 1 and somewhat larger for labs 2 and 3 (Table 1). Estimates of individual variance components and their SEs are provided in Table A1 of the "Appendix."

DHEA. No definite time trends were evident (figure not shown). There was some overlap of aliquot means for subjects in all groups. The ranks of the subjects' mean responses were highly correlated for all lab pairs. Correlations ranged from 0.98 to 0.99. The geometric means of all measurements of DHEA were 158.9, 199.5, 187.9, and 101.4 ng/dl at labs 1, 2, 3, and 4, indicating somewhat lower levels for lab 4.

The CVs were the smallest from lab 1, ranging from 8 to 9%, and somewhat larger from the other labs, ranging from 13 to 28% (Table 1). The CV was small for midluteal phase measurements from each lab. The ICCs were largest for measurements at lab 1, 97–99%, and somewhat smaller for other labs

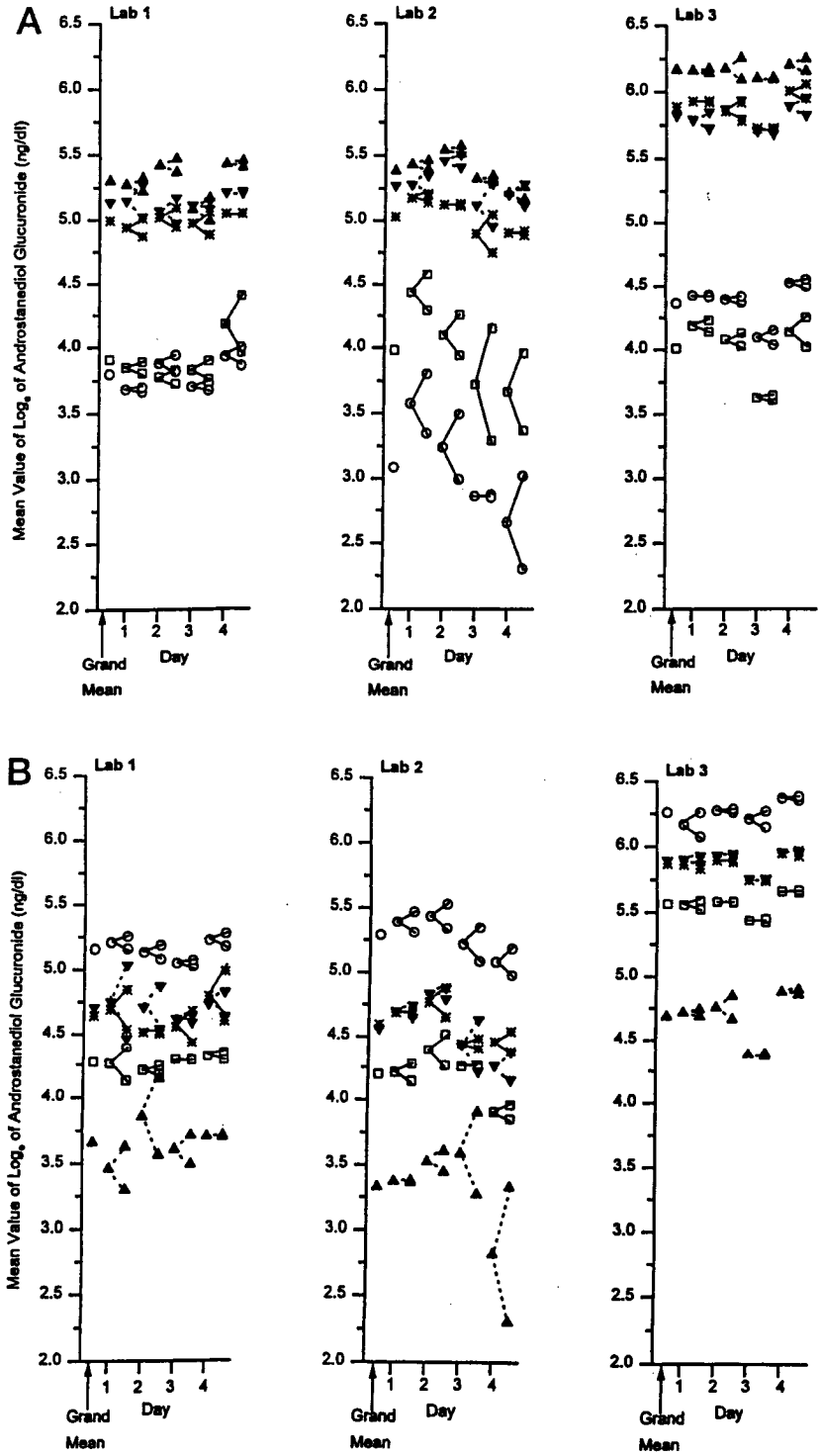

Fig. 1. *A*, ADIOL G measurements in midfollicular phase women. *B*, ADIOL G in midluteal phase women. *C*, ADIOL G postmenopausal women.

(87–94%). The MDRDs were smallest for midluteal phase measurements, 9–13%, and larger for other phases, 14–21%.

DHEA S. No time trends were evident in measurements from any lab in any phase (figure not shown). There was some overlap in assay measurements from each lab for women in each phase. The ranks of the subjects' mean responses were highly correlated (0.98–1.00) across labs. The mean levels of DHEA S, 85.3, 61.1, 95.1, and 76.6 μg/dl at labs 1, 2, 3 and 4, were statistically significantly different for all pairs of labs.

The CVs were lowest from lab 4 (7–10%), slightly higher from lab 1 (10–12%), and still higher at the other labs (11–19%; Table 1). The ICCs were >96% at labs 1 and 4 and slightly lower but still very high at labs 2 and 3 (92–95%). The MDRD ranged from 12 to 19%.

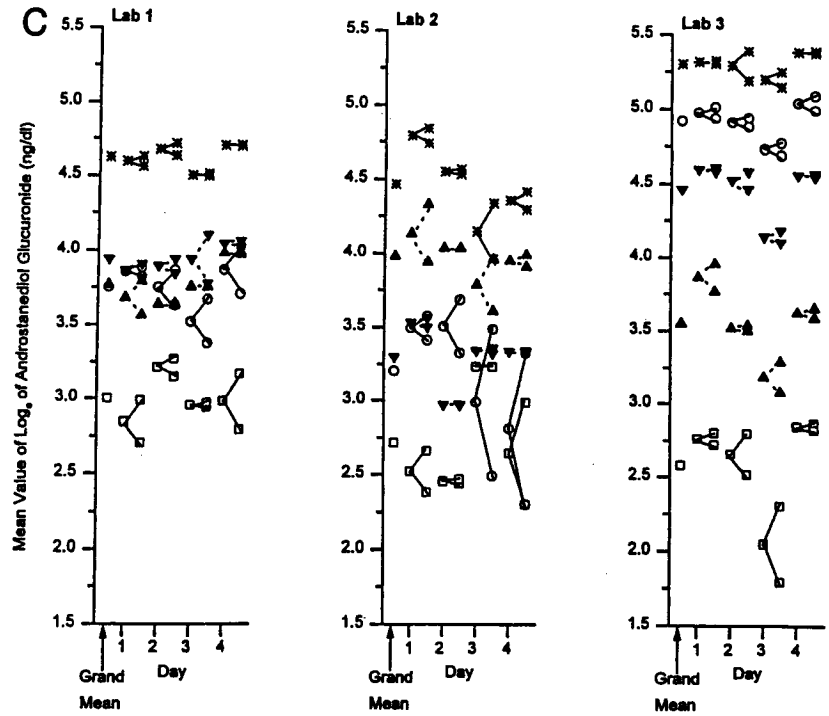

Fig. 1. C continued.

DHT. There were no consistent time trends (Fig. 2, *a-c*). There was some overlap of aliquot means of midfollicular women, but there was substantial overlap of aliquot means of midluteal and postmenopausal women. In particular, there were large differences between aliquot means of midluteal women at labs 2 and 3. For all laboratories, the ranks of the subjects' mean responses were highly correlated (0.97–0.99). The geometric mean levels of DHT were similar for labs 1 and 3 (8.53 and 8.22 ng/dl) and somewhat lower for lab 2 (6.56 ng/ml).

The CVs were smallest for midfollicular and midluteal measurements, 11–17%, and slightly larger for postmenopausal measurements, 18–21% (Table 1). The ICC was 93–99% for

Table 1 Estimated CVs, ICC coefficients, and MDRDs (assuming 300 cases and 600 controls) for androgens assayed at multiple laboratories[a]

	Lab 1			Lab 2			Lab 3			Lab 4		
	CV	ICC	MDRD	CV	ICC	MDRD	CV	ICC	MDRD	CV	ICC	MDRD
ADOL G												
Midfollicular	15.0	95.8	18.2	32.0	90.3	26.9	16.6	97.2	25.8			
Midluteal	17.1	91.4	14.4	28.7	85.8	19.2	13.1	95.4	14.9			
Postmenopausal	16.1	92.7	14.8	34.7	79.1	19.4	25.3	94.9	29.6			
DHEA												
Midfollicular	9.3	98.0	16.3	16.3	91.8	13.9	15.5	93.8	15.4	21.3	92.4	19.4
Midluteal	8.6	96.6	11.3	13.2	86.7	8.7	13.8	90.0	10.5	18.3	88.3	13.0
Postmenopausal	8.0	98.9	19.2	22.6	89.5	17.3	19.6	93.3	18.9	28.1	88.2	20.7
DHEA S												
Midfollicular	10.9	98.0	19.1	19.1	93.8	19.2	17.1	92.8	15.8	10.1	98.1	18.2
Midluteal	11.5	97.5	17.9	17.8	92.4	16.0	15.9	92.9	14.6	6.9	99.0	16.7
Postmenopausal	10.3	96.7	14.0	18.1	91.9	15.7	11.3	95.1	12.2	6.8	98.6	14.1
DHT												
Midfollicular	12.3	93.4	11.5	7.6	98.7	16.2	15.4	93.8	15.3			
Midluteal	11.4	22.9	3.0	13.7	0.0	3.2	16.5	0.0	3.9			
Postmenopausal	21.1	81.9	12.0	18.2	80.5	9.9	20.0	82.0	11.4			
TESTO												
Midfollicular	13.3	83.9	7.9	23.2	83.9	14.2	13.5	87.9	9.3	17.8	84.8	11.0
Midluteal	13.4	69.0	5.7	22.8	54.9	8.1	12.9	74.9	6.1	26.2	65.7	10.8
Postmenopausal	11.6	91.2	9.3	21.1	87.5	14.7	9.0	95.6	10.2	20.7	87.7	14.5
ADION												
Midfollicular	12.8	91.2	10.4	24.7	71.7	11.2	18.0	85.0	11.4	8.8	95.9	10.4
Midluteal	10.1	20.7	2.7	16.6	21.1	4.4	15.8	0.0	3.7	8.3	71.9	3.7
Postmenopausal	11.9	91.8	9.5	40.6	57.5	15.3	19.2	82.1	11.0	10.4	95.1	11.3

[a] All estimates assume that two replicates are averaged to produce a single measurement.

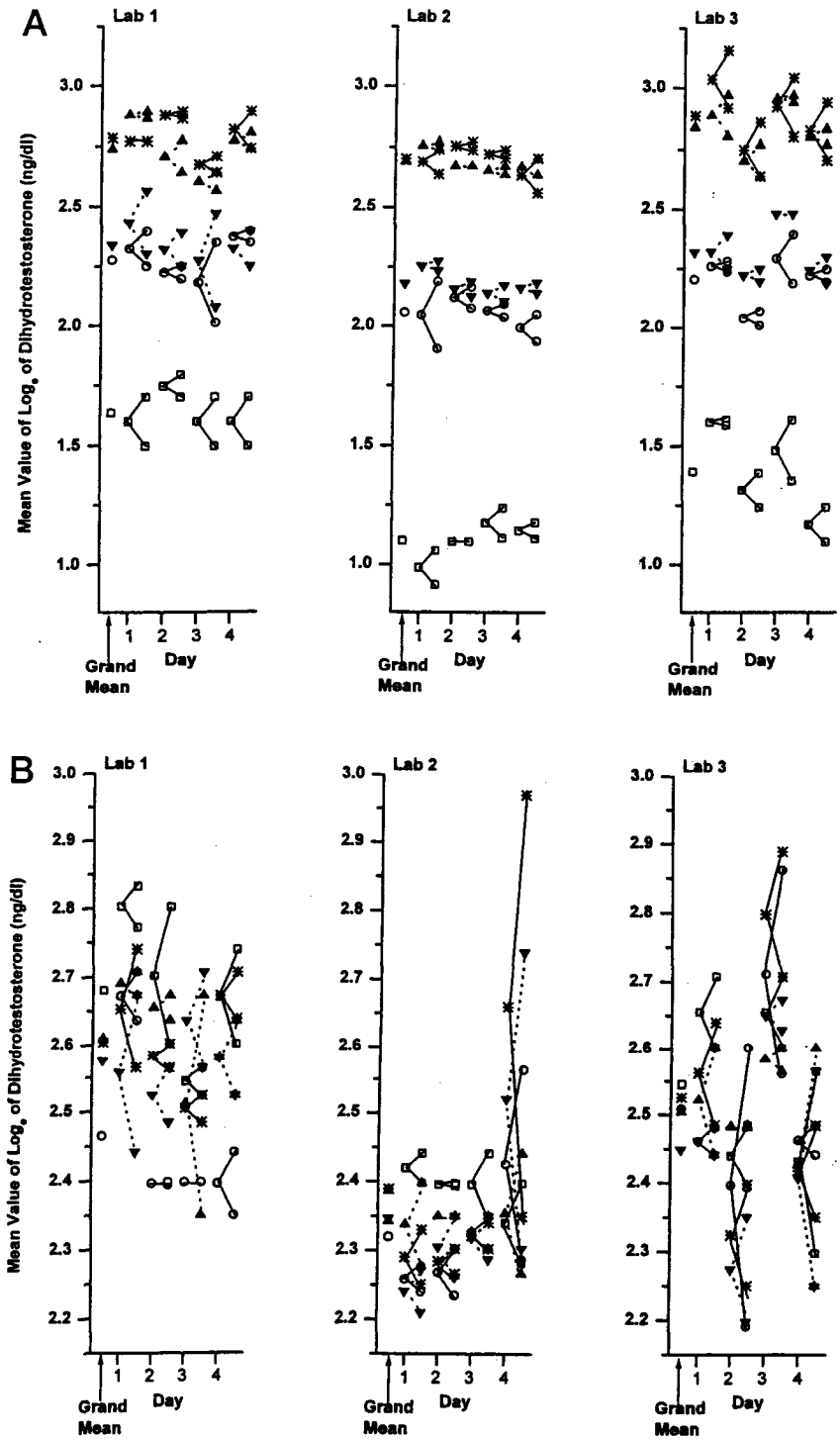

Fig. 2. A, DHT measurements in midfollicular phase women. *B,* DHT measurements in midluteal phase women. *C,* DHT measurements in postmenopausal women.

midfollicular measurements; slightly lower for postmenopausal measurements, 81–82%; and very low for midluteal measurements, 0–23%. The ICC was zero when the observed variability between subjects was very small compared to the variability associated with the lab, *i.e.,* when the variability between subjects could be completely explained by the within lab variability. The MDRDs were 10–16% for midfollicular and postmenopausal measurements, but only 3–4% for midluteal measurements. These very small MDRDs occurred because the total variability was quite small.

TESTO. There were no clear time trends in the measurements for TESTO (figure not shown). The aliquot means from lab 3 were clearly separated for postmenopausal women; there was some overlap in the aliquot means from other labs. There was

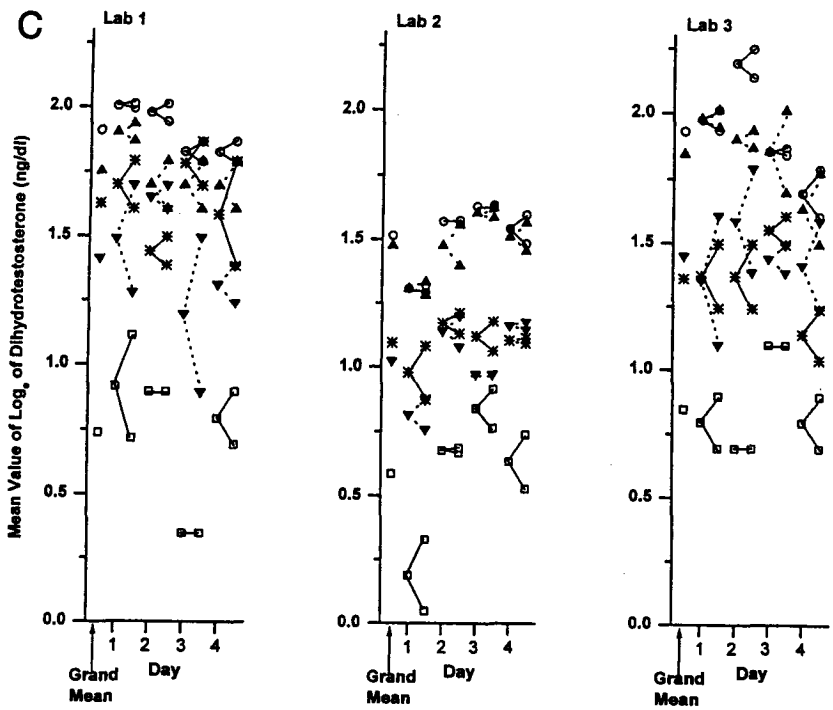

Fig. 2. C Continued.

extensive overlap of aliquot means from all labs for midfollicular and midluteal women. The ranks of the subjects' mean responses were highly correlated (0.90–1.00) for labs 1, 2, and 3. Correlations with lab 4 were smaller, −0.10 to 0.20 for midluteal and postmenopausal women and 0.80–0.90 for midfollicular women. Geometric mean levels of TESTO, 18.5, 16.0, 19.1, and 18.8 ng/dl, were significantly different at labs 1, 2, and 3.

The CVs differed by lab (Table 1). CVs for measurements from labs 1 and 3 were 9–14%, whereas those from labs 2 and 4 were 21–26%. ICCs ranged from 0.84 to 0.88 for midfollicular women, 0.55–0.75 for midluteal women, and 0.88–0.96 for postmenopausal women. The ICCs reflected the large variability of measurements between postmenopausal women and small variability of the measurements between midluteal women. The MDRDs were smaller for the midluteal measurements, 6–11%, and somewhat larger for midfollicular or postmenopausal measurements, 8–15%.

ADION. No time trends were evident in the data for lab 1 (figure not shown). Although there was some overlap of aliquot means of the midfollicular and postmenopausal women, measurements at the high levels were clearly separated from those at lower levels. The grand means of the midluteal women had a narrow range, and there was greater overlap with poor separation. There was an increasing trend with time in means for all menstrual phases in the data from lab 2. There was also overlap and poor separation for all phases. The daily means from lab 3 showed substantial variability because measurements on day 3 were consistently lower than comparable measurements on other days. Measurements from lab 4 showed no time trends. There was no overlap of aliquot means and clear separation for midfollicular phase women. However, there was substantial overlap for midluteal women and some overlap for postmenopausal women with only the highest and lowest measurements clearly separated.

The ranks of the subjects' mean responses for midluteal and postmenopausal women were highly correlated (0.90–1.00), but not the ranks for midfollicular women (0.30–0.90). The geometric means of ADION were 60.0, 62.9, 55.7, and 65.6 ng/dl and were significantly lower from labs 1 and 3 than from labs 2 and 4. Labs 2 and 3 had lower ICCs and higher CVs than labs 1 and 4 (Table 1). Estimated MDRDs were slightly larger at lab 2 than at other labs. At all labs, ICCs and MDRDs were much smaller for women in the midluteal menstrual phase than for women in the midfollicular and postmenopausal phases.

ANDRO G. One lab provided measurements for ANDRO G levels. There were no definite time trends (figure not shown). Although there was some overlap of aliquot means for midfollicular phase women, the separation was clear. There was some overlap among the aliquot means for midluteal women, and only those with highest and lowest measurements were clearly separated. For postmenopausal women, there was considerable overlap, and only the woman with the lowest measurements were separated.

The estimated CVs were around 20% for midfollicular and midluteal women and somewhat higher, 33%, for postmenopausal women (Table 2). The ICCs were high, >85% (Table 2). The MDRD was about 14% for midluteal samples, but >20% for postmenopausal and midfollicular samples. Estimates of individual variance components and their SEs are provided in Table A2 of the "Appendix."

ANDRO S. There was considerable overlap in aliquot means for all groups (figure not shown). The estimated CVs (Table 2) were 25–30%. The ICCs were 88–92% for midluteal and postmenopausal measurements, but they were much lower, 64%, for midfollicular measurements. The MDRDs for midluteal and postmenopausal women were 22%, but only 10% for midfollicular women. The component of variance for subjects, σ_a^2, was relatively small for these midfollicular women, resulting in small total variation and therefore a small MDRD.

Table 2 Estimated CVs, ICC coefficients, and MDRDs (assuming 300 cases and 600 controls) for androgens assayed only at lab 3[a]

	ANDRO G			ANDRO S			ADIOL		
	CV	ICC	MDRD	CV	ICC	MDRD	CV	ICC	MDRD
Midfollicular	20.9	95.9	26.7	24.8	63.6	9.9	45.9	12.1	11.9
Midluteal	17.9	89.9	13.7	29.3	88.4	21.8	42.7	2.0	10.4
Postmenopausal	33.3	86.4	23.0	24.7	91.7	21.8	79.2	11.3	21.3

[a] All estimates assume that two replicates are averaged to produce a single measurement.

ADIOL. Only one lab carried out assays for ADIOL. No definite time trends were evident for any phase (figure not shown). There was great overlap of aliquot means for all groups. The range of aliquot means for a particular woman on 1 day can be large. The ICCs were only 12.1, 2.0, and 11.3 for midfolicular, midluteal, and postmenopausal women, respectively (Table 2). The CVs for the ADIOL assay were large, 43–79% (Table 2). Nevertheless, because the total variability in the data were small, the MDRDs were 10–21% (Table 2).

Discussion

Epidemiological field studies that are planned to evaluate associations between serum hormone levels and risk of cancer may require that many samples be analyzed over a period of months or years. The degree of variability in hormone assay results should be small enough so that the assay is likely to detect differences between cases and controls. In this study, we have obtained data on components of variability in androgen assay results. Such data allow one to assess the reproducibility of these assays and the measurements for epidemiological studies.

ADIOL G, DHEA, DHEA S, DHT, TESTO, and ADION were assayed in several laboratories. There was variation in the mean assay levels among the participating labs, but the correlations of rankings of subjects' mean results among the labs were high. The CVs were fairly high and did not vary widely by menstrual or menopausal states. The CVs for measurements from lab 1 were usually <15% but ranged to 20%, whereas those from lab 2 were usually <20% but ranged as high as 40%.

The ICC was 100 times the ratio of the biological variability among women to the total variability, including sources of variation associated with lab procedures. Values of ICC near 100 indicated that lab variability was small compared to biological variability. ICC values for ADIOL G, DHEA, and DHEAS exceeded 90% for lab 1 and lab 3 and usually exceeded 85% for labs 2 and 4. At all labs, the ICCs usually exceeded 80% for DHT, TESTO, and ADION in postmenopausal and midfollicular phase women. For midluteal phase women, the biological variability among women was small, and there were lower values of the ICC; specifically, they were <70% for TESTO and <22% for DHT and ADION.

Another way to assess the utility of these assays is to determine the minimal detectable relative difference in percent, MDRD, that can be detected in a case-control study. The comparison in this report was based on 300 cases and 600 controls, approximately the size of the study we are contemplating. For ADIOL G, DHEA, DHEA S, DHT, TESTO, and ADION, the MDRDs for this design were <20% at labs 1 and 4, and <30% at labs 2 and 3. The MDRDs were smallest for DHT, TESTO, and ADION, the assays for which the biological variability among women was particularly small.

Lab 3 was the only lab that volunteered to assay ANDRO G, ANDRO S, and ADIOL. For ANDRO G and ANDRO S, the CVs ranged from 18% to 33%, whereas the CVs for ADIOL were very high (43–79%). Levels of the latter hormone were very low (usually <100 pg/ml) in both premenopausal and postmenopausal women. The ICCs ranged from 86–97% for ANDRO G and 64–92% for ANDRO S, but were <15% for ADIOL. For each of these assays, the MDRDs ranged from 10 to 25% for a study with 300 cases and 600 controls.

The CV is useful for lab quality control, whereas the ICC and MDRD are more important in determining the feasibility of an epidemiological study. If the variation among subjects is large, the ICC may be large even if the CV is large. If the ICC is large, estimates of the slope of the log relative risk on log (hormone) will suffer little attenuation from lab measurement error, and required sample sizes will be minimally inflated from lab measurement error. On the other hand, the MDRD depends on all sources of variability. If variation among subjects is small, the MDRD may be small enough to justify an epidemiological study even if the CV is large, provided the ICC is not too small. Conversely, a study can be impeded by small values of ICC and large values of MDRD, even when the CV is small.

Estimates of the components of variance allow one to identify the aspects of lab procedures that lead to increased variability and to learn how to efficiently allocate resources to improve assay reproducibility. For example, if there were more variation among aliquots than among replicates, one might increase the number of aliquots and decrease the number of replicates. The total effort would not change, but the CV would decrease, the ICC would increase, and the MDRD would decrease. For those interested in design issues, the estimated components of variance and their SEs for each of the androgen assays are given in the "Appendix."

Our study used aliquots from a woman whose blood was drawn on a single day, so our estimates of subject variation for a premenopausal or postmenopausal woman include both the between subject variation and the secular variation for a given woman in the middle of that phase. This reliability study design is entirely appropriate for the typical case-control study, which uses only one sample per subject. These data do not allow us to estimate the component of variation that corresponds to repeated blood samples taken over time from the same woman.

Lab 1 usually exhibited smaller CVs, higher ICCs, and smaller MDRDs than the other labs. Lab 3 also exhibited relatively small CVs, high ICCs, and small MDRDs while also providing results on more hormones than the other labs.

The present study used only five women in each men opausal or menstrual phase. Larger numbers of women would be desirable to estimate ICCs and other parameters with greater precision. This study provided valuable guidance, nonetheless, for designing epidemiological studies. These data suggest that a single sample (with two lab replicates per sample) of ADIOL G, DHEA, DHEA S, and ANDRO G can be used to discriminate reliably among women in a given menstrual phase or menopausal status. The results for DHT, TESTO, ADION, and

Appendix

Table A1 Estimated variance components and SEs for androgens assayed at multiple laboratories

	Midfollicular phase women				Midluteal phase women				Postmenopausal women			
	Lab 1	Lab 2	Lab 3	Lab 4	Lab 1	Lab 2	Lab 3	Lab 4	Lab 1	Lab 2	Lab 3	Lab 4
ADIOL G												
Subject (σ_a^2)	0.5076	0.9595	0.9657		0.3106	0.4975	0.3491		0.3299	0.4559	1.1960	
	(0.1571)	(0.3001)	(0.2985)		(0.0965)	(0.1571)	(0.1084)		(0.1026)	(0.1462)	(0.3716)	
Analysis day (σ_b^2)	0.0100	0.0394	0.0221		0.0000	0.0307	0.0144		0.0091	0.0403	0.0473	
	(0.0027)	(0.0121)	(0.0039)			(0.0097)	(0.0025)		(0.0030)	(0.0139)	(0.0089)	
Aliquot (σ_c^2)	0.0103	0.0441	0.0000		0.0268	0.0448	0.0012		0.0114	0.0566	0.0000	
	(0.0017)	(0.0089)			(0.0031)	(0.0072)	(0.0004)		(0.0024)	(0.0113)		
Replicates (σ^2)	0.0043	0.0384	0.0109		0.0051	0.0142	0.0029		0.0109	0.0476	0.0338	
	(0.0004)	(0.0037)	(0.0009)		(0.0005)	(0.0014)	(0.0003)		(0.0011)	(0.0046)	(0.0027)	
DHEA												
Subject (σ_a^2)	0.4264	0.2945	0.3648	0.5542	0.2108	0.1148	0.1718	0.2507	0.5798	0.4350	0.5331	0.5919
	(0.1313)	(0.0922)	(0.1133)	(0.1722)	(0.0650)	(0.0365)	(0.0541)	(0.0785)	(0.1783)	(0.1365)	(0.1659)	(0.1868)
Analysis day (σ_b^2)	0.0001	0.0190	0.0093	0.0065	0.0000	0.0150	0.0155	0.0055	0.0000	0.0241	0.0186	0.0506
	(0.0009)	(0.0036)	(0.0028)	(0.0049)	(0)	(0.0026)	(0.0027)	(0.0036)	(0)	(0.0062)	(0.0047)	(0.0105)
Aliquot (σ_c^2)	0.0070	0.0054	0.0103	0.0385	0.0059	0.0010	0.0017	0.0272	0.0047	0.0184	0.0113	0.0275
	(0.0012)	(0.0010)	(0.0021)	(0.0054)	(0.0008)	(0.0004)	(0.0005)	(0.0038)	(0.0007)	(0.0038)	(0.0028)	(0.0039)
Replicates (σ^2)	0.0029	0.0042	0.0092	0.0011	0.0032	0.0030	0.0035	0.0014	0.0033	0.0167	0.0167	0.0021
	(0.0003)	(0.0004)	(0.0009)	(0.0001)	(0.0003)	(0.0003)	(0.0003)	(0.0001)	(0.0003)	(0.0016)	(0.0016)	(0.0002)
DHEA S												
Subject (σ_a^2)	0.5669	0.5516	0.3787	0.5220	0.5036	0.3873	0.3282	0.4480	0.3140	0.3718	0.2394	0.3256
	(0.1746)	(0.1717)	(0.1183)	(0.1608)	(0.1553)	(0.1212)	(0.1025)	(0.1378)	(0.0969)	(0.1161)	(0.0743)	(0.1002)
Analysis day (σ_b^2)	0.0006	0.0226	0.0219	0.0024	0.0034	0.0261	0.0195	0.0015	0.0011	0.0174	0.0090	0.0021
	(0.0013)	(0.0048)	(0.0041)	(0.0011)	(0.0015)	(0.0046)	(0.0036)	(0.0005)	(0.0011)	(0.0041)	(0.0017)	(0.0006)
Aliquot (σ_c^2)	0.0089	0.0124	0.0054	0.0071	0.0064	0.0051	0.0013	0.0027	0.0065	0.0122	0.0013	0.0015
	(0.0016)	(0.0019)	(0.0010)	(0.0011)	(0.0014)	(0.0008)	(0.0009)	(0.0005)	(0.0013)	(0.0021)	(0.0005)	(0.0003)
Replicates (σ^2)	0.0048	0.0030	0.0039	0.0013	0.0066	0.0008	0.0088	0.0012	0.0058	0.0063	0.0042	0.0019
	(0.0005)	(0.0003)	(0.0004)	(0.0001)	(0.0006)	(0.0001)	(0.0009)	(0.0001)	(0.0006)	(0.0006)	(0.0004)	(0.0002)
DHT												
Subject (σ_a^2)	0.2118	0.4251	0.3606		0.0039	0.0000	0.0000		0.2014	0.1369	0.0579	
	(0.0656)	(0.1308)	(0.1121)		(0.0019)	(0)	(0)		(0.0642)	(0.0443)	(0.0579)	
Analysis day (σ_b^2)	0.0002	0.0009	0.0121		0.0048	0.0000	0.0088		0.0164	0.0258	0.0135	
	(0.0016)	(0.0006)	(0.0029)		(0.0015)	(0)	(0.0028)		(0.0052)	(0.0047)	(0.0046)	
Aliquot (σ_c^2)	0.0113	0.0025	0.0039		0.0069	0.0172	0.0157		0.0112	0.0034	0.0127	
	(0.0021)	(0.0007)	(0.0018)		(0.0011)	(0.0019)	(0.0026)		(0.0042)	(0.0011)	(0.0039)	
Replicates (σ^2)	0.0070	0.0049	0.0153		0.0028	0.0034	0.0056		0.0337	0.0078	0.0277	
	(0.0007)	(0.0005)	(0.0015)		(0.0003)	(0.0003)	(0.0005)		(0.0033)	(0.0008)	(0.0027)	
TESTO												
Subject (σ_a^2)	0.0921	0.2810	0.1326	0.1763	0.0399	0.0630	0.0494	0.1314	0.1383	0.3120	0.1723	0.3048
	(0.0291)	(0.0898)	(0.0416)	(0.0561)	(0.0129)	(0.0226)	(0.0161)	(0.0430)	(0.0431)	(0.0988)	(0.0532)	(0.0958)
Analysis day (σ_b^2)	0.0026	0.0373	0.0039	0.0197	0.0000	0.0320	0.0071	0.0000	0.0033	0.0333	0.0002	0.0150
	(0.0019)	(0.0073)	(0.0020)	(0.0041)	(0)	(0.0068)	(0.0020)	(0)	(0.0015)	(0.0062)	(0.0008)	(0.0050)
Aliquot (σ_c^2)	0.0125	0.0131	0.0024	0.0110	0.0150	0.0161	0.0039	0.0676	0.0078	0.0055	0.0059	0.0264
	(0.0021)	(0.0023)	(0.0023)	(0.0016)	(0.0019)	(0.0027)	(0.0014)	(0.0071)	(0.0014)	(0.0016)	(0.0011)	(0.0038)
Replicates (σ^2)	0.0051	0.0072	0.0238	0.0018	0.0059	0.0073	0.0110	0.0019	0.0045	0.0114	0.0039	0.0027
	(0.0005)	(0.0007)	(0.0023)	(0.0002)	(0.0006)	(0.0007)	(0.0011)	(0.0002)	(0.0004)	(0.0011)	(0.0004)	(0.0003)
ADION												
Subject (σ_a^2)	0.1700	0.1550	0.1896	0.1769	0.0027	0.0073	0.0000	0.0178	0.1411	0.2229	0.1691	0.2059
	(0.0531)	(0.0520)	(0.0605)	(0.0548)	(0.0013)	(0.0040)	(0)	(0.0058)	(0.0439)	(0.0807)	(0.0541)	(0.0638)
Analysis day (σ_b^2)	0.0067	0.0532	0.0263	0.0040	0.0019	0.0155	0.0149	0.0007	0.0000	0.1479	0.0200	0.0030
	(0.0020)	(0.0091)	(0.0047)	(0.0010)	(0.0011)	(0.0035)	(0.0029)	(0.0007)	(0.0000)	(0.0248)	(0.0046)	(0.0012)
Aliquot (σ_c^2)	0.0084	0.0000	0.0030	0.0029	0.0071	0.0087	0.0071	0.0056	0.0121	0.0082	0.0090	0.0066
	(0.0013)	(0)	(0.0009)	(0.0005)	(0.0012)	(0.0017)	(0.0014)	(0.0009)	(0.0015)	(0.0025)	(0.0024)	(0.0011)
Replicates (σ^2)	0.0027	0.0158	0.0060	0.0016	0.0026	0.0066	0.0057	0.0014	0.0044	0.0173	0.0158	0.0023
	(0.0003)	(0.0013)	(0.0006)	(0.0002)	(0.0002)	(0.0006)	(0.0006)	(0.0001)	(0.0004)	(0.0017)	(0.0015)	(0.0002)

Table A2 Estimated variance components and SEs for androgens assayed only by laboratory 3

	ANDRO G			ANDRO S			ADIOL		
	Midfollicular phase	Midluteal phase	Postmenopausal	Midfollicular phase	Midluteal phase	Postmenopausal	Midfollicular phase	Midluteal phase	Postmenopausal
Subject (σ_a^2)	1.0224	0.2827	0.7073	0.1076	0.6553	0.6764	0.0290	0.0037	0.0797
	(0.3162)	(0.0880)	(0.2215)	(0.0360)	(0.2048)	(0.2112)	(0.0222)	(0.0143)	(0.0604)
Analysis day (σ_b^2)	0.0142	0.0000	0.0005	0.0294	0.0000	0.0294	0.1102	0.1245	0.2430
	(0.0050)		(0.0116)	(0.0068)		(0.0075)	(0.0264)	(0.0247)	(0.0739)
Aliquot (σ_c^2)	0.0227	0.0192	0.1026	0.0393	0.0639	0.0248	0.0974	0.0460	0.3805
	(0.0041)	(0.0035)	(0.0152)	(0.0069)	(0.0093)	(0.0044)	(0.0139)	(0.0081)	(0.0528)
Replicates (σ^2)	0.0138	0.0254	0.0155	0.0194	0.0439	0.0134	0.0069	0.0241	0.0082
	(0.0013)	(0.0025)	(0.0015)	(0.0019)	(0.0043)	(0.0013)	(0.0007)	(0.0023)	(0.0008)

ANDRO S are more problematic and suggest that the present measurement techniques should be used with care, especially with midluteal women. The results for ADIOL suggest that this assay is not yet ready for use in epidemiological studies.

References

1. Bernstein, L., and Ross, R. K. Endogenous hormones and breast cancer risk. Epidemiol. Rev., *15:* 48–65, 1993.

2. Pike, M. C., Spicer, D. V., Dalmoush, L., and Press, M. F. Estrogens, progestogens, normal breast cell proliferation, and breast cancer risk. Epidemiol. Rev., *15:* 17–35, 1993.

3. Kuller, L. H. The etiology of breast cancer—from epidemiology to prevention. Public Health Rev., *23:* 157–213, 1995.

4. Key, T. J. A., and Pike, M. C. The role of oestrogens and progestogens in the epidemiology and prevention of breast cancer. Eur. J. Cancer Clin. Oncol., *24:* 29–43, 1988.

5. Stoll, B. A., and Secreto, G. New hormone-related markers of high risk to breast cancer. Ann. Oncol., *3:* 435–438, 1992.

6. Toniolo, P. G. Endogenous estrogens and breast cancer risk: the case for prospective cohort studies. Environ. Health Perspect., *105* (Suppl. 3): 587–592, 1997.

7. Dorgan, J. F., Stanczyk, F. Z., Longcope, C., Stephenson, H. E., Jr., Chang, L., Miller, R., Franz, C., Falk, R. T., and Kahle, L. Relationship of serum dehydroepiandrosterone (DHEA), DHEASulfate, and 5-androstene-3β, 17β-diol to risk of breast cancer in postmenopausal women. Cancer Epidemiol. Biomark. Prev., *6:* 177–181, 1997.

8. Gail, M. H., Fears, T. R., Hoover, R. N., Chandler, D. W., Donaldson, J. L., Hyer, M. B., Pee, D., Ricker, W. V., Siiteri, P. K., Stanczyk, F. Z., Vaught, J. B., and Ziegler, R. G. Reproducibility studies and interlaboratory concordance for assays of serum hormone levels: estrone, estradiol, estrone sulfate and progesterone. Cancer Epidemiol. Biomark. Prev., *5:* 835–844, 1996.

9. Ziegler, R. G., Rossi, S. C., Fears, T. R., Bradlow, H. L., Adlercreutz, H., Sepkovic, D., Kiuru, P., Wahala, K., Vaught, J. B., Donaldson, J. L., Falk, R. T., Fillmore, C. M., Siiteri, P. K., Hoover, R. N., and Gail, M. H. Quantifying estrogen metabolism: An evaluation of the reproducibility and validity of enzyme immunoassays for 2-hydroxyestrone and 16α-hydroxyestrone in urine. Environ. Health Perspect., *105* (Suppl. 3): 607–614, 1997.

10. Horton, R., Hawks, D., and Lobo, R. 3-α, 17-β androstanediol glucuronide in plasma: a marker of Androgen action in idiopathic hirsutism. J. Clin. Invest., *69:* 1203–1206, 1982.

11. Horton, R., Imperator-McGinley, J., and Peterson, R. Androstanediol glucuronide (3 a-diol) in plasma is a unique marker in disorders of peripheral Androgen production and action in male pseudohermaphroditism. *In:* M. Serio and L. Martini (eds.), Sexual Differentiation, pp. 261–273. New York: Raven Press, 1984.

12. Gives, J. R. Hirsutism and hyperandrogenism. Adv. Int. Med., *21:* 221–247, 1976.

13. Abraham, G. E., Maroulis, G. B., and Buster, J. E. Effect of dexamethasone on serum cortisol and Androgen levels in hirsute patients. Obstet. Gynecol., *47:* 395–402, 1976.

14. Korth-Schutz, S., Virdis, R., Saenger, P., Chow, D. M., Levine, L. S., and New, M. I. Serum Androgens as a continuing index of adequacy of treatment of congenital adrenal hyperplasia. J. Clin. Endocrinol. Metab., *46:* 452–458, 1978.

15. Buster, J. E., and Abraham, G. E. Radioimmunoassay of plasma dehydroepiandrosterone sulfate. Ann. Lett., *5:* 543–547, 1972.

16. de Peretti, E., and Forest, M. G. Unconjugated dehydroepiandrosterone plasma levels in normal subjects from birth to adolescence in humans: the use of a sensitive radioimmunoassay. J. Clin. Endocrinol. Metab., *43:* 982–991, 1976.

17. Nieschlag, E., Loriaux, D. L., Ruder, H. J., Zucker, I. R., Kirschner, M. A., and Lipsett, M. B. The secretion of dehydroepiandrosterone and dehydroepiandrosterone sulfate in man. J. Endocrinol., *57:* 123–134, 1973.

18. Korth-Schutz, S., Levine, L. S., and New, M. I. Dehydroepiandrosterone sulfate levels, a rapid test for abnormal adrenal Androgen secretion. J. Clin. Endocrinol. Metab., *42:* 1005–1013, 1976.

19. de Peretti, E., and Forest, M. G. Pattern of plasma dehydroepiandrosterone sulfate levels in humans from birth to adulthood: evidence for testicular production. J. Clin. Endocrinol. Metab., *47:* 572–577, 1978.

20. Kinouchi, T., Pages, L., and Horton, R. A. A specific radioimmunoassay for testosterone in peripheral plasma. J. Lab. Clin. Med., *82:* 309–316, 1973.

21. Kirschner, M., and Bardin, W. Androgen production and metabolism in normal and virilized women. Metabolism, *21:* 667–688, 1972.

22. Rao, P. N., and Moore, P. H. Synthesis of new steroid haptens for radioimmunoassay. I. 15-β-carboxyethylmercapto-testosterone-bovine serum albumin conjugate. Measurement of testosterone in male plasma without chromatography. Steroids, *28:* 101–109, 1976.

23. Rao, P. N., Khan, A. H., and Moore, P. H. Synthesis of new steroid haptens for radioimmunoassay. III. 15-β-carboxyethlymercaptosteroid-bovine serum albumin conjugates. Specific antisera for radioimmunoassay of 5-α-dihydrotestosterone, 5-α-androstane-3-β, 17 β-diol and 5-α-androstane-3-α, 17-β-diol. Steroids, *29:* 171–184, 1977.

24. Furuyama, S., Mayes, D. M., and Nugent, C. A radioimmunoassay for plasma testosterone. Steroids, *16:* 415–428, 1970.

25. Lobo, R. A., Paul, W., and Goebelsmann, U. Dehydroepiandrosterone sulfate as an indicator of adrenal Androgen function. Obstet. Gynecol., *57:* 69–73, 1981.

26. Lobo, R. A., Goebelsmann, U., Brenner, P. F., and Mishell, D. R., Jr. The effects of estrogens on adrenal Androgens in oophorectomized women. Am. J. Obstet. Gynecol., *142:* 471–478, 1982.

27. Goebelsmann, U., Horton, R., Mestman, J. H., and Arce, J. J. Male pseudoherm-aphroditism due to testicular 17-β-hydroxysteroid dehydrogenase deficiency. J. Clin. Endocrinol. Metab., *36:* 867–879, 1973.

28. Goebelsmann, U., Bernstein, G. S., Gale, J. A., Kletzky, O. A., Nakamura, R. M., Coulson, A. H., and Korelitz, J. J. Serum gonadotropin, testosterone, estradiol and estrone levels prior to and following bilateral vasectomy. *In:* I. H. Lepow and R. Crozier (eds.), Vasectomy: Immunologic and Pathophysiologic Effects in Animals and Man, pp. 165–175. New York: Academic Press, 1979.

29. Serafini, P., Ablan, F., and Lobo, R. A. 5-α-Reductase activity in the genital skin of hirsute women. J. Clin. Endocrinol. Metab., *60:* 349–355, 1985.

30. Narang, R., Rao, J., Savjani, G., Peterson, J., Gentzschein, E., and Stanczyk, F. Z. Radioimmunoassay kit for the quantitative measurement of androstanediol glucuronide in unextracted serum. Presented at the 17th National Meeting of the Clinical Ligand Assay Society, April 10–13, 1991, Chicago, IL.

31. Matteri, R. K., Stanczyk, F. Z., Gentzschein, E. E., Delgado, C., and Lobo, R. A. Androgen sulfate and glucuronide conjugates in nonhirsute women with polycystic ovarian syndrome. Am. J. Obstet. Gynecol., *161:* 1704–1709, 1989.

32. SAS Institute Inc. SAS/STAT User's Guide, Version 6, 4th ed., Vol. 2, Cary, NC: SAS Institute Inc., 1989.

33. Rao, C. R. (ed.). Linear Statistical Inference and Its Applications, 2nd ed., New York: John Wiley, 1965.