Short Communication

Reproducibility Studies and Interlaboratory Concordance for Androgen Assays of Male Plasma Hormone Levels

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., M. H. G.], Epidemiology and Biostatistics Program [R. G. Z., R. N. H.], Radiation Epidemiology Branch [J. L. D.], Environmental Epidemiology Branch [R. T. F.] and Division of Cancer Epidemiology and Genetics, [J. B. V.], National Cancer Institute, Bethesda, Maryland 20892, and University of Southern California, Keck School of Medicine, Los Angeles, California 90033 [F. Z. S.]

Abstract

To help us identify appropriate techniques and laboratories for measuring hormones, we studied the variability and reproducibility of assay measurements of androstanediol glucuronide, androstenedione, dehydroepiandrosterone (DHEA), DHEA sulfate, dihydrotestosterone, testosterone, androstanediol, androsterone glucuronide, and androsterone sulfate for five men. Four sets of two aliquots from each sample were sent to participating laboratories, and one set was used for analyses monthly for four consecutive months. For each assay, estimates of components of variance were then used to estimate the coefficient of variation, the intraclass correlation between measurements on different days from a given individual, and the minimum detectable relative difference for a standard design. These data indicate that for at least one of the laboratories a single sample with two laboratory replicates per sample of androstanediol glucuronide, androstenedrone, DHEA, DHEA sulfate, and dihydrotestosterone yields an intraclass correlation coefficient exceeding 0.80 and can be used to discriminate reliably among men. The results for testosterone, androstanediol, androsterone, glucuronide, and androsterone sulfate do not meet this test. These data do not allow us to estimate the component of variation that corresponds to repeated blood samples taken over time from the same man. This reliability study design is, however, entirely appropriate for the typical case-control study which utilizes only one sample per subject.

Introduction

The NCI[2] obtained plasma samples as part of several field studies to evaluate hormone levels and the risk of cancer. We have studied the reproducibility of hormone assays used by laboratories with the capability of performing large numbers of tests. Gail *et al.* (1) estimated the variability and reproducibility of assay measurements of estrone, estradiol, estrone sulfate, and progesterone using plasma from women. Fears *et al.* (2) reported results for nine androgens obtained using plasma samples from women: ADIOL G; ADION; DHEA; DHEA S; DHT; TESTO; ADIOL; ANDRO G; and ANDRO S. In this final report of the series, we present results for the androgen assays using plasma samples from men.

Materials and Methods

Each of the four participating laboratories was asked to use their standard assay procedures and to perform only those assays with which they had experience.

Laboratory 1. For ADIOL G, unconjugated steroids were removed by organic extraction, followed by incubation with β-glucuronidase, enzyme hydrolysis, and celite chromatography and, finally, measurement of ADIOL by RIA. ADION, DHEA, TESTO, and DHT were measured by extracting plasma with ethyl acetate (20%) in hexane, celite column chromatography, and RIA. DHEA S was measured by RIA.

Laboratory 2. For ADIOL G, plasma was extracted with a polar solvent that was then subjected to complete enzymatic hydrolysis, followed by extraction with hexane:ethyl acetate, purification by high performance liquid chromatography, and measurement by RIA. ADION, DHEA, and TESTO were measured after extraction with hexane:ethyl acetate, followed by RIA. For DHEA S, the sulfate was removed by overnight hydrolysis with sulfatase, after which the procedures for measuring DHEA were followed. For DHT, extraction with hexane: ethyl acetate was followed by treatment with a strong oxidizer to destroy unsaturated steroids and purification on alumina columns.

Laboratory 3. DHEA S was quanitified by direct RIA after a 1000-fold dilution of the plasma sample with assay buffer. DHEA, ADION, TESTO, DHT, and ADIOL were measured by RIA after extraction of plasma with diethyl ether and purification by celite column chromatography. ADIOL G was quantified directly in plasma using a commercial kit. ANDRO S and ANDRO G were measured after unconjugated steroids were removed by extraction with diethyl ether. The conjugated steroids were hydrolyzed utilizing hydrochloric acid or glucuronidase followed by extraction with ethyl acetate and purification by celite column chromatography.

Received 2/6/01; revised 4/22/02; accepted 4/25/02.
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, Biostatistics Branch, National Cancer Institute, 6120 Executive Boulevard, EPS/8040, Bethesda, MD 20892.

[2] The abbreviations used are: NCI, National Cancer Institute; ADIOL G, androstanediol glucuronide; ADION, androstenedione; DHEA, dehydroepiandrosterone; DHEA S, DHEA sulfate; DHT, dihydrotestosterone; TESTO, testosterone; ADIOL, androstanediol; ANDRO G, androsterone glucuronide; ANDRO S, androsterone sulfate; CV, coefficient of variation; ICC, intraclass correlation coefficient; MDRD, minimum detectable relative difference; CI, confidence interval; SD, standard deviation.

Laboratory 4. ADION was measured by carbon tetrachloride extraction of plasma followed by a RIA kit. DHEA was measured by dichloromethane extraction and a RIA. DHEA S and TESTO were assayed directly using RIA. Further details of the assay procedures are provided in Fears *et al.* (2).

A blood sample was obtained from each of five male volunteers from NCI, ages 31, 45, 47, 50, and 67 years. Each volunteer was in good health, with no known hormonal abnormality. Within 24 h of draw, the plasma was separated, aliquoted, and stored at −70°C (see Ref. 1). Each participating laboratory received four batches of samples, with one batch to be assayed on each of four consecutive months. Each batch contained two aliquots from each of the five subjects. Each aliquot was assayed in duplicate. The order of the 10 aliquots within each batch was randomly assigned.

A nested components of variance analysis was performed using logarithmically transformed measurements to stabilize variances (1). Components were estimated for subjects, days, aliquots, and replicates. We used three measures of reproducibility derived from these components. The CV, which is the SD divided by the mean, is the usual measure of reproducibility. The sum of the components associated with day, aliquot, and replicate is a good estimate of the square of the CV. The ICC is the correlation between measurements on different days from a given sample. It is the ratio of the component associated with subjects to the sum of all components. The MDRD is the minimum difference that is reliably detected with a given number of cases and controls. The CV is of primary interest to the laboratory for quality control, whereas the ICC and MDRD are more important to the epidemiologist in determining the feasibility of an epidemiological study.

As in the earlier studies, for each hormone and laboratory, we examined graphs of grand means, daily means and aliquot means (available on request). The Friedman rank order statistic was used to compare subject means across laboratories, and Spearman rank correlations were used to measure concordance of the participants' grand means across laboratories. The estimated components of variance are given in the "Appendix." The components of variance (Table A1) were used to obtain estimates of the CVs, ICCs, and MDRDs that were compared among the laboratories. To calculate ICCs and CVs, we assumed that the measurement used was the mean of the two logarithmic-transformed replicates. To calculate MDRDs, we assumed $n_1 = 300$ cases and $n_2 = 600$ controls, as in our earlier reports. Further details of these statistical methods are presented in Fears *et al.* (2). The estimates and their 95% CIs are given in Table 1.

Results

ADIOL G. The subject means differed by laboratory ($P = 0.015$). The overall mean of Ln(ADIOL G) was 6.66 at laboratory 3, which was higher than the mean values at laboratories 1 and 2 (5.84 and 5.92, respectively). The ranks of the subjects' mean responses were highly correlated across laboratories (0.90–1.00). The ICC was 97% from laboratory 3 but only 81% from laboratory 1 and 74% from laboratory 2. The CVs were small at laboratories 1 and 3 (14% and 5% respectively) but were higher at laboratory 2 (21%). The MDRDs were 10% at laboratories 1 and 3 and less than 13% at laboratory 2.

ADION. The mean levels of Ln(ADION) differed across laboratories ($P = 0.002$). Grand means were 4.19, 4.46, 4.07, and 4.27 at laboratories 1, 2, 3, and 4, respectively. The ranks of the subjects' mean responses were, however, highly correlated (0.90–1.00). The CVs for ADION were 20% or less at all four laboratories (9% and 11% at laboratories 3 and 4 and 18% and 20% at laboratories 1 and 2). The ICCs were 92% and 94% at laboratories 3 and 4 and 80% and 66% at laboratories 1 and 2. The MDRDs ranged from 10% to 13%.

DHEA. The mean levels of Ln(DHEA) from these laboratories differed across laboratories ($P = 0.002$). Grand means were 5.25, 5.48, 5.63, and 4.69 at laboratories 1, 2, 3, and 4, respectively. The ranks of the subjects' mean responses were, however, perfectly correlated. The ICC was above 93% at all the laboratories. The CV was 17% at laboratory 4 and less than 13% at the other laboratories. The MDRDs were between 15% and 17% at laboratories 2 and 3 but above 20% at laboratories 1 and 4.

DHEA S. The mean levels of Ln(DHEA S) differed by laboratory ($P = 0.004$). Grand means were highest at laboratories 1 and 3 (5.08 and 5.14) and were quite different from the mean levels at laboratories 2 and 4 (4.62 and 4.97, respectively). The ranks of the subjects' means were perfectly correlated. The ICC was 98% or higher at all laboratories. The CV at laboratory 1 was 10%, and other laboratories had CVs of 8% or less. The MDRD ranged from 20% to 25%.

DHI. There was weak evidence ($P = 0.07$) that the subject mean levels of Ln(DHT) differed by laboratory. The grand mean was lowest at laboratory 1 (3.68) *versus* 3.77 and 3.76 at laboratories 2 and 3, respectively. The ranks of the subjects' means were highly correlated (0.90–1.00). The ICC was only 76% at laboratory 3 and 92–95% at laboratories 1 and 2. Similarly, the CV was 16% at laboratory 3 and 8–10% at laboratories 1 and 2. The MDRD ranged from 10% to 11%.

TESTO. The mean levels of Ln(TESTO) were comparable ($P = 0.36$). The grand means were 5.99, 6.02, 6.01, and 5.91 at laboratories 1, 2, 3 and 4, respectively. The correlations of the ranks of the subjects' mean responses ranged from 0.70 to 1.00. The ICC was 85% and 89% at laboratories 1 and 2, respectively, and less at laboratories 3 and 4 (71% and 57%, respectively). The CVs were less than 13% at laboratories 1 and 2 but were 21% and 35% at laboratories 3 and 4, respectively. The MDRDs were 10–12% at laboratories 1, 2, and 3 and 17% at laboratory 4.

ADIOL, ANDRO G, and ANDRO S. Only laboratory 3 conducted these assays. For ADIOL, the ICC was not large (63%), the CV was relatively high (22%), and, because the total variation was small, the MDRD was small (11.4%). The ICC for ANDRO G was 84%, the CV was 21%, and the MDRD was 17%. For ANDRO S, the ICC was low (59%), and the CV was high (36%); however, the MDRD was only 18%.

Discussion

ADIOL G, ADION, DHEA, DHEA S, DHT, and TESTO were assayed in several laboratories. There was variation in the mean assay levels among the participating laboratories, but the correlations of rankings of mean subjects' results among the laboratories were high. The ICCs were greater than 90% at all laboratories for DHEA and DHEA S. They were less than 90% at all laboratories for TESTO. All ICCs were greater than 80% at laboratory 1, greater than 70% at laboratory 3, greater than 65% at laboratory 2, and greater than 55% at laboratory 4. Most CVs were less than 15%, and all were less than 21%, except TESTO at laboratory 4. All CVs were less than 18% at laboratory 1. All MDRDs were less than 20% at laboratory 3 and less than 25% at laboratories 1 and 2. No laboratory used

Table 1 Estimates of CVs, ICCs, MDRDs, and 95% CIs for androgen assays using male plasma

Hormone	Lab 1		Lab 2		Lab 3		Lab 4	
	Estimate	CI	Estimate	CI	Estimate	CI	Estimate	CI
ADIOL G (ng/dl)								
ICC	81.4	58.0–100.0	73.8	43.4–100.0	97.3	93.4–100.0		
CV	13.6	8.0–19.3	20.7	12.4–29.1	5.2	3.7–6.8		
MDRD	9.8	3.7–15.9	12.7	5.3–20.1	9.8	2.9–16.7		
ADION (ng/dl)								
ICC	79.8	54.2–100.0	66.0	29.0–100.0	94.4	86.3–100.0	92.3	81.4–100.0
CV	17.5	9.5–25.5	19.5	10.5–28.5	9.0	5.4–12.5	11.1	6.0–16.2
MDRD	12.2	4.6–19.8	10.4	4.7–16.1	11.8	3.6–20.1	12.6	3.9–21.4
DHEA (ng/dl)								
ICC	98.5	96.4–100.0	96.0	90.2–100.0	92.8	82.9–100.0	95.9	90.1–100.0
CV	7.6	4.9–10.3	10.5	5.8–15.2	12.9	8.3–17.4	17.0	11.1–22.9
MDRD	20.4	5.1–35.7	16.9	4.6–29.2	15.3	4.6–25.9	28.3	6.9–49.7
DHEA S (mcg/dl)								
ICC	98.0	95.2–100.0	98.8	97.1–100.0	98.6	96.5–100.0	99.1	97.8–100.0
CV	10.4	6.9–13.9	8.2	4.6–11.8	7.3	4.9–9.8	6.3	4.0–8.6
MDRD	24.7	5.9–43.4	25.2	5.9–44.6	19.9	5.0–34.8	21.9	5.3–38.5
DHT (ng/dl)								
ICC	91.9	80.7–100.0	94.5	86.7–100.0	75.7	47.1–100.0		
CV	9.6	6.1–13.2	7.9	4.9–10.9	16.2	9.8–22.6		
MDRD	10.5	3.4–17.6	10.5	3.2–17.8	10.2	4.2–16.2		
TESTO (ng/dl)								
ICC	85.2	66.2–100.0	88.7	72.9–100.0	70.7	36.8–100.0	56.7	14.1–99.3
CV	12.9	8.3–17.5	11.5	6.1–16.9	21.4	11.1–31.8	34.5	17.6–51.3
MDRD	10.4	3.8–17.1	10.6	3.5–17.6	12.4	5.3–19.6	16.8	8.1–25.4
ADIOL (pg/ml)								
ICC					63.4	27.5–99.3		
CV					22.0	14.7–29.3		
MDRD					11.4	5.6–17.1		
ANDRO G (ng/ml)								
ICC					83.5	61.7–100.0		
CV					21.3	11.2–31.4		
MDRD					16.8	5.8–27.8		
ANDRO S (ng/ml)								
ICC					58.8	17.2–100.0		
CV					35.9	18.3–53.4		
MDRD					18.0	8.5–27.5		

assays that were uniformly superior to those used by other laboratories.

Laboratory 3 was the only laboratory that volunteered to assay ANDRO G, ANDRO S, and ADIOL. The ICCs were relatively low (84% for ANDRO G but less than 65% for ANDRO S and ADIOL). The CVs were fairly high, ranging from 21% to 36%. Nevertheless, the MDRDs were low, ranging from 12% to 18%.

A larger number of men would be desirable to estimate the components of variance, especially the subject component, with greater precision (Table A1). However, even five men are sufficient to yield usefully precise estimates in many instances, as indicated by the confidence intervals in Table 1. For example, for DHEA in laboratory 4, the estimated ICC was 95.9% with a 95% CI of 90.1–100.0%. The CIs are larger for assays with lower ICC point estimates, *e.g.*, ADIOL, ICC = 63.4% and CI, 27.5–99.3%. The results in Table 1 indicate that the ICC significantly exceeds 0.80 for at least one laboratory with ADIOL G, ADION, DHEA, DHEA S, and DHT, indicating that a single sample with two laboratory replicates/sample yields sufficiently high ICC values for field studies. In contrast ANDRO G, TESTO, ADIOL, and ANDRO S do not meet this test.

A separate issue from precision is whether samples from the NCI volunteers can form a basis for generalization to the population of participants in an etiologic field study. An ideal experiment would recruit a random sample from potential participants in a field study. Questions of generalization could arise even in that context, however, because data would be available only on those who agreed to provide blood samples for a reliability study. We are unaware of any evidence that volunteers for the NCI study are unrepresentative of the general population with respect to parameters such as ICC or MDRD, but this possibility cannot be dismissed.

Fears *et al.* (2) studied these same assays in women. Assay levels tended to be over 5-fold greater in men than in women. This may account for the smaller CVs for ADIOL G, DHEA, and DHEA S in men. Men have higher ICCs for ADIOL than women, and although the assay is not recommended for women, it might be used with caution in men. For ANDRO G, the ICC was lower for men than for women. Previous recommendations that ADIOL G, DHEA, and DHEA S are suitable for field studies in women also apply to men. In addition, both ADION and DHT may be appropriate for field studies in men but not for women in the midluteal menstrual phase. The previous conclusions that TESTO and ANDRO S can only be used with caution in women also hold for men.

Appendix

Table A1 Estimated variance components and sex for androgens assayed at multiple laboratories (Labs)

		Lab 1		Lab 2		Lab 3		Lab 4	
		Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
ADIOL G	Subject	0.0816	0.0604	0.1208	0.0913	0.0968	0.0687		
	Analysis day	0.0109	0.0055	0.0233	0.0125	0.0002	0.0006		
	Aliquot	0.0058	0.0025	0.0176	0.0062	0.0000	0.0000		
	Replicates	0.0037	0.0008	0.0040	0.0009	0.0050	0.0009		
ADION	Subject	0.1209	0.0904	0.0738	0.0584	0.1342	0.0960	0.1494	0.1076
	Analysis day	0.0242	0.0101	0.0312	0.0127	0.0046	0.0023	0.0099	0.0041
	Aliquot	0.0054	0.0021	0.0042	0.0022	0.0000	0.0000	0.0016	0.0008
	Replicates	0.0022	0.0005	0.0051	0.0011	0.0069	0.0013	0.0018	0.0004
DHEA	Subject	0.3881	0.2750	0.2675	0.1908	0.2136	0.1528	0.6799	0.4837
	Analysis day	0.0014	0.0015	0.0084	0.0036	0.0038	0.0042	0.0047	0.0072
	Aliquot	0.0024	0.0014	0.0011	0.0009	0.0085	0.0041	0.0236	0.0077
	Replicates	0.0039	0.0009	0.0032	0.0007	0.0084	0.0019	0.0013	0.0003
DHEA S	Subject	0.5443	0.3859	0.5713	0.4050	0.3716	0.2632	0.4436	0.3142
	Analysis day	0.0002	0.0026	0.0049	0.0022	0.0000	0.0000	0.0012	0.0010
	Aliquot	0.0078	0.0035	0.0011	0.0006	0.0039	0.0013	0.0022	0.0009
	Replicates	0.0058	0.0013	0.0015	0.0003	0.0030	0.0007	0.0010	0.0002
DHT	Subject	0.1046	0.0750	0.1071	0.0765	0.0818	0.0614		
	Analysis day	0.0032	0.0025	0.0026	0.0017	0.0132	0.0075		
	Aliquot	0.0049	0.0020	0.0031	0.0011	0.0111	0.0042		
	Replicates	0.0024	0.0005	0.0009	0.0002	0.0040	0.0009		
TESTO	Subject	0.0959	0.0696	0.1026	0.0747	0.1110	0.0863	0.1555	0.1306
	Analysis day	0.0039	0.0043	0.0113	0.0045	0.0415	0.0160	0.1102	0.0418
	Aliquot	0.0094	0.0041	0.0008	0.0006	0.0016	0.0015	0.0063	0.0028
	Replicates	0.0068	0.0015	0.0021	0.0005	0.0057	0.0013	0.0045	0.0010
ADIOL	Subject					0.0839	0.0636		
	Analysis day					0.0000	0.0000		
	Aliquot					0.0433	0.0116		
	Replicates					0.0102	0.0023		
ANDRO G	Subject					0.2299	0.1701		
	Analysis day					0.0391	0.0155		
	Aliquot					0.0045	0.0020		
	Replicates					0.0036	0.0008		
ANDRO S	Subject					0.1835	0.1521		
	Analysis day					0.1195	0.0453		
	Aliquot					0.0057	0.0030		
	Replicates					0.0067	0.0015		

References

1. 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.

2. Fears, T. R., Ziegler, R. G., Donaldson, J. L., Falk, R. T., Hoover, R. N., Stanczyk, F. Z., Vaught, J. B., and Gail, M. H. Reproducibility studies and interlaboratory concordance for androgen assays in female plasma. Cancer Epidemiol. Biomark. Prev., *9:* 403–412, 2000.