

 Open access • Journal Article • DOI:10.1093/CLINCHEM/41.11.1581

## Reliability of salivary testosterone measurements: a multicenter evaluation.

— [Source link](#) 

J.M. Dabbs, B. C. Campbell, B.A. Gladue, A. R. Midgley ...+5 more authors

**Institutions:** Georgia State University

**Published on:** 01 Nov 1995 - Clinical Chemistry (American Association for Clinical Chemistry)

**Topics:** Intraclass correlation

Related papers:

- [Salivary testosterone measurements: Reliability across hours, days, and weeks](#)
- [Salivary testosterone measurements: Collecting, storing, and mailing saliva samples](#)
- [Salivary testosterone in men: further evidence of a direct correlation with free serum testosterone.](#)
- [Direct radioimmunoassay \(RIA\) of salivary testosterone: correlation with free and total serum testosterone.](#)
- [Testosterone, and winning and losing in human competition.](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/reliability-of-salivary-testosterone-measurements-a-495n0b8mhi>

## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/21028>

Please be advised that this information was generated on 2021-02-02 and may be subject to change.

## Reliability of Salivary Testosterone Measurements: A Multicenter Evaluation

James M. Dabbs Jr.,<sup>1,10</sup> Ben C. Campbell,<sup>2</sup> Brian A. Gladue,<sup>3</sup> A. Rees Midgley,<sup>4</sup> Miguel A. Navarro,<sup>5</sup> Graham F. Read,<sup>6</sup> Elizabeth J. Susman,<sup>7</sup> Leon M. J. W. Swinkels,<sup>8</sup> and Carol M. Worthman<sup>9</sup>

The reliability of salivary testosterone assays was evaluated by nine laboratories in four countries. Each laboratory used its own RIA procedures to assay samples from a set of 100 male and 100 female subjects. Agreement among the laboratories on mean scores was within the range reported by Read (Ann N Y Acad Sci 1993;694:161-76). Overall agreement on individual scores, as indicated by the intraclass correlation coefficient computed within subjects across laboratories, was  $r = 0.87$  for men and  $r = 0.78$  for women. Mean agreement between each laboratory and the combined set of all other laboratories (via Fisher's Z-transformation) was  $r = 0.61$  for men and  $r = 0.58$  for women. We take these latter values to be the best estimates of the average reliability of laboratories in their ordering of individual samples.

**Indexing Terms:** radioimmunoassay/interlaboratory comparison/sex-related differences

The ease of sample collection has led to a growing number of studies involving salivary testosterone measurements (1). Subjects will participate readily, and research can be carried out in diverse settings and populations. Salivary measures are less well known than serum measures, however, and researchers and reviewers have been skeptical about their use. Salivary testosterone is a well-established marker for circulating free testosterone concentrations (2-5), but questions remain about the reliability of measurements. The present study addressed the issue of agreement among laboratories in assaying identical sets of saliva samples.

Identical saliva samples were sent to laboratories with established records of conducting RIAs of salivary testosterone. Each laboratory used its own procedure to conduct the assays. The study examined agreement among laboratories in their mean values for men and women and in their ordering of individual scores around the respective means.

<sup>1</sup> Department of Psychology, Georgia State University, Atlanta, GA.

<sup>2</sup> Carolina Population Center, Chapel Hill, NC.

<sup>3</sup> University of Cincinnati, Cincinnati, OH.

<sup>4</sup> University of Michigan, Ann Arbor, MI.

<sup>5</sup> Hospital Princeps d'Espanya, Barcelona, Spain.

<sup>6</sup> Tenovus Cancer Research Centre, Cardiff, UK.

<sup>7</sup> The Pennsylvania State University, University Park, PA.

<sup>8</sup> Katholieke Universiteit, Nijmegen, The Netherlands.

<sup>9</sup> Emory University, Atlanta, GA.

<sup>10</sup> Author for correspondence. Fax 404-651-1391; e-mail psyjmd@gsusgi2.gsu.edu.

Received December 20, 1994; accepted July 18, 1995.

### Subjects and Methods

Saliva samples were collected from 100 male and 100 female undergraduates at Georgia State University, following a protocol approved by the Institutional Review Board. Each subject chewed a stick of sugar-free gum to stimulate the flow of saliva and deposited 15-20 mL into a 20-mL polyethylene vial. The saliva was centrifuged to remove debris, and 1.5-mL volumes from each subject were pipetted into nine separate vials, producing a total of 1800 samples. Sets of 100 men's and 100 women's samples were frozen and shipped on frozen CO<sub>2</sub> to each of nine participating laboratories (associated with the authors of the present paper). Table 1 shows characteristics of the RIA procedure at each laboratory. The references include representative publications from the authors at these laboratories (6-17).

All nine laboratories assayed the men's samples, and six of them assayed the women's samples. Laboratories 2, 3, and 7 chose not to assay women's samples because they had no experience assaying them or needed larger sample volumes. Of the resulting 1500 assays, 13 were not completed for various reasons. Before statistical analysis, a log transformation was applied to normalize the testosterone distributions. Six scores were discarded because they fell >3 SD above the means for their laboratories. The procedure for discarding outliers was performed once, using means and SDs for the log-transformed data from each laboratory. Log-transformed scores were used in all statistical analyses, but means and variances are reported below in untransformed raw score units, because readers are more familiar with raw score units and find them easier to

**Table 1. General assay procedures at each laboratory.**

Laboratory	Procedure
1	<sup>125</sup> I from ICN (Costa Mesa, CA), ether extraction, charcoal separation.
2	<sup>125</sup> I coated-tube antibody kit from Diagnostic Products Corp. (Los Angeles, CA), no extraction.
3	<sup>125</sup> I coated-tube antibody kit from Diagnostic Products Corp., no extraction.
4	<sup>125</sup> I kit from Farnos Diagnostica (Turku, Finland), ether-ethylacetate extraction, and polyethylene glycol separation.
5	<sup>3</sup> H from Amersham (Bucks, UK), ether extraction, charcoal separation.
6	<sup>125</sup> I kit from ICN, no extraction.
7	<sup>3</sup> H from Amersham, ether extraction, followed by paper chromatography, charcoal separation.
8	<sup>125</sup> I modified coated-tube kit from Diagnostic Products Corp., double ether extraction.
9	<sup>125</sup> I modified kit from Binax (S. Portland, ME), ether extraction, 2nd antibody separation.

**Table 2. Testosterone measurements at each laboratory.**

Laboratory	Men		Women	
	Mean (SD), pmol/L	CV, %	Mean (SD), pmol/L	CV, %
1	370 (133) <sup>bc</sup>	3.8	55 (20) <sup>a</sup>	6.3
2	357 (167) <sup>b</sup>	8.5	—	—
3	410 (191) <sup>d</sup>	8.1	—	—
4	240 (95)	4.7	89 (21)	5.1
5	272 (99)	—	75 (25)	—
6	309 (111) <sup>a</sup>	2.8	101 (47)	6.9
7	393 (181) <sup>cd</sup>	6.3	—	—
8	324 (129) <sup>a</sup>	8.6	55 (24) <sup>a</sup>	12.9
9	382 (142) <sup>cd</sup>	3.9	48 (28)	8.5

Measurements are based on 94–100 cases each. SD refers to variation across subjects in each assay. CV refers to variation between pairs of sample duplicates within each assay. Superscript letters indicate results of statistical comparisons among means. Means within each sex that do not share a common superscript are significantly different from each other (Neuman-Keuls test,  $P < 0.05$ ).

interpret. Men's and women's scores were analyzed separately, because the two distributions do not overlap, and because three laboratories did not assay women's samples. In addition, many investigators want separate information on women's samples, because salivary testosterone from women is commonly believed to be especially difficult to assay.

## Results

The study examined two components of the agreement among laboratories in assaying salivary testosterone. One component is at the group level, reflecting agreement on group means and variances. The other is at the individual level, reflecting agreement in the ordering of individual samples.

### Agreement on Group Means and Variances

Mean testosterone concentration across all subjects, where each subject's concentration was defined as the mean of his or her scores across all laboratories, was 342 pmol/L (SD = 97) for men and 71 pmol/L (SD = 23) for women. Table 2 shows sample means, SDs among subjects, and CVs between assay duplicates at each laboratory. Repeated-measures analyses of variance indicated significant mean differences among the laboratories (for men,  $F(8, 696) = 26.35$ ,  $P < 0.001$ ; for women,  $F(5, 480) = 79.98$ ,  $P < 0.001$ ). Differences among pairs of means were tested by using the Newman-Keuls procedure; the results are indicated by superscripts in Table 2. Fig. 1 shows the distribution of men's and women's scores at each laboratory.

### Agreement on Ordering of Individual Samples

Overall reliability in the ordering of individual samples was evaluated by using the intraclass correlation coefficient, adjusted to remove mean laboratory differences. This coefficient was computed to indicate agreement within subjects across laboratories. A high value would indicate that scores from each subject from different laboratories were tightly clustered around the

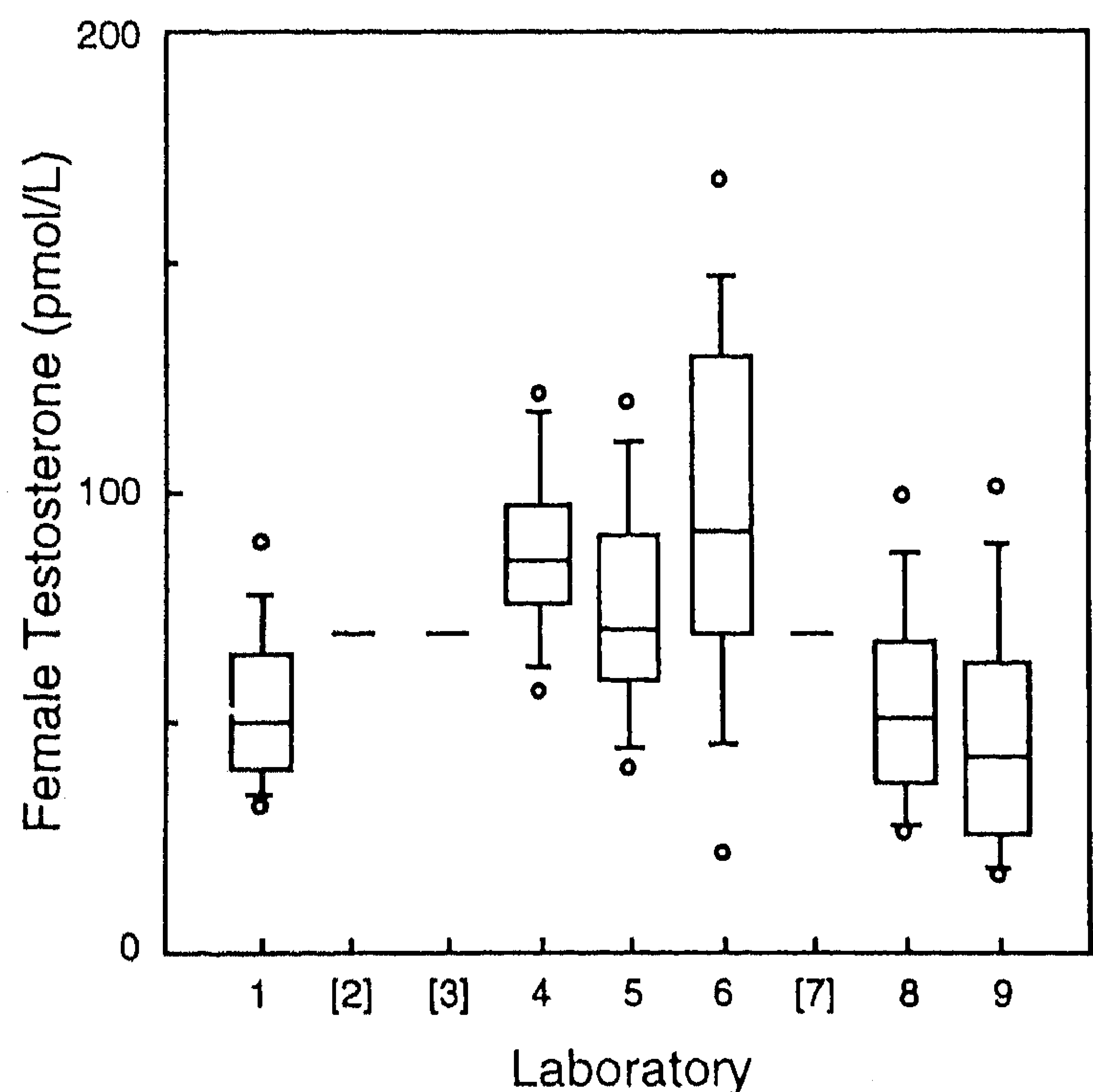
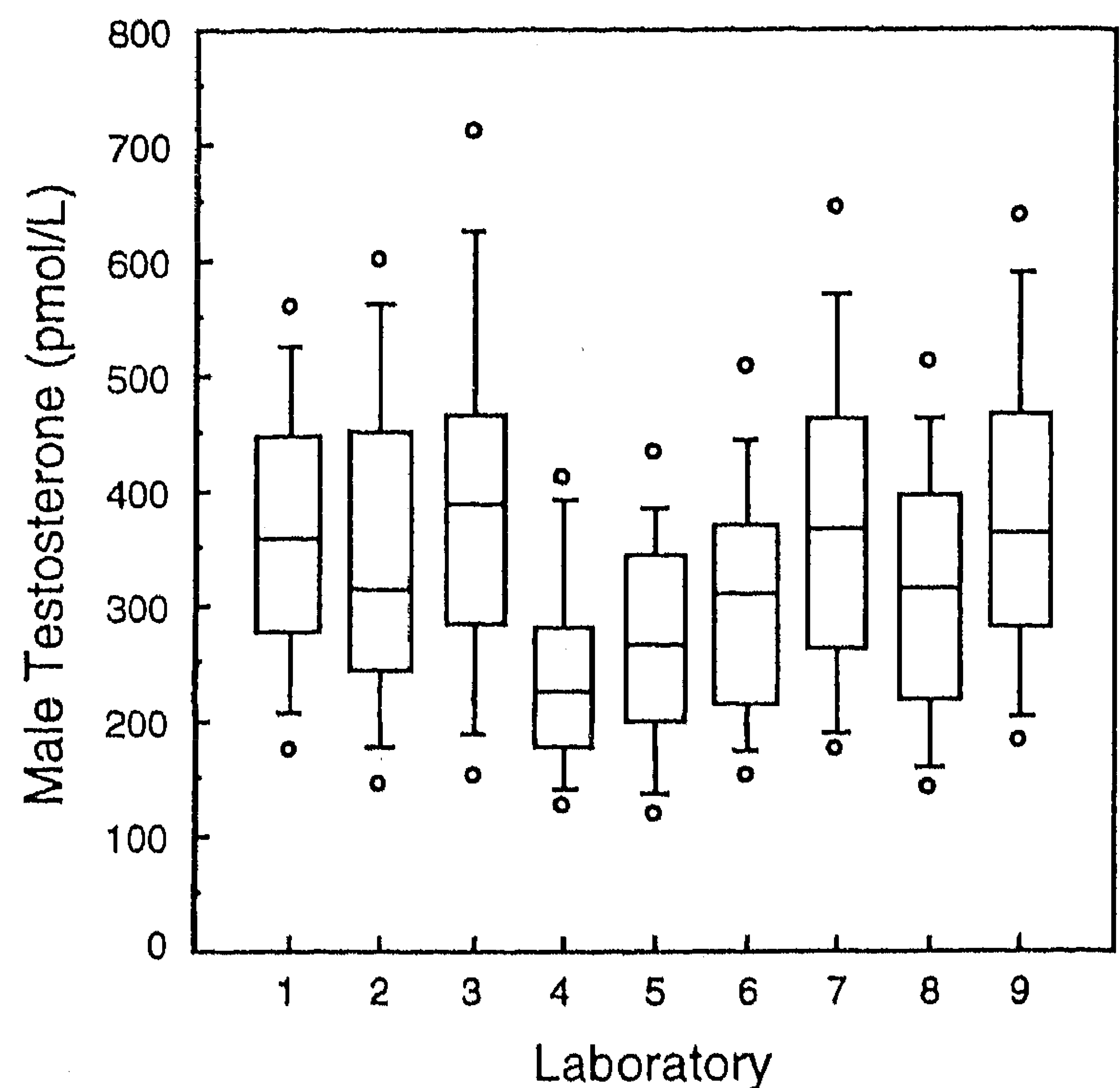


Fig. 1. Box plots showing distribution of men's and women's assay scores at nine laboratories.

Lines in boxes mark 25th, 50th, and 75th percentiles; lines outside boxes mark 10th and 90th percentiles; circles mark 5th and 95th percentiles. Laboratories are ordered the same way in the upper and lower figures. Two outliers from laboratory 1 and one outlier each from laboratories 4, 5, 8, and 9 are dropped from the data.

subject's mean, relative to the mean differences among subjects. Intraclass correlation coefficients were  $r = 0.87$  for men and  $r = 0.78$  for women.

The intraclass correlation coefficient reflects the combined reliability of assays at all laboratories. For practical purposes, it is more useful for a working investigator to know the reliability of an individual laboratory, rather than the reliability of all laboratories combined. Table 3 provides one view of this information, showing product-moment correlations among all pairs of laboratories, with men's data above the diagonal and women's data below. The mean correlation between pairs of laboratories (computed via Fisher's  $Z$ -transformation) was  $r = 0.44$  for men and  $r =$

**Table 3. Correlations among laboratories.**

Laboratory	1	2	3	4	5	6	7	8	9
1		0.61	0.52	0.43	0.24	0.42	0.59	0.45	0.54
2	—		0.64	0.36	0.19	0.32	0.53	0.47	0.34
3	—	—		0.37	0.26	0.52	0.47	0.42	0.46
4	0.39	—	—		0.23	0.41	0.35	0.26	0.21
5	0.42	—	—	0.39		0.36	0.29	0.64	0.37
6	0.44	—	—	0.26	0.28		0.40	0.45	0.51
7	—	—	—	—	—	—		0.56	0.64
8	0.40	—	—	0.33	0.44	0.38	—		0.54
9	0.74	—	—	0.39	0.61	0.37	—	0.42	

Correlations are based on 93–100 cases each; values above the diagonal are based on male subjects; values below the diagonal are based on female subjects.

0.46 for women. Inspection of Table 3 suggests that laboratories were relatively similar in their intercorrelations.

Table 3 compares individual laboratories with each other; our next step was to compare each laboratory with the mean of all the other laboratories combined. In the absence of an external standard, such as mass spectroscopy, our best estimate of a subject's true testosterone score comes from the combined scores of all laboratories together, in which random errors attributable to individual laboratories tend to cancel out. Table 4 shows product-moment correlations between scores from each single laboratory and mean scores from all other laboratories combined. Differences among correlations were tested with Fisher's Z-transformation; the results are indicated by superscripts in Table 4. For men, the two lowest laboratories differed significantly from the four highest. For women, the two lowest laboratories differed significantly from the two highest. Averaged across all laboratories (via Fisher's Z-transformation), the mean correlation between a single laboratory and the combined set of all other laboratories was  $r = 0.60$  for men and  $r = 0.61$  for women. These values provide the best estimate of the average reliability of the laboratories in assaying men's and women's samples.

### Discussion

Laboratory agreement on group means and variances was in the range reported by Read (18), although

there were statistically significant differences among the laboratories. Given that RIA results are notoriously sensitive to changes in reagents and operating procedures between assay batches within the same laboratory, differences among laboratories are not surprising. Presumably there would also be differences among laboratories in the results of serum testosterone assays, but we have been unable to find published information on this topic.

The laboratories agreed more on the relative ordering of individual scores than on group means, as indicated by fewer significant differences among the correlations in Table 4 than among the mean scores in Table 2. In examining the raw data we did not find any subject who scored high in one laboratory and low in another. Table 4 shows that the reliability of laboratories was similar for men's and women's scores. Although some investigators have been skeptical about the feasibility of measuring women's salivary testosterone, the present findings indicate that one could equally well use men or women in studying correlations between salivary testosterone and other variables, as long as all assays are performed in the same laboratory.

Each laboratory followed its customary assay procedures, and there is no obvious explanation for disagreements between pairs of laboratories in Table 3. We would expect less error in assays that involve extraction and chromatography than in direct assays. However, Table 3 suggests that whether or not a laboratory uses extraction makes little difference, and because only one laboratory used chromatography, we do not know whether laboratories using chromatography would agree better with each other. Consistency among the laboratories in their mean values was higher than in earlier comparisons reported by Baxendale and James (19). This suggests that investigators have improved their techniques over the years as they have gained experience with salivary assays. This point is supported by data we collected from a commercial laboratory. The laboratory had a good reputation for its work in assaying steroid hormones, but it did not have experience in assaying women's salivary testosterone. It failed to detect testosterone in any of our women's samples, and the correlation between its men's assay

**Table 4. Correlation between each laboratory and mean scores from other laboratories.**

Laboratory	Men	Women
1	0.67 <sup>b</sup>	0.71 <sup>b</sup>
2	0.62 <sup>ab</sup>	—
3	0.66 <sup>b</sup>	—
4	0.45 <sup>a</sup>	0.44 <sup>a</sup>
5	0.45 <sup>a</sup>	0.59 <sup>ab</sup>
6	0.60 <sup>ab</sup>	0.48 <sup>a</sup>
7	0.68 <sup>b</sup>	—
8	0.67 <sup>b</sup>	0.52 <sup>ab</sup>
9	0.64 <sup>ab</sup>	0.69 <sup>b</sup>
Mean	0.61	0.58

Correlations are based on 94–100 cases each. Correlation coefficients within each sex that do not share a common superscript are significantly different from each other (Fisher's Z-transformation test,  $P < 0.05$ ).

scores and mean men's scores from our nine laboratories was only  $r = 0.23$ .

Current assays of men's and women's salivary testosterone concentrations are reliable enough to be quite usable for research purposes. As the state of the art sharpens and technical improvements become more widespread, we can expect increases in reliability to lead to the establishment of generally accepted laboratory standards. This will further enhance the usefulness of salivary assays in studies requiring noninvasive techniques or repeated sampling from subjects.

This research was supported in part by grants from the National Science Foundation (J.M.D., DBS-9021393; B.A.G., EP-SCoR REU-NSF); the National Institute of Mental Health (E.J.S., MH46430); the National Institute of Child Health and Human Development (B.C.C., HD12806; A.R.M., U54HD29184, via National Cooperative Program on Infertility Research; E.J.S., HD26004); and Emory University (C.M.W., W. T. Grant Foundation Faculty Scholar award).

#### References

1. Malamud D, Tabak L. Saliva as a diagnostic fluid. *Ann N Y Acad Sci* 1993;694:1-342.
2. Wang C, Plymate S, Nieschlag E, Paulsen CA. Salivary testosterone in men: further evidence of a direct correlation with free serum testosterone. *J Clin Endocrinol Metab* 1981;53:1021-4.
3. Baxendale PM, Jacobs HS, James VHT. Salivary testosterone: relationship to unbound plasma testosterone in normal and hyperandrogenic women. *Clin Endocrinol* 1982;16:595-603.
4. Navarro MA, Juan L, Bonnin MR. Salivary testosterone: relationship to total and free testosterone in serum [Letter]. *Clin Chem* 1986;32:231-2.
5. Campbell BC, Udry JR, Halpern C. Salivary and serum testosterone in adolescent boys [Abstract]. *Am J Phys Anthropol* 1992;14:57.
6. Dabbs JM Jr. Salivary testosterone measurements in behavioral studies. *Ann N Y Acad Sci* 1993;694:177-83.
7. Dabbs JM Jr, Carr TS, Frady RL, Riad JK. Testosterone, crime, and misbehavior among 692 male prison inmates. *Person Individ Diff* 1995;18:627-33.
8. Gladue BA, Boechler M, McCaul KD. Hormonal response to competition in human males. *Aggr Behav* 1989;15:409-22.
9. Lopez-Calbet JA, Navarro MA, Barbany JR, Garcia MJ, Bonnin MR, Valero J. Salivary steroid changes and physical performance in highly trained cyclists. *Int J Sports Med* 1993;14:111-7.
10. Navarro MA, Villabona CM, Blanco A, Gomez JM, Bonnin MR, Soler J. Salivary excretory pattern of testosterone in substitutive therapy with testosterone enanthate. *Fertil Steril* 1994;61:125-8.
11. Walker RF, Wilson DW, Read GF, Riad-Fahmy D. Assessment of testicular function by the radioimmunoassay of testosterone in saliva. *Int J Androl* 1980;3:105-21.
12. Butler GE, Walker RF, Walker RV, Teague P, Riad-Fahmy D, Ratcliffe SG. Salivary testosterone levels and the progress of puberty in the normal boy. *Clin Endocrinol* 1989;30:587-96.
13. Reid JD, Intrieri RC, Susman EJ, Beard JL. The relationship of serum and salivary cortisol in a sample of healthy elderly. *J Gerontol* 1992;47:176-9.
14. Dorn LD, Susman SJ. Serum-saliva relations during pregnancy. *Biol Psychiatry* 1993;34:226-33.
15. Swinkels LMJW, Meulenberg PMM, Ross HA, Benraad TJ. Salivary and plasma free testosterone and androstenedione levels in women using oral contraceptives containing desogestrel or levonorgestrel. *Ann Clin Biochem* 1988;25:354-9.
16. Swinkels LMJW, Van Hoof HJC, Ross HA, Smals AGH, Benraad TJ. Concentrations of salivary testosterone and plasma total, non-sex-hormone-binding globulin-bound and free testosterone in normal and hirsute women during administration of dexamethasone/synthetic corticotropin. *Clin Chem* 1991;37:180-5.
17. Beall CM, Worthman CM, Stallings J, Strohl KP, Brittenham GM, Barragan M. Salivary testosterone concentration of Aymara men native to 3600 m. *Ann Hum Biol* 1992;19:67-78.
18. Read GF. Status report on measurement of salivary estrogens and androgens. *Ann N Y Acad Sci* 1993;694:161-76.
19. Baxendale PM, James VHT. Specificity of androgen measurements in saliva. In: Read GF, Riad-Fahmy D, Walker RR, Griffiths K, eds. Ninth Tenovus workshop: immunoassays of steroids in saliva. Cardiff, UK: Alpha Omega, 1984:228-38.