Journal of Steroid Biochemistry, 1975, Vol. 6, pp. 271–275. Pergamon Press. Printed in Great Britain

ASSESSMENT OF RELIABILITY OF STEROID RADIOIMMUNOASSAYS

Z. CEKAN

Swedish Medical Research Council, Reproductive Endocrinology Research Unit,
Karolinska Sjukhuset, 104 01 Stockholm 60, Sweden

SUMMARY

A reliable radioimmunoassay (RIA) has to have high precision and accuracy. The precision of the measurement of plasma samples is estimated by the within-assay and between-assay variations. The precision of the measurement of the standard curve is closely related to the sensitivity. The latter can be defined as the ability to distinguish a dose from zero or one dose from another.

A highly accurate RIA must have a high degree of specificity which can be achieved by efficient separation of the steroid prior to RIA and by using highly specific antisera. The specificity can be checked by a test of parallelism between increasing amounts of authentic and endogenous steroid. This test, however, has a low efficiency; it can only detect a considerable disturbance of specificity. Nevertheless, it should be applied whenever new biological materials are to be assayed. In order to attain high accuracy, systematic errors (blanks, neglected influence of internal standard, etc.) and biases (e.g., graphical construction of standard curve) have to be avoided.

Quality control should become an indispensable part of RIA.

A radioimmunoassay (RIA), in order to be reliable, should have a high precision and a high accuracy. The high precision means that repeated measurements (shots) (Fig. 1) are as close to each other as possible. The high accuracy means that the measured values describe, on the average, the true contents of a steroid in a sample as authentically as possible. It is the ambition of every radioimmunassayist to hit the bull's eye with every shot and to have the assay both highly precise and accurate. However, whereas a high precision can be achieved and measured, a high accuracy is something that can be struggled for but, unfortunately, cannot be measured and exactly proved.

First, a few words about the precision. When a new or modified RIA is published, the authors usually show figures for the precision of replicate measurements of the same plasma pool and for the precision of measurements of a pool assayed on several occasions. The former index of precision is called within-assay variation, the latter, between-assay variation. Both are expressed usually as coefficients of variation. Sometimes it is felt that it is more justified to measure the within-assay variation at different steroid levels. This is done either by assaying different pools or by assaying a number of plasma samples in duplicate [1].

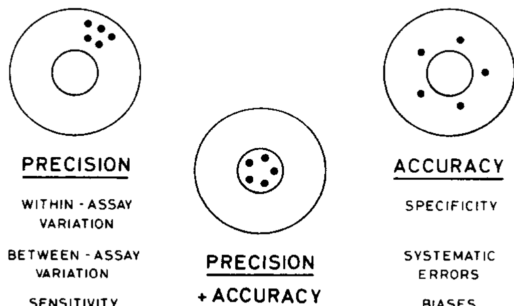

Fig. 1. Schematic representation of the assay reliability.

For a calculation of within-assay variance from duplicates, a simplified formula is available, based on differences between duplicates [2] (Table 1). From the variances within individual samples an average variance is calculated from which an average coefficient of variation *i.e.*, the within-assay variation is derived. If the latter is calculated in this way for several groups of samples differing in their steroid contents, it can be seen (Table 2) that the coefficient of variation remains approximately the same throughout the whole range of measurements [3]. We observed this phenomenon not only with estradiol but also with 7 other steroid assays in human plasma. Hence, it seems that a pool may yield the same information as many individual samples measured in

Table 1. Calculation of variance from duplicates

DUPLICATES		MEAN	d	d^2	VARIANCE $=d^2/2$	S.D.	C.V.
198	202	200	4	16	8	2.82	1.41
190	210	200	20	400	200	14.14	7.07
180	220	200	40	1600	800	28.28	14.14
170	230	200	60	3600	1800	42.43	21.21
MEANS		200			702		
S.D.					26.5		
C.V.					13.2		

S.D. = Standard deviation.
C.V. = Coefficient of variation = (S.D./mean). 100.

Table 2. Within-assay variation in 11 groups of 300 plasma samples collected from normally menstruating women

Group	1	2	3	4	5	6	7	8	9	10	11
No. of samples	4	22	29	44	60	50	36	35	10	4	6
Geometrical mid-points (pg/ml)	44.9	56.8	71.7	90.5	114	144	182	230	290	366	462
S.D.[a]	3.39	5.28	4.95	7.26	7.67	10.6	12.6	16.0	18.4	39.4	38.6
C.V.[b]	7.55	9.30	6.91	8.02	6.73	7.37	6.90	6.97	6.33	10.08	8.36
Mean C.V.[c] (weighted)						7.36					

All samples were measured in duplicate. In distributing the plasma samples into groups, a lognormal distribution was assumed.

[a] S.D. = Standard deviation $=\sqrt{(\sum d^2/2n)}$, where d = difference between duplicates and n = No. of samples.

[b] C.V. = Coefficient of variation = (S.D./geometrical mid-point) 100.

[c] Mean C.V. was obtained by weighting the individual coefficients of variation against the number of samples in the corresponding groups.

duplicate. The pool seems to be more convenient when the within-assay variation of a new assay is to be established, the duplicates are suitable for quality control of running assays. In this function the duplicates offer an additional possibility for quality control.

All duplicates giving an individual coefficient of variation which is higher than a predetermined value, *e.g.*, 20% (cf. Table 1), may be excluded as unreliable. The number of samples excluded is a direct reflection of the quality of the assay.

Precision connected with the determination of a standard curve is closely related to the sensitivity of an assay. One can generally say that the higher the precision, the higher the ability to distinguish a dose from zero or one dose from another. In the majority of cases, only the "zero" sensitivity is considered. In Fig. 2 the concept of Ekins and Newman[4] for zero sensitivity has been applied to a standard curve linearized by a logit-log transformation. The lowest dose distinguishable from zero is indicated by intercept of the standard curve with the lower confidence limit calculated for the zero tubes.

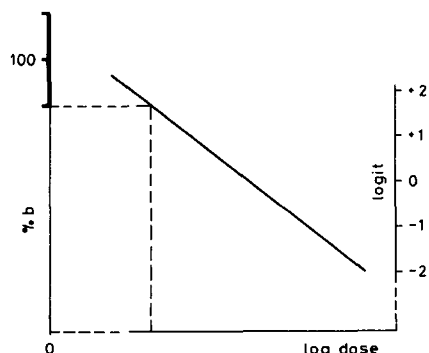

Fig. 2. Schematic representation of the assay sensitivity. The lowest dose distinguishable from zero is given by the intercept of the standard curve with the lower confidence limit of zero tubes.

The sensitivity, in the sense of the distinguishing power of one dose from another, is also clearly dependent on precision. It is obvious that in the case where the error indicated by confidence limits of the standard curve is high, the distinguishing power for one dose from another is low. This type of sensitivity is called "precision" by Ekins and Newman[4]. This overlapping of terms shows, *inter alia*, how closely interlinked the precision and sensitivity are.

A radioimmunoassay of a high accuracy must have a high degree of specificity. In order to achieve high specificity one usually tries to combine a high specificity of antisera with the best possible separation of the assayed steroids from other steroids and plasma constituents. One tries to create a radioimmunoassay in which all significantly cross-reacting compounds are separated prior to radioimmunoassay. Sometimes this procedure is successful, as can be seen

Table 3. Specificity for dihydrotestosterone.

Steroid	In the chromatographic fraction[a]	Outside the fraction[b]
Dihydrotestosterone	100	
Adrenosterone	< 0.1	
20α-Dihydroprogesterone	< 0.1	
20β-Dihydroprogesterone	< 0.1	
Deoxycorticosterone	< 0.1	
Testosterone		88.3
17α-Testosterone		0.3
19-Nortestosterone		7.3
Androstenedione		0.7
5α-Androstane-3α, 17β-diol		13.6
4-Androstene-3β, 17β-diol		6.4
5α-Androst-1-ene-3, 17-dione		0.8
3β-Hydroxy-5α-androstan-17-one		0.3

[a] In this column cross-reactions (in percent) of all steroids are indicated which were located in the chromatographic fraction (celite chromatography).

[b] In this column only those steroids are mentioned (out of the 57 tested) which exhibited a cross-reaction higher than 0·1%.

from Table 3, where the apparent specificity of a dihydrotestosterone assay is demonstrated.

Out of 57 steroids tested, only 8 had a cross-reaction higher than 0·1%. All these compounds could be separated from dihydrotestosterone by means of celite chromatography, and only 4 non-significantly cross-reacting compounds could be detected in the chromatographic fraction containing dihydrotestosterone [5]. But can we be absolutely sure that there are no other cross-reacting compounds, so far not known or not tested, present in the chromatographic fraction? It can be expected that in future this uncertainty will decrease progressively as we become more adept at producing antibodies of far higher specificity and as we develop steroid purification techniques of greater efficiency than are presently available.

It is possible to test the presence of foreign substances by studying parallelism of the amounts of endogenous steroid in increasing volumes of plasma with increasing amounts of authentic steroid. However, as Ekins *et al.* [6] showed both theoretically and practically, such a parallelism can be disturbed only by the presence of a compound or compounds which have an equilibrium constant and/or concentration very different from that of the compound measured. So the parallelism demonstrated, for example, in the testosterone assay in seminal plasma [7] (Fig. 3) cannot, unfortunately, be taken as proof of an absolute specificity. A simple experiment may clarify this statement (Fig. 4). Estriol cross-reacts with a specific estradiol antiserum to about 5 per cent and its apparent equilibrium constant is lower by about one order. If increasing doses of estradiol are assayed together with 8 times higher increasing doses of estriol, statistics cannot find any evidence which speaks against parallelism with pure estradiol. So that if we find a non-parallelism as in Fig. 5, the contamination must, indeed, be very great.

In Figure 5 an example of non-parallelism is shown when estrone was assayed in seminal plasma [7]. On

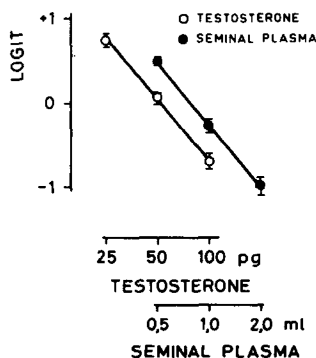

Fig. 3. Relationship between increasing doses of authentic testosterone and increasing amounts of endogenous testosterone extracted from various volumes of seminal plasma. The logits of individual samples were corrected for a 100% recovery. Each plasma volume and each dose of the standard hormone were assayed in 4 replicates. Mean values ± standard deviations are indicated. F-value for parallelism = 1·97; F-value for linearity = 0·03; $F_{0\cdot95(1,15)}$ = 4·54.

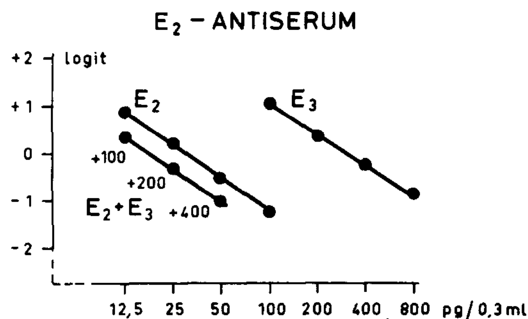

Fig. 4. Parallelism of increasing doses of estradiol with increasing doses of a mixture estradiol:estriol. E_2 = estradiol, E_3 = estriol, $E_2 + E_3$ = mixture of estradiol and estriol, consisting of 12·5 pg/0·3 ml E_2 + 100 pg/0·3 ml E_3, etc. F-value for parallelism of E_2 with $E_2 + E_3$ was 3·96, F-value for linearity was 0·56; $F_{0\cdot95(1,15)}$ = 4·54.

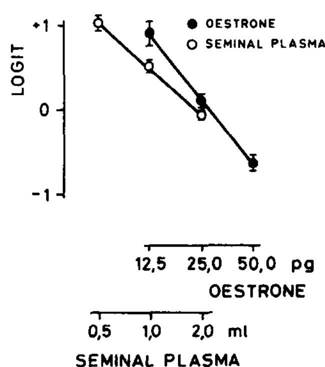

Fig. 5. Relationship between increasing doses of authentic estrone and increasing amounts of endogenous estrone extracted from various volumes of seminal plasma. For explanations cf. Fig. 3. F-value for parallelism = 11·58, F-value for linearity = 0·13; $F_{0\cdot95(1,15)}$ = 4·54.

the other hand, in the estrone assay in systemic plasma the parallelism was obtained [8]. This contrast indicates that the test of parallelism, despite its drawbacks, is of value and ought to be applied whenever new biological materials are to be assayed.

Another method which is stated by many authors to be a check of accuracy is the assay of increasing amounts of standard hormone added to the same amount of plasma. This experiment, however, cannot show by any means whether the assay is accurate and/or specific. It shows only that the extraction is complete at several dose levels, that the extraction is adequately monitored by the internal standardization and that the system is working properly in general. Or as Reeves and Calhoun[9] put it: "Proportionality between added and found increments tests the system for biases of scale, but not for zero bias." Consequently, the previously mentioned test of parallelism, despite its low efficiency, remains the only internal check of the accuracy and specificity. If one wishes to have an external judgement of the accuracy one has to resort to a comparison with other methods, for example double isotope dilution. It is obvious, however, that such a comparison carries along its own errors.

One of the major systematic errors which may influence the accuracy is the blank originating from solvents and chromatographic material. The presence of solvent blank can be estimated by *t*-test; cpm of zero tubes are compared with c.p.m. of zero tubes containing blank. If the *t*-test indicates a significant presence of blank, great pains should be taken to remove it. It is very difficult, if not impossible, to make corrections for the blank, e.g., subtract some value from the results. The following example may substantiate this statement. When 1 ml of an ether sample was evaporated and the residue assayed together with testosterone, a Scatchard plot was obtained (Fig. 6) indicating that the ether residue (blank) decreased both the molar concentration of binding sites and the apparent equilibrium constant. The standard curve resulting from this assay (Fig. 6) significantly deviated from the "pure" standard curve.

If, in spite of all efforts, the blank cannot be removed, the unknowns should be calculated from a standard curve containing blank, which may be a rather tedious procedure. Solvent blank may become a serious systematic error and a check of its presence should be included in the quality control.

In a number of publications, the test for plasma blank is described. In contrast to the solvent blank, which is a summary term for organic compounds entering the system from "outside," the plasma blank is actually a sum of cross-reacting compounds present in plasma. Plasma blank is usually measured in plasma which has been deprived of steroids either by extensive extraction with solvents or by "stripping" with charcoal. However, if the steroids are really completely extracted, what makes us think that the plasma blank is not extracted too, and that the blank eventually found is not a solvent blank? The significance of blanks remaining after adsorbing with charcoal is uncertain. For example, is the difference between amounts added and recovered shown in Fig. 7 due to blank, unadsorbed steroid or both? Obviously, the present methods used for testing the plasma blank are not adequate enough, and the only way to demonstrate the presence of significant amounts of plasma blank, i.e., cross-reacting compounds in general, is the test of parallelism.

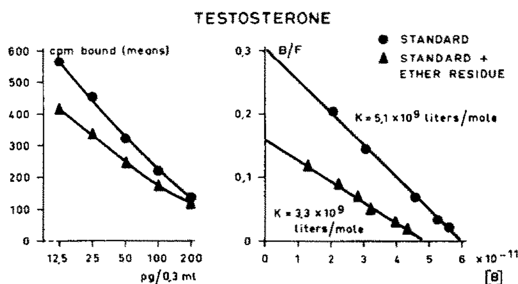

Fig. 6. Standard curves and Scatchard plots for authentic testosterone alone and in the presence of an ether blank. In all test tubes containing blank the residue obtained after evaporation of 1 ml ether was present.

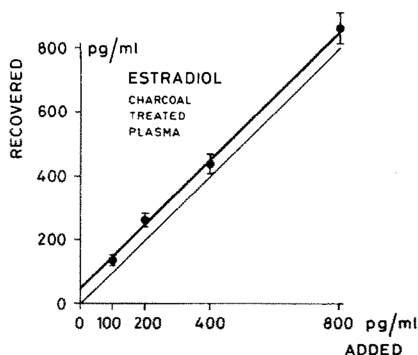

Fig. 7. Relationship between added and recovered estradiol. Increasing doses of estradiol were added to a charcoal treated plasma.

Another systematic error is the neglected influence of the internal standard which is added to plasma samples to monitor extraction and isolation losses. This error can be easily coped with by modifying the calculation of unknowns (Fig. 8). This calculation is based on the fact that the number of total counts in the tube is increased by a portion of the internal standard present in the assay tube. This portion is in fact specific for each plasma sample depending on the recovery of the steroid. Therefore, for each sample, an individual absolute percent binding (B/T), relative percent binding (B/B_0) and logit have to be computed. The mass of the portion of the internal standard is then substracted from the result.

Biases are unavoidable when standard curves are drawn by hand and unknowns are extrapolated from them. My guess is that in 40–50% of the radioimmunoassay laboratories in the world the results are still calculated in this way. I would like to submit that at least a programmable desk-top calculator (Wang, Hewlett–Packard or similar) should become as indispensable to the radioimmunoassay as is a beta- or gamma-counter. It is again a guess that about 40–50% of the laboratories use the desk-top calculators, whereas some 2–5% employ big computers. There is no doubt that big computers offer the most qualified evaluation of radioimmunoassays. However, for practical reasons most laboratories will

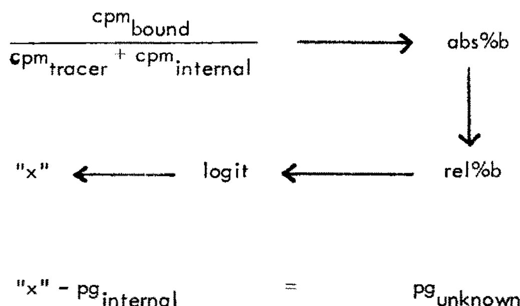

Fig. 8. Schematic representation of a correction for the mass and counts of internal standard used for monitoring procedural losses in radioimmunoassays of biological materials. $Cpm_{internal}$ is a portion of internal standard cpm present in the assay tube, $pg_{internal}$ is a portion of the mass of the internal standard present in the assay tube.

be restricted to desk-top calculators for some time to come.

Whereas the big computers can handle a curvilinear standard curve, small desk-top calculators require a linearization. One of the best and most popular linearization methods is the logit-log transformation developed by Rodbard *et al.* [10]. This linearization has its drawbacks and merits. Both have been discussed, *e.g.*, by Rodbard and Hutt[11] and Ekins[12]. In my opinion the merits prevail very much. The statistics of a straight line are relatively simple, can be acquired by every radioimmunosssayist and can be easily programmed on desk-top calculators. Such a statistical evaluation includes the calculation of the best-fit straight line, either in the weighted or unweighted manner, test of linearity, estimation of the over-all error, mathematical calculation of unknowns and their confidence limits, and tests of parallelism. Only such a statistical evaluation prevents personal and inter-personal biases.

From the practical point of view, the logit-log transformation works very well. The linearized standard curves are reproducible, especially as far as their slopes are concerned. As can be seen from Table 4, the slopes of standard curves in all our radioimmunoassays approach very closely the ideal value $-2{\cdot}303$ theoretically derived by Rodbard *et al.* [10]. and the variation around this ideal value is acceptable. The slope can thus be used as one of the parameters for quality control.

Table 4. Slopes of logit-log transformed standard curves

Theory: b = -2.303 (Rodbard)

Assay	n	Slope (b)	± S.D.
Progesterone	16	- 2.392	0.216
Pregnenolone	18	- 2.408	0.102
17-OH-Progesterone	14	- 2.327	0.086
17-OH-Pregnenolone	10	- 2.372	0.164
20α-Dihydroprogesterone	9	- 2.411	0.150
Testosterone	18	- 2.451	0.101
Dihydrotestosterone	16	- 2.678	0.273
Androstenedione	14	- 2.613	0.188
Dehydroepiandrosterone	18	- 2.394	0.065
Estrone	17	- 2.427	0.104
Estradiol	22	- 2.409	0.102

S.D. = Standard deviation.

A proper quality control should become a rule in every RIA laboratory. It should consist, at least, of regular checks of within-assay and between-assay variation. The procedures for quality control have been clearly defined by Rodbard[13].

Concluding this short review, I would say that the RIA of steroids is an analytical tool of high precision and sensitivity. The accuracy can be worked out to a high degree, absolute accuracy, however, can probably never be achieved and small differences in measured values between different laboratories have to be expected.

REFERENCES

1. Wide L., Nillius S. J., Gemzell C. and Roos P.: *Acta Endocr., Copenh. Suppl.* **174** (1973) 5–58.
2. Kimball A. W.: *Biometrics* **8** (1952) 85–88.
3. Guerrero R., Aso T., Brenner P. F., Cekan Z., Landgren B.-M., Hagenfeldt K. and Diczfalusy E.: to be published.
4. Ekins R. P. and Newman G. B.: In *Karolinska Symposia on Research Methods in Reproductive Endocrinology, 2nd Symposium, Steriod Assay by Protein Binding* (Edited by E. Diczfalusy). Karolinska Institutet, Stockholm (1970) pp. 11–30.
5. Purvis K., Brenner P. F., Landgren B.-M., Cekan Z. and Diczfalusy E.: *Clin. Endocr.* **4** (1975) in press.
6. Ekins R. P., Newman G. B. and O'Riordan J. L. H.: In *Radioisotopes in Medicine*: IN VITRO *Studies. Proceedings of a Symposium* (Edited by R. L. Hayes). U.S. Atomic Energy Commission, Oak Ridge (1967) pp. 59–100.
7. Purvis K., Landgren B.-M., Cekan Z. and Diczfalusy E.: *Clin. Endocr.* **4** (1975) in press.
8. Brenner P. F., Guerrero R., Cekan Z. and Diczfalusy E.: *Steriods* **22** (1973) 775–794.
9. Reeves B. D. and Calhoun D. W.: In *Karolinska Symposia on Research Methods in Reproductive Endocrinology, 2nd Symposium, Steriod Assay by Protein Binding* (Edited by E. Diczfalusy). Karolinska Institutet, Stockholm (1970) pp. 61–77.
10. Rodbard D., Bridson W. and Rayford P. L.: *J. Lab. clin Med.* **74** (1969) 770–781.
11. Rodbard D. and Hutt D. M.: In *Radioimmunoassay and Related Procedures in Clinical Medicine and Research*, Proceedings of a Symposium, International Atomic Energy Agency, Vienna (1974) 165–192.
12. Ekins R. P.: *Brit. med. Bull.* **30** (1974) 3–11.
13. Rodbard D.: In *Principles of Competitive Protein-Binding Assays* (Edited by W. D. Odell and W. H. Daughaday). Lippincott Co., Philadelphia (1971) pp. 204–259.