Europ. J. Cancer Vol. 17, No. 6, pp. 643–649, 1981
Printed in Great Britain

0014-2964/81/060643-07 $02.00/0

Estrogen Receptor Assay: Interlaboratory and Intralaboratory Variations in the Measurement of Receptors Using Dextran-coated Charcoal Technique: A Study Sponsored by E.C.O.G.*

SHANTHI RAAM,†|| REBECCA GELMAN§ and JOSEPH L. COHEN,† with the assistance of ALFRED BACHARACH,¶ ANDREW J. FISHCHINGER,¶ HERBERT I. JACOBSON,¶ ALBERT A. KESHGEGIAN,¶ STANLEY J. KONOPKA¶ and JAMES L. WITTLIFF¶

†*Oncology Laboratory, Steroid Receptor Lab, and* ‡*Oncology Department, Tufts University Medical Cancer Unit, Lemuel Shattuck Hospital, Boston, Massachusetts, U.S.A.*
§*Biostatistics Department, Sidney Farber Cancer Institute, Boston, Massachusetts, U.S.A.*

Abstract—*Six laboratories performing estrogen receptor analysis in breast cancer for patients included in clinical protocols of the Eastern Cooperative Oncology Group participated in a quality control study for estrogen receptor assay. Three tissue powders containing* 10–50 femtomoles *of estrogen receptors/*mg *cytosol protein were distributed and the results analyzed for interlaboratory and intralaboratory variations. Frequency of misclassification of the receptor-positive powders as receptor-negative was related to the quantity of receptors. All labs except one were able to distinguish "low content" from "medium content" and the "medium" from the "high content" powders despite the variability in the numerical values reported for the powders. Two of the six labs reported consistently low values for all samples and the results of one lab were consistently higher than the average of the values obtained for the powders. These results could not be explained by the variations in protein content reported by these laboratories for different powders. Implications of the degree of inter-lab and intra-lab variations in the assay results are discussed with reference to their effect on selecting patients for hormonal vs chemotherapy. Based on the results of this study, tentative suggestions for quality control for the receptor assay are also made.*

INTRODUCTION

SINCE the introduction of dextran-coated charcoal technique for quantifying estrogen receptors in mastectomy specimens of human breast cancer by Korenman *et al.* [1], the clinical usefulness of this assay has been assessed and defined. Clinical relationships between the estrogen receptor status of the breast biopsies and response to different therapeutic modalites have been established. These included an observed failure of response to hormonal therapy in patients whose tumors are defined as "receptor-negative" in contrast to an apparent increasing likelihood of response to hormonal therapy associated with increasing content of estradiol receptors in the malignant tissues. In general, while 60% of the "receptor-positive" patients respond to hormonal therapies, fewer than 10% of the "receptor-negative" patients do. These observations establish a firm role for this assay in the management of breast cancer patients (McGuire *et al.* [2]).

The assay used for the measurement of receptors present in tissues involves certain problems. Fluctuations have been observed by several investigators in the values obtained for receptor content among the replicate assays for the same tissue sample (DeSombre *et al.* [3]). The radioligand binding technique is more complex than most of the biochemical

Accepted 17 December 1980.
*This study was supported by a grant from the Eastern Cooperative Oncology Group (Paul P. Carbone M.D.) Grant No.CA2 115 and Public Health Service Grants from NCI, NIH and DHEW.
||To whom reprint requests should be addressed.
¶The names of the responsible persons of the participating laboratories are arranged alphabetically.

assays involving several steps. In addition, the receptors are highly thermolabile. Therefore, current methodologies require careful tissue sampling and handling, precise homogenization techniques to avoid overheating of the receptors, sophisticated separation of cell cytosol fraction and accurate measurement of radiolabelled estradiol used for incubation. The complexity of the procedure thus increases the probability of errors at every step. Moreover, there exists amongst the laboratories considerable variations in the technique in the composition of buffers, range of ^3H estradiol concentrations used, duration and temperature of incubation of cytosol with estradiol, methodology of protein determinations and the separation of free estradiol from the receptor-bound estradiol. Such a procedure beset with technical variations requires the establishment of a method for interlaboratory and intralaboratory normalization so as to render comparable the values that are obtained in different laboratories.

There have been several attempts to analyze the interlaboratory and intralaboratory variations in the results of receptor analysis either by using whole tissue sectioned for distribution or cytosol fractions aliquoted and distributed to several laboratories (King *et al.* [4]; NCI Consensus Development Conference in Breast Cancer, 1979–1980; King [5]). Although these experiments revealed considerable interlab and intralab variations in the assay results, these variations could be attributed to factors other than the assay conditions such as heterogeneity of the tissue and lability of receptors in the cytosol.

The purpose, therefore, of the present study was to analyze the quality of the single most commonly used technique for ER measurement, namely DCC technique, using multiple point Scatchard analysis of the data. Only laboratories which had performed at least ten assays on E.C.O.G. patients by June 1979 were included in the study. All the participating labs were experienced in using the technique (each lab performed at least 300 assays per annum). The assay was tested for its accuracy in measuring ER content in the lower range, 10–50 fmol/mg cytosol protein. The laboratory which prepared the tissue powders (reference lab) and pre-determined the ER levels also participated in the study by reassaying the powders that were blindly coded. The study was conducted using the premise that the assay would be subjected to the same degree of variation in the hands of the reference lab as in the other participating labs and the data analyzed assuming the quality of performance (expertise) of all the participating labs was equal.

MATERIALS AND METHODS

(1) *Choice of laboratories*

Based on the following criteria five laboratories were chosen to participate in the study:

(a) Proximity to north east part of U.S.A.;

(b) Laboratories engaged in performing ER assays for patients included in E.C.O.G. clinical protocols;

(c) Laboratories using DCC technique with multiple concentrations of $*E_2$ (at least five) and analyzing the data by Scatchard plot; expressing the results as fmol/mg cytosol protein;

(d) Those that agreed to perform 12 assays within a period of two weeks for the quality control study.

(2) *Production and distribution*

The reference lab, Dr. J. W., Louisville, Kentucky, was contracted to prepare powders (72 vials/batch 0.8 ± 0.1 g/vial) (low, medium, high) each containing different levels of estrogen receptors ranging from 5 to 50 fmol/mg cytosol protein. The classification as low, medium and high was originally based on the ER values obtained by the reference lab after repeated ER determinations. This was done not only to ascertain the level of receptors but also to test the consistency of the level of ER measured before the powders were mailed to LSH (central laboratory) for coding and distribution as well as to test the stability of the receptors during the course of the study. The fmol/mg values for each powders and the coefficient of variations obtained in the reference lab are given in Table 1.

Upon arrival at the central lab (LSH) the powders were transferred to another vial, prechilled in liquid nitrogen and pre-labelled with the appropriate code number. Each lab received four replicates of each of the three powders. These were divided into three batches, each batch having two different powders represented among four samples. Each batch was to be analyzed on a single day with all three batches to be completed within two weeks. No mishaps such as thawing of the powders, etc., occurred. All the labs received the powders within 24 hr of mailing. The labs were requested to complete all three sets in two weeks after receipt and to freeze an aliquot of cytosol from each sample at a concentration used for incubation and mail

Table 1. *Results of continuous analysis of the receptor content* (fmol/mg *cytosol protein*) *of the distributed tissue powders by The Reference Lab*

	Low	Medium	High
participant labs 8/1/79–8/31/79	*17±5 (29%)	26±8 (31%)	60±10 (17%)
After completion of ECOG study by the labs 9/5/79–10/19/79	16±8 (50%)	22±4 (18%)	66±6 (9%)

*Fmol/mg ± S.D. (coefficient of variation).

them back to the central lab along with the data.

(3) *Procedural details of the ER assay*

The details pertinent to the assay conditions in each lab are tabulated in Table 2.

(4) *Data analysis*

The predetermined ER values of the reference lab obtained for each powder were not used as "true" values. Instead, the values reported by the reference lab for the "coded" powders were included along with the values reported by the other five laboratories for data analysis.

The data were analyzed for interlaboratory and intralaboratory variations in order to determine:

(a) Whether the values obtained by each lab in fmol/mg protein can be used as a continuous prognostic indicator; i.e., does a tissue reported by lab X as containing 20 fmol/mg cytosol protein for example, have essentially the same amount of receptors as a tissue reported as 20 fmol/mg by Lab Y?

(b) Whether the ER values reported by the labs be used for classifying a sample as + or − and for determining treatment. In other words, although there is variation in the numerical values reported by each lab per sample, is there a qualitative agreement between the labs as to the presence/absence of ER? (reference lab was the only one which knew ahead of time that no negative powders were distributed).

(c) Assuming there are differences in ER values reported by different labs, whether it is possible to identify a factor/factors contributing to such a variation. In order to answer this question, factors influencing the results such as interlab variation in protein quantification methods, time and temperature of incubation and range of estradiol concentration were examined.

(5) *Protein determination*

Since the laboratories differed in the methodology used to quantify the proteins, aliquots of all the cytosol returned to the central lab were reassayed using a single technique [6]. The protein quantity expressed by each lab was compared with the value obtained by the central lab; moreover, the ER values (femtomoles) reported by each lab were recalculated using the central lab's protein value. Data reported and data "corrected" were

Table 2. *Experimental conditions*

	Q	R	S	T	U	V
No. assays/month	40	80	70	20	50	75
Positive standard	Calf uterus	Calf uterus	Rabbit uterus	Rabbit uterus	Human uterus	Calf uterus
Negative standard	None	Beef kidney	Rabbit kidney	Rabbit kidney	Skeletal muscle	None
Estradiol conc. Range (nM)	0.5–5.0	0.1–4.6	0.2–2.0	0.19–1.12	0.04–8.00	0.5–5.0
E_2 concentration (No.)	5	8	7	5	5	5
Incubation (hr)	5	18	16	16–24	1	16
Temperature (°C)	0–4	0–4	4	0–4	25	4
Method of protein determination	Biuret	u.v.	Lowry	u.v.	Lowry	Waddell
Diethylstilbesterol measure of non-sp. binding?	Yes	No	Yes	Yes	Yes	Yes

Table 3. *Receptor content*

Labs	Low	Medium	High
Q	11±8 (0.69)	26±2 (0.09)	47±5 (0.11)
R	11±7 (0.70)	29±14 (0.48)	47±17 (0.37)
S	0	7±0.3 (0.05)	11±1 (0.12)
T	2±1 (0.66)	6±7 (1.16)	17±7 (0.39)
U	47±4 (0.08)	55±22 (0.39)	96±11 (0.12)
V	21±8 (0.38)	20±7 (0.36)	59±3 (0.05)
Average of labs	15±6 (0.42)	25±11 (0.42)	46±9 (0.19)
Average of 24 powders	15±17 (1.13)	25±19 (0.78)	46±30 (0.65)

Mean±S.D. (correlation coefficient).

compared for the extent of variation in the results.

(6) *Concentration of *E_2 used for incubation*

After the study was completed and initial analysis was performed, it was decided to classify the six laboratories into three groups based on the assay conditions. Group I—labs S and T—used ≤2 nmol as the highest estradiol concentration. Group II—labs Q, R and V—used 5 nmol as the highest concentration. Group III—Lab U—differed from the other two groups not only in the *E_2 concentrations (8 nM: highest) but also in the time and temperature of incubation (see Table 2). The data obtained within each group were compared with those between the groups for all the three different tissue powders and analysis of variance was used to see if one set of assay conditions gives consistently lower/higher results for all the three powders.

RESULTS

Nature of variabilities

(1) *Qualitative.* Since each of the test powders contained ER and the majority of the assays measured ER in all the three tissue powders, it can be assumed that all powders were + for ER. If 10 fmol/mg cytosol is taken as a cut-off value for categorizing a tissue as +, frequencies of misclassification of the powders as negative were as follows: low powder, 42%; medium, 25%; high, 4%. If the cut-off level is dropped to 3 fmol/mg, the frequencies of misclassification were 29% for low powder, 8% for medium and 0 for high powder.

The data were analyzed to see if the fmol/mg values reported for the three powders increased in the order L<M<H and if the differences between powders were significant. The results revealed that in all but one lab (Lab V) the average value of ER for the low powder was lower than that obtained for the medium powder. The difference was statistically significant for four labs out of five (not statistically significant for Lab T). For all the labs the average value for the medium powder was significantly lower than the average for high powder (Table 3).

(2) *Quantitative. Variations* (interlaboratory and intralaboratory) *in the quantity of ER measured.* Table 3 displays the average of the values reported by each lab and the coefficient of variation (ratio of standard deviation to the mean) of assay values for each powder. With one exception (Lab T) for the medium powder, all the coefficients of variation were less than 60% of the overall variation (average of 24 powders taken as samples) for the corresponding powder. Table 4 illustrates the ratio of average ER value obtained by each lab ($n=4$) to the overall mean value ($n=24$) obtained by all the six labs for each tissue powder. Labs S and T show a ratio lower than 0.4 while Lab U shows a ratio of higher than 2 for all three powders. Labs Q, R and

Table 4. *Receptor content (ratio)**

Participant lab	Low	Medium	High
Q	0.733	1.04	1.020
R	0.733	1.16	1.020
S	NEG	0.28	0.239
T	0.133	0.24	0.370
U	3.130	2.20	2.080
V	1.400	0.80	1.280

*Ratio = $\frac{\text{fmol/mg value reported by individual labs}}{\text{mean value of all six labs.}}$

V, on the other hand, show a range of ratios >0.5 and closer to 1.0.

(3) *Variation in protein quantitation.* The total protein yield as well as the degree of disparity between the protein values obtained by the central lab and those reported by individual labs were analyzed in order to find out if the differences in ER values between the labs could be a results of variations in the protein determinations. Table 5 illustrates the ratio of

Table 5. *Protein analysis*

Participants	Ratio*: Mean ± S.D.
Q	0.94 ± 0.05
R	0.79 ± 0.20
S	1.18 ± 0.34
T	1.35 ± 0.27
U	0.88 ± 0.13
V	0.98 ± 0.12

$$*\text{Ratio} = \frac{\text{Protein content of participant (mean)}}{\text{Protein content measured by control lab}}.$$

protein values obtained by each lab to that measured by the central lab. For no labs were these ratios significantly different from 1.00 when all the twelve protein determinations were analyzed as a single group. If the protein ratios were analyzed by batch (i.e., four assays done on a single day) there were two labs (Q and R) which had one batch with the ratio significantly less than 1.00 and one greater than 1.00 (Lab T). But since separate 0.05 level tests were performed on the 18 such batches (3 batches/lab/6 labs), one would expect to find two or three significant differences just by chance.

Comparative analysis of ER content reported by six labs divided into three groups

Since the division of six labs into three groups represents a *post-hoc* hypothesis, Scheffe's methods [7] were used so that the multiplicity of possible comparisons (there are 191 ways of dividing six labs into two or three groups) would not produce spuriously low significance levels. Table 6 shows the means, standard deviations and coefficients of variations of each group together with the results of testing the significance for differences in the ER values. Three groups differed from each other significantly for high level tissue powder with $P<0.001$. Group I obtained consistently lower values than Group II and Group III higher values than Group II. The differences in the ER values obtained by the three groups for medium level powder did not show statistical significance ($P>0.23$). On the other hand, the values obtained by Group III for the low powder were significantly higher than obtained by Group II ($P<0.005$). The values obtained for low level powder by Group I and II did not differ significantly ($P>0.25$).

DISCUSSION

There is a marked variability both in the qualitative (+ or −) and quantitative (fmol/mg cytosol protein) results of the estrogen receptor assay and hence the results reported by one laboratory are not comparable to those of the other. Significant conclusions that can be drawn from the study are:

(1) In this study the qualitative variability leading to misclassification of tumors is related to the quantity of receptors present in the samples, i.e., the higher the amount of re-

Table 6. *ER content*

(*a*) *Analysis of grouped labs*

	Group 1 S+T	Group 2 Q+R+V	Group 3 U
Low	1.1 ± 1.5 (1.36)	13 ± 8.8 (0.66)	46.8 ± 3.6 (0.08)
Medium	6.8 ± 4.9 (0.72)	24.3 ± 10.7 (0.44)	55.0 ± 21.5 (0.39)
High	13.9 ± 5.5 (0.40)	51.0 ± 11.3 (0.22)	95.5 ± 11.4 (0.12)

(*b*) *Values for tests of equality means*

Group 1 vs Group 2		Group 2 vs Group 3	
Low	>0.25	Low	0.005
Medium	>0.25	Medium	>0.25
High	<0.001	High	<0.001

ceptors, the less the frequency of misclassification.

(2) Despite the variability in numerical value reported by the labs for any particular sample, all the labs except V (Lab V measured the same amount of receptors in low and medium level powders) measured increasing quantity of receptors in the three samples, thus not misclassifying the "low level" samples as "high" or vice versa.

(3) Although this study was not designed to investigate the cause of the variability of results, an analysis of data after regrouping them based on similarity of assay conditions such as incubation time and temperature, range of estradiol concentration, etc. (Table 6), reveals certain interesting trends. When one compares the ratio of fmol/mg reported by Labs S and T over the mean values reported by all the labs, the ratios for all three samples are consistently below 0.5. Similarly, for Lab U they are consistently higher than 2.0 (Table 5). Labs S and T employed similar incubation time and temperature as Q, R and V but used a lower range of E_2 concentration. Lab U which used highest concentration of E_2 and high incubation temperature measured highest amount of receptors. It cannot be concluded whether it is the estradiol concentration or the incubation time and temperature that contributed to a higher content of receptors measured by Lab U. Theoretically, the differences in the concentration of estradiol used by different laboratories should have no significant influence in the measurement of absolute quantity of receptors. It could be argued that the assay conditions employed by Lab U were optimal leading to least denaturation of the receptor protein (i.e., longer incubation time may lead to proteolysis). It is to be emphasized that the contribution of the homogenization step to the observed variability in the assay results could not be ascertained from the data of our study.

Although a method for standardizing ER values reported by different laboratories is highly desirable, more quality control studies have to be conducted in order to understand the "reasons" for variabilities in the measurement and subsequently devise a protocol for standardization. If major factors that contribute to variation could be identified, then they could perhaps be eliminated.

It is highly desirable to institute a quality control program based on the "clinical response" data of patients whose tumors were assayed in a particular laboratory, but it is impractical at present, given the variability in treatment procedures as well as in parameters used for scoring responders. It is to be emphasized that the ultimate test for the laboratories' performance of the assay, however, lies in the reproducibility of the clinical response data of the patients whose tumors were assayed by the lab.

If a large scale quality control study based on a design similar to the present study were to be implemented, it might be possible to "normalize" the values reported by different labs, *especially* if a particular lab reports values that are constantly lower than (example S and T) or higher than (example U) the average of all the values obtained for a particular standard powder distributed by a central lab. One possible procedure for "normalizing" the data which seems to work well for the labs included in this study is described below.

A tissue powder in which all the labs measure some receptor protein is selected (in this study, the "high level" powder which is never classified as receptor-negative) and an arbitary receptor content close to the mean value is attributed to this tissue powder (i.e., 50 fmol/mg cytosol protein). An adjustment factor for each lab is then calculated, which is the ratio of 50 fmol to the mean values actually reported by the individual lab for that tissue powder. All other values reported by that lab will then be multiplied by the adjustment factor. Using this procedure, the "adjusted values" are given along with the reported

Table 7. Adjusted values

	Low	Medium	High
Average of all 24 samples prior to adjustment	15±17 (1.13)	25±19 (0.78)	46±30 (0.65)
Average of samples after adjustment	12±9 (0.75)	26±12 (0.46)	50±10 (0.20)
Range of values after adjustment (fmol/mg)	0–27	0–62	29–73

values (Table 7). It can be seen that the scatter of values is reduced; coefficient of variation also approximates the average intralab variation. A quality control program could be based on the standard deviation of the adjusted values. More study of this plan is needed to see if the proposed adjustment factor is fairly constant over time and range of ER values, and to determine what quality control measures could be most helpful.

The usefulness of any quality control program depends on its ability to screen out values which are too unreliable to use in making clinical decisions; however, it is not known how much "true" ER values vary within a tumor or whether any amount of care either in obtaining tissues for the ER assay or in the assay procedure itself would result in being able to predict with great accuracy hormonal responders among patients reported to have tumors with ER values between 3 and 100 fmol. For this reason, we have made only tentative suggestions for quality control.

REFERENCES

1. KORENMAN SJ, DUKES BA. Specific estrogen binding by the cytoplasm of human breast carcinoma. *J Clin Endocrinol* 1970; **30:** 639.
2. MCGUIRE WL, CARBONE PP, VOLLMER EP, eds. *Estrogen Receptors in Human Breast Cancer* New York: Raven Press, 1975.
3. DESOMBRE ER, CARBONE PP, JENSON EV, MCGUIRE WL, WELLS SA, WITTLIFF JL, LIPSETT MB. Steroids in breast cancer. *N Engl J Med* 1979; **301:** 1011.
4. KING RJB, ed. *Steroid Receptor Assays in Human Breast Tumors: Methodological and Clinical Aspects* Cardiff: Alpha Omega, 1979.
5. KING RJB, WITTLIFF JL, JUNGBLUT P, SMITH R. *NIH Consensus Conference on Steroid Receptors in Breast Cancer*, Bethesda, MD. (1978–79).
6. LOWRY OH, ROSEBROUGH NJ, FARR AL, RANDALL RJ. Protein measurements with Folin-phenol reagent. *J Biol Chem* 1951; **193:** 265.
7. SCHEFFE H. *The Analysis of Variance* New York: John Wiley, 1959.