Eur J Cancer Clin Oncol, Vol. 24, No. 8, pp. 1263–1269, 1988.
Printed in Great Britain

0277-5379/88$3.00+0.00

Impact of Standardization of Estrogen and Progesterone Receptor Assays of Breast Cancer Biopsies in Denmark

S.M. THORPE,* H.S. POULSEN,† K.O. PEDERSEN‡ and C. ROSE§
on behalf of the DBCG Receptor Committee

**University Department of Clinical Physiology and Nuclear Medicine, The Finsen Institute, Rigshospitalet, DK-2100, Copenhagen, Denmark, †The Danish Cancer Society, Receptor Laboratory, Department of Experimental Clinical Oncology, Radiumstationen, University Hospital of Aarhus, DK-8000, Aarhus, Denmark, ‡Department of Clinical Chemistry, Aalborg Sygehus, DK-9000, Aalborg, Denmark and §Department of Oncology R, Odense University Hospital, DK-5000, Odense, Denmark*

Abstract—*Estrogen and progesterone receptor (ER and PgR, respectively) data for the three laboratories participating in the Danish Breast Cancer Cooperative Group (DBCG) project for treatment of primary breast cancer are presented for the period* 1979–1986. *The frequency of ER positivity remained constant for one laboratory throughout this period, while this value changed significantly in the other two laboratories. Inter-laboratory reproducibility (evaluated as the frequency of ER positivity) was poor at the onset of the project* ($P = 0.0003$) *but, due to standardization procedures, improved after* 1982.

Slight but significant differences in the composition of the patient populations at the three centers (menopausal status and tumor size) may account for some of the differences observed both in frequency of PgR positivity as well as PgR concentrations determined.

Greater intra- and inter-laboratory differences were observed in all three laboratories for PgR than for ER. Part of this variation is believed to have been alleviated by the addition of 10 mM *sodium molybdate to the assay buffer in* 1983. *From having very divergent frequencies of PgR positivity in the three laboratories* (31–71%) *in* 1981, *this divergence has been reduced in* 1985 (62–78%).

While data regarding ER status significantly distinguish between patients with long versus short recurrence-free survival (irrespective of treatment) in one laboratory in both the 77 *and the* 82 *generations of clinical protocols within the DBCG program, ER status from the other two laboratories makes this distinction only in the* 82 *protocols. We attribute this inability of ER status to distinguish among patients in the* 77 *protocols to the suboptimal nature of the assays performed in these two laboratories at that time.*

The overall improvement in comparability of data from the three laboratories during the standardization procedures as well as the fact that ER status from all three laboratories is now capable of distinguishing different patient groups attest to the fact that standardization procedures are both necessary and useful.

INTRODUCTION

TWO DIFFERENT approaches can be applied to evaluate whether inter-laboratory methodology for estrogen and progesterone receptor (ER and PgR, respectively) determinations is comparable. The first, and most frequently applied, is a comparative analysis of standard, lyophilized cytosols within each laboratory. An alternative approach is to evaluate the frequency of receptor positivities as well as the distributions of receptor concentrations in primary breast cancer biopsies determined in each laboratory. However, since age, menopausal status and tumor size are related to ER and PgR concentrations [1, 2], the composition of the patient populations in each center must be similar with respect to all three of these factors before the latter type of evaluation is valid. Furthermore, there must be a reasonable number of patients in each center

Accepted 18 January 1988.
Correspondence to be addressed to: Susan M. Thorpe, Department of Clinical Physiology and Nuclear Medicine, The Finsen Institute, Strandboulevarden 49, 2100-DK Copenhagen Ø, Denmark.

to form the foundation for such a study.

In the nationwide Danish Breast Cancer Cooperative Group (DBCG) Project, ER and PgR analyses are now routinely performed in three laboratories. All three laboratories use the dextran-coated charcoal (DCC) technique with multipoint titration analysis as recommended by the EORTC Receptor Study Group [3]. Because of the large number of patients accrued and the uniformity of the program, a study of the above-mentioned type is potentially possible. A report of assay results using lyophilized standard cytosols in the three laboratories prior to instigation of standardization procedures has previously been published [4]. Inter-laboratory reproducibility in 1981 was poor: improvement of ER assay procedures was needed in both laboratories 2 and 3, while PgR analyses needed special attention mainly in laboratory 2. Efforts to standardize the methods in these three laboratories were initiated through the DBCG Receptor Committee in 1981. The present report describes both the consequences of the standardization efforts and illustrates the impact that differences in assay results can have upon the observed clinical correlations.

PATIENTS AND METHODS

The 5538 patients in the present study are patients with primary breast cancer who are registered in the Danish Breast Cancer Cooperative (DBCG) project. Approximately 62% of these patients (3433) have been entered into the DBCG protocols for patients at either low (protocol a) or high (protocol b or c) risk for recurrent disease. The organization, design, and follow-up of the DBCG program and the 77 protocols have been described in detail elsewhere [5]. Briefly, the primary surgical treatment was total mastectomy and axillary node sampling. In the present study 28% and 34% of the patients were allocated to the low and high risk protocols, respectively. High risk patients (lymph node involvement, tumor > 5 cm in diameter and/or fixation of tumor to skin or fascia) received various forms of adjuvant therapy (radio-, chemo- and/or endocrine therapy).

Women were defined as being postmenopausal when menostasia had persisted for at least 5 years; otherwise they were classified as being pre/perimenopausal.

Death/recurrence data were evaluated as of 31 December 1986. Recurrence is defined as the appearance of new lesion(s) in patients with no previous evidence of recurrent disease and is confirmed by physical examination, biopsies and/or other relevant diagnostic procedures. Cause of death is not recorded; the frequency of deaths due to causes other than breast cancer is assumed to be equal in different patient subgroups for patients of the same age.

Accrual of data from laboratory 1 commenced in 1979, while laboratories 2 and 3 commenced in 1980 and 1981, respectively. At the time of this analysis complete annual reports had been filed by laboratories 1 and 2 for 1986, while the most recent complete annual report from laboratory 3 is for 1985.

Receptor analyses

Tumor tissue was analyzed for ER and PgR concentrations using the dextran-coated charcoal (DCC) method recommended by the EORTC [3], which involves multipoint titration analysis and Scatchard analysis of the binding data [6]. Correction for non-specific binding was performed using the method described by Chamness and McGuire [7]. 2,4,6,7-[^{3}H]Estradiol (TRK.322, 85–110 Ci/mmol, Amersham) was used for the ER analyses and either 17-methyl[^{3}H]R5020 (RU 5020, 70–90 Ci/mmol, New England Nuclear) or 16-ethyl-21-hydroxy-10-nor-6,7[^{3}H]pregna-4-en-3,20-dione (ORG.2058, 40–60 Ci/mmol, Amersham TRK 629) was used for the PgR analyses.

Only patients with verified malignant tissue in the biopsy specimen sent for receptor analysis were included in this study.

A cut-off level of 10 fmol/mg cytosol protein has been employed to distinguish between receptor positive ($\geq$ 10 fmol/mg cytosol protein) and receptor negative biopsies.

Protein analysis

Protein determinations were performed using the Bio-Rad method (Coomassie Brilliant Blue) with Kabi Diagnostica human serum albumin protein standard as a reference. Receptor concentrations are expressed in relation to protein concentration in the DCC-treated cytosol.

Statistical methods

The nonparametric Kruskal–Wallis test has been employed to detect differences in distributions. Otherwise, conventional statistical methods have been applied to the data. The particular test used in the individual circumstance is indicated in the text or in the relevant table or legend. A value of $P < 0.05$ was considered to be significant.

Recurrence-free survival (RFS) was evaluated using the Kaplan–Meier estimates with recurrence or death as endpoints and compared by the log-rank test.

RESULTS

Composition of patient population

Patient age does not differ significantly in the three centers viewed overall (Table 1). Despite the similarity in distribution of patient age, a slight but highly significant difference in frequencies of

Table 1. Characteristics of patients enrolled in the DBCG project

Characteristic	Laboratory 1	Laboratory 2	Laboratory 3
Patient number	4126	730	682
Age* (year: median, first and last quartile)	61 (51–71)	60 (49–70)	61 (50–71)
Menopausal status† (% of each classification within each center)			
Pre-/peri-	31%	35%	37%
Post-	67%	64%	63%
Unknown	2%	1%	0%
Tumor diameter‡ (cm: median, first and last quartile)	2.5 (1.8–4.0)	2.8 (2.0–4.0)	2.0 (1.5–3.0)

*Kruskal–Wallis test, $P = 0.13$.
†$\chi^2 = 22.6$, $P = 0.0002$.
‡Kruskal–Wallis test, $P = 0.0001$.

menopausal states is found among the three centers (Table 1). Moreover, the difference in median tumor size (2.0–2.8 cm) is also highly significant.

Frequency of receptor positivity

The frequencies of E and PgR positivities observed within each laboratory is shown in Fig. 1a and b. While the frequency of ER positivity was relatively constant throughout 1979–1986 in laboratory 1 (median, range: 77.5%, 74–80%), it varied significantly in laboratories 2 and 3 (medians, ranges: 64%, 46–76% and 78%, 76–87%, respectively). The intra-laboratory changes in frequencies of ER positivity are significant for both laboratories 2 and 3 ($P = 0.002$ and 0.05, respectively), but not for laboratory 1 ($P = 0.09$). As can be seen in Table 2 significant inter-laboratory differences in frequencies of ER positivity are observed from 1980–1982, after which time they are similar (with the exception of 1985).

Differences in frequencies of PgR positivity (Fig. 1b) are more pronounced than are those for ER positivity. The inter-laboratory difference in frequency of PgR positivity has been significantly reduced from 40% in 1981 to 14% in 1985. While a general increase in frequency of PgR positivity was observed in laboratories 2 and 3, an overall decrease was observed in laboratory 1. This decrease in frequency of PgR positivity coincides in time with the addition of sodium molybdate (10 mM) to the assay buffer (1983). The intra-laboratory changes in frequency of PgR positivity are significant for laboratories 1 and 2 ($P = 0.001$ in both cases) but not for laboratory 3 ($P = 0.09$).

As can be seen in Table 2, there continues to be more disparity with regard to frequency of PgR positivity than ER positivity during the course of time.

Distribution of receptor concentrations in receptor positive biopsies

To examine whether standardization procedures affected absolute concentrations of receptor determined as well as the ability to detect receptors (reflected by frequency of receptor positivity), both intra- and inter-laboratory distributions of receptor values have been investigated. This has been accomplished by comparing the distributions of the logarithmically transformed concentrations of those biopsies classified as receptor positive (i.e. $\geqslant 10$ fmol/mg cytosol protein) from the year of 1981 with the distributions observed in the year most recently reported from each laboratory (Table 3).

Between 1981 and 1985 a significant increase in ER concentrations was noted in laboratory 3, while no significant differences were noted in the other two laboratories (Table 3). Moreover, the concentrations of PgR determined in 1985/86 are significantly higher in laboratories 1 and 3 than in 1981, while no changes were observed in laboratory 2.

Viewed in terms of inter-laboratory comparability, there is no significant difference in recorded ER concentrations among ER positive biopsies within the three laboratories in 1981, while there is a significant difference in 1985/86. Since there is no difference in distribution of patient age in the three centers in 1985/86 ($P = 0.40$), the reasons for this difference remain obscure. While inter-laboratory PgR concentrations differed significantly

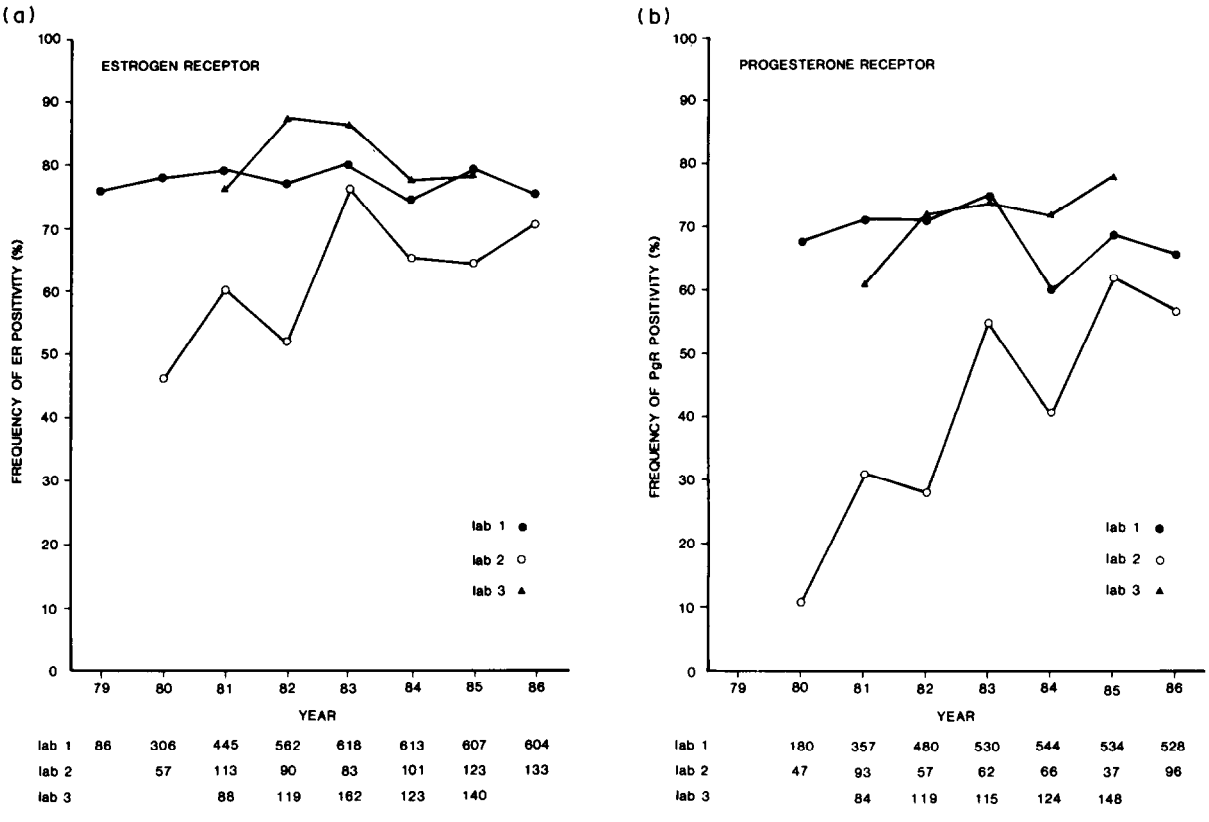

Fig. 1. (a) Frequency of ER positivity for primary breast tumors 1979–1986. Laboratories 1, 2 and 3 are indicated by closed circles (●), open circles (○), and closed triangles (▲), respectively. The number of biopsies for each year with each laboratory is shown at the bottom of the figure. (b) Frequency of PgR positivity for primary breast tumors 1979–1986. Symbols as in Fig. 1a. The number of biopsies for each year within each laboratory is shown at the bottom of the figure.

Table 2. Comparability of inter-laboratory frequency of receptor positivity

	P value for inter-laboratory comparability Year						
	1980	1981	1982	1983	1984	1985	1986*
Frequency ER+	0.001	0.0003	0.0001	0.099	0.12	0.001	0.26
Frequency PgR+	0.0001	0.0001	0.0001	0.003	0.0002	0.07	0.10

*Values for laboratories 1 and 2 only.
P values for Kruskal–Wallis test.

before standardization was begun, they were similar in 1985/86.

Clinical course of the disease in relation to receptor status

The distribution of patients from each of the three centers into the protocols for low and high risk patients can be seen in Table 4. While there is no difference in distribution of patients in low and high risk protocols in the 77-generation of clinical trials, a highly significant difference is observed with regard to the 82-generation. In the 1982 protocols exceptionally low (laboratory 2) and high (laboratory 3) proportions of patients are allocated to the high risk groups compared to the distribution observed in laboratory 1.

The ability of receptor status to distinguish between groups of patients with long versus short recurrence-free survival irrespective of treatment is shown in Table 5. Inter-laboratory results from the 77 protocols can be considered only cautiously

Table 3. Intra- and inter-laboratory distribution of receptor concentrations in receptor positive primary breast cancer tumors in 1981 and 1985/86

	ER* (fmol/mg cytosol protein)			PgR* (fmol/mg cytosol protein)		
	1981	1985/86	*P* value†	1981	1985/86	*P* value†
Laboratory 1	95	97	0.83	99	141	0.001
Laboratory 2	122	126	0.90	106	136	0.472
Laboratory 3	84	138	0.008	38	143	0.0001
Inter-laboratory comparability (*P* value)	0.23	0.01		0.0001	0.93	

*Mean of logarithmically transformed values.
†*P* values based on Wilcoxon rank sum test performed on logarithmically transformed receptor values.

Table 4. Distribution of patients in low and high risk protocols

	77 protocols*			82 protocols†		
	No. of patients	Low risk	High risk	No. of patients	Low risk	High risk
Laboratory 1	884	45%	55%	1610	46%	54%
Laboratory 2	174	46%	54%	310	52%	48%
Laboratory 3	172	45%	55%	283	39%	61%

*$\chi^2 = 0.099$; $P = 0.95$.
†$\chi^2 = 10.13$; $P = 0.006$.

Table 5. P values for lifetable analysis of patients enrolled in low or high risk DBCG protocols

	Laboratory 1			Laboratory 2			Laboratory 3		
	n	Frequency receptor positivity	*P* value	*n*	Frequency receptor positivity	*P* value	*n*	Frequency receptor positivity	*P* value
77 protocols									
ER	860	77%	0.0005	174	54%	0.344	172	82%	0.661
PgR	610	68%	0.0012	131	28%	0.931	170	67%	0.548
82 protocols									
ER	1471	74%	0.0001	295	66%	0.011	276	79%	0.0001
PgR	1289	68%	0.0001	184	55%	0.740	278	75%	0.005

because of the small number of patients in laboratories 2 and 3.

The ability of both ER and PgR statuses to significantly discriminate between patients with long versus short RFS irrespective of adjuvant treatment is clearly observed in data from laboratory 1 for both the 77 and 82 protocols. While such differences are not apparent in the 77 protocols in either laboratory 2 or 3, they appear in the 82 protocols for ER status in both laboratories, and for PgR status in laboratory 3. It is important to note that with regard to the 77 protocols inclusion of the data from 346 patients (i.e. 29% of the total available) from laboratories 2 and 3 together with data from laboratory 1 for lifetable analysis reduces rather than increases the overall ability to discriminate beween patients with long versus short RFS employing ER status as the discriminator (reduction of *P* value from 0.0005 to 0.0013). Equally important, in the 82 protocols—and at a time when intercomparability had improved—pooling of data from the three laboratories does *not* alter the significance of ER or PgR status as discriminators between long versus short RFS.

Despite a similar number of patients from laboratories 2 and 3 in the 82 protocols, the difference in

ability to differentiate patient groups with respect to RFS is greater in laboratory 3 than in laboratory 2 (see Discussion).

DISCUSSION

These results illustrate that the tenuous tasks involved in inter-laboratory standardization of assay methodology are constructive. An increase in frequency of receptor positivity is observed following initiation of standardization procedures for both ER and PgR in laboratory 2 and for PgR in laboratory 3. Furthermore, the highly significant inter-laboratory differences in frequency of receptor positivity for both ER and PgR in 1981 have been reduced or eliminated in 1985.

Provided that patient populations are similar, the frequency of receptor positivity can be used as an indicator of the sensitivity of the assay as performed in different centers. The frequency of receptor positivity is affected by numerous factors, several of which are associated with the thermolabile nature of ER and PgR. Another factor influencing the observed frequency of receptor positivity is the stabilizing effect of molybdate especially with regard to PgR. This influence of molybdate on the frequency of PgR but not ER positivity has previously been reported [8]. All three laboratories included molybdate (10 mM) in the assay buffer in 1983. The fact that the frequency of PgR positivity in laboratory 2 approaches that of laboratories 1 and 3 after 1983 may be due in part to the stabilizing effect of molybdate on the PgR receptor. A consistent decrease in frequency of PgR positivity has been noted following addition of molybdate to the assay buffer in laboratory 1. This can probably be attributed to an improved ability to discriminate between specific and nonspecific binding of the radioactive ligand. The fact that the K_d values for PgR binding decreased significantly following the addition of molybdate to the buffer [8] substantiate this conjecture. Moreover, despite the decrease in frequency of PgR positivity observed in laboratory 1, it is noteworthy that an overall increase in the absolute concentrations of PgR was simultaneously found in laboratory 1 as well as in laboratories 2 and 3 (Table 3).

Aside from the addition of molybdate, no other single factor(s) can be pointed out that have led to the increase in frequency of receptor positivities observed. During the standardization phase, a common assay protocol that is in accord with the recommendations put forth by the EORTC Receptor Study Group was adapted. Methods of computing binding results were discussed in detail. Working sessions encompassing the entire assay procedure were set up with laboratory technologists from the three centers as participants. Batches of lyophilized cytosols were prepared and distributed: these 'standard' cytosols are still routinely analyzed along with patient samples weekly in each of the three laboratories and results are discussed biannually. Finally, simple awareness of existing inter-laboratory discrepancies appears to be an important factor, which emphasizes the usefulness of continuous inter-laboratory comparisons using lyophilized tissue cytosols.

Although all three laboratories participate in the same national protocols, slight but highly significant differences are found in the composition of patient populations in the three centers. Biopsies received at laboratory 3 are generally smaller than those received at the other two laboratories, and they are derived from a greater proportion of pre/peri- as opposed to postmenopausal patients. Because of the associations between receptor concentrations and both menopausal status and tumor size [1, 2], the frequency of PgR positivity as well as PgR concentrations might be expected to be somewhat higher in laboratory 3 than in either laboratory 1 or 2. Such a tendency in frequency of PgR positivity is, indeed, observed, which emphasizes the necessity of considering the composition of patient populations when comparing results from different centers.

It is noteworthy that this study demonstrates that inclusion of data from suboptimally performed assays can reduce or mask clinical correlations. Neither laboratory 2 or 3 could significantly distinguish between patients with long versus short RFS in the 77 protocols using receptor status as the prognosticator, while both can make this distinction in the 82 protocols with regard to ER status. Furthermore, a distinction is also possible in laboratory 3 with regard to PgR status in the 82 protocols. Even when data for the 77 protocols from both laboratories 2 and 3 are pooled, which increases the number of patients two-fold and to a total comparable to that seen in each of the two laboratories in the 82 protocols, the P values for the lifetable analysis remain insignificant (ER, $P = 0.73$ and PgR, $P = 0.65$), indicating that it is the quality of the assay rather than the number of patients that influence the results.

While all three laboratories find significant correlations between ER status and RFS in the 82 protocols, it would appear that laboratory 2 might be less proficient in this respect than laboratory 3 ($P = 0.01$ vs. 0.001). However, laboratory 3 has an unusually large proportion of patients protocolled in high opposed to low risk protocols, while the opposite situation is observed in laboratory 2. Since events occur earlier among high than low risk patients, the apparent difference in ability of ER status to predict RFS in the two laboratories may in part be a reflection of the difference in the proportion of low and high risk patients in the two populations.

In conclusion, this study emphasizes that when data generated in different receptor laboratories are being interpreted in relation to clinical parameters, both inter-laboratory comparability of the assays performed as well as the composition of the patient population must be considered in the interpretation of the results.

Acknowledgements—The skillful technical assistance of Ane-Marie Hansen, Marianne Barfoed, Bente Egeblad, Birthe Hermansen and Mette Thomsen is gratefully acknowledged.

REFERENCES

1. Clark GM, Osborne CK, McGuire WL. Correlations between estrogen receptor, progesterone receptor, and patient characteristics in human breast cancer. *J Clin Oncol* 1984, **2**, 1102–1109.
2. Thorpe SM, Rose C. Oestrogen and progestrone receptor determinations in breast cancer: technology and biology. *Cancer Surv* 1986, **5**, 505–525.
3. EORTC Breast Cancer Cooperative Group. Revision of the standards for assessment of hormone receptors in human breast cancer: report of the second EORTC workshop held on 16–17 March, 1979 in the Netherlands Cancer Institute. *Eur J Cancer* 1980, **16**, 1513–1515.
4. Thorpe SM, Olsen K, Pedersen KO, Poulsen HS, Rose C. Quality control of estrogen and progesterone receptor analyses in Denmark. *Recent Results Cancer Res* 1984, **91**, 91–93.
5. Andersen KW, Mouridsen HT, Castberg TH *et al.* Organization of the Danish adjuvant trials in breast cancer. *Dan Med Bull* 1981, **28**, 102–106.
6. Scatchard G. The attractions of proteins for small molecules and ions. *Ann NY Acad Sci USA* 1949, **51**, 660–672.
7. Chamness GC, McGuire WL. Scatchard plots: common errors in correction and interpretation. *Steroids* 1975, **26**, 538–540.
8. Thorpe SM. Steroid receptors in breast cancer: sources of inter-laboratory variation in dextran-charcoal assays. *Breast Cancer Res Treat* 1987, **9**, 175–189.