

# Reliability of Serum Prolactin Measurements in Women<sup>1</sup>

Karen L. Koenig,<sup>2</sup> Paolo Toniolo, Peter F. Bruning,  
J. M. G. Bonfrer, Roy E. Shore, and  
Bernard S. Pasternack

Institute of Environmental Medicine and Kaplan Comprehensive Cancer Center, New York University Medical Center, New York, New York [K. L. K., P. T., R. E. S., B. S. P.]; and The Netherlands Cancer Institute, Amsterdam, The Netherlands [P. F. B., J. M. G. B.]

## Abstract

**Prolactin, a hormone indispensable for milk secretion, has been shown to enhance the development and growth of mammary tumors in rodents; however, its importance in human breast cancer is uncertain. Serum prolactin levels are known to fluctuate considerably under normal conditions, and lack of precision in the hormone measurements may have contributed to the largely negative findings in humans to date. The purpose of this study was to investigate the reliability of prolactin measurements in women using stored serum from an ongoing prospective study of breast cancer. Separate groups of postmenopausal and premenopausal women who donated multiple blood samples at approximately 1-year intervals were studied. The reliability of a single log prolactin determination, as measured by the intraclass correlation coefficient, was 0.76 for the postmenopausal women (95% confidence interval, 0.66–0.85) and 0.48 for the premenopausal women (95% confidence interval, 0.31–0.62). These findings suggest that a single measurement is sufficient to characterize the serum prolactin level of postmenopausal women for epidemiological research. For premenopausal women, however, multiple samples are desirable. Controlling for phase of the menstrual cycle does not appear to substantially improve the reliability of premenopausal measurements.**

## Introduction

Prolactin, a hormone indispensable for milk secretion, has been shown to enhance the development and growth of mammary tumors in rodents (1). The importance of prolactin in human breast cancer, however, remains to be demonstrated (2). We are investigating the relation between prolactin levels and subsequent breast cancer risk in the NYU<sup>3</sup> Women's Health Study, a case-control-within-a-cohort

study of endogenous hormones and breast cancer (3). An important consideration in this and other epidemiological studies involving prolactin measurements is the reliability of the measurements, since prolactin levels are known to fluctuate considerably under normal conditions. Among the factors affecting prolactin levels are stress, drugs, menopausal status, time in the menstrual cycle, time of day, and time since last meal (4).

The availability of repeat blood samples drawn at approximately 1-year intervals in a large subset of the study cohort gave us the opportunity to study the reproducibility of prolactin measurements in women. The goal of our investigation was to answer the following questions. (a) Is the variability in the level of prolactin within subjects so great that a single measurement reveals little about a woman's long-term exposure? (b) How many measurements are needed to characterize a woman's prolactin exposure status for the purposes of epidemiological research?

## Methods

**The NYU Women's Health Study.** Between March 1985 and June 1991, the NYU Women's Health Study enrolled a cohort of 14,290 women aged 34–65 years at the Guttman Breast Diagnostic Institute, a breast screening clinic in New York City. At the time of enrollment and at annual screening visits thereafter, subjects were asked to complete questionnaires and to provide 30 ml of peripheral venous blood. Fifty-one % of cohort members donated blood on more than one occasion, usually at 1-year intervals. Blood was collected before breast examination, between 9:00 a.m. and 3:00 p.m., and at any time during the menstrual cycle of the premenopausal women. No fasting was required. Blood specimens were kept at room temperature for approximately 1 h and at 4°C for 30 min. Samples were then centrifuged at 3500 rpm for 15 min, and serum was partitioned into 1-ml aliquots in airtight plastic vials and frozen at –80°C for long-term storage.

**Study Groups.** Since prolactin levels drop after the menopause, separate groups of premenopausal and postmenopausal women were studied. Three groups were selected: (a) postmenopausal women; (b) premenopausal women who gave repeat samples in any phase of the menstrual cycle; and (c) premenopausal women who gave multiple samples in the luteal phase. The third group was chosen to control for the additional variability in premenopausal prolactin levels caused by menstrual cycling. A sample was considered to be luteal phase if it was drawn within 11 days of the next menstruation. Date of next menstruation was obtained from mail-back calendars distributed at the time of blood drawing. Statistical analyses were carried out separately in each of the three study groups.

To control for variability of prolactin levels with age (5), age at sampling was restricted to ranges within which prolactin levels are relatively constant. Thus, postmenopausal samples had to have been drawn between the ages of 55 and

Received 11/18/92; revised 2/10/93; accepted 2/10/93.

<sup>1</sup> This research was supported by Grants CA 34588, CA 13343, and CA 16087 from the National Cancer Institute.

<sup>2</sup> To whom requests for reprints should be addressed, at Institute of Environmental Medicine, New York University Medical Center, 341 East 25th Street, New York, NY 10010.

<sup>3</sup> The abbreviations used are: NYU, New York University;  $\rho_i$ , intraclass correlation coefficient.

69 years, and premenopausal samples had to have been drawn between the ages of 34 and 49 years. Only women with at least one intact ovary were eligible. Any samples drawn in the 6 months following pregnancy were excluded. Also excluded were samples drawn while the subject was taking drugs that might affect serum levels of prolactin. These drugs included sex hormones; glucocorticoids; major tranquilizers, such as haldol and prolixin; certain blood pressure medications (reserpine,  $\alpha$ -methyldopa, pargyline); certain antidepressants, such as tofranil and pamelor; codeine; apomorphine; dilantin; L-dopa; metoclopramide; amphetamines; dopamine; bromocryptine; and cimetidine (4, 6–8).

**Laboratory Assays.** The serum samples were packed in dry ice and sent to the Netherlands Cancer Institute, where they were stored at  $-80^{\circ}\text{C}$  until they were assayed in a single batch. The laboratory assay was an automated “sandwich” enzyme immunoassay technique using two monoclonal antibodies highly specific to prolactin (9). Boehringer Mannheim (Germany), the manufacturer, reports no measurable cross-reaction with other anterior pituitary hormones (hGH, hCG, TSH, FSH, and LH). All sample values were above the 2 ng/ml lower limit of detection of the assay.

**Reliability.** Our goal was to determine how well a single serum measurement characterizes a woman’s long-term serum level of prolactin relative to other women. We assume that each woman has a long-term mean level of prolactin, about which her individual measurements fluctuate. Part of this fluctuation is due to error in the laboratory measurement and part to real biological variation in prolactin levels over time. We assumed the following model (model 1) for  $Y_{ijk}$ , the measured prolactin value for the  $k$ th aliquot from the  $j$ th blood sample donated by the  $i$ th subject:

$$Y_{ijk} = \mu + \alpha_i + \beta_{ij} + \epsilon_{ijk}$$

where  $\mu$  is the overall mean prolactin level,  $\mu + \alpha_i$  is the mean prolactin level for the  $i$ th subject,  $\beta_{ij}$  is the difference between the true prolactin level for subject  $i$  at donation  $j$  and her mean level, and  $\epsilon_{ijk}$  is the laboratory measurement error, or the difference between the measured and true prolactin values for the  $k$ th aliquot from the  $j$ th sample donated by subject  $i$ . The random variables  $\alpha_i$ ,  $\beta_{ij}$ , and  $\epsilon_{ijk}$  were assumed to be mutually independent with mean 0 and variance  $\sigma_{\alpha}^2$ ,  $\sigma_{\beta}^2$ , and  $\sigma_{\epsilon}^2$ , respectively. Under these conditions, the variance of a prolactin measurement ( $Y_{ijk}$ ) is the sum of the three variance components,  $\sigma_{\alpha}^2$ ,  $\sigma_{\beta}^2$ , and  $\sigma_{\epsilon}^2$ .  $\sigma_{\alpha}^2$  represents the variance of the distribution of the mean levels of the subjects (the between-subjects component), while  $\sigma_{\beta}^2$  and  $\sigma_{\epsilon}^2$  sum to the within-subjects component, the variance of the measured values of an individual subject over time. For the purpose of characterizing the long-term mean prolactin level of a woman, both  $\beta_{ij}$  and  $\epsilon_{ijk}$  can be considered to be “error,” and the sum of these two components will be referred to in this article as the within-subject error.

The extent of agreement among repeated measurements made on the same subject was assessed by means of the reliability coefficient (10). The reliability ( $\rho_1$ ) of a measurement is defined as the proportion of the total variability in the measurement which is due to variability between subjects. The reliability of a single prolactin measurement is therefore given by

$$\rho_1 = \frac{\sigma_{\alpha}^2}{\sigma_{\alpha}^2 + \sigma_{\beta}^2 + \sigma_{\epsilon}^2}$$

$\rho_1$  is also known as the intraclass correlation coefficient, since it is the correlation between repeated measurements on the same variable (11).

**Design of the Reliability Study.** Our laboratory had agreed to conduct 500 assays for this study. After reserving about 50 assays for the measurement of duplicate aliquots from the same sample, we allocated the remaining measurements equally between the three study groups. The criterion we used to decide how many samples to measure per subject was to choose the number of replicates which would minimize the variance of our statistic of interest, the intraclass correlation coefficient. The variance of this statistic (12, 13) is a function of  $\rho_1$ , the number of subjects and the number of replicates/subject. We computed this variance for different values of the intraclass correlation and different numbers of replicates, holding the total number of measurements/group (number of replicates  $\times$  number of subjects) constant at 150. For intraclass correlations around 0.5, which we expected to find for prolactin, the variance was minimized by use of three or four replicates.

We thus chose random samples of 40 women with four replicates for the postmenopausal and premenopausal (any phase) groups, and we used all 49 women available with three replicates in the luteal phase for the premenopausal (luteal-phase) group. Four of the subjects were selected for both premenopausal groups. Duplicate aliquots for 44 of the samples were also assayed, bringing the total number of assays to 500.

**Data Analysis.** The distributions of both the means of the subjects and the within-subject errors were more nearly normal when the measurements were log-transformed. Because of this, and because we plan to use the log of the measurements when investigating the relation between prolactin and breast cancer, we log-transformed the prolactin measurements for most analyses. Plots of the errors versus the mean gave no evidence that the errors increased with increasing mean, consistent with our assumption that the errors and the means are independent.

One-way random effects analysis of variance models (ignoring the 44 duplicate measurements and combining  $\beta_{ij}$  and  $\epsilon_{ijk}$  from model 1 into a single error term) were used to estimate the between- and within-subjects variance components in each of the three study groups. A separate one-way random effects model limited to the subset of samples with duplicate measurements was used to estimate the component of variance due to laboratory error ( $\sigma_{\epsilon}^2$ ). This was the only analysis in which the duplicate measurements were used. The samples with duplicates were taken from both premenopausal and postmenopausal women, and  $\sigma_{\epsilon}^2$  was assumed to be the same for both menopausal groups. Confidence intervals and statistical tests for the reliability coefficients were computed as described in Ref. 12.

The effects of the continuous covariates age and sample storage time were assessed using the analysis of covariance model (model 2)

$$Y_{ij} = \mu + \alpha_i + \delta X_{ij} + \omega_{ij}$$

where  $\alpha_i$  is the random subject effect,  $\omega_{ij}$  is the within-subject error (analogous to the sum of  $\beta_{ij}$  and  $\epsilon_{ijk}$  in model 1 but adjusted for the covariate  $X$ ), and  $Y_{ij}$  is the log of the measured prolactin value for subject  $i$  at donation  $j$ . In this model the fixed effect  $\delta$ , assumed constant across subjects, measures the change in log prolactin value for a 1-unit increase in the value of the covariate. The effect of categorical covariates was assessed using the analysis of variance model

Table 1 Serum prolactin values in the three study groups

	Postmenopausal		Premenopausal	
	All subjects	Minus subject 21	All phases	Luteal phase
Age at sampling (years)	55–69		34–49	34–49
Number of subjects	40		40	49
Number of replicates	4		4	3
Prolactin values (ng/ml)	2.9–43.2	2.9–32.5	3.5–80.4	3.6–57.2
Arithmetic mean	8.9	8.2	13.5	13.6
Geometric mean	7.8	7.5	12.1	12.1

(model 3)

$$Y_{ijk} = \mu + \alpha_i + \delta_j + \omega_{ijk}$$

where  $\alpha_i$  is the random subject effect,  $\omega_{ijk}$  is the within-subject error,  $\delta_j$  is the (fixed) effect on the log prolactin value of level  $j$  of the covariate, and  $Y_{ijk}$  is the measured log prolactin value for the  $k$ th sample donated by subject  $i$  at level  $j$  of the covariate. The categorical covariates investigated were time of day of the sample (before versus after noon), time since last meal (<3 versus  $\geq$ 3 h), and phase of the menstrual cycle. Phase of cycle, analyzed in the group of premenopausal subjects with samples taken throughout the menstrual cycle, was categorized as luteal (0–11 days prior to the next menstruation), ovulatory (12–16 days), late follicular (17–19 days), and early follicular ( $\geq$ 20 days). The covariate effects ( $\delta$ ) for models 2 and 3 were estimated using the Statistical Analysis System (SAS) General Linear Models procedure (14). Variance components were estimated using the SAS VARCOMP procedure, and reliability coefficients adjusted for covariates were computed as (see Refs. 10 and 15)

$$\frac{\sigma_a^2}{\sigma_a^2 + \sigma_w^2}$$

## Results

Serum prolactin values in the three study groups are given in Table 1. The prolactin values in the postmenopausal women ranged from 2.9 to 43.2 ng/ml with a geometric mean of 7.8. The premenopausal prolactin values were higher, ranging from 3.5 to 80.4 ng/ml with a geometric mean of 12.1. The geometric means in all groups were less than the arithmetic means, indicating that the distributions were skewed to the right. Values for the postmenopausal group are also shown with one influential subject (number 21) excluded. Four of the five highest measurements in the postmenopausal group belonged to this subject.

Variance components and intraclass correlations computed on the log-transformed data are shown in Table 2. Despite the fact that the luteal phase group had all samples taken in the same phase of the menstrual cycle, the intraclass correlation coefficient for this group was the same as in the premenopausal group in which samples were taken randomly throughout the cycle. The intraclass correlation coefficient for both premenopausal groups was 0.48 (95% confidence interval,  $\sim$ 0.31–0.62). The intraclass correlation coefficient in the postmenopausal women was 0.76 (95% confidence interval, 0.66–0.85), and this value was significantly higher ( $P < 0.01$ ) than the premenopausal estimates. The lower reliability in premenopausal women appears to be due to greater within-subject variability before menopause, since the within-subject variance component is twice as great in the premenopausal women as in the postmeno-

Table 2 Reliability of serum prolactin measurements (natural log scale)

	Postmenopausal		Premenopausal	
	All subjects	Minus subject 21	All phases	Luteal phase
Variance components				
Between subjects	0.166	0.107	0.101	0.105
Within subjects	0.052	0.052	0.110	0.116
Within samples (laboratory)			0.011	
Intraclass correlation	0.76	0.67	0.48	0.48
95% confidence interval	0.66–0.85		0.31–0.62	0.31–0.63
Coefficient of variation (laboratory)			4.8%	

pausal group. Although the between-subjects variance component in the postmenopausal group was also higher than the premenopausal estimate, it fell to the premenopausal level when subject 21 was removed, suggesting that this variance component may not really differ by menopausal status.

Variability in the results of the laboratory assay (the within-samples variance component in Table 2) was only a small proportion of the total variability within subjects. This implies that most of the within-subject variability was due to biological variation in prolactin levels within individuals over time. The coefficient of variation of the assay, based on the within-sample variability, was 4.8%.

In the group of premenopausal women who donated samples throughout the menstrual cycle, the estimated geometric mean prolactin values in different phases of the cycle were 10.8 ng/ml in the early follicular phase, 12.1 ng/ml in the late follicular phase, 14.8 ng/ml in the ovulatory phase, and 11.7 ng/ml in the luteal phase. This pattern is similar to that reported for prolactin levels throughout the menstrual cycle (16) and follows with lower amplitude the rise and fall in estradiol levels. Although the effect of phase of cycle on prolactin level was highly significant ( $P < 0.001$ ), analysis of variance adjustment for this factor increased the log-scale reliability coefficient only a very small amount, from 0.48 to 0.52. This negligible improvement in reliability after controlling for phase of cycle in the analysis corresponds to the lack of improvement observed when phase of cycle was controlled by design, by restricting all measurements to the luteal phase.

Age was unrelated to prolactin level, as expected since the age ranges were chosen to minimize the effect of age. Sample storage time, which ranged from 0.5 to 5.8 years, likewise did not affect the prolactin value: the direction of the storage time effect was not consistent in the three study groups nor was the effect statistically significant in any group, implying that prolactin did not undergo degradation during storage. Time of day and time since last meal influenced prolactin levels only in the postmenopausal group. With both time of day and time since last meal in the model (*i.e.*, with each variable controlled for the other), afternoon prolactin values in postmenopausal women were estimated to be 17% higher than morning values ( $P = 0.001$ ), and drawing the sample within 3 h of eating was estimated to raise the prolactin value 10.5% ( $P = 0.04$ ). In spite of the significance of these effects, adjustment for both time of day and time since last meal increased the log-scale reliability coefficient in postmenopausal women only marginally, from 0.76 to 0.78. It is likely that time of day would have been more influential if evening or night samples had been included.

## Discussion

These results indicate that the ability of a single measurement to characterize a woman's long-term serum prolactin level relative to other women depends on her menopausal status. For the postmenopausal subjects the reliability of a single log prolactin measurement was quite high (0.76), which implies that prolactin levels are quite stable after menopause. For the premenopausal subjects the reliability of a single log measurement was considerably worse (0.48), and negligible improvement was observed after controlling for phase of the menstrual cycle. Because prolactin levels are more variable in premenopausal women, several samples are needed to raise the reliability of the geometric mean to the level observed in postmenopausal women. Three premenopausal samples are needed to raise the reliability to 0.73 (10), four samples to raise the reliability to 0.79.

Variability due to the laboratory assay was reduced in this study by processing all samples in a single batch. If different samples were assayed at different times and if significant random batch-to-batch variation were present, the reliability of the measurements would be somewhat lower than the values calculated here. However, since we have observed systematic time trends in the assay results for several hormones under investigation in our case-control study, we believe that batch effects are likely to vary systematically over time and between laboratories. By assaying all samples in the same batch, we avoided any systematic time trends in the assay results. We are employing a similar approach in our case-control study by using a matched design and assaying all samples from a case and her matched controls in the same batch.

It should be noted that this study contained few premenopausal women. Our requirement that phase of cycle be known for all premenopausal measurements excluded most perimenopausal samples, which were likely to have phase of cycle unknown. Since control for phase of cycle had little influence on the premenopausal reliability estimates, we would expect the reliability of a prolactin measurement in menstruating perimenopausal women to be similar to the premenopausal estimate. We had no direct information on the reliability of prolactin measurements taken shortly after menopause, since only six measurements in the postmenopausal group were taken within 4 years of last period. However, it seems reasonable to assume that the reliability of a prolactin measurement taken shortly after menopause lies between the premenopausal and postmenopausal estimates.

We were surprised that the reliability of the premenopausal prolactin measurements was not improved when phase of the menstrual cycle was controlled. Analyses of the premenopausal, any-phase group demonstrated that geometric mean prolactin levels followed a predictable pattern through the cycle. But these cyclic changes were small compared with the range of the measurements. It is only during the brief ovulatory phase that the mean prolactin level rises appreciably. Most measurements were therefore taken during a time in the cycle when the mean level was relatively constant. The lack of reduction in the within-subjects variability in the luteal phase group compared with the unrestricted group suggests that changes in prolactin levels across the menstrual cycle are small compared with the variability in the prolactin level of an individual subject from one luteal phase to another.

The consequences of unreliability are a loss of statistical power and a bias toward unity in relative risk estimates. The effective sample size is reduced by a factor of  $\rho_i$ , and the log odds ratio in the logistic regression model is attenuated by

this same factor (17). The reliability of 0.76 for single postmenopausal prolactin measurements is close to the value of 0.80 which is considered acceptable for relative risk estimation in epidemiological studies (18). We therefore conclude that a single serum sample is sufficiently characteristic of the mean prolactin level of a postmenopausal woman to provide meaningful exposure data for epidemiological research. For premenopausal women, however, four samples are required to raise the reliability of the geometric mean to a target level near 0.80. Since controlling for phase of the menstrual cycle produced only slight improvement in the reliability of the premenopausal prolactin measurements, it appears that control for phase of cycle is unnecessary in epidemiological analyses of serum prolactin in premenopausal women.

## Acknowledgments

The authors thank Dr. Mimi Kim for statistical advice, Miriam Moseson for helpful information, and Frances Mastrotta, Anna Bassiri, Dorothe Linders, Loes de Groot-Beckers, and Tiny Korse for technical assistance. The assistance of the Guttman Breast Diagnostic Institute in all phases of data collection and of its Medical Director Emeritus, Dr. Philip Strax, is also gratefully acknowledged.

## References

1. Welsch, C. W., and Nagasawa, H. Prolactin and murine mammary tumorigenesis: a review. *Cancer Res.*, 37: 951-963, 1977.
2. Kwa, H. G., Bulbrook, R. D., and Wang, D. Y. An overall perspective on the role of prolactin in the breast. In: H. Nagasawa (ed.), *Prolactin and Lesions in Breast, Uterus, and Prostate*, pp. 3-22. Boca Raton: CRC Press, 1989.
3. Toniolo, P. G., Pasternack, B. S., Shore, R. E., Sonnenschein, E., Koenig, K. L., Rosenberg, C., Strax, P., and Strax, S. Endogenous hormones and breast cancer: a prospective cohort study. *Breast Cancer Res. Treat.*, 18: S23-26, 1991.
4. Yen, S. S. C. Prolactin in human reproduction. In: S. S. C. Yen, and R. B. Jaffe (eds.), *Reproductive Endocrinology*, Ed. 2, pp. 239-248. Philadelphia: W. B. Saunders, 1986.
5. Bulbrook, R. D., Wang, D. Y., Hayward, J. L., Kwa, H. G., and Cleton, F. Plasma prolactin levels and age in a female population: relation to breast cancer. *Int. J. Cancer*, 28: 43-45, 1981.
6. Winholz, M. (ed.). *The Merck Index*, Ed. 9. Rahway: Merck, 1976.
7. Del Pozo, E., and Brownell, J. Prolactin I. Mechanisms of control, peripheral actions and modification by drugs. *Horm. Res.*, 10: 143-172, 1979.
8. Gilman, A. G., Goodman, L. S., and Gilman, A. (eds.). *The Pharmacological Basis of Therapeutics*, Ed. 6. New York: MacMillan, 1980.
9. Babiak, R., Willnow, P., Baer, M., Van Gent, M., and Ehrhardt, V. A new enzyme immunoassay for prolactin in serum or plasma. *Clin. Chim. Acta*, 36: 76-80, 1990.
10. Fleiss, J. *The Design and Analysis of Clinical Experiments*, pp. 1-32. New York: Wiley, 1986.
11. Fisher, R. A. *Statistical Methods for Research Workers*, Ed. 13, pp. 211-247. New York: Hafner, 1958.
12. Donner, A. A review of inference procedures for the intraclass correlation coefficient. *Int. Stat. Rev.*, 54: 67-82, 1986.
13. Rosner, B., and Willett, W. C. Interval estimates for correlation coefficients corrected for within-person variation: implications for study design and hypothesis testing. *Am. J. Epidemiol.*, 127: 377-386, 1988.
14. SAS Institute Inc. *SAS/STAT User's Guide*, Release 6.03 Ed. Cary, NC: SAS Institute Inc, 1988.
15. Stanish, W. H., and Taylor, N. Estimation of the intraclass correlation coefficient for the analysis of covariance model. *Am. Stat.*, 37: 221-324, 1983.
16. Cole, E. N., Sellwood, R. A., England, P. C., and Griffiths, K. Serum prolactin concentrations in benign breast disease throughout the menstrual cycle. *Eur. J. Cancer*, 13: 597-603, 1977.
17. De Klerk, N. H., English, D. R., and Armstrong, B. K. A review of the effects of random measurement error on relative risk estimates in epidemiological studies. *Int. J. Epidemiol.*, 18: 705-712, 1989.
18. Willett, W. An overview of issues related to the correction of non-differential exposure measurement error in epidemiologic studies. *Stat. Med.*, 8: 1031-1040, 1989.