JOURNAL OF IMMUNOASSAY & IMMUNOCHEMISTRY
Vol. 25, No. 3, pp. 259–278, 2004

Statistical Correction for Non-parallelism in a Urinary Enzyme Immunoassay

Kathleen A. O'Connor,[1,2,*] Eleanor Brindle,[2] Jane B. Shofer,[2] Rebecca C. Miller,[2] Nancy A. Klein,[3] Michael R. Soules,[3] Kenneth L. Campbell,[4] Cori Mar,[2] and Mark S. Handcock[2,5]

[1]Department of Anthropology, [2]Center for Studies in Demography and Ecology, and [3]Department of Obstetrics and Gynecology, University of Washington, Seattle, Washington, USA
[4]Department of Biology, University of Massachusetts, Boston, Massachusetts, USA
[5]Department of Statistics, University of Washington, Seattle, Washington, USA

ABSTRACT

Our aim was to develop a statistical method to correct for non-parallelism in an estrone-3-glucuronide (E1G) enzyme immunoassay (EIA). Non-parallelism of serially diluted urine specimens with a calibration curve was demonstrated in an EIA for E1G. A linear mixed-effects analysis of 40 urine specimens was used to model the relationship of E1G

*Correspondence: Kathleen A. O'Connor, Department of Anthropology, Center for Studies in Demography and Ecology, University of Washington, Seattle, WA 98195, USA; E-mail: oconnork@u.washington.edu.

DOI: 10.1081/IAS-200028078
1532-1819 (Print); 1532-4230 (Online)
Copyright © 2004 by Marcel Dekker, Inc.
www.dekker.com

concentration with urine volume and derive a statistical correction. The model was validated on an independent sample and applied to 30 menstrual cycles from American women. Specificity, detection limit, parallelism, recovery, correlation with serum estradiol, and imprecision of the assay were determined. Intra-and inter-assay CVs were less than 14% for high- and low-urine controls. Urinary E1G across the menstrual cycle was highly correlated with serum estradiol ($r = 0.94$). Non-parallelism produced decreasing E1G concentration with increase in urine volume (slope = -0.210, $p < 0.0001$). At 50% inhibition, the assay had 100% cross-reactivity with E1G and 83% with 17β-estradiol 3-glucuronide. The dose–response curve of the latter did not parallel that of E1G and is a possible cause of the non-parallelism. The statistical correction adjusting E1G concentration to a standardized urine volume produced parallelism in 24 independent specimens (slope = -0.043 ± 0.010), and improved the average CV of E1G concentration across dilutions from 19.5% $\pm$ 5.6% before correction to 10.3% $\pm$ 5.3% after correction. A statistical method based on linear mixed effects modeling is an expedient approach for correction of non-parallelism, particularly for hormone data that will be analyzed in aggregate.

Key Words: E1G; EIA; Urinary reproductive steroids; 3F11 clone; Assay validation; Linear mixed effects model.

INTRODUCTION

A common problem in assay validation is non-parallelism between the dose–response curve of a set of calibrators and serially diluted specimens of the analyte of interest.[1] Non-parallelism may arise from one of several sources, including tracer heterogeneity,[2] interference from substances present in the diluent or specimen matrix,[1,3–11] inappropriate assay reaction times,[12] lack of specificity of antiserum,[13,14] or other sources of non-specific or specific cross-reaction.[1,11,15–19] In some cases, non-parallelism may not be clinically significant[12] but, when it is, a range of methods have been employed to mitigate its effects, including limiting assay use to the parallel range of dilution or concentration of the dose–response curve,[2,4,11,17,18] purification of the matrix used to dilute calibrators, or the matrix of the analyte,[3,17,19,20] the application of detergents, heating, extraction or charcoal absorption to specimen or calibrator matrix,[6–10,15,16,21] preparing calibrators in specimen matrix,[5,16,19] and statistical correction.[14]

We offer a new statistical approach for correcting assay non-parallelism. This approach is useful when it is impractical or ineffective to manipulate the assay, but the assay meets all other diagnostic criteria, and the non-parallelism is consistent across specimens. We are aware of only one other statistical approach for adjusting for non-parallelism. Andersen et al.[14] used non-linear

regression analysis to derive a correction method for non-parallelism in an insulin aspart assay that corrected the measured concentration to the true concentration determined from an alternative method. We propose a correction method that does not require knowing the true concentration, and uses linear mixed effects models that include estimates of the precision of the corrected concentrations. This approach is particularly useful for assay applications where the hormone data from groups of subjects are the focus of analytical interest.

Our original objective was to validate an enzyme immunoassay (EIA) for urinary E1G for population research on ovarian function. Metabolites of estradiol in urine, which are used for research in reproductive endocrinology, include free estrone (E1) and the estrone conjugates, estrone sulfate and E1G.[22] Urinary levels of these metabolites closely parallel serum levels of estradiol, after correction for hydration status.[23,24] We modified an existing fluoroimmunoassay (FIA) using the 3F11 anti-E1G monoclonal antibody[25] for an EIA format. We chose the EIA format because of its cost-effectiveness and efficiency for population and prospective research:[26] the equipment and reagents are affordable, no specimen preparation (e.g., extraction) is needed, and no hazardous or radioactive materials are used. The assay met all validation criteria except independence of specimen volume: measured concentration of E1G decreased as urine volume increased. The source of the non-parallelism is unknown, but may be from a high level of cross-reaction with a similar analyte, 17β-estradiol 3-glucuronide, which has a non-parallel dose–response relationship with the E1G calibration curve.

Because the assay met all other validation criteria, our next objective was to evaluate the significance of the non-parallelism in research use of the assay, and design a correction method. The non-parallelism was significant enough to produce erroneous results in research use, but it was consistent across a range of clinically normal specimens, including those from menopausal, cycling, pregnant, and male subjects. Simple manipulations of assay conditions, including dilution of calibrators and specimens in urine matrix, did not produce parallelism. These considerations led to the development of a statistical procedure to correct for non-parallelism.

EXPERIMENTAL

Subjects and Specimens

Urine specimens for the validation experiments and development and testing of the non-parallelism statistical model were collected from the US participants in clinical and home settings. Paired urine and serum specimens ($N = 808$) were

collected daily from thirty US women in 1997–1998. Thirteen women aged 20–25 years and seventeen women aged 40–45 years were recruited for a study on reproductive aging. Monetary compensation was provided for participation. All participants had regular 25–35-day menstrual cycles, were in good health, had a mean body mass index of 22.6 kg/m^2(SD = 2.36, range 18.9–27.7), and were using no medications or hormones. Daily blood specimens were obtained by venipuncture, beginning with the first day of menstrual bleeding and continuing until the first day of menstrual bleeding of the subsequent cycle. Daily transvaginal ultrasound was performed on all subjects from the mid- to late-follicular phase until evidence of ovulation was observed. Daily urine specimens were frozen at −20°C immediately after collection, and remained frozen until thawing 2 years later for aliquoting and assay. All subjects provided written informed consent, and all procedures were approved by the institutional review boards of the University of Washington. Single specimens were also obtained from 13 volunteers. Participants included one healthy normal adult male, a one month post-partum breastfeeding adult female, and healthy normal adult women who were cycling, post-menopausal, pregnant or on oral contraceptives. No monetary compensation was provided.

Assay Reagents and Protocol

A competitive microtiter plate solid phase EIA for E1G was developed using a rat-derived, monoclonal antibody (clone 3F11). The antibody has been well characterized[27] and used in an FIA.[25,28] Microtiter plates were pre-coated with 50 μL/well of 10 μg/mL rabbit anti-rat IgG (Jackson Immuno-research, West Grove, PA) in coating buffer (50 mmol/L bicarbonate buffer, pH 9.6). After the plates incubated overnight at 4°C, they were washed (0.15 M NaCl; 0.05% Tween 20) and the unpurified ascites fluid containing the monoclonal antibodies was diluted in coating buffer and added to the wells (50 μL/well). The plates were then washed and 50 μL/well of assay buffer (0.1 mol/L sodium phosphate buffer, pH 7.0, with 8.7 g NaCl and 1 g bovine serum albumin per liter) was added. After the incubation of plates for 0.5–3 hr at room temperature, standards, neat or pre-diluted specimens and pre-diluted controls (40 μL/well) were added to the wells still containing the 50 μL/well of assay buffer. The tracer, estrone 3-glucuronide conjugated to horseradish-peroxidase,[23] was diluted in assay buffer and then added at 50 μL/well exactly 30 min after the addition of standards, specimens, and controls. After an overnight incubation at 4°C, the plates were washed and developed in citrate buffer (50 mmol citrate, pH 4.0) combined with 0.4 mmol 2,2-azino-*bis*(3-ethylbenzthiazoline-6-sulfonic acid) (Sigma, St Louis, MO) and 1.6 mmol hydrogen peroxide (100 μL/well). Optical

density was measured with a Dynatech MR7000 Plate Reader (test wavelength 405 nm, reference wavelength 570 nm). Hormone concentrations were estimated from optical density using a four parameter logistic model[29] in Biolinx 1.0 Software (Dynex Laboratories, Inc., Chantilly, VA). Commercial standards (estrone-β-D-glucuronide, Sigma Catalog No. E1752) and in-house urine controls were used in all assays. The 8-point calibration curve covered the concentration range 0.67–85.38 nmol/L.

Standards, zero dose blanks, specimens, and controls were run in duplicate on every plate. Specimens were added to the assays neat or, for higher concentration specimens, pre-diluted. Standards, pre-diluted specimens, and in-house control dilutions were made in de-ionized, distilled water (ddH_2O) just prior to adding them to the plate wells. Urinary hormone concentrations were corrected for hydration status using specific gravity.[30,31] Specific gravity was measured by placing a drop of urine on a hand-held urine specific gravity refractometer (Uricon-PN, NSG Precision Cells, Inc., Farmingdale, NY). The correction formula[31] was applied to each hormone result using a population mean specific gravity of 1.020.

Assay Validation

Recovery for the E1G EIA was determined as percent of added mass (known standard dose) recovered from a urine matrix. Urine specimens low in endogenous steroids from five subjects were run neat and spiked with low, medium, and high doses of standard. Spikes were prepared in ddH_2O and added as 10% of the specimen volume. Each specimen/dose combination was run in four replicates and assayed at 10 separate times. Percent recovery was estimated by dividing the observed assay result by the expected result, with the latter defined as mean neat concentration plus added dose.

Parallelism was assessed using calibration curves and five different urine specimens serially diluted with ddH_2O. Results are expressed as percent of antibody sites bound by tracer and compared with a calibration curve. The standards and specimens were assayed in duplicate. Specificity was measured as the percent cross-reaction with commercially available steroids of similar molecular structure. The 50% inhibition point of respective dose–response curves was expressed as (nmol of E1G/nmol of steroid or steroid metabolite) $\times$ 100%.

Sensitivity, the minimum detectable dose determined from standards, including a zero dose blank, was examined across 20 microtiter plates. Imprecision was estimated by examining intra- and inter-assay variation of in-house

high- and low-hormone concentration urine control pools run in duplicate on 20 microtiter plates.

The performance of the urinary E1G EIA was also evaluated by comparing results from paired urine and serum specimens from 30 cycling US women for one complete menstrual cycle ($N = 808$ specimens). Serum estradiol (E2) was measured by an RIA (ICN Biomedicals, Costa Mesa, CA) that cross-reacts 20% with estrone, 1.5% with estriol, and <1% with all other steroids. The inter-assay and intra-assay CVs were 16% and 7%, respectively. Menstrual cycles were aligned by day of the midcycle serum luteinizing hormone (LH) peak (day 0). The day of ovulation was determined from ultrasound using specific criteria including follicle collapse. The mean day of follicle collapse was one day after the LH surge. Serum LH was measured by a solid phase two-site immunofluorometric assay (DELFIA, Pharmacia, Gaithersburg, MD), with intra- and inter-assay CVs of 2.8% and 4.7%.

Statistical Analyses

Parallelism was statistically evaluated by modeling the relationship between percent hormone bound and log dilution using a 3-compartment logistic model.[26,32] The three parameters estimated are the upper bound of the curve, the midpoint (point of inflection), and a scale parameter measuring the approximate slope of the curve at the midsection. A random effect term for the scale parameter, corresponding to the inverse of the slope of the curves, was estimated for each standard and specimen curve. The null hypothesis that the curves were parallel was tested by determining if the standard deviation of the random effects term for the scale parameter was close to zero, indicating that there was little variation in slope among specimens. Parallelism of the cross-reactants in this assay was also assessed using a 3-compartment logistic model. For this analysis, we estimated the logistic parameters using generalized non-linear least squares because the curves compared were specific standards and not random specimens. A significant difference in the scale parameter between E1G and any of the cross-reactant curves implied non-parallelism. The minimum detectable dose was estimated as the dose that produced a significantly different response ($p < 0.05$) from the zero dose–response.[33] A variance components model[29] was used to determine inter- and intra-assay variation for high- and low-urine controls. The paired urine and serum data were examined by Pearson correlation coefficient, using the averaged cycle days of the paired urine/blood data ($n = 34$ paired urine/serum cycle days from 30 cycles) with specific-gravity-corrected urinary hormone concentrations.

Development of Correction Method for Non-parallelism

Simple assay manipulations were tested for their effects on reducing or eliminating the non-parallelism, including experimenting with different preparations of the IgG pre-coat, adding blocking steps between the pre-coat and coat steps, and removing the pre-coat step and preparing the calibrators in specimen matrix. Both pre-pubertal and post-menopausal urines were tested as diluent for the standard curve and specimens.

To derive a statistical correction for non-parallelism we used the linear relationship between log concentration of E1G and log μL/well of urine to calculate a mean slope between these two measures across 40 subject specimens (27 specimens from the sample of 808 urine specimens and 13 specimens from the individual collections) run in five separate batches using linear mixed effects modeling. In this model, the slope defining the relationship between log concentration and log urine volume is a fixed effect, while inter- and intra-subject and inter-batch effects are modeled as random effects. The goal of the statistical correction was to flatten the mean slope to zero, achieving parallelism, i.e., the average log concentration is the same for any log μL/well of urine multiplied by dilution factor. From the above model we estimated Y_s, the E1G concentration corrected to a standard dilution, given an E1G concentration at any dilution using the equation:

$$\log Y_s = (a_r + a_m) + (b_r + b_m)\log X_s$$

where X_s represents the volume of urine (μL/well) at the standard dilution; a_m and b_m are the mean intercept and slope found using the linear mixed effects model; and a_r and b_r are random variation terms for the intercept and slope. The random variation terms are a function of the urine volume and E1G concentration and the standard deviations of the random components (across batch, across subject, within subject) estimated from the linear mixed effects model. The exact derivation of these terms is beyond the scope of this paper: interested readers may consult Ref.[34] From dilutions ranging from neat to 1 : 128, we chose to correct to 1 : 16 (2.5 μL/well urine) in model development. All of the specimens fell on the standard curve at this dilution, and as the approximate midpoint among dilutions, the errors at both high and low dilutions were equally minimized, given that more error is expected the farther the observed dilution is from the standard. In practice, assay results can be corrected to any of the dilutions.

Validation and Application of Correction Method for Non-parallelism

The validity of our statistical correction was assessed by applying it to 24 independent urine specimens (from different cycle days of the 808 specimen sample than those used in model development) assayed at a range of dilutions. We used linear mixed effects models to calculate the mean slope before and after statistical correction. A CV for the corrected concentration across dilutions within specimens was estimated using the standard deviation of the log E1G concentration from each specimen mean. This was compared with the CV for the uncorrected data.

Finally, we combined the 40 specimens used for model development and the 24 used for validation to derive our "best" model. We applied the correction method based on this model to 808 urine specimens of the paired urine/serum sample from 30 menstrual cycles. The estimates of the slope and standard deviations of the random variations from the "best" model are incorporated into a program written in R code that computes the corrected E1G concentration using any dilution as the standard (http://csde.washington.edu/endolab/E1G.Predict/run.pl). All the statistical analyses for the non-parallelism parts of this study were performed using S-PLUS 6, Release 1 (Insightful Corporation, Seattle, WA) and R 1.7.1 (The R Development Core Team).

RESULTS

Assay Validation

Analytical recovery for low, medium, and high doses of added mass are shown in Table 1. Average recovery across the three doses was 101%. The minimum detectable dose was 3.1 nmol/L. Measures of imprecision are shown in Table 2.

Averaged urinary and serum hormone profiles from the 30 US menstrual cycles are shown in Fig. 1. The highest correlation between the averaged

Table 1. Recovery of added metabolites in urine in 3F11 EIA.

Amount added (nmol/L)	Mean (SE) recovery (%)
4.2	97 (2.8)
12.8	102 (1.4)
21.3	102 (1.8)

Table 2. Imprecision (CV) in urinary 3F11 EIA.

	Urine pool mean (nmol/L)	CV (%)
Within-run	8.00	3.6
	3.23	12.2
Between-run	8.00	3.2
	3.23	13.7

serum and the specific gravity corrected urinary data was 0.95 ($N = 34$ cycle days, $p < 0.01$) at lag day 1, indicating a 1-day lag between serum and corrected urinary measures (Table 3).

Assay specificity is shown in Table 4 and Fig. 2. At the 50% inhibition point, there is 100% cross-reaction with estrone-3-glucuronide, 83% with 17β-estradiol 3-glucuronide, and 9% with estriol-3-glucuronide. The

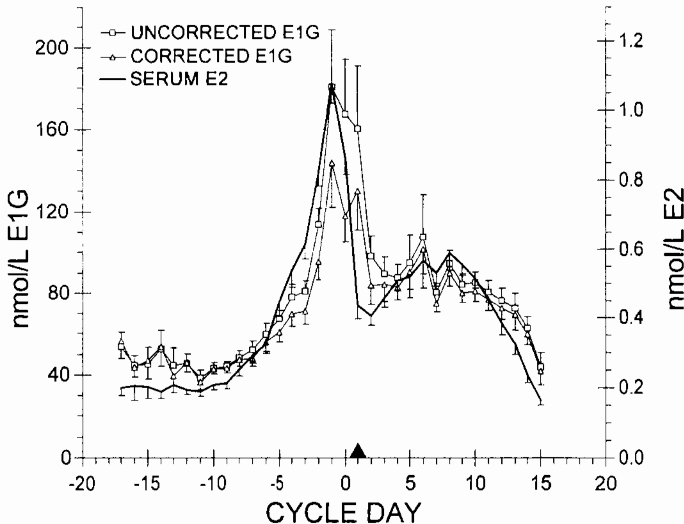

Figure 1. Mean ($\pm$ 1 SEM) serum E2 and urinary E1G of 30 cycles from daily paired urine and serum specimens. Cycles are aligned by day of the serum LH peak (day 0). The filled triangle indicates average day of follicle collapse. The number of observations varies by cycle day, with a minimum of four observations on cycle days −17 and 16 and a maximum of 28–30 observations for cycle days −10 through 11. Urinary hormone concentration adjusted for specific gravity. Urinary E1G corrected (open triangles) and not corrected (open squares) for non-parallelism.

Table 3. Pearson correlations between urine and serum with time lags.[a]

	E2-E1G	
Lag	Urinary data not corrected for non-parallelism	Urinary data corrected for non-parallelism
Urine 1 day before serum	0.64	0.68
None	0.89	0.90
Urine 1 day after serum	0.95	0.94
Urine 2 days after serum	0.83	0.83

[a]Specific gravity adjusted urine concentrations; $N = 34$ mean paired urine/serum cycle days from 30 cycles.

17β-estradiol 3-glucuronide dose–response curve does not parallel the E1G dose–response curve (Fig. 2), and a 3-compartment logistic analysis confirms that the E1G slope is significantly steeper ($p < 0.0001$).

The hypothesis of parallelism between an E1G standard curve and four serially diluted subject specimens was rejected. The standard deviation of the scale parameter was significantly different from 0 ($p < 0.0001$). Figure 3 illustrates that the lack of parallelism is evident in graphical inspection.

Correction of Assay Non-parallelism

Simple assay manipulations were not successful in mitigating the non-parallelism. Preparing E1G standards and serially diluted urine specimens in pre-pubertal male or post-menopausal urine resulted in the same non-

Table 4. Cross-reactivity in urinary 3F11 EIA.

Steroid	Cross-reactivity (%)
Estrone 3-glucuronide	100
Estrone	2.2
Estrone 3-sulfate	<0.2
17β-Estradiol 3-glucuronide	82.9
Estriol 3-glucuronide	9.0
17β-Estradiol	<0.2
Estriol	<0.1

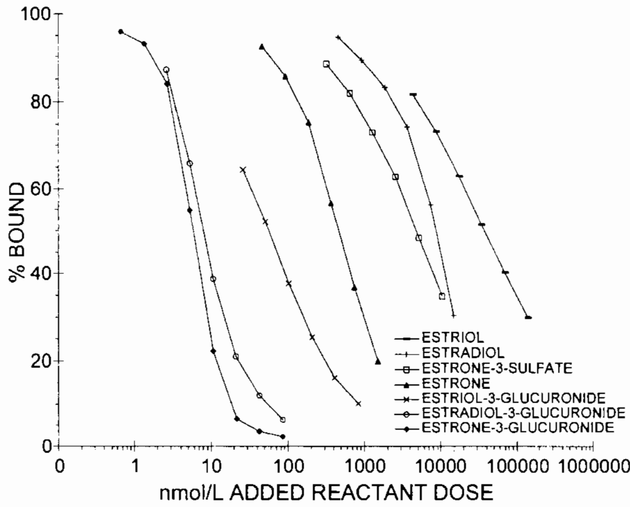

Figure 2. Dose–response curves of cross-reactants and E1G standard.

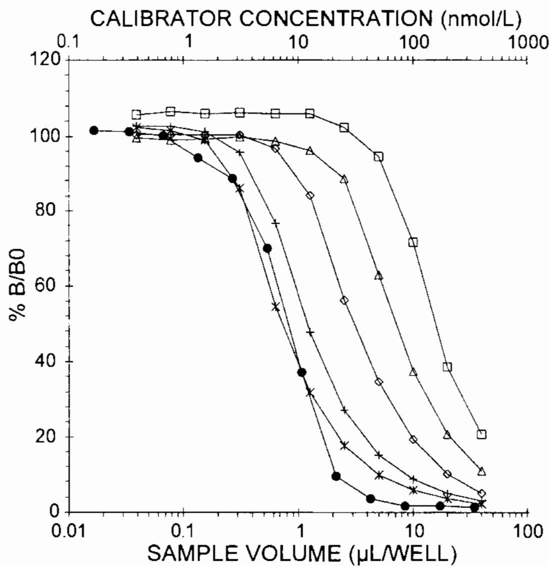

Figure 3. Dose–response curves of E1G standard (closed circles) and five subject specimens. Volumes of the specimens range from 40 to 0.04 μL/well. Specimens are from cycling, menopausal, and pregnant women.

parallelism as when both were diluted in distilled ddH_2O (data not shown). Other simple assay manipulations, including experimenting with different preparations of the IgG pre-coat, adding blocking steps between the pre-coat and coat steps, and removing the pre-coat step were also unsuccessful in eliminating or reducing the non-parallelism (data not shown). We therefore pursued a statistical method for correcting non-parallelism.

A plot of E1G concentration by urine volume on log scales (Fig. 4) for 40 specimens showed that the non-parallelism caused the measured concentration of E1G to increase as the volume of urine assayed decreased from a maximum of 40 μL/well to a minimum of 0.3125 μL/well. The non-parallelism was significant (mean slope = −0.210, $p < 0.0001$) and generally consistent across specimens (Fig. 4), although there was significant random variation in the slope among individual specimens. We also found significant random variation in the estimate of slope due to assay batch ($p = 0.0053$). Despite the variation due to individual and batch effects, the general consistency of the non-parallelism allowed for development of a statistical model for correction.

From the estimate of the average slope from the 40 specimens we derived the concentration corrected to a standard dilution (Y_s) for individual specimens. The model for correcting concentration standardized to 2.5uL of urine was:

$$\log Y_s = (a_r + 4.26) + (b_r + 0.210) \log 2.5$$

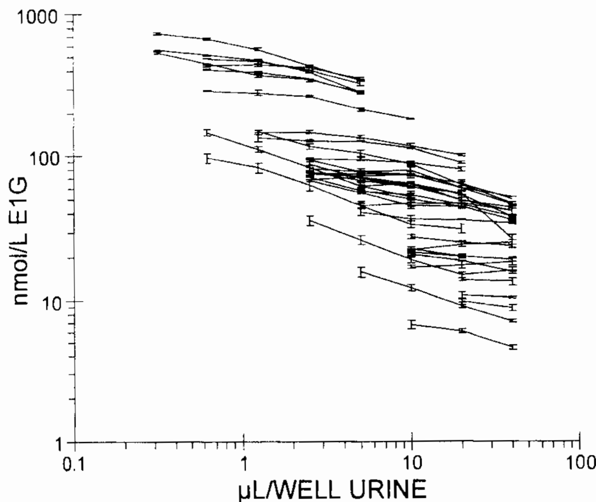

Figure 4. Log concentration of E1G plotted against log urine volume for 40 urine specimens. Error bars are ± 1 SEM.

The random effects terms for the intercept and the slope vary for each individual specimen and batch; the standard deviation for the intercept for between-batch variation was 0.69; for within-batch variation it was 0.86. The standard deviation for the slope was 0.076 for between-batch variation and 0.078 for within-batch variation.

Validation and Application of Non-parallelism Correction Method

Application of the correction model to 24 independent urine specimens, standardized to the concentration at a urine volume of 2.5 uL, resulted in a mean slope of 0.043 ± 0.010 compared with a mean of -0.176 ± 0.009 for the uncorrected log concentrations. Figure 5 shows the 24 specimens before and after the correction method was applied. The mean CV for Y_s across specimens was 10.3% (SD 5.3%), a marked improvement from the average CV of the uncorrected log concentrations of 19.5% (SD 5.6%).

Combining the 40 specimens used for model development, and the 24 used for validation, we developed a "best" model, with a mean slope of -0.206 ± 0.031. Using this model, we applied the statistical correction to the 808 urine specimens of the paired urine/serum sample (omitting 52 specimens that were used in model development). Because the vast majority of these were assayed at a 1 : 5 dilution (8 uL urine/well), the correction was standardized to this dilution. The resulting averaged menstrual cycle profile is plotted in Fig. 1, beside the serum E2 and uncorrected urinary E1G averaged profiles. Pearson correlations of serum E2 with the non-parallelism corrected urinary E1G data are shown in Table 3. The serum–urine correlations are essentially unchanged by application of the correction method, and the hormone profile is minimally changed. The most obvious effect is a reduction of the peri-ovulatory E1G peak.

DISCUSSION

We developed the E1G assay in the EIA format because it is cost-effective and efficient for population and prospective research.[26] The microtiter plate format and reagent assembly make the assay cost-effective for processing large numbers of specimens: costs are less than $0.50 per specimen (not including labor), whereas the E2 RIA kit costs over $1.00 per specimen. The EIA uses a monoclonal antibody, ensuring its long-term availability for clinical, epidemiological, and field research in reproductive biology.

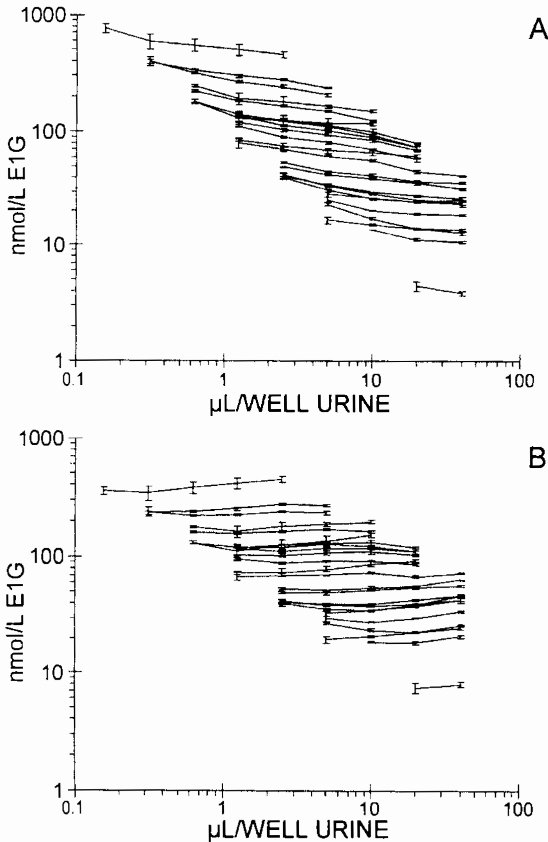

Figure 5. Log concentration of E1G plotted against log urine volume for 24 validation urine specimens. Panel A, before statistical correction for non-parallelism and Panel B, after statistical correction. Error bars are ± 1 SEM.

The E1G EIA showed acceptable recovery, imprecision, specificity, and sensitivity. Hormone profiles were highly correlated with and parallel to serum E2 profiles. Urinary E1G lagged behind serum by 1 day on average, and should be accommodated in algorithms that estimate ovulation day from urinary data.[35] Non-independence of specimen volume was demonstrated: serially diluted urines were not parallel to a calibrator curve made in ddH_2O, and concentration increased as dilution factor increased.

Given that the validation criteria supported the assay's usefulness, and the non-parallelism was highly consistent across a range of specimen types, we

attempted to mitigate its effects. One common approach is to limit assay use to one or two dilutions,[18] but this is not feasible for research on the ovarian cycle where variation in concentration across the cycle and across individuals is too broad to be accommodated with only one or two dilutions. Another common approach is to prepare the calibrators in specimen matrix.[5,16,19] This was, however, not successful in our case: E1G standards diluted in pre-pubertal male or post-menopausal urine exhibited non-parallelism with similarly treated urine specimens. Other simple assay manipulations were also unsuccessful in eliminating or reducing the non-parallelism. Further manipulations of the assay, such as extraction or purification of specimens or diluents were not pursued because they would add considerable processing time and expense to an assay designed for efficiency in large scale population research. Consequently, we adopted a less common approach for correcting non-parallelism, based on statistical modeling. This is an appropriate approach for assays that will be used in studies focused on measurements in groups of subjects, and where the average effects are of interest. In practice, the range of dilutions used in a study will be relatively narrow, and will rarely span the range of dilutions tested in our parallelism experiments. For example, in the sample of 808 paired urine/serum specimens, although dilutions ranged from neat to 1 : 50, 86% of the specimens were successfully characterized at 1 : 5, with only 7% needing to be run neat and another 7% needing to be run at 1 : 50.

Our statistical correction method was successful in reducing the dependency between E1G concentration and urine volume. The average slope of the validation sample of 24 specimens (−0.176) was reduced to essentially zero after correction (0.043), and the average CV of concentration across dilutions was reduced from 19.5% to 10.3%. The latter CV is within the acceptable range of variation in an EIA.[1] In the validation sample, the specimens were from 24 of the same subjects used in model development, but from different cycle days. We do not believe that this favorably biases the validation results because the non-parallelism was quite consistent across specimens from a range of subjects used in model development.

When we applied our "best" model to the paired urine/serum sample, the serum–urine correlations remained essentially unchanged (Table 3), and the only observable effect was a reduction in the magnitude of the peri-ovulatory E1G peak (Fig. 1). It is clear that the non-parallelism has little effect on reproducing the ovarian cycle, and thus would have little effect on applications examining cycle patterns. However, the effect of the non-parallelism was to decrease the concentration of less diluted specimens and increase the concentration of more diluted specimens. This could potentially affect analyses comparing concentration levels across subjects or within subjects because the non-parallelism magnifies the difference between low- and high-concentration

specimens. Thus, for these applications it is important to correct for the non-parallelism.

Evaluation of assay specificity revealed that there was high cross-reaction (83%) with 17β-estradiol 3-glucuronide. We hypothesize that this cross-reactant is a possible source of the non-parallelism of the assay. Evidence supporting this hypothesis includes observations that the cross-reactant: (1) was not parallel to the E1G dose–response curve (Fig. 2); (2) showed a pattern similar to urine specimens measured against an E1G curve (Fig. 3); and (3) is excreted in urine across the menstrual cycle in the same pattern as E1G, with a clear mid-cycle peak, but at concentrations five times lower than E1G.[36] The uniform pattern of non-parallelism across E1G concentrations and subjects is also consistent with the hypothesis, that is, if the interfering substance varied in amount randomly from specimen to specimen, the non-parallelism would not be so consistent in form. Finally, women vary in the pattern of urinary estrogen glucuronide metabolite excretion, with a CV of 40% based on variation among five metabolites,[36] which is consistent with the slight inter-subject differences in slopes that we found (Fig. 4). Despite these suggestive observations, further lab work is necessary to determine the source of the non-parallelism.

Our approach offers both a new application of linear mixed-effects modeling, and a new statistical approach for correcting assay non-parallelism. We conclude that the statistical correction is a useful approach for mitigating the effects of moderate non-parallelism, especially for assay data that will be used in studies focusing on group comparisons, rather than prediction of individual hormone concentrations. In addition to the benefit of expediency over added sample preparation steps or assay manipulations, our statistically smoothed estimate of concentration takes into account the random variation resulting from batch and subject effects that was measured in model development ($N = 40$ sample). These sources of variation exist in most assays, but are rarely directly incorporated into estimates of hormone concentration.

A limitation of our approach is that we are unable to ascertain the true value of a given specimen, i.e., we have no way of knowing which standard dilution yields the correct concentration. An additional limitation is the relatively large standard deviations of the corrected concentrations. These two limitations make the assay inappropriate for clinical applications. However, in research situations where we are interested in comparing concentrations across individuals or groups, or determining the relationship between estrogen levels and other factors, the statistical correction allows relative comparisons.

An R-code program, *E1G.Predict*, is available for public use at no charge (http://csde.washington.edu/endolab/E1G.Predict/run.pl). The user inputs the logged dilution-adjusted E1G concentration, urine volume used for each

specimen, and the logged standard urine volume to which the concentration should be corrected. The program returns the logged estimated corrected value, logged standard deviation, and logged 90% confidence interval.

Although we were concerned with developing a statistical correction specifically for the non-parallel E1G assay discussed in this paper, the method can be applied to other situations of non-parallelism, if an assay meets other validation criteria, the non-parallelism is consistent, and the assay is targeted toward applications where the hormone data will be analyzed in aggregate. The primary advantage of a statistical approach is its expediency; it reduces time and labor expense that might otherwise be invested in preparing specimens or diluents or other assay manipulations. Application of the correction method using the R program takes a matter of minutes.

ACKNOWLEDGMENTS

This research was supported by NIH (NIA 5 T32 AG00208, NIA RO1 AG15141, NIA RO1 AG14579, NICHD 2 P30 HD28263, NICHD R24 HD42828), the Population Research Institute of Pennsylvania State University and the Center for Studies in Demography and Ecology of the University of Washington. Thanks to J. Aranda, D. Clifton, D. Holman, F. Kohen, D. Schechter, C. Stroud, and G. Tomblingson for their contributions.

REFERENCES

1. O'Fegan, P. Validation. In *Immunoassays: A Practical Approach*; Gosling, J.P., Ed.; Oxford University Press: Oxford, 2000; 211–238.
2. Spencer, C.A.; Platler, B.W.; Nicoloff, J.T. The effect of [^{125}I]thyroglobulin tracer heterogeneity on serum Tg RIA measurement. Clin. Chim. Acta **1985**, *153* (2), 105–115.
3. Thomas, C.M.G.; van den Berg, R.J.; de Koning Gans, H.J.; Lequin, R.M. Radioimmunoassays for prostaglandins. I. Technical validation of prostaglandin F2alpha measurements in human plasma using Sephadex G-25 gelfiltration. Prostaglandins **1978**, *15* (5), 839–847.
4. McCready, J.; Braunstein, G.D.; Helm, D.; Wade, M.E. Modification of the choriogonadotropin beta-subunit radioimmunoassay for determination of urinary choriogonadotropin. Clin. Chem. **1978**, *24* (11), 1958–1961.
5. Linnet, K. Effect of the biological matrix on the urinary testosterone/epitestosterone ratio measured by gas chromatography/mass spectrometry in doping analysis. Biol. Mass Spectrom. **1993**, *22* (7), 412–418.

6. Gilfillan, C.P.; Robertson, D.M. Development and validation of a radioimmunoassay for follistatin in human serum. Clin. Endocrinol. **1994**, *41* (4), 453–461.
7. Lichtenberg, V.; Pahnke, V.G.; Graesslin, D.; Bettendorf, G. Biological and immunological potencies of lutropin (LH) in human serum: comparative studies using different standard preparations. Horm. Metab. Res. **1982**, *14* (1), 39–45.
8. Leidenberger, F.A.; Willashek, R.; Pahnke, V.G.; Reichert, L.E. Application of a radioligand receptor assay for determination of luteinizing hormone in human serum. Acta Endocrinol. **1976**, *81* (1), 54–72.
9. Rajalakshmi, M.; Robertson, D.M.; Choi, S.K.; Diczfalusy, E. Biologically active luteinizing hormone (LH) in plasma: validation of the in vitro bioassay when applied to male plasma and the possible role of steroidal precursors. Acta Endocrinol. **1979**, *90* (4), 585–598.
10. O'Rorke, A.; Kane, M.M.; Gosling, J.P.; Tallon, D.F.; Fottrell, P.F. Development and validation of a monoclonal antibody enzyme immunoassay for measuring progesterone in saliva. Clin. Chem. **1994**, *40* (3), 454–458.
11. Chard, T. *An Introduction to Radioimmunoassay and Related Techniques*; Elsevier: Amsterdam, 1995.
12. Reader, S.C.J.; Daly, J.R.; Robertson, W.R. Effect of human plasma on the bioactivity of adrenocorticotrophin. J. Endocrinol. **1982**, *92* (3), 449–455.
13. Bayer, M.F.; McDougall, I.R. Differences in radioimmunoassay results for thyroglobulin that affect management of patients with thyroid cancer. Clin. Chem. **1984**, *30* (1), 81–86.
14. Andersen, L.; Vølund, A.; Olsen, K.J.; Plum, A.; Walsh, D. Validity and use of a non-parallel insulin assay for pharmacokinetic studies of the rapid-acting insulin analogue, insulin aspart. J. Immunoassay Immunochem. **2001**, *22* (2), 147–163.
15. Ding, Y.Q.; Huhtaniemi, I. Human serum LH inhibitor(s): behaviour and contribution to in vitro bioassay of LH using dispersed mouse leydig cells. Acta Endocrinol. **1989**, *121* (1), 46–54.
16. Romaní, P.; Robertson, D.M.; Diczfalusy, E. Biologically active luteinizing hormone (LH) in plasma: validation of the in vitro bioassay when applied to plasma of women. Acta Endocrinol. **1976**, *83* (3), 454–465.
17. Tomada, Y.; Miwa, T.; Ishizuka, N. Radioligand receptor assay for urinary and serum hCG. J. Clin. Endocrinol. Metab. **1975**, *40* (4), 644–651.
18. Hunter, W.M.; Bennie, J.G. Reduction of non-specific serum responses in human pituitary gonadotrophin radioimmunoassays. J. Endocrinol. **1979**, *80* (1), 59–68.

19. Dufau, M.L.; Pock, R.; Neubauer, A.; Catt, K.J. In vitro bioassay of LH in human serum: the rat interstitial cell testosterone (RICT) assay. J. Clin. Endocrinol. Metab. **1976**, *42* (5), 958–969.
20. Hitzler, W.; Schmidt-Gayk, H.; Spiropoulos, P.; Raue, F.; Hüfner, M. Homologous radioimmunoassay for human parathyrin (residues 53–84). Clin. Chem. **1982**, *28* (8), 1749–1753.
21. Lichtenberg, V.; Pahnke, V.G. Measurement of biologically active LH/HCG by in vitro testosterone production assay (TPA) and problems when applied to serum. Acta Endocrinol. Suppl. **1976**, *202*, 54–57.
22. O'Malley, B.W.; Strott, C.A. Steroid hormones: metabolism and mechanism of action. In *Reproductive Endocrinology: Physiology, Pathophysiology and Clinical Management*; Yen, S.S.C., Jaffe, R.B., Barbieri, R.L., Eds.; WB Saunders Co.: Philadelphia, 1999; 110–133.
23. Munro, C.J.; Stabenfeldt, G.H.; Cragun, J.R.; Addiego, L.A.; Overstreet, J.W.; Lasley, B.L. Relationship of serum estradiol and progesterone concentrations to the excretion profiles of their major urinary metabolites as measured by enzyme immunoassay and radioimmunoassay. Clin. Chem. **1991**, *37* (6), 838–844.
24. Denari, J.H.; Farinati, Z.; Figueroa Casas, P.R.; Oliva, A. Determination of ovarian function using first morning steroid assays. Obstet. Gynecol. **1981**, *58* (1), 5–9.
25. Barnard, G.; Kohen, F.; Mikola, H.; Lovgren, T. Measurement of estrone-3-glucuronide in urine by rapid, homogeneous time-resolved fluoroimmunoassay. Clin. Chem. **1989**, *35* (4), 555–559.
26. O'Connor, K.A.; Brindle, E.; Holman, D.J.; Klein, N.A.; Soules, M.R.; Campbell, K.L.; Kohen, F.; Munro, C.J.; Shofer, J.B.; Lasley, B.L.; Wood, J.W. Urinary estrone conjugate and pregnanediol-3-glucuronide enzyme immunoassays for population research. Clin. Chem. **2003**, *49* (7), 1139–1148.
27. Kohen, F.; Lichter, S. Monoclonal antibodies to steroid hormones. In *Monoclonal Antibodies: Basic Principles, Experimental and Clinical Applications in Endocrinology*; Fortie, G., Lipsett, M.B., Serio, M., Eds.; Raven Press: New York, 1986; 87–95.
28. Kesner, J.S.; Knecht, E.A.; Krieg, E.F.; Barnard, G.; Mikola, H.; Kohen, F.; Gani, M.M.; Coley, J. Validations of time-resolved fluoroimmunoassays for urinary estrone-3-glucuronide and pregnanediol-3-glucuronide. Steroids **1994**, *59* (3), 205–211.
29. Rodbard, D. Statistical quality control and routine data processing for radioimmunoassays and immunoradiometric assays. Clin. Chem. **1974**, *20* (10), 1255–1270.

30. Alfthan, H.; Bjorses, U.; Tiitinen, A.; Stenman, U. Specificity and detection limit of ten pregnancy tests. Scand. J. Clin. Lab. Invest. Suppl. **1993**, *216*, 105–113.
31. Miller, R.C.; Brindle, E.; Holman, D.J.; Shofer, J.B.; Klein, N.A.; Sooles, M.R.; O'Connor, K.A. Comparison of specific gravity and creatinine methods for normalizing urinary reproductive hormone concentrations. Clin. Chem. **2004**, *50* (5), 924–932.
32. Pinheiro, J.C.; Bates, D.M. *Mixed-Effects Models in S and S-PLUS*; Springer: New York, 2000.
33. Rodbard, D. Statistical estimation of the minimal detectable concentration ("Sensitivity") for radioligand assays. Anal. Biochem. **1978**, *90* (1), 1–12.
34. Vonesh, E.F.; Chinchilli, V.M. *Linear and Nonlinear Models for the Analysis of Repeated Measurements*; Dekker: New York, 1997; 252–253.
35. Baird, D.D.; McConnaughey, R.; Weinberg, C.R.; Musey, P.I.; Collins, D.C.; Kesner, J.S.; Knecht, E.A.; Wilcox, A.J. Application of a method for estimating day of ovulation using urinary estrogen and progesterone metabolites. Epidemiology **1995**, *6* (5), 547–550.
36. Baker, T.E.; Jennison, K.M.; Kellie, A.E. The direct radioimmunoassay of oestrogen glucuronides in human female urine. Biochem. J. **1979**, *177* (2), 729–738.

Received January 22, 2004
Accepted February 9, 2004
Manuscript 3126