

# Liquid Chromatography–Tandem Mass Spectrometry Assay for Simultaneous Measurement of Estradiol and Estrone in Human Plasma

ROBERT E. NELSON, STEFAN K. GREBE, DENNIS J. O'KANE, and RAVINDER J. SINGH\*

**Background:** Estradiol (E2) and estrone (E1) measurements form an integral part of the assessment of female reproductive function and have expanding roles in other fields. However, many E1 and E2 immunoassays have limited functional sensitivity, suffer from cross-reactivity, and display poor intermethod agreement. To overcome these problems, we developed a sensitive liquid chromatography–tandem mass spectrometry (LC-MS/MS) assay for the simultaneous measurement of E1 and E2.

**Methods:** After dansyl chloride derivatization, samples were separated by fast gradient chromatography and injected into a tandem mass spectrometer after formation of positive ions with atmospheric pressure chemical ionization. The limits of detection and quantification, recovery, linearity, precision, and reference intervals were determined, and performance was compared with several immunoassays.

**Results:** Total run time per sample was 5 min. The multiple-reaction monitoring ion pairs were  $m/z$  506/171 for 3-dansyl-estradiol and  $m/z$  504/171 for 3-dansyl-estrone. The limits of detection for E1 and E2 were 12.9 pmol/L (3.5 ng/L) and 10.3 pmol/L (2.8 ng/L), respectively. Interassay imprecision (CV) was 4–20% ( $n = 20$ ). The limits of quantification (functional sensitivities) for E1 and E2 were 44.1 pmol/L (11.9 ng/L) and 23.2 pmol/L (6.3 ng/L), respectively. The assay was linear to >2200 pmol/L (~600 ng/L) for each analyte. Recoveries were 93–108% for E1 and 100–110% for E2. No cross-reactivity was observed. Method comparison with several immu-

noassays revealed that the latter were inaccurate and prone to interferences at low E1 and E2 concentrations. **Conclusions:** LC-MS/MS allows rapid, simultaneous, sensitive, and accurate quantification of E1 and E2 in human serum.

© 2004 American Association for Clinical Chemistry

Estrogens are involved in development and maintenance of the female phenotype, germ cell maturation, and pregnancy. They are also important for many other, non-gender-specific processes, including growth, nervous system maturation, bone metabolism/remodeling, and endothelial responsiveness (1–3). The two major biologically active estrogens in nonpregnant humans are estrone (E1)<sup>1</sup> and estradiol (E2). E2 is inactivated by 17 $\beta$ -hydroxy steroid dehydrogenase type II to E1, which can be reduced back to E2 by type I of the same enzyme. E1 has 20–80% of the bioactivity of E2, depending on the assay used (4–6). E2 is also the predominant bioactive estrogen in premenopausal, nonpregnant women, circulating at 1.5–4 times the concentration of E1. E2 concentrations in men and postmenopausal women are one-tenth of those in premenopausal women, whereas E1 concentrations differ less, which leads to a reversal of the premenopausal E2:E1 ratio (7).

Measurement of serum E2 forms an integral part of the assessment of female reproductive function, including studies of infertility, oligo-amenorrhea, and menopausal status. In addition, it is widely used for monitoring of ovulation induction as well as during preparation for in vitro fertilization (8, 9). These applications account for the bulk of clinical estrogen analyses; E2 is measured frequently and E1 rarely. The E2 assays used are optimized

Department of Laboratory Medicine and Pathology, Mayo Clinic and Foundation, Rochester, MN.

\*Address correspondence to this author at: Hilton 730, Department of Laboratory Medicine and Pathology, Mayo Clinic and Foundation, 200 First St. SW, Rochester, MN 55905. Fax 507-284-9758; e-mail singh.ravinder@mayo.edu.

Received August 4, 2003; accepted November 11, 2003.

Previously published online at DOI: 10.1373/clinchem.2003.025478

<sup>1</sup> Nonstandard abbreviations: E1, estrone; E2, estradiol; GC-MS, gas chromatography–mass spectrometry; and LC-MS/MS, liquid chromatography–tandem mass spectrometry.

for these clinical scenarios, which place only modest demands on assay sensitivity but require fast assay times and high throughput. However, a more sensitive E2 assay, simultaneous measurement of E1, or both are needed in many other clinical situations. These include inborn errors of sex-steroid metabolism, disorders of puberty, estrogen deficiency in men, and increasingly, therapeutic drug monitoring, either in the context of low-dose female hormone replacement therapy or antiestrogen treatment (10–17). There is also an increasing research-driven demand for high-sensitivity E1 and E2 assays, e.g., to study breast cancer, male osteoporosis, Alzheimer disease, and cardiovascular disorders (2, 3, 10, 14, 18–20).

High-sensitivity E2 immunoassays are challenging because physiologic serum concentrations of E2 are typically <140 pmol/L (40 ng/L) in adult men and postmenopausal women and in both sexes during infancy and childhood. None of the commercially available automated direct E2 assays appears to have sufficient sensitivity for the evaluation of E2 in the sera of children and men. Assays with higher sensitivity are available, but they have traditionally been manual RIAs. Although some of these E2 and E1 RIAs provide better sensitivity, they have several important drawbacks: They require handling of radioactive materials, organic extraction, chromatography, and prolonged incubation. They are sometimes difficult to quality control in a routine laboratory and are very susceptible to artifacts caused by nonspecific binding of radioactivity (21). They may also be subject to cross-reactivity and other analytical interferences. Most importantly, there is often very poor agreement among the results obtained by different RIAs, sometimes even assays from the same manufacturer (21), making patient follow-up over time or between laboratories, as well as longitudinal studies, extremely difficult. Similar issues are also relevant with regard to automated, chemiluminescence-based, direct E2 immunoassays. College of American Pathologists survey results for the past few years confirm that the performance of direct E2 immunoassays needs to improve with respect to analytical accuracy and detection limit.

E1 and E2 assays based on gas chromatography–mass spectrometry (GC-MS) address many of the shortcomings of automated immunoassays and RIAs, and they are considered to be the most accurate methodology (22). However, sensitivity is often less than what can be achieved by sensitive RIAs, and run times may be longer than 30 min/sample, limiting throughput (22–25). Liquid chromatography with tandem mass spectrometry detection (LC-MS/MS) has been shown to be superior to GC-MS in many scenarios, in terms of both sensitivity and sample throughput, and, in combination with isotope dilution, is also considered as a reference methodology (26–28). In our laboratory, LC-MS/MS-based methods have successfully replaced immunoassays for biological

amines and many steroid analytes (29, 30). We therefore aimed to develop a sensitive LC-MS/MS assay for the simultaneous measurement of E1 and E2.

## Materials and Methods

### MATERIALS AND SPECIMENS

We prepared working standards of E1, E2 (both >98% purity; Sigma-Aldrich),  $d_4$ -E1, and  $d_5$ -E2 (both 98% purity; C/D/N Isotopes, Inc.) by diluting 1 g/L stock solutions in phosphate-buffered saline containing 1 g/L bovine serum albumin (Sigma-Aldrich) and 1 g/L sodium azide (Sigma-Aldrich). Calibrators containing E1 and E2 in concentrations of 0, 10, 25, 50, 100, 200, 400, and 600 ng/L were prepared in the same matrix. HPLC-grade acetonitrile and methylene chloride were obtained from EM Science, and dansyl chloride, estriol, estrone sulfate, and ethinylestradiol were obtained from Sigma-Aldrich.

We analyzed all samples on an API 3000 triple-quadrupole mass spectrometer (ABI-Sciex) equipped with an atmospheric pressure chemical ionization source operated at 400 °C. The system includes two Perkin-Elmer 200 series micropumps and a Perkin-Elmer 200 series autosampler.

We collected separate serum, EDTA-plasma, and sodium heparin-plasma samples from five healthy volunteers to assess specimen-type suitability and storage stability. We also obtained serum samples from 122 female and 32 male healthy donors to determine reference intervals. These latter samples were either assayed immediately or frozen at –20 °C until they could be tested. Finally, we studied 197 clinical samples received for routine E1 or E2 measurements. Our studies were approved by the Mayo Clinic Institutional Review Board.

### SAMPLE PREPARATION

Earlier LC-MS/MS multiple-reaction monitoring experiments with negative ion pairs corresponding to underivatized E1 ( $m/z$  269/145) and E2 ( $m/z$  271/145) gave a method with good precision, linearity, and functional sensitivities of 37 pmol/L (10 ng/L) and 370 pmol/L (100 ng/L) for E1 and E2, respectively. We considered this functional sensitivity insufficient for our intended clinical applications, particularly with regard to E2, and therefore pursued derivatization to improve ionization efficiency of E2 and consequently experimental sensitivity. Derivatization with dansyl chloride, as described below, was highly effective and was therefore used for all studies described for the LC-MS/MS method (31).

We aliquoted 0.5 mL of standards, controls, or patient serum into 16 × 100 mm glass screw-cap tubes. To each of these, we added 25  $\mu$ L of a working stock mixture of  $d_4$ -E1 and  $d_5$ -E2, at 5  $\mu$ g/L each, as an internal standard. We then extracted the samples with 6 mL of methylene chloride, removing the upper aqueous layer and transferring the remaining solvent to clean 13 × 100 mm glass tubes. The solvent was evaporated under nitrogen in a

45 °C water bath, and the dried residue was redissolved in 50  $\mu$ L of sodium bicarbonate buffer (100 mmol/L, pH 10.5). To derivatize the samples, we then added an equal volume of 1 g/L dansyl chloride in acetone, vortex-mixed each sample for 1 min, and incubated the samples in a heating block at 60 °C for 3 min. Immediately thereafter, the samples were transferred to autosampler glass vials, sealed, and assayed.

#### LC-MS/MS

The chromatographic separation was performed on 150  $\times$  2.0 (i.d.) mm Phenomenex Synergi 4 $\mu$ Max-RP columns. The mobile phase consisted of two eluents, solvent A (950 mL/L acetonitrile–50 mL/L H<sub>2</sub>O containing 1 mL/L formic acid) and solvent B (50 mL/L acetonitrile–950 mL/L H<sub>2</sub>O containing 1 mL/L formic acid), delivered at a flow rate of 0.6 mL/min. We injected 15  $\mu$ L of sample into the LC-MS/MS, after which the injector was washed with three postinjection washes with 750 mL/L methanol–250 mL/L H<sub>2</sub>O. The total run time was 5 min. During the first 0.1 min, we held the initial conditions at 50% solvent A. This was followed by a fast linear gradient to 95% solvent A over the next 0.6 min. These conditions were maintained for 1.8 min, before the system was returned to the initial proportion of 50% solvent A over the following 1.5 min and maintained for the final 1 min of each run.

Once we had selected the ion pairs for each of the derivatized estrogens and internal standards, we optimized the instrument settings for maximum signal, using the autotune feature of the instrument software. Results were generated in positive-ion mode with the following optimized voltages: declustering potential, 60 V; focusing potential, 220 V; entrance potential, 10 V; collision energy potential, 50 V; and exit potential, 10 V. Gas settings were as follows: nebulizer gas, 8; curtain gas, 8; and CAD, 10. Nebulizer current was optimum at 3  $\mu$ A. Nitrogen was used as the collision gas in the Q<sub>2</sub> collision cell. Before MS/MS experiments, mass calibration and resolution adjustments were performed on both quadrupoles by an infusion of polypropylene glycol.

All data were acquired and processed with Analyst Software, Ver. 1.2 (ABI-Sciex). A seven-point calibration curve was included with each assay, using calibrator concentrations from 10 to 600 ng/L for either analyte, corresponding to 37–2222 pmol/L for E1 and 36.7–2205 pmol/L for E2, and was used in conjunction with the internal standards to calculate unknown analyte concentrations.

#### BASIC ASSAY PERFORMANCE CHARACTERISTICS

We determined the critical limit and the detection limit of our assay as suggested by the IUPAC guidelines based on statistical calculations (32). The basic assay performance characteristics are defined as critical limit, detection limit, limit of quantification (functional sensitivity), and upper limit of the detectable range. Ten replicate measurements

each of blanks were used to determine the critical concentrations for E1 and E2, defined as the measurement values below which there is >95% certainty that no analyte is present that might be detected by this method. Similarly, the detection limits were established as the lowest analyte concentrations that could be distinguished from the absence of analyte with >95% certainty, based on 10 replicate experiments each on low amounts of added E1 and E2. The limits of quantification (functional assay sensitivities) were, in agreement with common clinical laboratory practice, arbitrarily set at the lowest analyte concentration that could be measured in actual patient samples with an interassay CV <20%. The upper limits of the reportable ranges were arbitrarily defined as 2222 pmol/L (600 ng/L) for E1 and 2205 (600 ng/L) for E2, the values of the highest calibrators for either E1 or E2 included in the assay.

#### EXTRACTION EFFICIENCY, RECOVERY, PRECISION, LINEARITY, AND INTERFERENCES

We assessed the extraction efficiency by comparing the peak areas of derivatized d<sub>4</sub>-E1 and d<sub>5</sub>-E2 internal standards in extracted samples with internal standard of the same concentration that was not extracted but was simply dried down and derivatized.

To determine recoveries, we added E1 and E2 at four concentrations [74, 148, 370, and 1480 pmol/L (20, 40, 100, and 400 ng/L)] to each of three serum samples containing endogenous amounts of both estrogens. We then calculated the percentage recovery as a measure of accuracy as follows: (observed value – baseline value)/amount added  $\times$  100.

For assessment of assay precision, serum pools were prepared at various concentrations, aliquoted, and stored at –20 °C until ready to assay. The intraassay precision was determined by extracting and assaying 18 separate aliquots of each pool in a single batch. Interassay precision was determined by testing a single aliquot of each pool in 20 consecutive assays.

We used two methods to determine assay linearity. In the first method, we used linear regression to compare the E1 and E2 calibrator concentrations plotted against the analyte:internal standard area ratio for five consecutive assays. In the second method, we assayed different amounts of sample material (500, 200, 100, 50, and 25  $\mu$ L) from three specimens. The E1 and E2 values obtained from the 500- $\mu$ L sample were used as the basis to calculate expected values for the respective lower sample volumes. The percentage of observed/expected was then calculated for each dilution and used as a measure of linearity.

To determine cross-reactivity, we added various amounts of E1, E2, estriol, ethinylestradiol, and E1 sulfate to a sample pool with previously measured E1 and E2 concentrations. More than 100 unselected samples from patients, as opposed to healthy volunteers, were also assayed to detect potential drug- or disease-related inter-

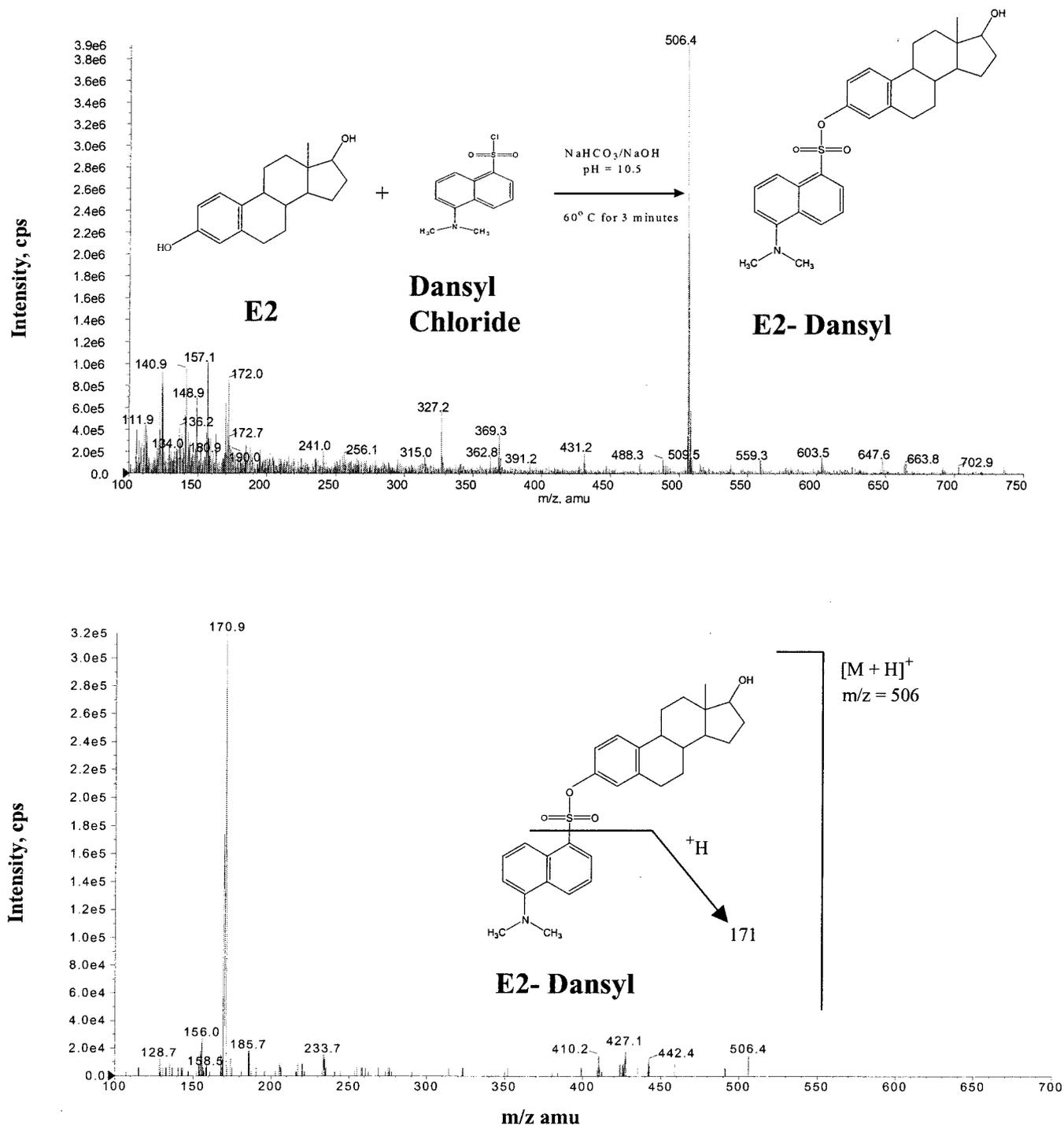


Fig. 1. Positive-ionization atmospheric pressure chemical ionization mass (*top*) and collisionally induced dissociation (*bottom*) spectra for E2. Similar scan spectra were obtained for E1, d<sub>5</sub>-E2, and d<sub>4</sub>-E1 with parent ions at *m/z* 504, 511, and 508, respectively. The daughter ions *m/z* 171 were the same because the cleavage site is the same for all the dansyl derivatives of estrogens. Derivatized estrogens were injected on the column under HPLC conditions rather than infusion to circumvent the dansyl chloride background signal in the scanning mode.

ferences. Any additional peaks, broadening of the main peak, significant increase in background, or significant difference between the heights of the E1 and E2 internal standards were interpreted as potential interferences.

#### SPECIMEN-TYPE SUITABILITY AND STABILITY

We divided each sample from five volunteers into two sets: one set was stored at ambient temperature and the other at 4 °C. On days 0, 3, and 7, an aliquot of each

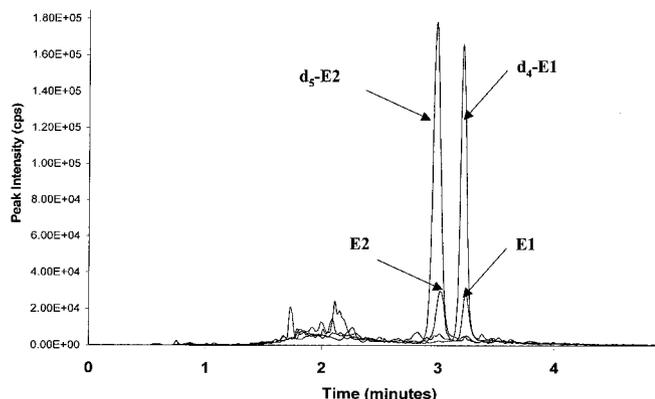


Fig. 2. LC-MS/MS ion chromatograms of E2 and E1 in an extracted patient sample.

The sample was injected by the LC-MS/MS method. The E1 and E2 concentrations of 222.2 pmol/L (60 ng/L) and 220.6 pmol/L (60 ng/L), respectively, were calculated for this specimen.

sample type from each temperature set was removed and stored at  $-20^{\circ}\text{C}$ . All samples from an individual donor were then thawed and assayed at the same time. Freeze-thaw studies were also performed on serum, EDTA plasma, and heparin plasma from the five donors, comparing aliquots that had been through one, two, or three freeze-thaw cycles with a refrigerated aliquot of the same sample.

#### ESTIMATION OF REFERENCE INTERVALS

We measured E1 on E2 in 122 healthy female and 32 healthy male volunteers without regard to age, gender, menopausal status, or phase within the menstrual cycle. We excluded women on hormonal contraceptives and those receiving female hormone replacement therapy.

#### METHOD COMPARISON

LC-MS/MS E2 measurements were compared with (a) an automated direct immunoassay of moderate sensitivity (E2-Assay1, ADVIA Centaur; Bayer Corporation), (b) a high-sensitivity automated immunoassay (E2-Assay2, Vitros ECi; Ortho-Clinical Diagnostics), (c) an enhanced-sensitivity automated immunoassay (E2-Assay3, in-house-modified ACS:180; Bayer Corporation, which includes organic extraction and concentration), and (d) an outside E2 RIA, which includes preassay organic extraction and Sephadex LH-20 chromatography (Esoterix Endocrinology).

For method comparison with E2-Assay1, we determined the E2 values by LC-MS/MS for 42 patient samples that had previously been assayed on E2-Assay1. For the comparison between LC-MS/MS and E2-Assay2, we used serum samples from 118 of the "healthy donor" volunteers. The comparison between LC-MS/MS and E2-Assay3 was performed on 59 patient samples. In 50 of these 59 samples, E2 was  $<150$  pmol/L (40.5 ng/L). We used 50 patient samples for the comparison of LC-MS/MS and the

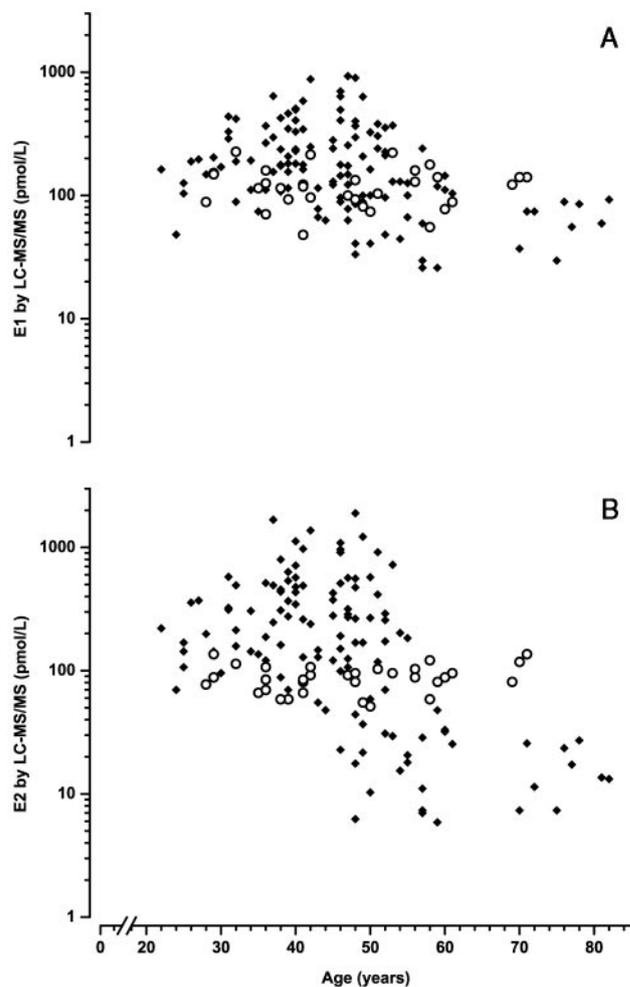


Fig. 3. Scatter plots for E1 (A) and E2 (B) concentrations used to determine the reference intervals, measured by LC-MS/MS.

The x axis shows individuals' ages in years; the y axes (log scale) correspond to E1 and E2 values, respectively, in pmol/L.  $\blacklozenge$ , females;  $\circ$ , males.

RIA with an extraction step. Ten of these had also been assayed with E2-Assay3.

E1 measurements by LC-MS/MS were compared with a direct RIA (Estrone; Diagnostic Systems Laboratories Inc.) for 36 patient samples that spanned the E1 reference interval and included some moderately increased E1 concentrations.

We also compared E1 measurements by LC-MS/MS for 20 patient samples with an extraction-based immunoassay performed in an outside laboratory (Esoterix Endocrinology), which performs individual E1 assays by RIA after extraction and Sephadex LH-20 chromatography.

**Table 1. Reference intervals for E1 and E2 in adult males and pre- and postmenopausal females.**

Group	n	E1, ng/L	E2, ng/L
Adult males	32	10–60	10–40
Premenopausal females	92	17–200	15–350
Postmenopausal females	30	7–40	$<10$

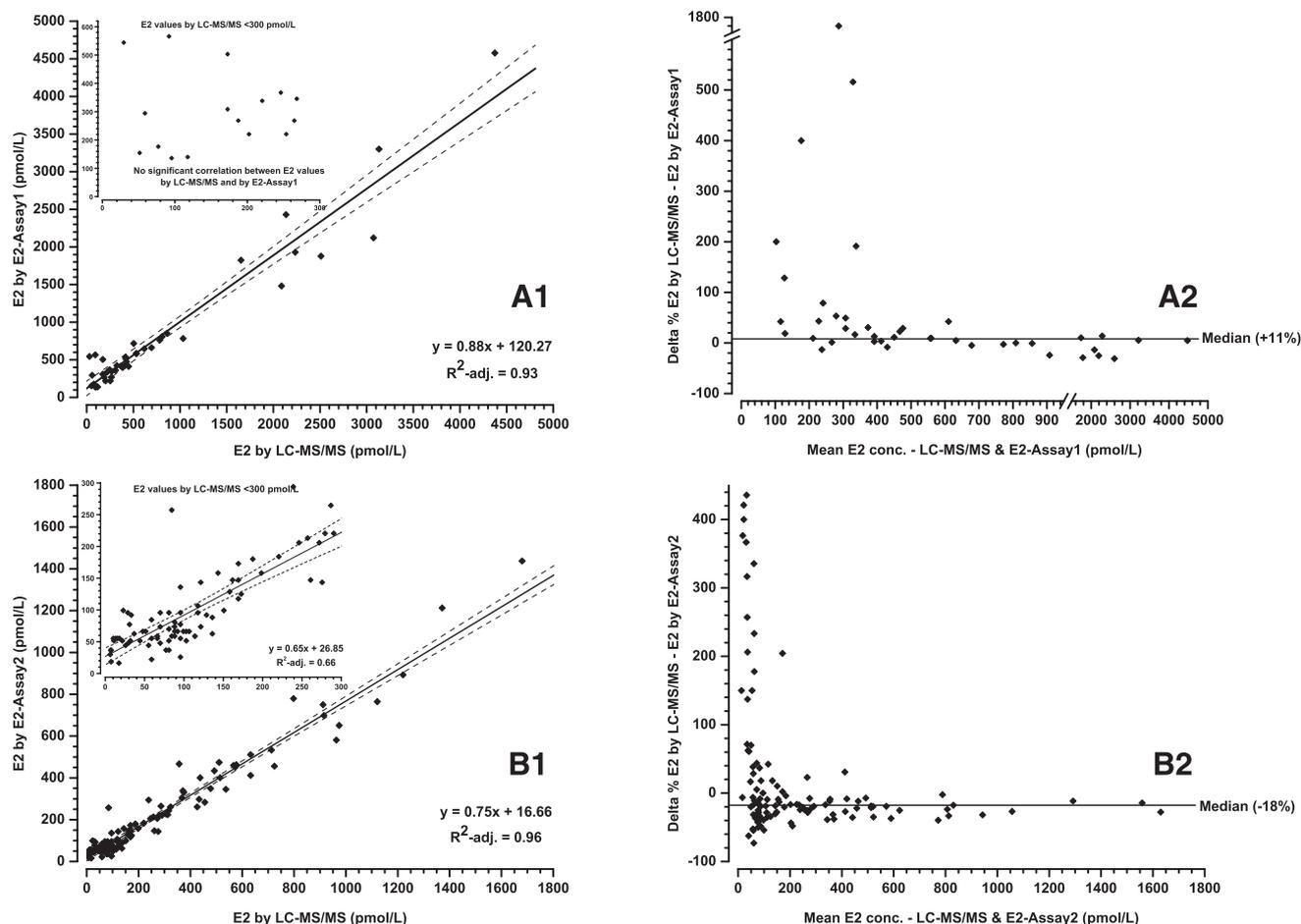


Fig. 4. Method comparison of E2 measurements by LC-MS/MS and immunoassays.

(A1 and A2), moderate-sensitivity automated direct immunoassay (E2-Assay1, ADVIA Centaur); (B1 and B2), high-sensitivity automated immunoassay (E2-Assay2, Vitros ECi); (C1 and C2; see next page), automated immunoassay (E2-Assay3, ACS:180) that had undergone in-house modification for enhanced sensitivity (organic extraction and concentration of the sample before assaying); (D1 and D2; see next page), outside RIA involving extraction and Sephadex LH-20 chromatography (Esoterix Endocrinology). The *left-hand panels* (A1, B1, C1, D1) depict scatter plots of E2 measurements by LC-MS/MS (*x* axes) vs E2 measurements by the different immunoassays (*y* axes). The *insets* show the scatter plots for E2 concentrations (as measured by LC-MS/MS) at lower concentrations. Regression lines (*solid lines*), 95% confidence intervals for the regressions (*dashed lines*), and linear fit functions with adjusted  $r^2$  {adjusted  $r^2 = 1 - [(1 - r^2) \times (n - 1)/(n - 2)]$ } are included on each scatter plot. The *right-hand panels* (A2, B2, C2, D2) show matched modified Bland–Altman plots for the scatter plots in A1, B1, C1, and D1. The *x* axes represent the mean values of E2 measurements by LC-MS/MS and the respective immunoassays (note axis break in A2). The *y* axes correspond to the percentage differences (expressed as percentages of the LC-MS/MS results) between E2 concentrations measured by LC-MS/MS and E2 concentrations measured by the various immunoassays (note axis breaks in A2 and C2). Positive percentage deviations indicate a higher E2 value in the respective immunoassay compared with LC-MS/MS, whereas negative percentages indicate the reverse. Because of the highly skewed data distribution, the median percentage deviations, rather than the mean percentage deviations, are depicted on each plot. For the same reason, parametric measures of data distribution (e.g., SD lines) are omitted from the plots.

We performed full regression (Deming) and Bland–Altman analyses for all method comparisons. For the lower portions of the analytical ranges, those parts that showed marked differences in the Bland–Altman plots were subjected to additional, separate regression analysis.

## Results

### LC-MS/MS CHARACTERISTICS OF E1 AND E2

The full-scan MS spectrum in positive-ion mode for the E2 peak on a column was obtained by injecting a 250  $\mu\text{g/L}$  solution of 3-dansyl-estradiol and is shown in Fig. 1. The main 3-dansyl-estradiol primary ion was at  $m/z$  506 in the first-quadrupole scan ( $Q_1$ ) and corresponded to  $[\text{M} + \text{H}]^+$  of the dansyl derivative of E2 ( $M_r$  505). After fragmentation of this primary 3-dansyl-estradiol ion in the collision

cell, scanning in the second quadrupole ( $Q_2$ ) yielded an intense peak at  $m/z$  171. This product ion corresponds to the 5-(dimethylamino)-naphthalene moiety of dansyl and is formed by cleavage of the C–S bond of the derivatized estrogens (Fig. 1B). The  $Q_1$  and  $Q_2$  scans of 3-dansyl-estrone also revealed the expected primary ion at  $m/z$  504 and product ion at  $m/z$  171. Similarly, the deuterated internal standards gave the expected transitions of  $m/z$  508 to 171 and  $m/z$  511 to 171 for the dansyl derivatives of  $d_4$ -E1 and  $d_5$ -E2, respectively.

A LC-MS/MS chromatogram of an extracted serum sample containing  $\sim 222.2$  pmol/L (60 ng/L) E1 and 220.6 pmol/L (60 ng/L) of E2 shown in Fig. 2. The derivatized E1 and E2 are separated chromatographically, which is necessary to achieve sensitivity by preventing ion sup-

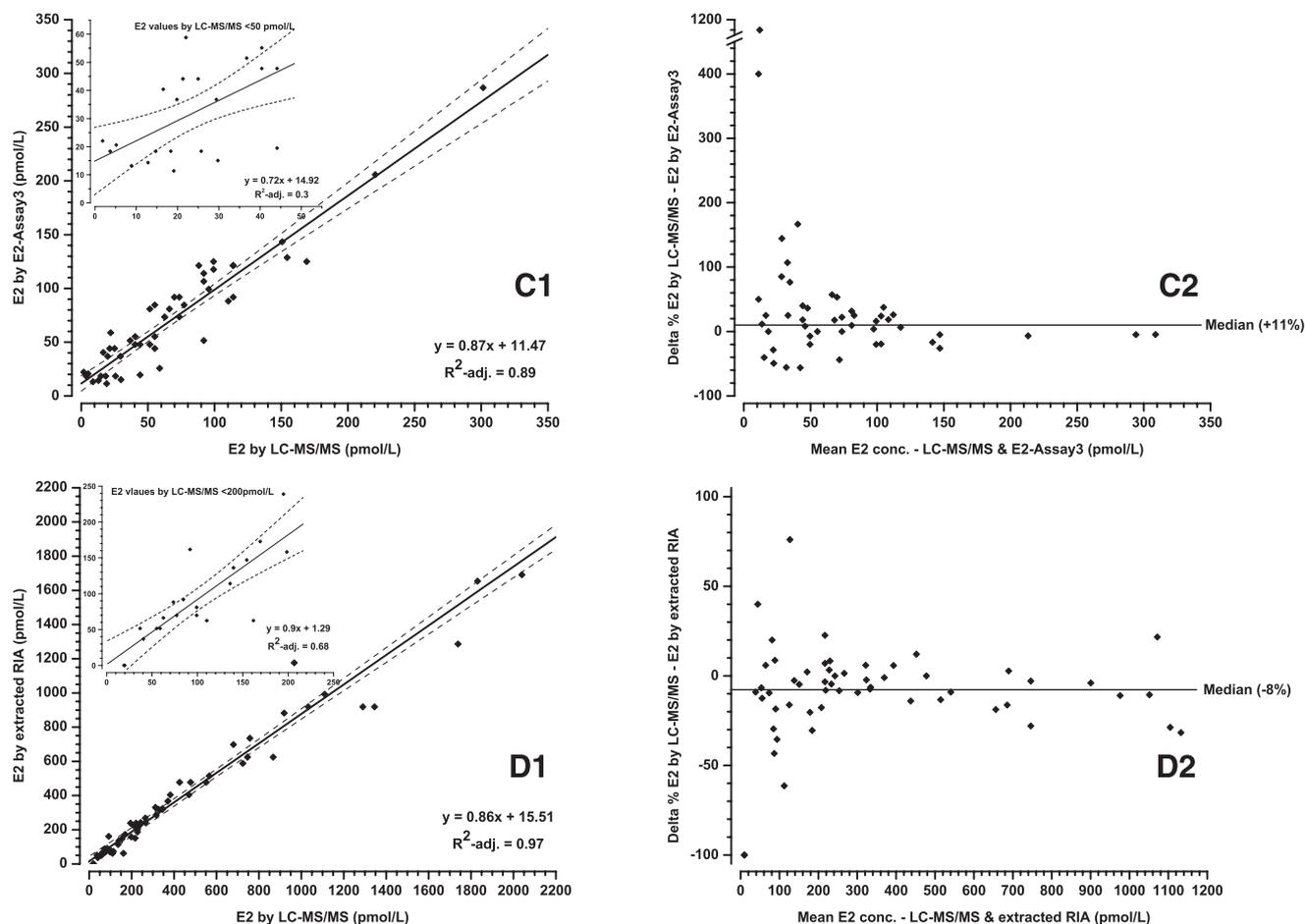


Fig. 4. Continued.

pression and any cross-talk from impurities (33). There was no indication of ion suppression at 3 min of elution: the peak area of the serum-extracted and unextracted internal standards and analytes were equal within the experimental error limits. The chosen column, flow rate, and gradient optimized the separation, while still allowing for a reasonable run time of 5 min. Retention times were 3.0 and 3.2 for the derivatives of E2 and E1, respectively. A signal-to-noise ratio of 10:1, as defined by the Analyst software as (analyte peak intensity)/(2 × SD of the noise), was observed for both E2 and E1 in an extract of a calibrator with a concentration of 37 pmol/L (10 ng/L).

#### BASIC ASSAY PERFORMANCE CHARACTERISTICS

The critical analyte concentration for E1 was 4.4 pmol/L (1.2 ng/L), and the detection limit was 12.9 pmol/L (3.5 ng/L), giving a detection range of 12.9–2222 pmol/L (3.5–600 ng/L). The CV of the replicates with added analyte at the detection limit was 27%. The functional assay sensitivity for E1 was <44.1 pmol/L (11.9 ng/L), but could not be determined more precisely because the lowest E1 concentration in an actual patient sample available was 44.1 pmol/L (11.9 ng/L; interassay CV = 12%).

For E2 the critical concentration was 6.2 pmol/L (1.7 ng/L), the detection limit was 10.3 pmol/L (2.8 ng/L; CV for replicates = 13%), the detection range was 10.3–2205 pmol/L (2.8–600 ng/L), and the functional sensitivity was 23.2 pmol/L (6.3 ng/L).

#### EXTRACTION EFFICIENCY, RECOVERY, PRECISION, LINEARITY, AND INTERFERENCES

The sample extraction efficiency was essentially 100%, as estimated by comparing the mean peak areas of extracted and unextracted internal standards, which agreed within the margins of experimental error.

The recoveries of known amounts of analyte added to samples ranged from 93% to 108% (mean, 100%) for E1 and from 100% to 110% (mean, 105%) for E2.

Intraassay CV for E1 ranged from 9% at a concentration of 42.6 pmol/L (11.5 ng/L) to 2% at a concentration of 1.35 nmol/L (365 ng/L), whereas interassay CV were between 12% [44.1 pmol/L (11.9 ng/L)] and 4% [1.36 nmol/L (368 ng/L)]. The corresponding intraassay CV for E2 ranged from 15% at a concentration of 24.3 pmol/L (6.6 ng/L) to 2.0% at a concentration of 1.3 nmol/L (357 ng/L), whereas the interassay CVs for E2 were between

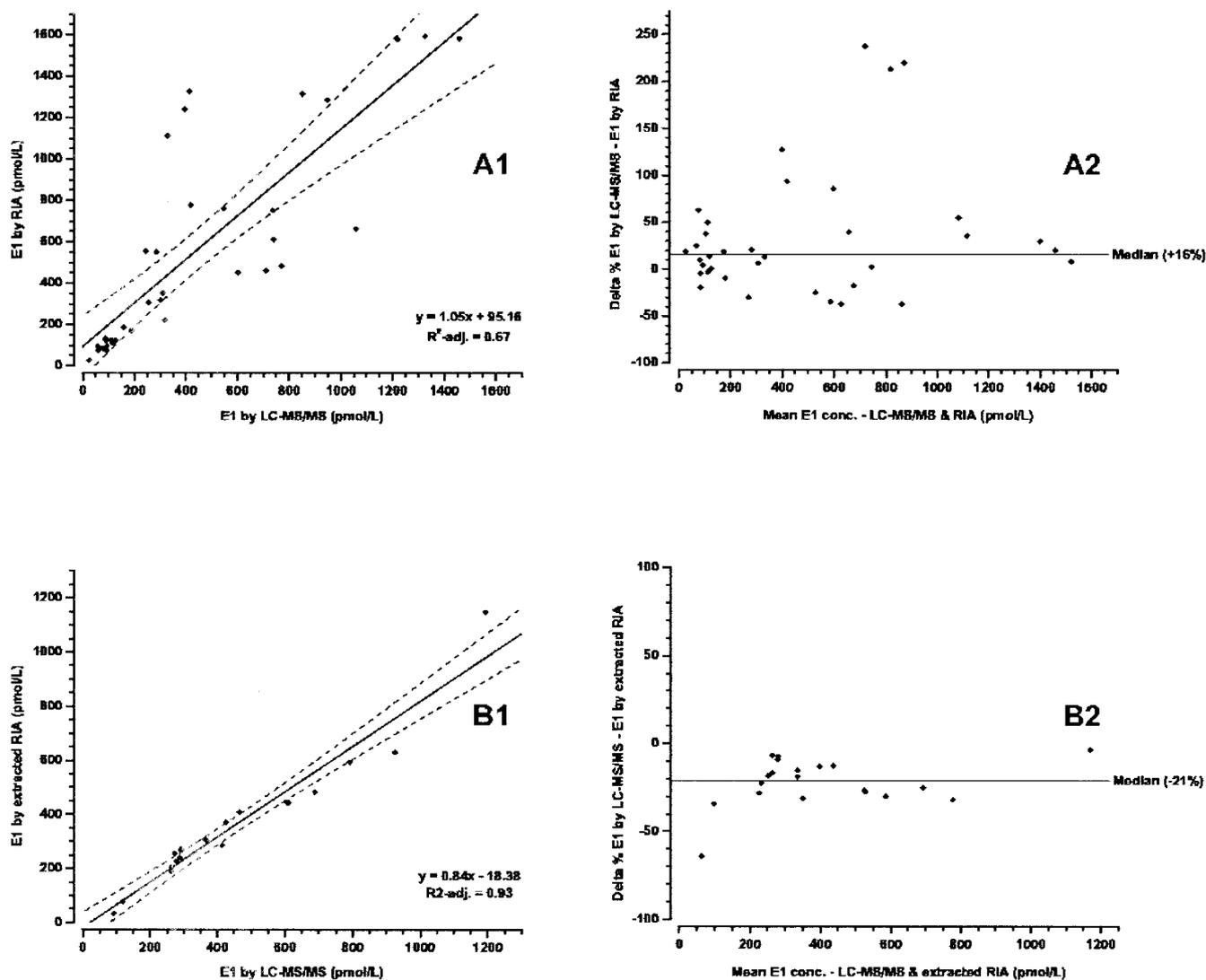


Fig. 5. Method comparison of E1 measurements by LC-MS/MS and with immunoassays.

(A1 and A2), direct RIA (Estrone; Diagnostic Systems Laboratories) compared with LC-MS/MS; (B1 and B2), outside RIA involving extraction and Sephadex LH-20 chromatography (Estrone; Esoterix Endocrinology) compared with LC-MS/MS. The *left-hand panels* (A1 and B1) are scatter plots of E1 measurements by LC-MS/MS (*x axes*) vs the corresponding measurements by the different immunoassays (*y axes*). Regression lines (*solid lines*), 95% confidence intervals for the regressions (*dashed lines*) and linear fit functions with adjusted  $r^2$  {adjusted  $r^2 = 1 - [(1 - r^2) \times (n - 1)/(n - 2)]$ } are included on each scatter plot. The *right-hand panels* (A2 and B2) show matched modified Bland-Altman plots for the scatter plots in A1 and B1. The *x axes* represent the mean values of E1 measured by LC-MS/MS and the respective immunoassays. The *y axes* correspond to the percentage differences (expressed as percentages of the LC-MS/MS results) between E1 concentrations measured by LC-MS/MS and the concentrations measured by the matched immunoassays. Positive percentage deviations indicate higher values in the respective immunoassays compared with LC-MS/MS, whereas negative percentages indicate the reverse. Because of the skewed data distribution, the median percentage deviations, rather than the mean percentage deviations, are depicted on each plot. For the same reason, parametric measures of data distribution (e.g., SD lines) are omitted from the plots.

20% [23.2 pmol/L (6.3 ng/L)] and 5% [1.29 nmol/L (352 ng/L)].

The seven-point calibrator concentrations (10–600 ng/L for either analyte, corresponding to 37–2222.4 pmol/L for E1 and 36.7–2205.6 pmol/L for E2), plotted against the ratio of analyte/internal standard areas for five consecutive assays, showed a linear and reproducible curve for each estrogen ( $r^2 = 0.995$  for E1 and 0.996 for E2). The sample-dilution series showed linear recoveries between 72% and 108% (mean, 95.1%) of expected values for E1 and between 83% and 111% (mean, 98%) for E2.

No E2 was detected when pure E1 was assayed, and vice versa. We observed no effect on the measured endogenous E2 concentration after addition of 7.408 nmol/L (2  $\mu\text{g/L}$ ) E1 to samples before extraction. Likewise, addition of 7.352 nmol/L (2  $\mu\text{g/L}$ ) E2 had no effect on the measured E1 concentration, whereas addition of 6.936 nmol/L (2  $\mu\text{g/L}$ ) estriol had no effect on the measurement values of either E1 or E2. We did observe a small effect of adding 6.748 nmol/L (2  $\mu\text{g/L}$ ) of ethinylestradiol to a sample containing a low E2 concentration of 70.4 pmol/L (19 ng/L), which increased to 92.6

pmol/L (25 ng/L). E1 sulfate added at an extremely high concentration of 546 nmol/L (200  $\mu$ g/L) had no effect on the measured E2 concentrations but did increase the measured E1 by 368 pmol/L (100 ng/L). We suspect that this is most likely attributable to contamination of the E1 sulfate source material by small amounts of E1. We had observed rare random contamination of samples by internal standard or E2 from high-concentration stocks used and prepared in the laboratory. Revised work practices of preparing stocks in isolation and in the hood have easily overcome this problem, and we have not observed any similar events in many hundreds of samples tested since then.

#### SPECIMEN-TYPE SUITABILITY AND STABILITY

We observed no significant differences among serum, EDTA plasma, and sodium heparin plasma. The mean differences between EDTA samples and serum samples were  $-5.6\%$  for E1 and  $-5.9\%$  for E2, whereas the corresponding values for the comparison of heparin samples with serum samples were  $-4.5\%$  for E1 and  $-0.9\%$  for E2. The samples were stable at 4 °C or ambient temperature for at least 7 days, with E1 values at day 7 for samples stored at ambient temperature differing from day 0 values by 4.0%, 5.7%, and 5.8% for serum, EDTA plasma, and heparin plasma, respectively. Day 7 E2 values differed from day 0 values by 2.5%, 4.0%, and 1.8% for serum, EDTA plasma, and heparin plasma, respectively. E1 and E2 were stable for at least three freeze-thaw cycles, with values after three cycles differing from baseline by a mean of  $-5.3\%$  for E1 and 2.6% for E2.

#### ESTIMATION OF REFERENCE INTERVALS

Age-adjusted estimated reference intervals for E1 and E2 in men and women are shown in Fig. 3. The intervals of 37–222.4 pmol/L (10–60 ng/L) for adult men, 63–740.8 pmol/L (17–200 ng/L) for premenopausal women, and 26–148.2 pmol/L (7–40 ng/L) for postmenopausal women for E1 are similar to published values (11). Likewise, the corresponding ranges for E2 of 36.7–147 pmol/L (10–40 ng/L) for men, 55.1–1286.6 pmol/L (15–350 ng/L) for premenopausal women, and  $<36.7$  pmol/L (10 ng/L) for postmenopausal women (Table 1) are in agreement with previously published values (34).

#### METHOD COMPARISON

All three automated E2 immunoassays (E2-Assay1, E2-Assay2, and E2-Assay3) demonstrated good overall correlation with the LC-MS/MS E2 assay ( $r^2 = 0.89$ – $0.96$ ), and slopes were between 0.75 and 0.88 (Fig. 4, A1, B1, and C1). However, at low E2 concentrations, the two direct E2 assays showed poor agreement with LC-MS/MS E2 measurements. E2-Assay1 showed no significant correlation with the LC-MS/MS E2 values for all concentrations  $<300$  pmol/L (81 ng/L), whereas results obtained with E2-Assay2, although still correlated with LC-MS/MS E2 results, differed increasingly from LC-MS/MS for mea-

surements  $<200$  pmol/L (54 ng/L; adjusted  $r^2 = 0.39$ ; insets in Fig. 4, A1 and B1). The divergent results at the lower E2 concentrations were predominately attributable to substantially higher E2 concentrations measured by the immunoassays compared with the LC-MS/MS assay (Fig. 4, A2 and B2). The in-house immunoassay (E2-Assay3), with enhanced sensitivity based on E2 extraction, agreed well with the LC-MS/MS E2 assay down to E2 concentrations of  $\sim 50$  pmol/L (13.5 ng/L), but correlated weakly at lower concentrations (Fig. 4, C1 and C2). The RIA (with a sample extraction step) from another reference laboratory showed excellent overall correlation (adjusted  $r^2 = 0.97$ ), with a slope of 0.86. At concentrations  $<200$  pmol/L (54 ng/L) the two assays correlated only modestly (Fig. 4, D1 and 4D2).

E1 results obtained by direct RIA correlated moderately well with E1 measurements by LC-MS/MS (adjusted  $r^2 = 0.67$ ; Fig. 5, A1 and A2), whereas E1 measurements by the RIA method that used a sample extraction step showed good to excellent agreement with LC-MS/MS (adjusted  $r^2 = 0.93$ ), although with a modest systematic downward bias of  $\sim 20\%$  across the measurement range (Fig. 5, B1 and B2).

#### Discussion

We describe a LC-MS/MS assay for simultaneous measurement of E1 and E2 with sensitivities that exceed those of most immunoassays, including many high-sensitivity immunoassays. The inclusion of stable-isotope internal standards in each sample allows accuracy and reproducibility similar to those of a reference method. It appears to be largely free of interference or cross-reactivity, a problem that continues to affect even estrogen immunoassays that incorporate sample extraction, as demonstrated in our method-comparison study. Unlike high- and ultra-high-sensitivity manual immunoassays, which achieve comparable or better sensitivities but require substantial manual labor and prolonged incubation times, our LC-MS/MS assay is fast, with little hands-on time despite the need for sample extraction and derivatization. Moreover, many of the front-end sample-processing steps can be automated by use of robotic workstations, making this assay suitable for high-throughput, short turnaround time applications that require high-sensitivity E1 or E2 measurements.

By contrast, most commercial estrogen immunoassays are unsuitable for clinical applications that require a low detection limit, e.g., E2 measurements in men or children or monitoring of gonadotropin-analog-mediated pituitary-gonadal axis down-regulation before in vitro fertilization and embryo transfer (35). The predominant E2 assays used in the modern clinical laboratory, direct automated immunoassays, lack sensitivity and are subject to interference by cross-reacting substances (36, 37). Many of these assays claim detection limits for E2 of as low as 40 pmol/L (10.8 ng/L), but as our method comparison data show, the results obtained by at least some immunoassays

are inaccurate, and the assays lack specificity at these low concentrations. In daily practice, most of these assays also are imprecise at E2 concentrations <200–250 pmol/L (~60 ng/L). Manufacturers of immunoassays often cite only assay detection limits in package inserts and do not mention functional sensitivities, which can be 4–10 times higher than the detection limit. Furthermore, our method-comparison data suggest that there is a systematic tendency for immunoassays to be down-calibrated, presumably to correct somewhat for the higher values observed at low estrogen concentrations because of interference and limited assay sensitivity (Figs. 4 and 5). Unfortunately, this does little to correct the fundamental problems of direct automated E2 immunoassays at low analyte concentrations, but simply superimposes a systematic downward bias on the results. Because this bias varies from manufacturer to manufacturer, E2 results obtained with different immunoassays seldom agree with each other. To address this issue, isotope-dilution GC-MS has been recommended as a reference method (38).

Immunoassays that incorporate sample extraction can overcome some of these problems, in particular the lack of sensitivity. Indeed, the most sensitive commercial E2 assays are immunoassays that use manual extraction with prolonged incubation times. Some of these assays have reported detection limits of ~1 pmol/L (39). Extraction and column chromatography can also reduce cross-reactivity, but as our method-comparison data show, they clearly do not eliminate it. High-sensitivity manual E2 immunoassays also share the problems of poor standardization that affect automated assays. Finally, high-sensitivity E2 immunoassays are typically labor-intensive, slow, and involve radioisotopes, characteristics that are all deemed undesirable in modern laboratories.

Chromatographic methods have been considered as an alternative to immunoassays, but these methods also have various limitations. Conventional HPLC-based methods have been reported, but they perform poorly in terms of turnaround time and sensitivity (40). GC-MS methods for the quantification of E2 in human serum samples represent the recognized gold-standard method, but their value in a clinical situation is restricted (25, 41, 42). The complexity of the serum matrix requires purification and concentration of estrogens before derivatization for GC-MS analysis. The common methods for purification are solvent-solvent and solid-phase extractions. Some other approaches use affinity columns, purification with Sephadex LH-20, strong ionic exchangers, weak ionic exchanger and reversed-phase HPLC, or various combinations of these methods. In addition to the extraction and derivatization process, reported run times for GC-MS-based estrogen assays may be as long as 45 min/sample and are rarely <15 min/sample. Finally, the reported GC-MS methods for E2 not only lack high-throughput capabilities but also have insufficient sensitivity to challenge modern immunoassays for routine clinical use.

LC-MS/MS has been recognized for some time as a

method that, when used in conjunction with isotopic calibrators, combines the reference methodology nature of GC-MS with higher throughput and improved sensitivity (29, 30). However, E1 and E2 are poor ionizers, seriously limiting the sensitivity of LC-MS/MS for underivatized samples (31). As far as we are aware, there has only been a single report in the literature that suggested that acceptable assay sensitivity could be achieved for E2 in a serum matrix (43). Unfortunately, in our laboratory we have not been able to reproduce these results, despite using superior instrumentation. Although the lowest achievable sensitivity for underivatized E1 of 37 pmol/L (10 ng/L) was acceptable, E2 ionization was so poor that we were unable to exceed a functional sensitivity of 370 pmol/L (100 ng/L) for this analyte. Derivatization with dansyl chloride overcame this problem without substantially increasing the complexity or time requirements of preanalytical sample preparation. The described method does not have any known cross-reactivity, the high specificity being attributable to monitoring of parent/daughter ions. Although chromatography prolongs the run time to 5 min, it provides additional specificity, prevents ion suppression (33), and allows fractionation of estrogens into E1 and E2 with their simultaneous detection.

The additional information that is provided by the high-sensitivity E1 measurement might be useful in certain clinical situations. For example, in obese postmenopausal women, the fraction of androstenedione converted to E1 increases, yielding enough estrogen to produce endometrial proliferation and bleeding. In men, E1 concentrations may be of value in the investigation of gynecomastia or the detection of estrogen-producing tumors. With regard to E2 measurement, our assay is suitable for most high-sensitivity applications, such as accurate measurement of E2 in men, most children, and the majority of postmenopausal women.

The sensitivity of the LC-MS/MS E2 assay is currently not sufficient with the present instrumentation for situations requiring determination of E2 concentrations between 1 and 10 pmol/L (15). There are some clinical and research applications falling into this category. These include some pediatric applications and monitoring of complete medical estrogen suppression as is used, for example, in some experimental adjuvant breast cancer therapies. However, because each successive generation of LC-MS/MS instruments has improved in sensitivity by a factor of 5–10, we would expect that these applications should become feasible within the next few years. Finally, turnaround time, although reasonably quick, may also be insufficient for some *in vitro* fertilization applications, but front-end automation might be able to overcome this limitation.

In addition to its potential clinical usefulness, estrogen measurement by LC-MS/MS could be used in quality-assurance programs to accurately determine the E2 value of the survey material, allowing improved assessment of the accuracy of E2 immunoassays used in participating

laboratories. The LC-MS/MS method is not limited to a particular MS/MS instrument, but is portable to other LC-MS/MS platforms. Indeed, our method had sensitivity and specificity comparable to those for the Quattro micro tandem mass spectrometer (Micromass).

In summary, we have developed a highly sensitive method for the simultaneous measurement of E1 and E2 in human serum by LC-MS/MS after derivatization with dansyl chloride. Measurement of E1 and E2 by LC-MS/MS is precise and accurate down to 15–20 pmol/L (5–10 ng/L). With a 5-min run time, the assay is a good alternative to immunoassays in the clinical reference laboratory. Compared with other reference methodologies, such as isotope-dilution GC-MS methods, the LC-MS/MS method has substantially higher throughput and sensitivity while offering at least equal specificity.

### References

1. Yasui T, Uemura H, Tezuka M, Yamada M, Irahara M, Miura M, et al. Biological effects of hormone replacement therapy in relation to serum estradiol levels. *Horm Res* 2001;56:38–44.
2. Ettinger B, Pressman A, Sklarin P, Bauer DC, Cauley JA, Cummings SR. Association between low levels of serum estradiol, bone density, and fractures among elderly women: the study of osteoporotic fractures. *J Clin Endocrinol Metab* 1998;83:2239–43.
3. Amin S, Zhang Y, Sawin CT, Evans SR, Hannan MT, Kiel DP, et al. Association of hypogonadism and estradiol levels with bone mineral density in elderly men from the Framingham Study. *Ann Intern Med* 2000;133:951–63.
4. Owens JW, Ashby J. Critical review and evaluation of the uterotrophic bioassay for the identification of possible estrogen agonists and antagonists: in support of the validation of the OECD uterotrophic protocols for the laboratory rodent. Organisation for Economic Co-operation and Development. *Crit Rev Toxicol* 2002;32:445–520.
5. Fang H, Tong W, Perkins R, Soto AM, Prechtl NV, Sheehan DM. Quantitative comparison of in vitro assays for estrogenic activities. *Environ Health Perspect* 2000;108:723–9.
6. Le Guevel R, Pakdel F. Assessment of oestrogenic potency of chemicals used as growth promoter by in-vitro methods. *Hum Reprod* 2001;16:1030–6.
7. O'Connell MB. Pharmacokinetic and pharmacologic variation between different estrogen products. *J Clin Pharmacol* 1995;35:18S–24S.
8. Valbuena D, Jasper M, Remohi J, Pellicer A, Simon C. Ovarian stimulation and endometrial receptivity. *Hum Reprod* 1999;14:107–11.
9. Kligman I, Rosenwaks Z. Differentiating clinical profiles: predicting good responders, poor responders, and hyperresponders. *Fertil Steril* 2001;76:1185–90.
10. Argente J. Diagnosis of late puberty. *Horm Res* 1999;51:95–100.
11. Bidlingmaier F, Wagner-Barnack M, Butenandt O, Knorr D. Plasma estrogens in childhood and puberty under physiologic and pathologic conditions. *Pediatr Res* 1973;7:901–7.
12. Feuillan PP, Foster CM, Pescovitz OH, Hench KD, Shawker T, Dwyer A, et al. Treatment of precocious puberty in the McCune-Albright syndrome with the aromatase inhibitor testolactone. *N Engl J Med* 1986;315:1115–9.
13. Ikegami S, Moriwake T, Tanaka H, Inoue M, Kubo T, Suzuki S, et al. An ultrasensitive assay revealed age-related changes in serum oestradiol at low concentrations in both sexes from infancy to puberty. *Clin Endocrinol* 2001;55:789–95.
14. lughetti L, Predieri B, Ferrari M, Gallo C, Livio L, Milioli S, et al. Diagnosis of central precocious puberty: endocrine assessment. *J Pediatr Endocrinol Metab* 2000;13:709–15.
15. Klein KO, Baron J, Colli MJ, McDonnell DP, Cutler GB Jr. Estrogen levels in childhood determined by an ultrasensitive recombinant cell bioassay. *J Clin Invest* 1994;94:2475–80.
16. Lebrethon MC, Bourguignon JP. Management of central isosexual precocity: diagnosis, treatment, outcome. *Curr Opin Pediatr* 2000;12:394–9.
17. Traggiai C, Stanhope R. Delayed puberty. *Best Pract Res Clin Endocrinol Metab* 2002;16:139–51.
18. Cummings SR, Duong T, Kenyon E, Cauley JA, Whitehead M, Krueger KA. Multiple Outcomes of Raloxifene Evaluation (MORE) Trial. Serum estradiol level and risk of breast cancer during treatment with raloxifene [Comment]. *JAMA* 2002;287:216–20.
19. Lonning PE, Geisler J, Johannessen DC, Ekse D. Plasma estrogen suppression with aromatase inhibitors evaluated by a novel, sensitive assay for estrone sulphate. *J Steroid Biochem Mol Biol* 1997;61:255–60.
20. Hogervorst E, Williams J, Combrinck M, David Smith A. Measuring serum oestradiol in women with Alzheimer's disease: the importance of the sensitivity of the assay method. *Eur J Endocrinol* 2003;148:67–72.
21. Carlstrom K. Low endogenous estrogen levels—analytical problems and tissue sensitivity. *Acta Obstet Gynecol Scand Suppl* 1996;163:11–5.
22. Siekmann L. Determination of oestradiol-17 $\beta$  in human serum by isotope dilution-mass spectrometry. *Definitive methods in clinical chemistry, II. J Clin Chem Clin Biochem* 1984;22:551–7.
23. Choi MH, Kim KR, Chung BC. Determination of estrone and 17 $\beta$ -estradiol in human hair by gas chromatography-mass spectrometry. *Analyst* 2000;125:711–4.
24. Xiao X, McCalley D. Quantitative analysis of estrogens in human urine using gas chromatography/negative chemical ionisation mass spectrometry. *Rapid Commun Mass Spectrom* 2000;14:1991–2001.
25. Biancotto G, Angeletti R, Traldi P, Silvestri M, Saccon M, Guidugli F. Determination of 17 $\beta$ -estradiol in bovine plasma: development of a highly sensitive technique by ion trap gas chromatography-tandem mass spectrometry using negative chemical ionization. *J Mass Spectrom* 2002;37:1266–71.
26. Griffiths WJ, Jonsson AP, Liu S, Rai DK, Wang Y. Electrospray and tandem mass spectrometry in biochemistry. *Biochem J* 2001;355:545–61.
27. Marquet P. Progress of liquid chromatography-mass spectrometry in clinical and forensic toxicology. *Ther Drug Monit* 2002;24:255–76.
28. Rashed MS, Rahbeeni Z, Ozand PT. Application of electrospray tandem mass spectrometry to neonatal screening. *Semin Perinatol* 1999;23:183–93.
29. Taylor RL, Machacek D, Singh RJ. Validation of a high-throughput liquid chromatography-tandem mass spectrometry method for urinary cortisol and cortisone. *Clin Chem* 2002;48:1511–9.
30. Taylor RL, Singh RJ. Validation of liquid chromatography-tandem mass spectrometry method for analysis of urinary conjugated metanephrine and normetanephrine for screening of pheochromocytoma. *Clin Chem* 2002;48:533–9.
31. Anari MR, Bakhtiar R, Zhu B, Huskey S, Franklin RB, Evans DC. Derivatization of ethinylestradiol with dansyl chloride to enhance electrospray ionization: application in trace analysis of ethinylestradiol in rhesus monkey plasma. *Anal Chem* 2002;74:4136–44.
32. Inczedy J, Lengyel T, Ure AM. Detection and quantification capa-

- bilities, section 18.4.3.7. In: Compendium of analytical nomenclature—definitive rules 1997. [http://www.iupac.org/publications/analytical\\_compendium/Cha18sec437.pdf](http://www.iupac.org/publications/analytical_compendium/Cha18sec437.pdf) (Accessed August 25, 2003).
33. Annesley TM. Ion suppression in mass spectrometry. *Clin Chem* 2003;49:1041–4.
  34. Yasui T, Yamada M, Kinoshita H, Uemura H, Yoneda N, Irahara M, et al. Combination of automatic HPLC-RIA method for determination of estrone and estradiol in serum. *J Clin Lab Anal* 1999;13:266–72.
  35. Kovacs P, Witt BR. Day 6 estradiol level predicts cycle cancellation among poor responder patients undergoing in vitro fertilization—embryo transfer cycles using a gonadotropin-releasing hormone agonist flare regimen. *J Assist Reprod Genet* 2002;19:349–53.
  36. Sinicco A, Raiteri R, Rossati A, Savarino A, Di Perri G. Efavirenz interference in estradiol ELISA assay. *Clin Chem* 2000;46:734–5.
  37. Leung YS, Dees K, Cyr R, Schloegel I, Kao PC. Falsely increased serum estradiol results reported in direct estradiol assays. *Clin Chem* 1997;43:1250–1.
  38. Thienpont LM, De Leenheer AP. Efforts by industry toward standardization of serum estradiol-17 $\beta$  measurements. *Clin Chem* 1998;44:671–4.
  39. Worthman CM, Stallings JF, Hofman LF. Sensitive salivary estradiol assay for monitoring ovarian function. *Clin Chem* 1990;36:1769–73.
  40. Games DE. High-performance liquid chromatography/mass spectrometry (HPLC/MS). *Adv Chromatogr* 1983;21:1–39.
  41. Ding WH, Chiang CC. Derivatization procedures for the detection of estrogenic chemicals by gas chromatography/mass spectrometry. *Rapid Commun Mass Spectrom* 2003;17:56–63.
  42. Wu H, Ramsay C, Ozaeta P, Liu L, Aboleneen H. Serum estradiol quantified by isotope dilution-gas chromatography/mass spectrometry [erratum published in *Clin Chem* 2002;48:686]. *Clin Chem* 2002;48:364–6.
  43. Draisci R, Palleschi L, Ferretti E, Marchiafava C, Lucentini L, Cammarata P. Quantification of 17 $\beta$ -estradiol residues in bovine serum by liquid chromatography-tandem mass spectrometry with atmospheric pressure chemical ionization. *Analyst* 1998;123:2605–9.