

# Biological and Technical Variables Affecting Immunoassay Recovery of Cytokines from Human Serum and Simulated Vaginal Fluid: A Multicenter Study

Raina N. Fichorova,<sup>\*,†</sup> Nicola Richardson-Harman,<sup>‡</sup> Massimo Alfano,<sup>§</sup> Laurent Belec,<sup>||</sup> Cedric Carbonneil,<sup>||,×</sup> Silvia Chen,<sup>⊥</sup> Lisa Cosentino,<sup>#</sup> Kelly Curtis,<sup>∇</sup> Charlene S. Dezzutti,<sup>#,∇</sup> Betty Donoval,<sup>○</sup> Gustavo F. Doncel,<sup>◆</sup> Melissa Donaghay,<sup>◆</sup> Jean-Charles Grivel,<sup>⊥</sup> Esmeralda Guzman,<sup>¶</sup> Madeleine Hayes,<sup>+</sup> Betsy Herold,<sup>¶,∇</sup> Sharon Hillier,<sup>#</sup> Carol Lackman-Smith,<sup>‡</sup> Alan Landay,<sup>○</sup> Leonid Margolis,<sup>⊥</sup> Kenneth H. Mayer,<sup>∞</sup> Jenna-Malia Pasicznyk,<sup>†</sup> Melanie Pallansch-Cokonis,<sup>‡</sup> Guido Poli,<sup>§</sup> Patricia Reichelderfer,<sup>⊥</sup> Paula Roberts,<sup>‡</sup> Irma Rodriguez,<sup>∞</sup> Hela Saidi,<sup>||,ε</sup> Rosaria Rita Sassi,<sup>†</sup> Robin Shattock,<sup>+</sup> and James E. Cummins, Jr.<sup>‡</sup>

Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, Microbicides Quality Assurance Program, Southern Research Institute, Frederick, Maryland 21701, Vita-Salute San Raffaele University and Scientific Institute, Milan, Italy, Centre de Recherche des Cordeliers, Paris, France, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, Magee-Womens Research Institute, Pittsburgh, Pennsylvania 15213, Centers for Disease Control and Prevention, Atlanta, Georgia 30333, Rush University Medical Center, Chicago, Illinois 60612, CONRAD, Eastern Virginia Medical School, Norfolk, Virginia 23507, Mount Sinai School of Medicine, New York, New York 10029, St. George's University of London, London, U.K., and Brown University, Providence, Rhode Island 02912

The increase of proinflammatory cytokines in vaginal secretions may serve as a surrogate marker of unwanted inflammatory reaction to microbicide products topically applied for the prevention of sexually transmitted diseases, including HIV-1. Interleukin (IL)-1 $\beta$  and IL-6 have been proposed as indicators of inflammation and increased risk of HIV-1 transmission; however, the lack of information regarding detection platforms optimal for vaginal fluids and interlaboratory variation limit their use for microbicide evaluation and other clinical applications. This study examines fluid matrix variants relevant to vaginal sampling techniques and proposes a model for interlaboratory comparisons across current cytokine detection technologies. IL-1 $\beta$  and IL-6 standards were measured by 12 laboratories in four countries, using 14 immunoassays and four detection platforms based on absorbance, chemiluminescence, electrochemiluminescence, and fluorescence. International reference preparations of cytokines with defined biological activity were

spiked into (1) a defined medium simulating the composition of human vaginal fluid at pH 4.5 and 7.2, (2) physiologic salt solutions (phosphate-buffered saline and saline) commonly used for vaginal lavage sampling in clinical studies of cytokines, and (3) human blood serum. Assays were assessed for reproducibility, linearity, accuracy, and significantly detectable fold difference in cytokine level. Factors with significant impact on cytokine recovery were determined by Kruskal–Wallis analysis of variance with Dunn's multiple comparison test and multiple regression models. All assays showed acceptable intra-assay reproducibility; however, most were associated with significant interlaboratory variation. The smallest reliably detectable cytokine differences ( $P < 0.05$ ) derived from pooled interlaboratory data varied from 1.5- to 26-fold depending on assay, cytokine, and matrix type. IL-6 but not IL-1 $\beta$  determinations were lower in both saline and phosphate-buffered saline as compared to vaginal fluid matrix, with no significant effect of pH. The (electro)chemiluminescence-based assays were most discriminative and consistently detected <2-fold differences within each matrix type. The Luminex-based assays were less discriminative with lower reproducibility between laboratories. These results suggest the need for uniform vaginal sampling techniques and a better understanding of immunoassay platform differences and cross-validation before the biological significance of cytokine variations can be validated in clinical trials. This investigation provides the first standardized analytic approach for assessing differences in mucosal cytokine levels and may improve strategies for monitoring immune responses at the vaginal mucosal interface.

\* To whom correspondence should be addressed. E-mail: rfichorova@rics.bwh.harvard.edu. Fax: +617-713-3018.

† Harvard Medical School.

‡ Southern Research Institute.

§ Vita-Salute San Raffaele University and Scientific Institute.

|| Centre de Recherche des Cordeliers.

⊥ National Institutes of Health.

# Magee-Womens Research Institute.

∇ Centers for Disease Control and Prevention.

○ Rush University Medical Center.

◆ Eastern Virginia Medical School.

¶ Mount Sinai School of Medicine.

+ St. George's University of London.

∞ Brown University.

× Current address: French Ministry of Health, Paris, France.

∇ Current address: Albert Einstein College of Medicine, New York, NY.

ε Current address: Antiviral Immunity, Biotherapy and Vaccine Unit, Pasteur Institute, Paris, France.

Topical microbicides are considered a leading preventive strategy to reduce the sexual transmission of HIV-1 worldwide. To be effective, topical microbicides should have minimal or no impact on the natural structural and functional integrity of the human cervicovaginal mucosa.<sup>1-3</sup> Cytokines such as interleukin (IL)-1 $\beta$  and IL-6 have emerged as sensitive indicators of compound-induced mucosal toxicity and are elevated with bacterial vaginosis and sexually transmitted infections (STI), conditions associated with inflammation known to increase the risk of acquiring or transmitting HIV-1 infection.<sup>4-7</sup> IL-1 $\beta$  and IL-6 can up-regulate HIV-1 replication,<sup>8</sup> and concentrations in vaginal secretions correlate with proviral HIV-1 DNA and cell-associated/cell-free HIV-1 RNA levels.<sup>9-12</sup> Therefore, cytokines have been proposed to be part of screening algorithms for microbicide safety evaluation, and to this purpose, an increasing number of clinical studies have collected cervicovaginal secretions, usually sampled by lavage with normal saline or phosphate-buffered saline (PBS).<sup>4-6,12-15</sup> However, the detection of cytokines in the complex biological background (matrix) of these fluids may be affected by multiple biological factors (e.g., abundance of high molecular weight proteins such as the mucins and pH variations that depend on microflora, menstrual cycle, and sexual intercourse) as well as technical performance and assay-related parameters (e.g., linearity range, sensitivity to cytokine variants). The lack of information on the relative importance of these factors and the comparative limitations of available detection methods make it difficult to interpret the significance of the detected variations in cytokine levels and to compare results generated in different laboratories. Currently available technologies for cytokine assessment include (1) immunoassays using an ELISA principle and either absorbance, luminescence, or fluorochrome-tagged detection to extrapolate concentrations from calibrator-based standard curves, (2) proteomic mass spectrometry analysis of peptide composition,

and (3) functional assays, in which a test sample is assigned arbitrary units of biological activity. Immunoassays are usually the method of choice because they are relatively easy to perform and standardize, and they are cytokine-specific, in contrast to bioassays that may be influenced by cytokine redundancy and synergistic effects.<sup>16,17</sup>

In this study, we evaluated current formats of cytokine immunoassays, which utilize either microplate- or bead-bound antibodies to extract the target cytokine concentration from a complex biological background (matrix). We applied a standardized approach to IL-1 $\beta$  and IL-6 quantification by cross-validating immunoassays against international cytokine standards of known biological activity. A defined medium simulating vaginal fluid content, saline and PBS, which are commonly used for vaginal lavage, and human blood serum, were analyzed with respect to their effect on cytokine recovery and reproducibility both within and between assays and laboratories. Our objectives were (1) to identify potential sources of analytical variations of cytokine measurement within and among independent laboratories based on biological matrix and assay type and (2) to establish recommendations for cytokine determination, data analysis, and interpretation. Our results provide novel information on the role of cytokines as biomarkers of mucosal immunity and are a step toward their future clinical utilization for screening and monitoring the bioactivity of safe anti-HIV-1 microbicides.

## MATERIALS AND METHODS

**Cytokines and Immunoassays.** *International Cytokine Standards and Matrix Types.* Recombinant human cytokine reference preparations have been established as international standards with the authorization of the Expert Committee on Biological Standardization (ECBS) of the World Health Organization (WHO).<sup>18</sup> The following panels were available through the U.K. National Institute for Biological Standards and Control (NIBSC): recombinant IL-1 $\beta$ , NIBSC code 86/680 and code 86/552, both derived from *E. coli*, and human recombinant IL-6, NIBSC code 89/548, expressed in Chinese hamster ovary (CHO) cells.<sup>19,20</sup> Freeze-dried preparations of the NIBSC cytokine standards were obtained by Southern Research Institute (SRI). Each NIBSC standard was reconstituted with distilled water (Cellgro Mediatech, Herndon, VA) to a concentration corresponding to 100 000 IU of biological activity/mL. The corresponding concentrations were 1  $\mu$ g/mL for IL-6 and IL-1 $\beta$  86/680 and 0.75  $\mu$ g/mL for IL-1 $\beta$  86/552.<sup>19,20</sup> Samples of these preparations were diluted 20 $\times$  to a final concentration of 5000 IU/mL in the following five matrices: saline with pH 5–6.5 (AmTech/Phoenix Scientific, Inc., St. Joseph, MO), Dulbecco's PBS with Ca<sup>2+</sup> and Mg<sup>2+</sup>, pH 7.2 (GIBCO Invitrogen, Carlsbad, CA), human blood serum AB blood type (Valley Biomedical, Inc., Winchester, VA), and vaginal fluid simulant (VFS) prepared in the Channing Laboratory, Harvard Medical

- (1) Hillier, S. L.; Moench, T.; Shattock, R.; Black, R.; Reichelderfer, P.; Veronese, F. J. *Acquired Immune Defi. Syndr.* **2005**, *39*, 1–8.
- (2) Stone, A. *Nat. Rev. Drug Discovery* **2002**, *1*, 977–985.
- (3) Turpin, J. *Expert Opin. Invest. Drugs* **2002**, *11*, 1077–1097.
- (4) Fichorova, R. N.; Bajpai, M.; Chandra, N.; Hsiu, J. G.; Spangler, M.; Ratnam, V.; Doncel, G. F. *Biol. Reprod.* **2004**, *71*, 761–769.
- (5) Doncel, G. F.; Chandra, N.; Fichorova, R. N. *J. Acquired Immune Defi. Syndr.* **2004**, *37*, S174–S179.
- (6) Fichorova, R. N.; Zhou, F.; Ratnam, V.; Atanassova, V.; Jiang, S.; Strick, N.; Neurath, A. R. *Antimicrob. Agents Chemother.* **2005**, *49*, 323–335.
- (7) Basso, B.; Gimenez, F.; Lopez, C. *Infect Dis. Obstet. Gynecol.* **2005**, *13*, 207–211.
- (8) Alfano, M.; Poli, G. *Mol. Immunol.* **2005**, *42*, 161–182.
- (9) Zara, F.; Nappi, R. E.; Berra, R.; Migliavacca, R.; Maserati, R.; Spinillo, A. *Sex. Transm. Infect.* **2004**, *80*, 108–112.
- (10) Al-Harthi, L.; Kovacs, A.; Coombs, R. W.; Reichelderfer, P. S.; Wright, D. J.; Cohen, M. H.; Cohn, J.; Cu-Uvin, S.; Watts, H.; Lewis, S.; Beckner, S.; Landay, A. *AIDS* **2001**, *15*, 1535–1543.
- (11) Coombs, R. W.; Reichelderfer, P. S.; Landay, A. L. *AIDS* **2003**, *17*, 455–480.
- (12) Cummins, J. E.; Christensen, L.; Lennox, J. L.; Bush, T. J.; Wu, Z.; Malamud, D.; Evans-Strickfaden, T.; Siddig, A.; Caliendo, A. M.; Hart, C. E.; Dezzutti, C. S. *AIDS Res. Hum. Retroviruses* **2006**, *22*, 788–795.
- (13) Schwartz, J. L.; Mauck, C.; Lai, J. J.; Creinin, M. D.; Brache, V.; Ballagh, S. A.; Weiner, D. H.; Hillier, S. L.; Fichorova, R. N.; Callahan, M. *Contraception* **2006**, *74*, 133–140.
- (14) Cheshenko, N.; Keller, M. J.; MasCasullo, V.; Jarvis, G. A.; Cheng, H.; John, M.; Li, J. H.; Hogarty, K.; Anderson, R. A.; Waller, D. P.; Zaneveld, L. J.; Profy, A. T.; Klotman, M. E.; Herold, B. C. *Antimicrob. Agents Chemother.* **2004**, *48*, 2025–2036.
- (15) Keller, M. J.; Zerhouni-Layachi, B.; Cheshenko, N.; John, M.; Hogarty, K.; Kasowitz, A.; Goldberg, C. L.; Wallenstein, S.; Profy, A. T.; Klotman, M. E.; Herold, B. C. *J. Infect. Dis.* **2006**, *193*, 27–35.

- (16) Bienvenu, J. A. D.; Monneret, G.; Gutowski, M. C.; Fabien, N. *Toxicology* **1998**, *129*, 55–61.
- (17) Wang, W.; Zhou, H.; Lin, H.; Roy, S.; Shaler, T. A.; Hill, L. R.; Norton, S.; Kumar, P.; Anderle, M.; Becker, C. H. *Anal. Chem.* **2003**, *75*, 4818–4826.
- (18) Mire-Sluis, A. R.; Das, R. G.; Padilla, A. J. *Immunol. Methods* **1998**, *216*, 103–116.
- (19) Gaines Das, R. E.; Poole, S. J. *Immunol. Methods* **1993**, *160*, 147–153.
- (20) Poole, S.; Gaines Das, R. E. *J. Immunol. Methods* **1991**, *142*, 1–13.

**Table 1. Cytokine Immunoassay Parameters**

assay name	kit manufacturer	sensitivity <pg/mL <sup>a</sup>	assay range pg/mL <sup>b</sup>	NIBSC (IU per pg) <sup>c</sup>	number (ID) of test laboratories	NIBSC spikes (IU)
<b>IL-1<math>\beta</math></b>						
BioSource ELISA	Invitrogen	1	3.9–250	86/680 (0.1)	2 (A, C)	0.25, 2.5, 25
Endogen EH2 ELISA	Pierce Biotechnology	1	10.24–400	86/680 (0.2)	1 (D)	0.8, 8, 80
Quantikine ELISA	R&D Systems	1	3.9–250	86/552 (0.098)	4 (B, E, F, H)	0.25, 2.5, 25
QuantiGlo ELISA	R&D Systems	0.4	0.5–5000	86/680 (0.1128)	2 (A, J)	5, 50, 500
Beadlyte Luminex	Millipore	0.5	10–7500 <sup>d</sup>	86/680 (0.244)	2 (A, K)	5, 50, 500
in-house Luminex <sup>e</sup>			4.57–3333	–(0.2) <sup>f</sup>	2 (G, I)	5, 50, 500
MSD Small Spot	Meso Scale Discovery	0.1	2.44–10 000	–(0.2) <sup>f</sup>	2 (A, J)	5, 50, 500
<b>IL-6</b>						
BioSource ELISA	Invitrogen	2	19–2024 <sup>d</sup>	89/548 (0.1)	2 (A, C)	2, 20, 200
Endogen EH2 ELISA	Pierce Biotechnology	1	10.24–400	89/548 (0.3)	2 (D, J)	1.2, 12, 120
Quantikine ELISA	R&D Systems	0.7	3.12–300	89/548 (0.131)	4 (B, E, F, L)	0.4, 4, 40
QuantiGlo ELISA	R&D Systems	0.4	0.48–1500	89/548 (0.131)	2 (A, J)	1.6, 16, 160
Beadlyte Luminex	Millipore	1	11–8000 <sup>d</sup>	89/548 (0.128)	2 (A, K)	5, 50, 500
in-house Luminex			4.57–3333	–(0.11) <sup>f</sup>	2 (G, I)	3.5, 35, 350
MSD Small Spot	Meso Scale Discovery	0.1	0.6–2500	–(0.1) <sup>f</sup>	2 (A, J)	2.5, 25, 250

<sup>a</sup> Value of the smallest dose that is not zero with 95% CI, typically calculated as the mean signal +2 SD of the zero standard read from standard curve. <sup>b</sup> Range provided by manufacturer, if commercially available assay kit, or by testing laboratory, if in-house assay. <sup>c</sup> WHO-recommended standard, provided by the U.K. National Institute of Biological Standards and Control (NIBSC), was used by manufacturer to calibrate the assay. <sup>d</sup> May vary by lot as per manufacturer instructions; however, it did not vary in this study. <sup>e</sup> Assay was designed in-house using separate commercially available reagents. <sup>f</sup> Manufacturer calibrated against recombinant human IL-1 $\beta$  or IL-6 standards from R&D Systems, which have been calibrated by R&D Systems against NIBSC IU.

School at pH 4.5 and 7.2, as previously described.<sup>21</sup> These preparations were distributed to the participating laboratories for analysis following the study design described below.

**Immunoassays.** The cytokine assays tested in this study included traditional solid-phase sandwich enzyme-linked immunosorbent assays (ELISA) utilizing colorimetry or luminometry platforms, immunobead-based assays utilizing the fluidics Luminex fluorometry platform, and a newer class of solid-phase microspot assays based on ultra-low-noise charge-coupled device (CCD) cameras represented by the Meso Scale Discovery (MSD) electrochemiluminescence platform. ELISA assays were obtained from Invitrogen BioSource (Carlsbad, CA), R&D Systems (Minneapolis, MN), and Pierce Endogen (Rockford, IL). Luminex assay kits were either purchased from Millipore Upstate (Charlottesville, VA), or were assembled in-house using reagents from R&D Systems. For comparison purposes in this paper, these two assays are labeled as Beadlyte Luminex and in-house Luminex. MSD assay kits were purchased from Meso Scale Discovery (Gaithersburg, MD). A detailed description of immunoassays is provided in Table 1.

**Study Design.** Aliquots of the five matrix variants of each NIBSC cytokine standard and the five matrices (saline, PBS, serum, and VFS at pH 4.5 or 7.2) were frozen at –70 °C and distributed to 12 participating laboratories in four countries (listed in alphabetical order at the end of the paper). The participating laboratories were instructed to follow a standardized protocol distributed to each laboratory. Each laboratory performed at least three immunoassays, i.e., one immunoassay for each NIBSC cytokine (IL-1 $\beta$  86/680, IL-1 $\beta$  86/552, and IL-6 89/548) (Table 1). The stock solutions (5000 IU/mL) were diluted by the participating laboratory in assay-specific buffer (provided by the manufacturer of each immunoassay) to generate three final NIBSC concentrations (spikes) spanning the full assay range, as follows: a spike centered around

the middle of the assay calibrator range and a low and a high spike with a log<sub>10</sub> difference below and above the middle spike (Table 1, column 7). Each of the three dilutions (spikes) was tested in quadruplicate with one exception (laboratory E performed the Quantikine assay in duplicate). The curve fit analysis was based on averaged duplicate measurements of assay calibrators provided with each assay kit and prepared in assay buffer as per manufacturer's instructions. If multiple buffer choices were provided by the manufacturer (as in the case of MSD and R&D Systems), the diluents recommended for serum/plasma samples were used. A calibrator spike (equivalent to one-fifth of top calibrator concentration for each assay) was tested in duplicate in the chemical and biological matrices described above in each assay run and used as an independent internal control for intra- and interassay variation. One-half of the value corresponding to the assay sensitivity (in pg/mL) (indicated in Table 1) was assigned to test samples that did not produce detectable readings. The raw readings and pg/mL values were submitted to the laboratories at Brigham and Women's Hospital (BWH) and SRI for data analysis. In this report each participating laboratory is identified by a letter from A to L that is not related to alphabetical order of names.

**Data Analysis and Statistics.** *Reliable Detection of Fold Difference between Cytokine Measurements.* Intra-assay reproducibility was evaluated using the coefficient of variation (% CV = 100[intra-assay standard deviation/intra-assay mean]) where an intra-assay % CV less than 20 met reliability criteria based on precision commonly accepted for immunoassays.<sup>16</sup> An additional criterion for intra-assay reproducibility was based on assessing the fold difference detectable within replicate measurements with a power of 95% ( $P < 0.05$ ) as previously described.<sup>22</sup> For this analysis, measured cytokine concentrations were converted to log<sub>10</sub> pg/mL followed by calculation of SD (standard deviation)

(22) Yen-Lieberman, B.; Brambilla, D.; Jackson, B.; Bremer, J.; Coombs, R.; Cronin, M.; Herman, S.; Katzenstein, D.; Leung, S.; Lin, H. J.; Palumbo, P.; Rasheed, S.; Todd, J.; Vahey, M.; Reichelderfer, P. *J. Clin. Microbiol.* **1996**, *34*, 2695–2701.

(21) Geshnizgani, A. M.; Onderdonk, A. B. *J. Clin. Microbiol.* **1992**, *30*, 1323–1326.

from replicate measures. The fold difference in cytokine concentration, for which the assay would have 95% power to detect, was calculated based on the intra-assay SD ( $\alpha = 0.05$ , two-tailed unpaired  $t$  test, degrees of freedom = 4, StatMate, version 2, 2006; GraphPad software). According to this analysis, an intra-assay  $\log_{10}$  SD < 0.12 allows the power to detect at least 1.5-fold difference between measured values, for all assays evaluated. Interassay differences were graphically compared using linear regressions fitted to the nominal and measured data provided by laboratory A, which performed a test with each of the four assay platforms assessed in this study—colorimetric and luminometric ELISA, electrochemiluminescence-based MSD, and fluorescence-based Luminex. Interlaboratory differences were assessed both graphically, using linear regressions fitted to the nominal and measured data provided by four laboratories running the same assay (Quantikine ELISA) with each NIBSC cytokine, and by testing for interlaboratory reproducibility where a nonsignificant effect of laboratory was used as an indicator of reproducibility ( $\alpha = 0.01$ , Kruskal–Wallis, PROC NPAR1WAY, SAS, version 9.1.3). The pooled data from all laboratories running the same assay were used to determine the smallest detectable interlaboratory fold difference and to test for the effects of assay factors on cytokine recovery.

**Assay Linearity and Accuracy.** Assay linearity was measured by linear regression, where the coefficient,  $r^2$ , was derived from the relationship between nominal values (WHO IU/mL) of the low, middle, and high spike levels of each NIBSC standard and their measured concentrations (pg/mL) in each immunoassay and matrix. An  $r^2 = 1$  indicates a straight linear relationship. The accuracy of each assay, e.g., its ability to recover the gravimetric concentration of each cytokine, was tested by comparison of measured to nominal weight/volume (w/v) concentrations of NIBSC cytokines. Because the linearity ranges of most immunoassays do not overlap (Table 1), a nominal value of 25 IU/mL NIBSC standard that represents a common overlap point of all assays regardless of assay platform or cytokine was chosen for interassay comparisons. An assay-specific measured pg/mL value for the 25 IU/mL nominal spike was determined for each matrix–cytokine variant by interpolation from the linear regression lines generated from each assay (PROC REG, SAS, version 9.1.3). The conversion from IU/mL to w/v concentrations was performed using ratios provided by NIBSC, which were 1 pg = 0.1 IU for IL-1 $\beta$  86/680 and IL-6 89/548, and 1 pg = 0.10933 IU for IL-1 $\beta$  86/552. Thus the expected w/v values corresponding to a 25 IU/mL spike were 250 pg/mL for IL-1 $\beta$  86/680 and IL-6 89/548 and 228.67 pg/mL for IL-1 $\beta$  86/552, respectively. The null hypothesis that assay-measured values were not different from these nominal concentrations was tested, i.e.,  $t = (b_1 - b_2)/SE_{(b_1, b_2)}$  at  $P < 0.01$  (two-tailed  $t$  test), where  $b_1$  and  $b_2$  were, respectively, the estimated and nominal spiked concentrations to be compared, and SE was the pooled standard error of the slope calculated using linear regression (PROC REG, SAS, version 9.1.3).

**Factors with Significant Impact on Analytic Cytokine Recovery.** The assay-specific measured values were further used to calculate % recovery from each matrix in each assay as  $100(\text{measured pg/mL})/(\text{nominal pg/mL})$ . This approach allowed comparisons between middle spike level NIBSC cytokines and assay calibrators regardless of differences in nominal values. Recovery greater than

100% indicates that the measured values for a matrix were higher than the nominal value of the spike, and a recovery less than 100% indicated that the measured values for a matrix were lower than the nominal value of the spike. The effect of biological matrices on % recovery from the middle NIBSC spike and the calibrator spike were tested using Kruskal–Wallis analysis of variance and Dunn's multiple comparisons test (PROC NPAR1WAY, SAS, version 9.1.3).

To examine the impact of experimental variables on assay measurement, a multiple regression analysis was performed where the dependent variable was % recovery, and the independent variables were number of replicates, pH of the vaginal fluid simulant, biological complexity of the matrix, and assay type. The hypothesis tested by a multivariate regression is that there is a joint linear effect of the set of independent variables on the dependent variable. Hence, the null hypothesis is that slope of all coefficients is simultaneously zero. A significant  $t$  value indicates that the slope for that independent variable (i.e., an assay variable) makes an independent contribution to account for variance in the dependent variable (i.e., % recovery). The data used in this analysis were the % recovery of the low, middle, and high NIBSC cytokines and the % recovery of the one-fifth top calibrator. A stepwise multiple regression model (PROC REG, SAS, version 9.1.3) was undertaken using a stringent  $P$  value for entry ( $P < 0.05$ ). Contrast variables were created to include the categorical variables of assay, matrix pH, and complexity in the multiple regression model for each cytokine. For all other statistical tests an  $\alpha$  of 0.01 was used, as the aim of this study was to identify robust effects in model systems that could be predicted to play a role in the measurement of cytokines for clinical trials that involve heterogenic populations and multiple laboratories.

## RESULTS

**Assay Performance and Reliable Detection of Fold Difference in Cytokine Levels.** Assay performance was assessed by intra-assay variation (% CV), fold difference reliably detectable with  $P = 0.05$  (based on  $\log_{10}$  SD), assay linearity (slope and  $r^2$  values derived from the linear regression for the low, medium, and high spiked samples), and detection accuracy (recovery of nominal gravimetric concentrations).

As shown in Table 2, most measurements (75.56–100%) across laboratories and assays met the <20% CV criterion for intra-assay reproducibility. The range in  $r^2$  values (0.79–1) indicated a strong linear relationship between measured and nominal concentrations for the spiked samples. The median intra-assay  $\log_{10}$  SDs were below 0.12, and the fold difference in cytokine concentration that would be detected with 95% power ( $P = 0.05$ ) was between 1.45 and 1.65 (Table 2). Thus, the intra-assay measurements met the set criterion for sensitive reliable comparison of cytokine levels.

The reliability of cytokine comparisons was reduced when the data were pooled from multiple laboratories ( $\log_{10}$  SD = 0.04–0.65, detectable fold difference = 1.63–16.37, Table 2) or from multiple biological matrices ( $\log_{10}$  SD = 0.07–0.78, detectable fold difference = 1.84–26.07, Table 2). The in-house Luminex assay met intra-assay but not interlaboratory reliability criteria. Thus, if data were pooled from multiple laboratories using the in-house Luminex assay, a difference much higher than 2-fold (>6.31–16.37-fold) would be required to achieve 95% power of statistical

**Table 2. Assay Linearity and Precision for the WHO Spiked Samples across All Participating Laboratories<sup>a</sup>**

cytokine/assay	log <sub>10</sub> SD			detectable fold difference <sup>b</sup>			CV < 20% (%) <sup>c</sup>	linearity <i>r</i> <sup>2</sup>
	intra-assay	interlaboratory	intermatrix	intra-assay	interlaboratory	intermatrix		
					IL-1β			
BioSource	0.04 (0.05)	0.07 (0.19)	0.15 (0.20)	1.65	1.84	2.46	75.56	0.99 (0)
Endogen	0.03 (0.03)	NA <sup>d</sup>	0.07 (0.08)	1.61	NA	1.84	96.67	0.98 (0.02)
Quantikine	0.01 (0.04)	0.13 (0.27)	0.14 (0.26)	1.5	2.35	2.39	86.67	1.00 (0.00)
QuantiGlo	0.01 (0.02)	0.08 (1.24)	0.08 (1.17)	1.45	1.9	1.89	98.33	1.00 (0.01)
Beadlyte Luminex	0.02 (0.02)	0.07 (0.29)	0.37 (0.28)	1.54	1.84	5.76	90	0.99 (0.01)
in-house Luminex	0.03 (0.03)	0.65 (0.08)	0.62 (0.07)	1.59	16.37	14.68	90	0.98 (0.03)
MSD Small Spot	0.03 (0.03)	0.05 (0.04)	0.07 (0.05)	1.56	1.68	1.88	96.67	0.99 (0.02)
					IL-6			
BioSource	0.04 (0.05)	0.30 (0.31)	0.51 (0.18)	1.64	4.4	9.49	76.67	0.99 (0.01)
Endogen	0.03 (0.05)	0.48 (0.58)	0.66 (0.43)	1.59	8.48	16.82	76.67	0.98 (0.04)
Quantikine	0.02 (0.02)	0.07 (0.05)	0.21 (0.01)	1.54	1.86	3.12	91.67	1.00 (0.00)
QuantiGlo	0.03 (0.03)	0.08 (0.1)	0.28 (0.04)	1.56	1.89	4.06	93.33	0.99 (0.01)
Beadlyte Luminex	0.02 (0.03)	0.41 (0.53)	0.78 (0.33)	1.54	6.5	26.07	96.67	0.79 (0.51)
in-house Luminex	0.02 (0.03)	0.40 (0.16)	0.48 (0.18)	1.54	6.31	8.62	90	1.00 (0.01)
MSD Small Spot	0.03 (0.02)	0.04 (0.01)	0.21 (0.06)	1.56	1.63	3.1	100	0.99 (0.01)

<sup>a</sup> Comparisons within (i) replicate samples (intra-assay), (ii) samples pooled from multiple laboratories (interlab), and (iii) samples pooled for the five different biological matrices (intermatrix). SD (log<sub>10</sub> pg/mL) and linearity coefficient *r*<sup>2</sup> are expressed as median (interquartile range).

<sup>b</sup> The fold difference in cytokine concentration for which the assay would have 95% power to detect with *P* < 0.05. <sup>c</sup> The proportion (%) of measurements with an intra-assay CV < 20%. <sup>d</sup> NA: not applicable (only one laboratory ran this assay).

significance (Table 2). It was predicted that the commercial assays would be able to detect 1.68–8.48-fold differences in cytokine concentration if measurements were made by multiple laboratories, with (electro)chemiluminescence-based assays (MSD and QuantiGlo) being consistently most discriminative, detecting <2-fold differences (Table 2).

Assay accuracy varied considerably depending on assay and cytokine tested (Figure 1). The pg/mL concentrations returned by the linear regression model for a 25 IU/mL equivalent of NIBSC spike were either higher or lower than the nominal w/v spike value (*P* < 0.01, *t* test) in most IL-1β runs [92% of 86/552 (Figure 1A) and 86% of 86/680 assay runs (Figure 1B)] and in all IL-6 runs (Figure 1C). The Quantikine, QuantiGlo, and MSD immunoassay extracted IL-1β 86/552 levels above the nominal values across all matrices (*P* < 0.01, Figure 1A). In contrast, the BioSource and Luminex Upstate Beadlyte assays were found to underestimate nominal IL-1β 86/552 levels across all matrices (Figure 1A). Neither of the two laboratory results for in-house Luminex IL-1β 86/552 measurements were close to the expected cytokine level; laboratory G results were extremely low and laboratory I results were extremely high (Figure 1A). Endogen-measured values for IL-1β 86/552 were higher in PBS pH 7.2 but lower in all other matrices as compared to nominal values (Figure 1A). All IL-1β 86/680 measurements were lower than nominal values with the notable exceptions of in-house Luminex (laboratory D) and Beadlyte Luminex (laboratory A) assays (Figure 1B). Most measured levels of IL-6 89/548 were below the expected nominal concentration (*P* < 0.01, *t* test, Figure 1C).

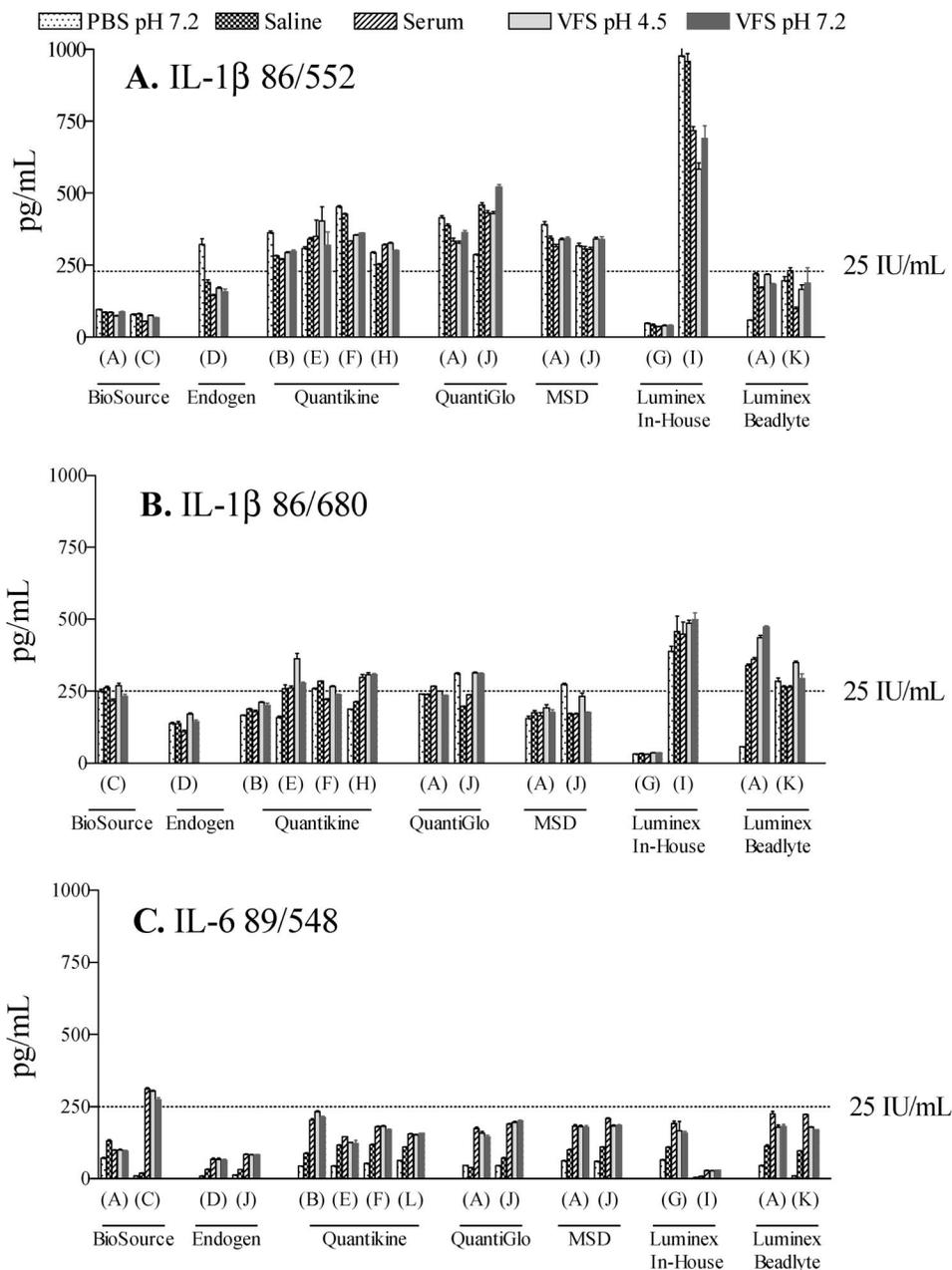
**Interlaboratory and Interassay Reproducibility.** There was a general consistency in the results obtained from four laboratories using the same Quantikine assays with greatest interlaboratory differences found in the lower end of the assay detection range (Figure 2A). A nonsignificant (*P* > 0.01, Kruskal–Wallis) effect of laboratory on cytokine recovery was used as an indicator of interlaboratory reproducibility when at least two laboratories ran the same assay (Table 3). Reproducible interlaboratory measurement of all cytokines was found only for the MSD assays (Table

3). In the case of the Quantikine assays, although the linear regression model showed interlaboratory consistency (Figure 2A), the high intra-assay reproducibility in each laboratory (% CV < 20 and log<sub>10</sub> SD < 0.12, Table 2) contributed to a statistically significant interlaboratory difference (*P* < 0.01, Kruskal–Wallis).

Each assay type was assessed for significant intercytokine differences in spike recovery (Kruskal–Wallis, Mann–Whitney, and Dunn's test, *P* < 0.01). The different performance of each assay by cytokine preparation (midrange spikes of NIBSC and assay calibrator) is illustrated in Figure 3, and probability values for all interlaboratory and intercytokine differences in % recovery measurements are summarized in Table 3. As expected, the median (interquartile range) % recovery for all assay calibrators was optimal for both IL-1β [108.00 (53.44)%, Figure 3C] and IL-6 [95.02 (26.53)%, Figure 4B]. However, the recovery of the NIBSC spikes differed (Mann–Whitney *P* < 0.001). The recovery of IL-1β 86/680 was lower [93.76 (53.22)%, Figure 3B] than the IL-1β 86/552 recovery [125.25 (90.11)%, Figure 3A]. The recovery of IL-6 89/548 was below the nominal concentrations in all IL-6 assays tested [41.70 (48.78)%, Figure 4A].

Interassay reproducibility was assessed when one laboratory tested multiple assays representing the four major cytokine detection platforms, e.g., absorbance ELISA (BioSource), chemiluminescence ELISA (QuantiGlo), fluorimetric immunobead assay (Beadlyte Luminex), and electrochemiluminescence assay (MSD). For all IL-6 89/548 and for IL-1β 86/680 measurements (Figure 2B) interassay disparities were greater in the low-concentration range of the assays coinciding with the lowest standard spikes (0.25–5 IU/mL). For IL-1β 86/552 measurements (Figure 2B, left panel) interassay disparities were seen over the entire cytokine detection range.

**Impact of Matrix on Interlaboratory Variation of Cytokine Measurements.** The effect of matrix on assay measurements was tested using the % recovery of the middle NIBSC spikes (Figure 3, parts A and B, and Figure 4A) and the assay calibrator spikes (Figure 3C and Figure 4B). Percent recovery was calculated in each assay, and results were pooled when multiple laboratories

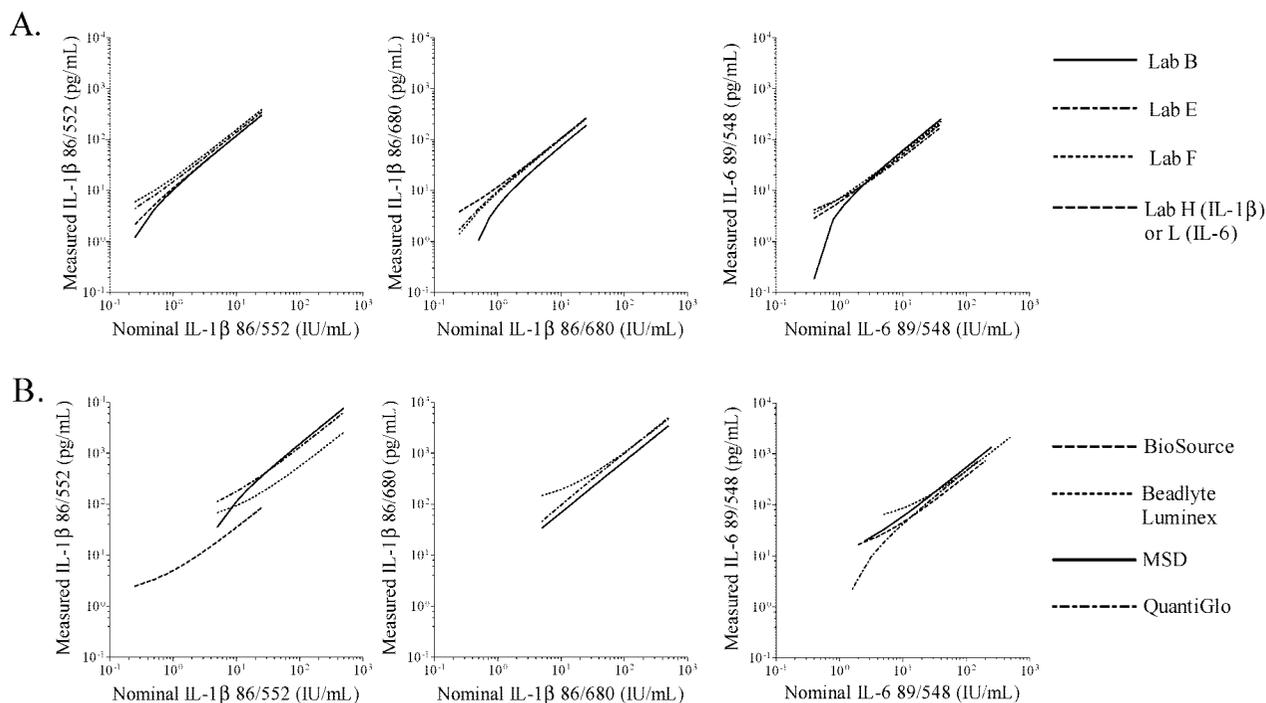


**Figure 1.** Assay accuracy for predicted measurements of 25 IU spikes of IL-1 $\beta$  86/552 (A), IL-1 $\beta$  86/680 (B), and IL-6 89/548 (C) across laboratories and kits. Values were derived from a linear regression curve with the standard error of the mean slope represented by error bars. The horizontal dashed line indicates nominal weight/volume concentrations of 25 IU/mL, which is 228.67 pg/mL for IL-1 $\beta$  86/552 and 250 pg/mL for IL-1 $\beta$  86/680 and IL-6 89/548.

used the same assay kit. As can be seen from Table 2, when pooling results from multiple laboratories a 1.6- to 8.5-fold difference would be required for any effect to be determined significant with the exception of the in-house Luminex assay where reliable interlaboratory detection required more than 16-fold difference. Therefore, only matrix effects consistent across laboratories could be shown by this type of analysis.

Matrix effects on IL-1 $\beta$  recovery varied depending on cytokine variant and immunoassay variant. No effects of matrix on recovery of both IL-1 $\beta$  86/552 and 86/680 were found by Biosource, Quantikine, MSD, or in-house Luminex assays. The Beadlyte Luminex IL-1 $\beta$  assay was most sensitive to matrix, showing lowest recovery in PBS (Figure 3, parts A and B,  $P < 0.01$ , Dunn's test). The QuantiGlo assay showed matrix effects only on the recovery of 86/

680 being lowest in saline (Figure 3B,  $P < 0.01$ , Dunn's test). The Endogen assay also showed matrix-sensitive recovery of NIBSC cytokines, but the significance of these findings could not be confirmed with pooled data since only one laboratory ran this assay. The recovery of IL-1 $\beta$  calibrator spikes was significantly affected by matrix only in the in-house Luminex assay, being lowest in serum and comparable between PBS, saline, and VFS, and in the Quantikine assay, being lower in PBS than in VFS at pH 7.2 ( $P < 0.01$ , Dunn's Test, Figure 3C). Unlike IL-1 $\beta$ , NIBSC IL-6 recovery was consistently affected by matrix type in most (six of seven) assays tested, with markedly lower recovery in PBS and saline as compared to VFS and serum (Figure 4A,  $P < 0.01$ , Dunn's test). In contrast, matrix had no significant effect on recovery of IL-6 assay calibrator spikes regardless of assay platform (Figure 4B).



**Figure 2.** Interlaboratory (A) and interassay (B) comparison of results obtained for IL-1 $\beta$  86/552, IL-1 $\beta$  86/680, and IL-6 89/548 cytokines from four laboratories using the same Quantikine ELISA (A) and from laboratory A using four assay types (B). Each line represents linear regression analysis between nominal values (WHO IU/mL) of the low, middle, and high spike of each NIBSC standard and their measured concentrations (pg/mL) in one assay run.

**Table 3. Probability of Interlaboratory and Intercytokine Differences in Recovery of Cytokines across Assays**

assay	interlaboratory			IL-6		intercytokine	
	86/552	86/680	assay calibrators	89/548	assay calibrators	IL-1 $\beta$	IL-6
BioSource	<0.0001 <sup>a</sup>	NA	<0.0001	0.280	0.007	<0.0001	0.005
Endogen	NA <sup>b</sup>	NA	NA	<0.0001	0.023	<0.0001	<0.0001
Quantikine	<0.0001	<0.0001	0.018	0.054	<0.0001	<0.0001	<0.0001
QuantiGlo	<0.0001	0.003	<0.0001	0.465	0.002	<0.0001	<0.0001
Beadlyte Luminex	0.552	0.001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
in-house Luminex	<0.0001	<0.0001	0.014	<0.0001	0.199	0.054	<0.0001
MSD Small Spot	0.094	0.049	0.028	0.814	0.049	<0.0001	<0.0001

<sup>a</sup>  $P < 0.01$  indicates that the % recovery was different between laboratories or between cytokines (Mann–Whitney, Kruskal–Wallis). <sup>b</sup> NA = not applicable; the analysis was not applicable when <2 laboratories provided data from the same kit.

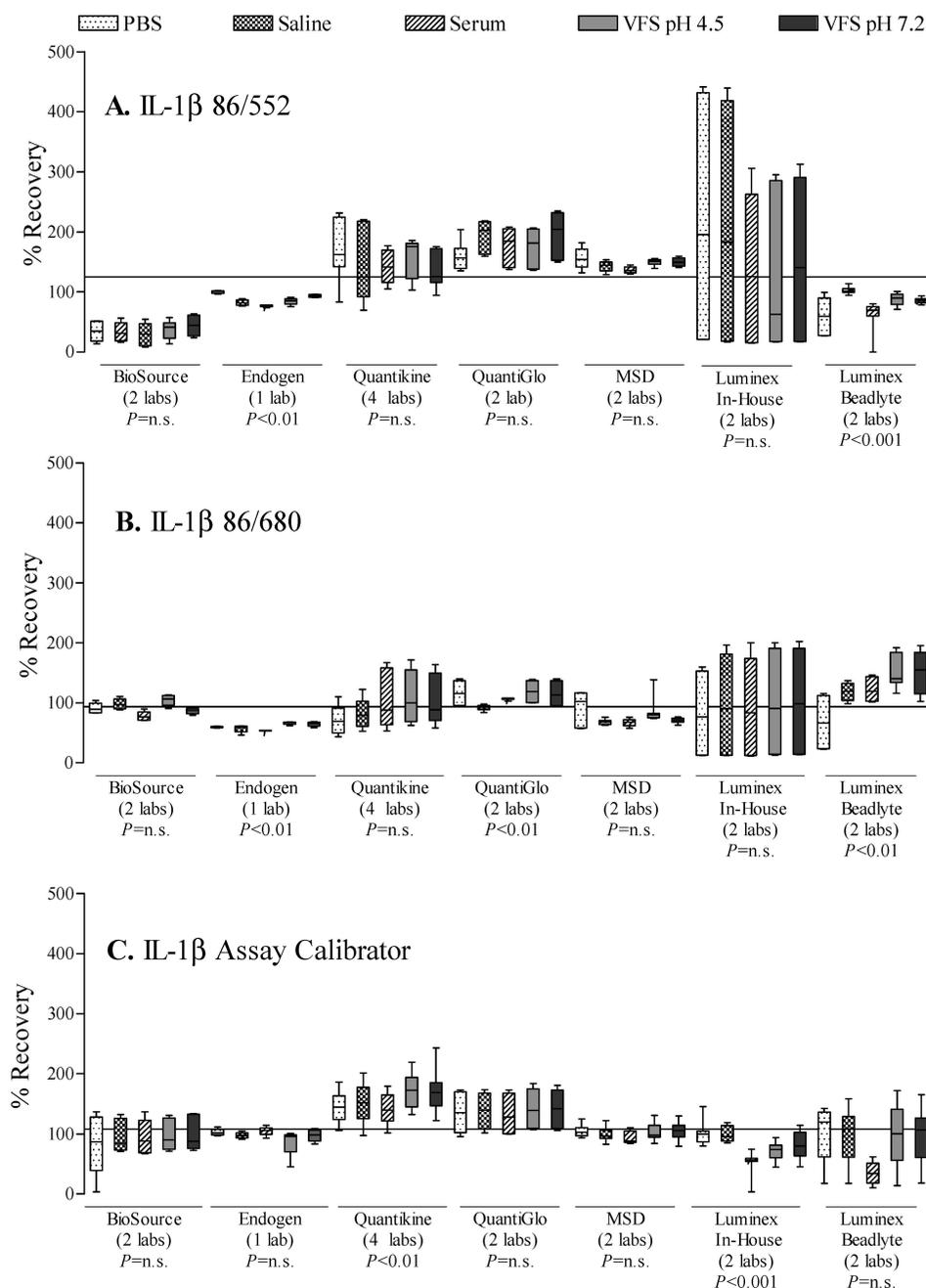
### Relative Importance of Biological and Technical Variables.

A multiple regression model was used to evaluate the relative importance of major parameters tested in this study such as pH of the vaginal fluid simulant, the relative complexity of the biological matrix, the assay type, and technical performance parameters reflecting interassay and inter-replicate (intra-assay) reproducibility of cytokine recovery. Multiple regression analysis was used to determine which, if any, of these factors tested could explain the variation found in cytokine recovery. The final multiple regression models were found to significantly explain recovery of the cytokines tested ( $r^2 = 0.10$ – $0.45$ ,  $P < 0.001$ , Table 4). The nonsignificant effect of replicate indicated the consistency of intra-assay results and is further evidence of assay reliability across all cytokine measurements reported in this study (Table 4). The variation in recovery of IL-1 $\beta$  86/552 and IL-6 89/548 cytokines was influenced by the assay type for all cytokines tested (Table 4) with the considerable

interassay differences illustrated in Figure 3A–C. Variation in recovery of IL-1 $\beta$  86/680, IL-1 $\beta$  kit calibrator, and IL-6 89/548 was explained by the complexity of the biological matrices [(PBS and saline) < VFS < serum, Table 4]. There was no effect of pH in the multiple regression models for any of the cytokines tested when VFS at pH 7.2 was compared to VFS buffered at pH 4.5 in the context of the vaginal fluid matrix (Table 4).

### DISCUSSION

This is the first study to assess the relative importance of technical and biological variables affecting multicenter measurement of proinflammatory cytokines in the context of the vaginal mucosal fluid matrix. The study compared complex biological fluids (e.g., human serum or surrogate vaginal fluid) versus simple protein-free matrix (e.g., buffered or nonbuffered saline both widely used to sample vaginal cytokines by lavage).

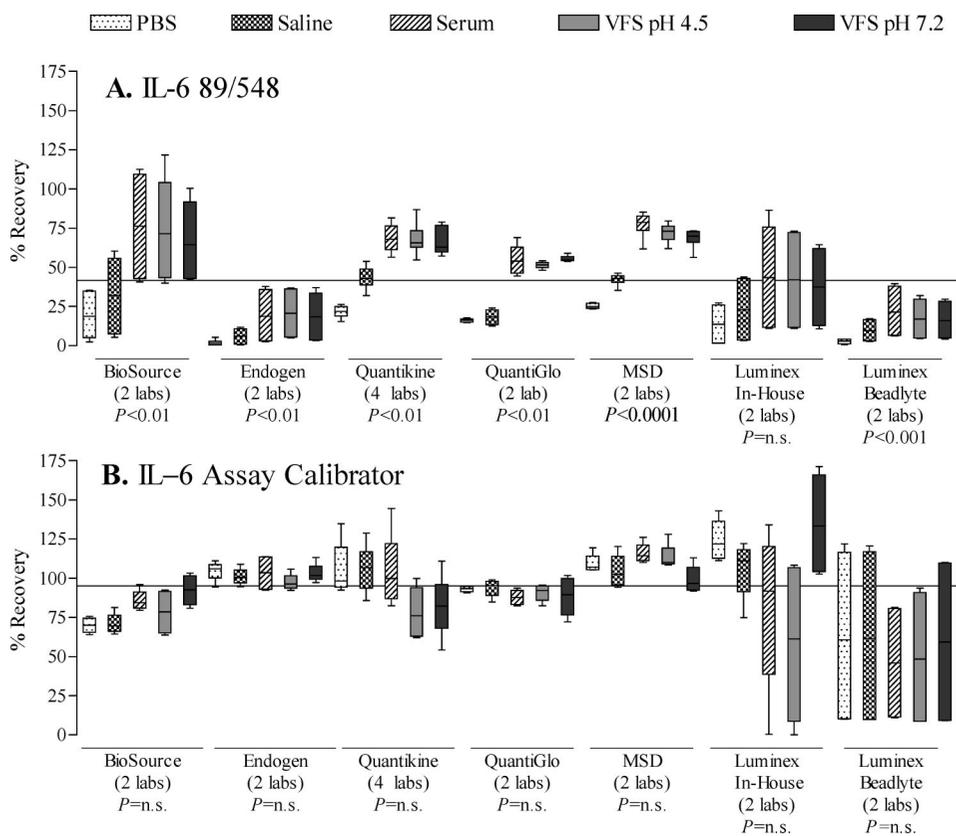


**Figure 3.** Effect of matrix on percent recovery of the middle spike of IL-1 $\beta$  NIBSC cytokines (A and B) and assay-specific calibrators (C). Box plots represent medians and interquartile range of each assay. The horizontal solid line represents the median from all assays. The *P* values indicate probability of matrix effects for each assay type where n.s. is a nonsignificant ( $P > 0.01$ ) effect of matrices (Kruskal–Wallis ANOVA, Dunn’s Post Hoc tests).

It was determined that matrix complexity unequally affected cytokine measurement, depending on individual cytokine preparation as well as detection platform or assay. In the case of IL-1 $\beta$ , the matrix effect varied by detection platform. In contrast, matrix complexity affected IL-6 measurements regardless of assay or detection platform, with saline and PBS consistently returning lower IL-6 values. In general, the analytic recovery of spiked cytokine assay calibrators was less sensitive to matrix variation than that of recombinant biologically active NIBSC cytokine standards, which should more closely mimic native endogenous cytokines. These results suggest that biologically active or endogenous cytokines rather than assay calibrators should be spiked into relevant biological fluids to

estimate possible effects of fluid matrix on cytokine measurements and to establish a standard procedure for future interlaboratory measurement of cytokines as biomarkers in clinical trials. Our results also suggest that the media used for vaginal washing or dilution of samples may play a critical role in the recovery of cytokine concentrations, and therefore, uniform vaginal sampling should be applied across clinical trials. The possibility for underestimating concentrations of some cytokines should be considered, and errors should be estimated by assessing each cytokine recovery in the medium of choice.

Differences in measured cytokines across assay platforms could not be explained solely by different nominal values of assay



**Figure 4.** Effect of matrix on percent recovery of the middle spike of the IL-6 NIBSC cytokine (A) and assay-specific calibrators (B). Box plots represent medians and interquartile range of each assay. The horizontal solid line represents the median from all assays. The *P* values indicate probability of matrix effects for each assay type where n.s. is a nonsignificant ( $P > 0.01$ ) effect of matrices (Kruskal–Wallis ANOVA, Dunn’s Post Hoc tests).

**Table 4. Multiple Regression Analysis of Relative Effects of Assay Factors on Cytokine Recovery**

variables	prob( <i>t</i> ) <sup>a</sup>				
	IL-1 $\beta$ 86/552	IL-1 $\beta$ 86/680	IL-1 $\beta$ (kit)	IL-6 89/548	IL-6 (kit)
	Biological Matrix				
matrix pH <sup>b</sup>	0.756	0.432	0.796	0.836	0.103
matrix complexity <sup>c</sup>	0.127	<0.0001 <sup>d</sup>	<0.0001 <sup>d</sup>	<0.0001 <sup>d</sup>	0.0675
	Technical Parameters				
assay platform	<0.0001 <sup>d</sup>	<0.0001 <sup>d</sup>	<0.0001 <sup>d</sup>	<0.0001 <sup>d</sup>	<0.0001 <sup>d</sup>
replicate	0.13	0.995	0.695	0.829	0.309
final model adjusted $r^2$	0.1	0.41	0.45	0.13	0.23
prob( <i>F</i> ) <sup>e</sup>	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

<sup>a</sup> Probability of the *t* test for the null hypothesis that the regression coefficient of each variable = 0. <sup>b</sup> For the vaginal fluid simulant only where pH 4.5 was compared to pH 7.2. <sup>c</sup> Where matrix complexity was determined to increase from PBS and saline < vaginal fluid simulant < serum. <sup>d</sup> Variables included in final model, where prob(*t*) < 0.05. <sup>e</sup> Probability of the *F* test of the significance of *R*, where a prob(*F*) < 0.05 indicates that the model is significantly better than would be expected by chance.

calibrators provided by each manufacturer. Similarly to our study, previous studies of human blood have reported significant inter-assay variations.<sup>23,24</sup> In those studies, various assays generated similar patterns of IL-6 blood plasma levels; however, absolute concentrations differed between ELISA kits produced by different manufacturers.<sup>23</sup> Neither in those nor in our study did conversion of cytokine concentrations from pg/mL to NIBSC IU/mL eliminate interassay differences. Our study confirms the need for use of

internal calibrator controls along with matrix-specific conversions to allow interassay and interlaboratory meta-analysis.

Antibody-based detection of cytokines in complex biological fluids can be complicated by denaturation of epitopes by fluid components as well as by the physiological variance of cytokine polymerization, glycosylation, soluble receptor binding, or degradation that may not be equally recognized by the detection antibodies used in different assays. Recombinant proteins used for calibration and generation of antibodies for various immunoassays may have various degrees of glycosylation and other posttranscriptional modifications depending on bacterial, yeast, or mammalian rDNA expression systems<sup>19</sup> which may explain

(23) Khan, S. S.; Smith, M. S.; Reda, R.; Suffredini, A. F., Jr. *Cytometry, Part B* **2004**, *61B*, 35–39.

(24) Aziz, N.; Nishanian, P.; Mitsuyasu, R.; Detels, R.; Fahey, J. L. *Clin. Diagn. Lab. Immunol.* **1999**, *6*, 89–95.

some differences found between assay calibrators and NIBSC standards. For example, IL-1 $\beta$  is initially translated as a biologically inactive precursor molecule that is subsequently processed into a mature protein associated with IL-1 bioactivity.<sup>25</sup> Both may be available in biological fluids. In the case of vaginal fluid, the presence of mucus can introduce another layer of variability since it can bind some cytokines and thus affect their recognition by capturing antibodies or may nonspecifically attach to immunobeads or assay plates thus affecting downstream assay events.

In this study, the two biologically active international reference preparations of IL-1 $\beta$  differed in terms of sensitivity to matrix complexity. IL-1 $\beta$  calibrators were also sensitive to matrix complexity but only in two out of seven assays. In contrast, the recovery of the NIBSC, but not the calibrator IL-6 variants, differed significantly in terms of their sensitivity to matrix types, regardless of assay type. Although the recovery of NIBSC IL-6 was lower in PBS and saline as compared to serum and the defined vaginal simulant, assay calibrators were equally recoverable from all matrices across all assays. In analogy to results obtained with natural IL-6 present in blood, different assays provided different recovery of IL-6 89/548 in the vaginal fluid simulant. This could be dependent upon the fact that recombinant IL-6 89/548 is glycosylated similarly to the natural cytokine<sup>19</sup> and its detection would depend on the level of recognition of protein or carbohydrate epitopes by the antibody pairs of each assay.

The interlaboratory differences observed in this study were not due to sample handling or technical skills since variation of sample handling was minimized by following a standardized operational procedure developed for each assay listed in Table 1 and since good intralaboratory and intra-assay reproducibility was achieved within each matrix type and within each assay platform. The small intra-assay variation was confirmed using multiple indexes (i.e., % CV, log<sub>10</sub> SD, and estimation of detectable fold differences).

The interlaboratory variation was most significantly associated with assay type and platform as demonstrated by the bar sizes in Figures 3 and 4 and by the fold differences reliably detectable in each assay (Table 2). Thus, the sources of interlaboratory variation appear inherent in cytokine–platform or matrix–platform interactions. The complex matrices (serum and vaginal fluid simulant) were consistently associated with higher interlaboratory variation in all assays except the QuantiGlo and MSD (Figures 2 and 3), which detect a luminescent signal at 620 nm thus eliminating the problem with color quenching by matrix components that may occur in regular absorbance ELISA or fluorimetric Luminex assays. In addition, MSD uses labels that emit light when electrochemically stimulated where the stimulation event (electricity) is decoupled from the signal (light) thus minimizing background. Nonuniform matrix interactions with immunobeads in the Luminex platform may add an additional layer of variation, which is complicated by the fluidics nature and dependence on efficient washing of the beads between immunoassay steps. Interlaboratory variability has been found previously for noncommercial assays for quantification of HIV-1 RNA in seminal plasma where commercial assays were found to demonstrate highly reproducible (intra-assay log<sub>10</sub> SD 0.11–0.32) and noncommercial (in-house)

assays less reproducible results (intra-assay log<sub>10</sub> SD 0.12–0.75).<sup>26</sup> A multicenter study suggested that the lack of reproducibility found for noncommercial assay measurements could be due to the fact that reagents for commercial kits are produced in “large and well-characterized” lots.<sup>26</sup> Thus, variation of immunobead and antibody lots obtained from different manufacturers that have not been as rigorously subjected to quality control and cross-calibration as commercially assembled kits may be an additional explanation for high interlaboratory variation in in-house assembled Luminex assays.

In our study the commercial assay kits as well as the assays compiled in-house from different commercially available reagents showed high inter-replicate reproducibility allowing reliable detection of 1.5-fold intra-assay difference in concentration levels regardless of cytokine type. However, interlaboratory variation significantly decreased the level of confidence in detectable difference. Our data indicate that IL-6 levels, for example, would have to vary by at least 6-fold in order for multiple laboratories using the Luminex platform to detect reliable differences, provided a uniform biological fluid (matrix) was used. With matrix variation, the threshold for reliable detection of IL-6 differences increased to >26-fold in Luminex Beadlyte and >8-fold in several other assays. Matrix variation affected less the statistical power to detect differences in IL-1 $\beta$  levels. These data emphasize the importance of uniform biological sampling techniques and careful analysis of matrix interferences. Because of high interlaboratory reproducibility regardless of specific cytokine tested, the chemiluminescence-based assays (QuantiGlo and MSD) appear more suitable for multicenter study comparisons. Our study suggests that if uniform platform use is not feasible for financial or technical reasons, each multicenter study should appoint a reference laboratory to perform interassay comparisons to determine the confidence of fold-difference detection in each specific biological fluid and for each cytokine assessed.

In summary, the cytokine measurement methods evaluated here were found to be reproducible for laboratories using commercial assay platforms; however, the measured values of identical samples varied across laboratories and assay platforms. Our data indicate that variability introduced by the sampling fluid and assay type or detection platform may confound interlaboratory or interstudy comparisons. The discrepancy between matrix recovery of assay calibrators and biologically active cytokine variants, which was especially obvious for IL-6 assays, may lead to a significant under- or overestimation of cytokine concentrations in complex biological fluids such as vaginal fluids and blood. It is important to understand that these effects may occur in every day analytical practice not limited to the case of cytokine immunoassays. To avoid their impact, any laboratory should prepare their own normality ranges in the relevant types of clinical samples. Our study emphasizes the need for assay choices based on criteria for reliable interlaboratory comparisons, which goes beyond the scope of vaginal cytokine measurement and microbicide research. Along with parameters commonly provided by manufacturers (e.g., assay sensitivity, lower limit of detection, and assay precision determined by reproducibility of replicate measurements), immunoassay utility should be evaluated also by matrix-specific fold

(26) Fiscus, S. A.; Brambilla, D.; Coombs, R. W.; Yen-Lieberman, B.; Bremer, J.; Kovacs, A.; Rasheed, S.; Vahey, M.; Schutzbank, T.; Reichelderfer, P. S. *J. Clin. Microbiol.* **2000**, *38*, 2348–2353.

(25) Arend, W. P. *Cytokine Growth Factor Rev.* **2002**, *13*, 323–340.

change in cytokine concentration reliably detectable by multiple comparisons methods. Of all the assays evaluated in this study, the (electro)chemiluminescent platform appeared to be the most reliable platform for detection of a  $\leq 1.5$ -fold change in IL-6 and IL-1 $\beta$  concentrations. However, future validation studies are needed to determine the biological significance of the magnitude of change for each biomarker candidate. It is recommended that prior to adoption of a standardized methodology, cytokine measurements from clinical trials be reported not only in terms of absolute ranges but also in terms of fold differences between samples with defined clinical characteristics.

#### ACKNOWLEDGMENT

This work was part of the Microbicide Quality Assurance Program (MQAP) supported by the following contracts from the U.S. National Institutes of Health: NICHD N01-HD-3-3350 (September 2003–January 2007) and NIAID N01-AI-33350 (February 2007–present). The views of the authors do not necessarily reflect those of the funding agencies. The authors have no conflicts of interest or financial interests regarding this work presented. Investigators using cytokine assays in their respective microbicide-related research were invited to participate in the design of the studies described in this paper. Technical staff participated in regular conference calls and discussions of the standardized procedures as the studies were implemented and data were acquired. The authors thank Dr. A. Onderdonk from the Channing Laboratory, Brigham and Women's Hospital and Harvard Medical School for the preparation of the vaginal fluid simulating media. The authors also thank the following NIH Program Staff for their scientific expertise and guidance in relation to the MQAP studies: Drs. Kailash Gupta (NIAID), Patricia Reichelderfer (NICHD), Fulvia Veronese (Office for AIDS Research to NIAID), Jonathan Glock (NIAID), and James Turpin (NIAID). The author list includes the Principal Investigators from the participating laboratories and their respective technical representatives. Since participation in the MQAP is voluntary, it should be noted that the cytokine studies described in this paper were facilitated by the infrastructure, resources, and microbicide-related grant support available at each of the following participating laboratories

as listed by alphabetical order (this order does not correspond to randomly assigned ID letter in Table 1): (1) Brigham and Women's Hospital, Laboratory of Genital Tract Biology, Department of Obstetrics, Gynecology and Reproductive Biology, Harvard Medical School, 221 Longwood Avenue RF468, Boston, MA 02115 (SRI subcontract COA no. 06 to NIAID N01-HD-3-3350; P01HD041760, Project 3; and CONRAD MSA-02-304 and -05-427 to R.N.F.). (2) Brown University, Miriam Hospital Immunology Research Laboratory, 164 Summit Avenue, Providence, RI 02912 (P30 AI-42853, NIAID to K.H.M.). (3) Centers for Disease Control and Prevention, Division of HIV/AIDS Prevention, Laboratory Branch, Atlanta, GA 30333. (4) Centre de Recherche des Cordeliers, Unité INSERM Internationale U743, Equipe "Immunité et Biothérapie Muqueuse", Paris, France 75270. (5) CONRAD, Microbicide and Contraception Research Laboratory, Department of Obstetrics and Gynecology, Eastern Virginia Medical School, 601 Colley Avenue, Norfolk, VA 23507 (CONRAD intramural funds from USAID and The Bill and Melinda Gates Foundation to G.F.D.). (6) Mount Sinai School of Medicine, New York, NY 10029. (7) National Institute of Child Health and Human Development, National Institutes of Health, 10 Center Drive 10/9D58 Bethesda, MD 20892 (Intramural Program). (8) Rush University Medical Center, Department of Immunology/Microbiology, 735 W. Harrison Street, Room 616 Cohn Chicago, Illinois 60612. (9) San Raffaele Scientific Institute, AIDS Immunopathogenesis Unit, Via Olgettina 58, Milan, Italy 20132 (Europrise, Network of Excellence, EC, to G.P.). (10) Southern Research Institute, Drug Development Division, Infectious Disease Research Department, Frederick, MD 21701 (contract NIAID N01-AI-33350 to J.E.C.). (11) St. George's, University of London, London, U.K. SW17 0RE. (12) University of Pittsburgh, School of Medicine, Department of Obstetrics, Gynecology, and Reproductive Sciences, Magee-Womens Research Institute, 300 Halket Street, Pittsburgh, PA 15213 (NIH 5U01AI068633 to S.H.)

Received for review December 28, 2007. Accepted April 7, 2008.

AC702628Q