



Multiplex assay reliability and long-term intra-individual variation of serologic inflammatory biomarkers



Heather S. McKay^{a,*}, Joseph B. Margolick^b, Otoniel Martínez-Maza^{c,d,e,f}, Joseph Lopez^b, John Phair^g, Giovanna Rappocciolo^h, Thomas N. Dennyⁱ, Larry I. Magpantay^c, Lisa P. Jacobson^a, Jay H. Bream^b

^a Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

^b Department of Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

^c Departments of Obstetrics & Gynecology, UCLA, Los Angeles, CA, USA

^d Department of Microbiology, Immunology & Molecular Genetics, UCLA, Los Angeles, CA, USA

^e David Geffen School of Medicine at UCLA, Los Angeles, CA, USA

^f Department of Epidemiology, UCLA Fielding School of Public Health, University of California, Los Angeles, CA, USA

^g Northwestern University Feinberg School of Medicine, Chicago, IL, USA

^h University of Pittsburgh, School of Public Health, Pittsburgh, PA, USA

ⁱ Duke Human Vaccine Institute, Department of Medicine, Duke University School of Medicine, Duke University, Durham, NC, USA

ARTICLE INFO

Article history:

Received 18 April 2016

Received in revised form 16 September 2016

Accepted 21 September 2016

Keywords:

Inflammatory biomarkers
Multiplex assay reliability
Intraclass correlation
Reliability

ABSTRACT

Background: Circulating cytokines, chemokines, and soluble cytokine receptors can serve as biomarkers of inflammation and immune dysregulation. Good reliability of multiplex platforms, which allow for simultaneous, comprehensive biomarker assessment, is critical for their utility in epidemiologic studies. We examined the reliability of the Meso-Scale Discovery (MSD) platform to simultaneously quantitate 15 cytokines and chemokines and the Luminex platform (R&D Systems) to quantitate 5 soluble receptors and 2 chemokines and cytokines and evaluated long-term within-person correlation of these biomarkers. **Methods:** The detectability and reliability of these assay systems were assessed using the same external controls across plates and archived sera from 250 HIV⁻ men in the Multicenter AIDS Cohort Study. Using up to four visits per person from 1984 to 2009, age-adjusted intraclass correlation coefficients (ICC) of biomarkers with >80% detectability (CCL11, CXCL8, CXCL10, CCL2, CCL4, CCL13, CCL17, CXCL13, IL-10, IL-12p70, IL-6, TNF- α , BAFF, sCD14, sCD27, sgp130, sIL-2R α , and sTNF-R2) were obtained using linear mixed models.

Results: Most biomarkers were detectable in 80% of control samples; IFN- γ , GM-CSF, and IL-2 were undetectable in >20% of samples. Among the HIV-uninfected men, most biomarkers showed fair to strong within-person correlation (ICC > 0.40) up to 15 years. The ICC for CXCL8 was good in the short term but decreased with increasing time between visits, becoming lower (ICC < 0.40) after 8 years.

Conclusions: These multiplexed assays showed acceptable reliability for use in epidemiologic research, despite some technical variability and limitations in cytokine quantitation. Most biomarkers displayed moderate-to-excellent intra-individual variability over the long term, suggesting their utility in prospective studies investigating etiologic associations with diverse chronic conditions.

© 2016 Elsevier Ltd. All rights reserved.

Abbreviations: BAFF, B-cell activating factor; BMI, body mass index; CI, confidence interval; CCL11, C-C motif ligand 11; CV, coefficient of variation; CXCL8, C-X-C motif ligand 8; GM-CSF, granulocytemacrophage colony-stimulating factor; HCV, hepatitis C virus; HIV, human immunodeficiency virus; ICC, intraclass correlation coefficient; IFN, interferon; IL, interleukin; IQR, interquartile range; LLOD, lower limit of detection; sIL-2R α , soluble IL-2 receptor alpha; MACS, Multicenter AIDS Cohort Study; sCD14, soluble cluster of differentiation 14; sgp130, soluble glycoprotein 130; TNF- α , tumor necrosis factor-alpha; sTNF-R2, soluble tumor necrosis factor-receptor 2.

* Corresponding author at: Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, E7006, Baltimore, MD 21205, USA.

E-mail address: hmckay4@jhu.edu (H.S. McKay).

1. Introduction

Inflammation and immune activation are associated with chronic health conditions, including malignancies [1,2], cardiovascular disease [3], kidney and liver dysfunction [4,5], and AIDS [6]. Chronic inflammation is characterized by persistent activation of the innate and adaptive immune systems. Circulating serologic biomarkers, such as chemokines, cytokines, soluble cytokine receptors, and acute phase proteins are commonly used in epidemiologic

studies to understand underlying inflammatory mechanisms associated with disease risk and progression. Most prior work has focused on single or small numbers of biomarkers, such as TNF- α , IL-6, and C-reactive protein (CRP). While informative, this approach offers an incomplete picture of the complex inflammatory response comprised of multiple interacting circulating mediators. Multiplex technologies permit concurrent testing of large numbers of analytes using minimal sample volume, allowing for rapid, cost-effective quantitation of a more comprehensive panel of biomarkers. While it is important to capitalize on these emerging methods, formal assessment of assay reliability is warranted. Prior studies of multiplex reliability have been restricted by small sample sizes or a limited number of biomarkers [7,8].

Furthermore, studies often use a single blood sample to quantify circulating concentrations of inflammatory biomarkers and characterize a participant's risk, assuming that a single measurement represents the individual's long-term state of inflammation. High within-person variability can result in measurement error, biasing risk estimates towards the null and attenuating the likelihood of identifying valid exposure-disease associations. The intraclass correlation coefficient (ICC) assesses inter-individual (between people) variability relative to total variability (between and within individuals) and provides a measure of the extent to which a biomarker tracks within a person over time. Low ICCs necessitate multiple measurements over time to more accurately capture the inflammatory state, while low variability within a person, or high tracking, improves the precision of estimates from longitudinal studies [9]. Biomarkers exhibiting constant within-person correlation over time may also suggest a lack of immunological response to transient or acute exposures, potentially offering an important insight into the relationship between biomarkers and disease [10]. Previous studies of within-person biomarker variability have been limited by small sample sizes, a narrow range of biomarkers, or relatively short periods of time between biomarker measurements (≤ 5 years) [11–20].

This study had two aims: (1) to determine the detectability and reliability of the Meso-Scale Discovery (MSD) and Luminex platforms; and (2) to evaluate the long-term within-person correlation of 22 biomarkers of inflammation in a well-characterized, long-standing prospective cohort study. Results from this study will aid in designing future epidemiologic studies on the role of inflammatory biomarkers in disease etiology.

2. Materials and methods

2.1. Study design and population

This analysis was conducted within the Multicenter AIDS Cohort Study (MACS), a prospective cohort study of men who sex with men (MSM) enrolled at four U.S. locations (Baltimore/Washington D.C., Chicago, Los Angeles, and Pittsburgh) to examine the natural and treated histories of HIV-1 infection. Since 1984, 6972 participants have been enrolled: 4954 in 1984–1985, 668 in 1987–1991, and 1350 in 2001–2003. Institutional review boards at each center approved the MACS protocols and informed consent was obtained from all participants. Descriptions of the MACS protocol have been published previously [21,22]. Study highlights, including data collection forms, may be found at <https://statepi.jhsph.edu/mac/macs.html>. Briefly, study participants are evaluated every six months with standardized interviews, physical examinations, and laboratory analysis of collected blood. Serum, plasma, and peripheral blood mononuclear cells are frozen and stored in local and national repositories. Serum samples used in this study were previously thawed.

Control plasma samples isolated from an HIV-uninfected donor and a HIV-infected donor during a single blood draw were aliquoted and frozen. These external control plasma samples were obtained from Thomas N. Denny at Duke Human Vaccine Institute, Immunology and Virology Quality Assessment Center. Duke University IRB approvals were in place for these activities. To evaluate the reliability of the MSD platform, previously thawed control plasma samples were run in duplicate on each plate over the course of the study.

To assess the ICCs, the concentrations of 22 inflammatory biomarkers were measured in serum samples from 250 HIV-uninfected men. Four study visits per individual were selected to represent the age and race distributions of the underlying cohort population from 1984 to 2009. All HIV-uninfected MACS men who were hepatitis C (HCV)-infected were included, to provide sufficient numbers for comparative analyses. All longitudinal specimens per individual were run on the same plate.

2.2. Laboratory methods

2.2.1. MSD platform

Serum concentrations of 9 cytokines and 7 chemokines were determined using the Meso-Scale Discovery (MSD) platform (Meso-Scale Diagnostics, LLC, Rockville, MD). The MSD system is an electrochemiluminescence-based, 96-well format solid-phase multiplex assay. Two separate kits, the Human Proinflammatory 9-Plex Ultra-Sensitive kit and Human Chemokine 7-Plex Ultra-Sensitive kit, were used to determine concentrations of IL-1 β , IL-2, IL-6, IL-10, IL-12p70, IFN- γ , GM-CSF, TNF- α , and CCL11, CXCL10, CCL2, CCL13, CCL4, and CCL17, respectively. CXCL8 was included in both kits. MSD assays were performed according to the manufacturer's instructions. All MSD testing was conducted in a single laboratory by a single technician at the Johns Hopkins Bloomberg School of Public Health.

Cytokine limits of detection ranged from 0.8 to 1.2 pg/ml, while those for chemokines ranged from 39 to 158 pg/ml. Standard curves were done on each plate in duplicate. The lower limit of detection (LLOD) for each plate-specific analyte was the concentration 2.5 standard deviations above the background.

2.2.2. Luminex platform

Serum concentrations of the soluble receptors were determined using the multiplexed Luminex xMAP system (Fluorokine[®] MAP) using assays produced by R & D systems (Minneapolis, MN) following the manufacturer's instructions, and a Bio-Plex 200 Luminex instrument and Bio-Plex software (Bio-Rad, Hercules, CA). Concentrations of four soluble receptors (sCD14, sgp130, sIL-2R α , sTNF-R2), plus a cytokine (BAFF) and the chemokine CXCL13, were measured in a single panel (Human Biomarker Custom Premix Kit A). All testing for these markers was conducted in a single laboratory at the University of California, Los Angeles. One external serum control from a normal donor (no spiked values) was run in duplicate on each assay. This control sample was from a single blood draw that was aliquoted multiply and stored at -80°C . For each plate tested, a biomarker- and plate-specific LLOD was defined. All the samples tested in the multiplex Kit A had concentrations above the lowest standard of the standard curve; therefore, the LLOD was defined as the observed concentration of the lowest standard.

2.3. Statistical methods

The detectability and reliability of each platform were calculated using the external control samples. Detectability was defined as the proportion of samples above the assay's LLOD. The plate-specific intra-assay coefficient of variation (CV) for each

marker was calculated as follows: (standard deviation (SD) of the duplicate values/mean of the duplicate values) \times 100. The median and IQR (25%, 75% percentiles) intra-assay CV were obtained using all the plate-specific intra-assay CVs from both controls. Each control's inter-assay CV (across plates) was obtained by calculating the SD of all the means across the plates and dividing it by the overall mean value for each analyte. An overall inter-assay CV for each marker was calculated by averaging inter-assay CVs for both controls. For markers with >20% of samples below the LLOD, Cohen's kappa was used to assess detectability agreement between the duplicate control concentrations per plate [23].

2.3.1. Within-person temporal reliability

ICCs (the proportion of total variability due to between-person variability) were calculated for different intervals of time between visits. Biomarker concentrations were natural-logarithm-transformed prior to analysis to account for non-normally distributed residuals. Between-person and within-person variance components and 95% confidence intervals (CI) were determined using linear mixed models, adjusting for age at visit. Age was kept as a continuous variable and centered at the median age across all person-visits. Categories of time between visits were: 0–1.9 years, 2–3.9 years, 4–7.9 years, 8–11.9 years, 12–14.9 years, and \geq 15 years. All combinations of visit pairings were used in the analysis, *i.e.*, consecutive and non-consecutive pairs for the same individual. For example, if time between the first and second visits for ID X was 3.9 years, the time contributed by the visit-pair was allocated to the 2–3.9 year interval; if time between the first and third visits was 16.0 years, it was allocated to the \geq 15 year interval, etc. Parameters were estimated using the maximum likelihood estimate (MLE) method in Stata 13. We established *a priori* that ICCs would be calculated only for cytokines that were detectable in >80% of samples from the 250 HIV-uninfected men. Rosner's interpretation of ICCs was utilized: ICCs between 0.40 and 0.75 reflect fair to good reliability and ICCs \geq 0.75 reflect strong reliability [24].

All analyses were conducted using SAS statistical software, version 9.3 (SAS Institute, Inc., Cary, North Carolina) and Stata, version 13 (College Station, Texas).

3. Results

3.1. Multiplexed assay reliability – MSD platform

3.1.1. Proportion detectable in control samples

For the MSD platform, the 2 external control samples were tested in duplicate across 146 assay runs from September 2010–April 2012. Table 1 provides the proportion detectable, median (IQR) of the detectable concentrations, the proportion of plates with intra-assay CVs \geq 15%, control-specific inter-assay CVs, and the median intra- and overall inter-assay CVs across all 292 samples. On the chemokine plates, all analytes were 100% detectable, as were two analytes on the cytokine plates (TNF- α and CXCL8). IL-10, IL-12p70, IL-1 β , and IL-6 were detectable in \geq 85% of the samples, with IL-10 and IL-6 detectable in nearly all samples.

The proinflammatory cytokines GM-CSF, IFN- γ , and IL-2 were detectable in <80% of control samples (Table 2).

3.1.2. Coefficients of variation

The estimated median intra-assay CVs for the 7 chemokines ranged from 3.3% (CCL17) to 7.6% (CCL11) and for the 5 cytokines from 5.0% (TNF- α) to 14.5% (IL-12p70) (Table 1). The median intra-assay CVs for CXCL8, run on both the chemokine and cytokine kits, were 4.3% and 4.9%, respectively. The inter-assay CVs were higher for the cytokines compared to the chemokines.

There was poor agreement (kappas \leq 0.4) between control duplicates for the 3 cytokines with <80% detectability (Table 2). To account for the possibility this resulted from low values lying close to the LLOD, we created new cutoffs by adding $0.5 \times$ SD of the detectable samples to the LLOD, and assigning all values between the LLOD and the new cutoff as undetectable. This modification did not change the kappa statistics.

Table 1

Reliability characteristics of the Meso-Scale Discovery platform using the external control plasma samples.

Biomarker	Control 1 (n = 146 duplicates)						Control 2 (n = 146 duplicates)						Median intra-assay CV (%)	Overall inter-assay CV (%) ^b
	Detectable (%)	Median, pg/mL	25%	75%	CVs \geq 15% (%) ^a	Inter-assay CV (%)	Detectable (%)	Median, pg/mL	25%	75%	CVs \geq 15% (%) ^a	Inter-assay CV (%)		
Chemokines^d														
CCL11	100	3033	2640	3356	24.7	16.6	100	846	724	951	27.4	16.2	7.6	16.4
CXCL8-chemo ^c	100	7.3	6.7	8.1	6.2	12.2	100	4.9	4.4	5.3	9.6	13.5	4.3	12.8
CXCL10	100	474	418	536	5.5	17.4	100	88.0	78.0	103	9.6	19.4	3.7	18.4
CCL4	100	149	140	159	4.1	9.8	100	36.6	33.3	39.2	7.5	11.1	4.0	10.4
CCL2	100	318	287	337	5.5	11.2	100	165	149	178	6.9	12.2	4.2	11.7
CCL13	100	798	726	867	13.0	12.5	100	969	892	1056	8.9	11.7	4.2	12.1
CCL17	100	733	674	810	9.6	12.6	100	706	648	761	8.2	12.0	3.3	12.3
Cytokines^e														
IL-10	97.9	1.7	1.2	2.2	47.9	35.1	97.6	1.4	1.0	1.8	49.3	34.9	13.3	35.0
IL-12p70	87.3	1.4	1.1	1.9	61.6	33.0	95.2	1.7	1.3	2.4	49.3	36.8	14.5	34.9
IL-6	97.6	0.6	0.5	0.7	30.8	20.0	98.3	0.8	0.7	1.0	18.5	21.5	8.3	20.7
CXCL8-pro ^c	100	7.6	6.9	8.3	8.9	13.0	100	5.2	4.7	5.8	9.6	16.4	4.9	14.7
TNF- α	100	6.9	5.9	7.9	11.6	19.8	100	5.6	4.7	6.2	18.5	20.3	5.0	20.1
IL-1 β	87.3	0.6	0.5	0.7	50.7	67.3	87.3	0.8	0.6	1.0	52.7	119.9	13.2	93.6

NOTE: CXCL8-pro denotes CXCL8 tested on the MSD proinflammatory cytokine kit; CXCL8-chemo denotes CXCL8 tested on the MSD chemokine kit.

^a Percent of plates with intra-assay CVs \geq 15%.

^b Mean of Control 1 and Control 2 inter-assay CVs.

^c CXCL8 was included in both the chemokine and cytokine kits.

^d 14 of the samples were missing chemokine measurements (total number of measurements per control = 284).

^e 2 of the samples were missing cytokine measurements (total number of measurements per control = 290).

Table 2
Agreement among cytokines with >20% undetectable among the external controls tested on the Meso-Scale Discovery platform.

Cytokines	Control 1			Control 2		
	% undetectable	κ^a	κ^b	% undetectable	κ^a	κ^b
GM-CSF	50.3	0.32	0.36	55.1	0.25	0.30
IFN- γ	53.1	0.30	0.28	61.3	0.20	0.33
IL-2	38.7	0.40	0.40	60.6	0.30	0.30

^a Kappa statistic. No adjustment to the lower limit of detection.

^b Kappa statistic. Modified cutoff by 0.5*SD of marker distribution.

Table 3
Reliability characteristics of the Luminex platform using the external control samples.

Biomarkers	Median (pg/mL)	25%	75%	Plates with CVs \geq 15% ($\%^b$)	Median intra-assay CV (%)	Inter-assay CV (%)
BAFF	2425	2192	2717	5.2	4.5	14.7
CXCL13	570	447	662	7.2	3.9	26.1
sCD14 ^a	1480	1387	1574	9.2	5.2	10.5
sCD27	9304	8514	10,290	3.9	4.7	13.9
sgp130 ^a	253	239	266	4.6	4.1	8.8
sIL-2R α	1402	1262	1540	4.6	4.5	29.0
sTNF-R2	2624	2479	2821	6.5	4.0	9.7

^a Expressed as ng/mL.

^b Percent of plates with intra-assay CVs \geq 15%.

3.2. Multiplexed assay reliability – Luminex platform

Table 3 shows the median (IQR) concentrations and median intra- and inter-assay CVs for the 7 biomarkers run on the Luminex platform. For this analysis, one external control was run in duplicate on 153 plates, resulting in 306 biomarker measurements. All 7 biomarkers showed acceptable intra-assay reliability (median CVs < 10%). With the exception of CXCL13 and sIL-2R α , all inter-assay CVs were less than 15%.

3.3. Within-person temporal reliability

For this analysis, we used 971 samples contributed by 250 HIV-uninfected MACS participants from June 1984 to September 2009; 89.6% of the participants ($n = 224$) had 4 samples, 9.2% ($n = 23$) had 3 samples, and 3 had 2 samples. The median (IQR) age across all person-visits was 45.6 years (39.7, 52.5), ranging from 18.7 to 74.5 years. Black men comprised 38.7% of the sample, and approximately 24% of the men had chronic HCV (RNA positive) at the first available measurement. Across all person-visits, nearly 55.0% had a BMI classified as either overweight (25–29.9 kg/m²) or obese (\geq 30 kg/m²).

The median time from first to last sample was 18.3 years (IQR: 5.5–19.8). Across all possible pairs of visits contributed by the same person, the median time between visits was 4.5 years (IQR: 2.6, 12.8). The median time from first to second sample was 2.5 years (IQR: 1.5–4.0), from second to third was 4.5 years (IQR: 2.0–12.4), and from third to fourth was 4.0 years (IQR: 2.0–4.5). Table 4 shows the proportion of detectable samples and the median (IQR) concentrations for the 250 participants. Over 25% of the samples had undetectable GM-CSF, IFN- γ , IL-2, and IL-1 β concentrations. The remaining analytes were detectable in >98% of the samples, except for IL-12p70 (93.5%). Initially, we anticipated calculating ICCs for 6 separate time intervals between blood draws: 0–1.9, 2–3.9, 4–7.9, 8–11.9, 12–14.9, and \geq 15 years. Due to sparse numbers of visit-pairs in 8–11.9 and 12–14.9 years, we combined these into one interval, 8–14.9 years. The numbers of visit-pairs contributing to the intervals were 223, 355, 358, 183, and 297, respectively.

Fig. 1 illustrates the ICCs for each analyte by time between measurements, with estimates plotted at the median of each visit-pair

Table 4

Proportion detectable and median (IQR) of the detectable concentrations of chemokines, cytokines, and soluble cytokine receptors for the 250 HIV-uninfected men (number of person-visits = 971), Multicenter AIDS Cohort Study.

	N	%	Median (pg/mL)	25%	75%
<i>Chemokines</i>					
CCL11	971	100	1696	1219	2419
CXCL8	971	100	13.3	8.9	22.9
CXCL10	971	100	140	92.0	234
CCL2	971	100	508	362	657
CCL4	967	99.6	146	98.4	205
CCL13	971	100	749	580	991
CCL17	970	99.9	532	368	836
CXCL13	960	98.9	298	246	349
<i>Cytokines</i>					
GM-CSF	620	63.9	1.1	0.7	2.1
IFN- γ	587	60.5	1.3	0.9	2.1
IL-10	969	99.8	3.2	2.1	6.3
IL-12p70	908	93.5	2.6	1.4	7.4
IL-1 β	541	55.7	0.5	0.3	0.9
IL-2	697	71.8	0.7	0.5	1.4
IL-6	970	99.9	1.0	0.7	1.5
TNF- α	970	99.9	8.5	6.9	10.6
BAFF	971	100	1970	1731	2263
<i>Soluble cytokine receptors</i>					
sCD14 ^a	970	99.9	2100	1800	2500
sCD27	971	100	9067	7491	11,442
sgp130 ^a	971	100	255	230	289
sIL-2R α	971	100	1382	1139	1739
sTNF-R2	971	100	2307	1880	2917

^a Expressed as ng/mL.

interval. All ICCs were significantly ($p < 0.05$) > 0. (See Table 5 for ICC estimates with 95% confidence intervals.) Tracking was strong (ICC \geq 0.75) for the majority of chemokines measured in samples spanning 1.9 years or less; only CXCL8 exhibited good tracking (>0.55), with statistically equivalent ICCs on both the chemokine and cytokine kits (0.57 and 0.59, respectively). Among the cytokines, IL-10 and IL-12p70 had strong tracking (ICC = 0.85 and 0.84, respectively), with IL-6, TNF- α , and BAFF exhibiting good ICCs (0.60, 0.54, and 0.66, respectively) during this shortest interval. The tracking of the soluble receptors was also good to strong, with the exception of sCD14, which exhibited moderate within-person correlation (ICC = 0.52).

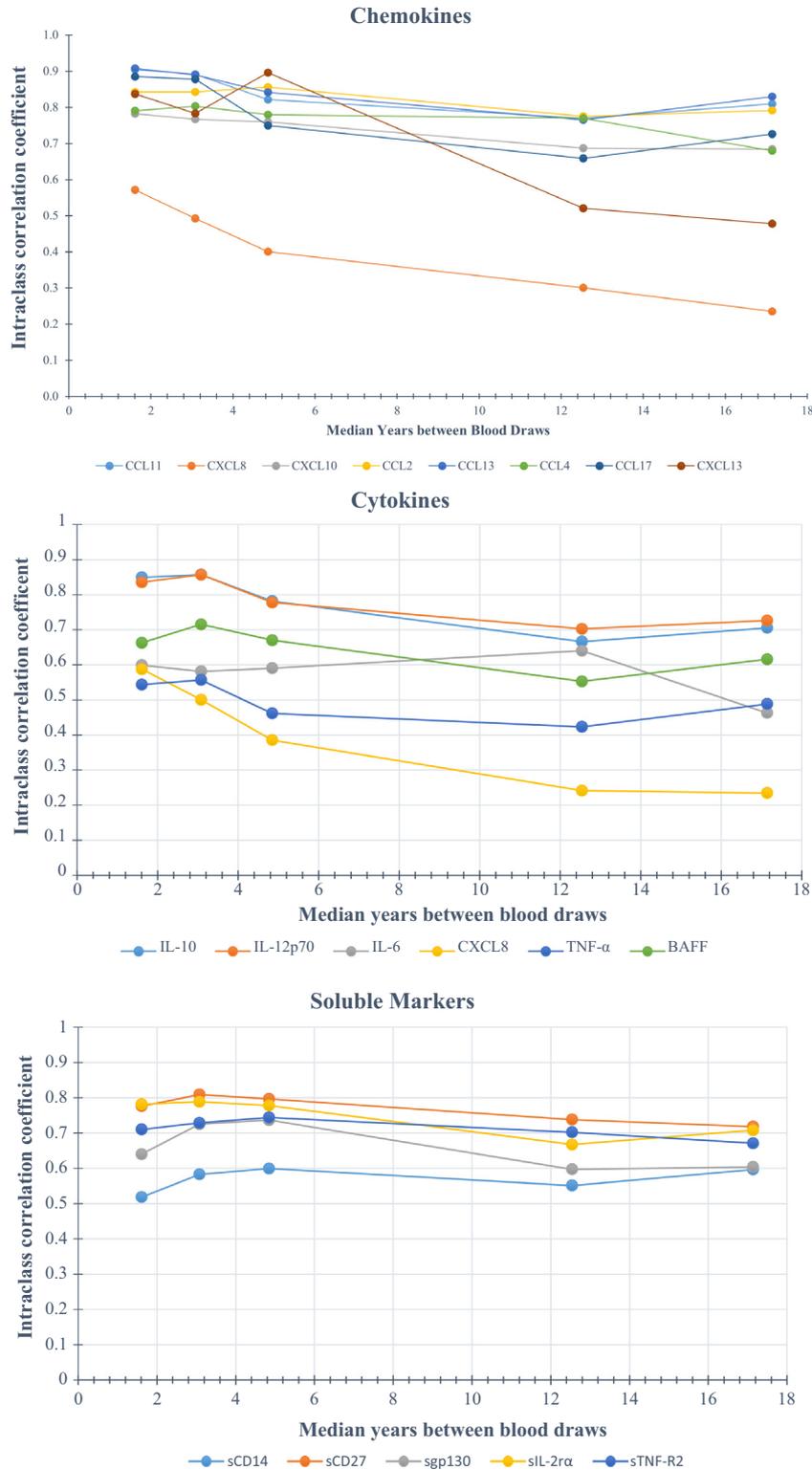


Fig. 1. Age-adjusted intraclass correlation coefficients (ICC) for each analyte across 5 visit-pair time intervals (0–1.9, 2–3.9, 4–7.9, 8–14.9, and ≥ 15 years) between blood draws for 250 HIV-uninfected men in the Multicenter AIDS Cohort Study (MACS). Estimates are plotted at the median of each visit-pair interval.

Although there was some degradation in the ICCs as time between measurements increased, most remained moderate to high, with the exception of CXCL8 and CXCL13. For example, with measurements 8–14.9 years apart, the ICCs for all the soluble receptors decreased to 0.50–0.74. Similarly, the ICCs for the cytokines decreased to <0.75 ; only those for IL-10, IL-12p70, and IL-6 remained above 0.60. Even when the time between sample mea-

surements was ≥ 15 years, several markers remained highly correlated within a person, including CCL11, CCL2, and CCL13 (ICCs >0.75), and CXCL10, CCL4, CCL17, IL-10, IL-12p70, BAFF, sCD14, sCD27, sgp130, sIL-2R α , and sTNF-R2 (ICCs >0.50). In contrast, CXCL8 exhibited good tracking for measurements taken within 4 years but was poorly correlated within a person for longer intervals between measurements. The ICC for CXCL13 was strong

Table 5
Age-adjusted intraclass correlation coefficients (ICC) and 95% confidence intervals for each biomarker across five categories of time between blood draws for 250 HIV-uninfected men, Multicenter AIDS Cohort Study (MACS), 1984–2009.

Biomarkers ^b	Interval 1 0–1.9 years n = 223 ^a		Interval 2 2–3.9 years n = 355		Interval 3 4–7.9 years n = 358		Interval 4 8–14.9 years n = 183		Interval 5 ≥ 15 years n = 297	
	ICC ^c	95% CI	ICC	95% CI	ICC	95% CI	ICC	95% CI	ICC	95% CI
<i>Chemokines</i>										
CCL11	0.90	(0.87–0.93)	0.89	(0.87–0.91)	0.82	(0.78–0.85)	0.77	(0.70–0.82)	0.81	(0.76–0.85)
CXCL8-chemo	0.57	(0.48–0.66)	0.49	(0.42–0.57)	0.40	(0.33–0.48)	0.30	(0.20–0.42)	0.24	(0.16–0.33)
CXCL10	0.78	(0.72–0.83)	0.77	(0.72–0.81)	0.76	(0.71–0.80)	0.69	(0.60–0.76)	0.68	(0.62–0.74)
CCL2	0.84	(0.80–0.88)	0.84	(0.81–0.87)	0.86	(0.82–0.88)	0.77	(0.71–0.83)	0.79	(0.74–0.83)
CCL13	0.91	(0.88–0.93)	0.89	(0.86–0.91)	0.84	(0.81–0.87)	0.76	(0.70–0.82)	0.83	(0.79–0.86)
CCL4	0.79	(0.73–0.84)	0.80	(0.76–0.84)	0.78	(0.73–0.82)	0.77	(0.70–0.82)	0.68	(0.61–0.74)
CCL17	0.89	(0.85–0.91)	0.88	(0.85–0.90)	0.75	(0.70–0.79)	0.66	(0.57–0.74)	0.73	(0.67–0.78)
CXCL13	0.84	(0.79–0.88)	0.78	(0.74–0.82)	0.90	(0.87–0.92)	0.52	(0.42–0.62)	0.48	(0.40–0.56)
<i>Cytokines</i>										
IL-10	0.85	(0.81–0.88)	0.86	(0.82–0.89)	0.78	(0.74–0.82)	0.67	(0.58–0.74)	0.71	(0.64–0.76)
IL-12p70	0.84	(0.79–0.87)	0.86	(0.82–0.88)	0.78	(0.73–0.82)	0.70	(0.62–0.77)	0.73	(0.67–0.78)
IL-6	0.60	(0.51–0.68)	0.58	(0.51–0.65)	0.59	(0.52–0.65)	0.64	(0.55–0.72)	0.46	(0.38–0.55)
CXCL8-pro	0.59	(0.50–0.67)	0.50	(0.43–0.57)	0.39	(0.31–0.47)	0.24	(0.15–0.37)	0.23	(0.16–0.33)
TNF- α	0.54	(0.46–0.63)	0.56	(0.49–0.63)	0.46	(0.39–0.54)	0.42	(0.32–0.53)	0.49	(0.40–0.57)
BAFF	0.66	(0.58–0.73)	0.72	(0.66–0.77)	0.67	(0.61–0.73)	0.55	(0.45–0.65)	0.62	(0.54–0.68)
<i>Soluble cytokine receptors</i>										
sCD14 ^d	0.52	(0.43–0.61)	0.58	(0.51–0.65)	0.60	(0.53–0.66)	0.55	(0.45–0.65)	0.60	(0.52–0.67)
sCD27	0.78	(0.72–0.83)	0.81	(0.77–0.84)	0.80	(0.75–0.83)	0.74	(0.67–0.80)	0.72	(0.66–0.77)
sgp130	0.64	(0.56–0.72)	0.73	(0.67–0.77)	0.74	(0.68–0.78)	0.60	(0.51–0.68)	0.60	(0.53–0.68)
sIL-2R α	0.78	(0.72–0.83)	0.79	(0.74–0.83)	0.78	(0.73–0.82)	0.67	(0.58–0.74)	0.71	(0.65–0.76)
sTNF-R2	0.71	(0.64–0.77)	0.73	(0.67–0.78)	0.74	(0.69–0.79)	0.70	(0.62–0.77)	0.67	(0.60–0.73)

NOTE: CXCL8-pro denotes CXCL8 tested on the MSD proinflammatory cytokine kit; CXCL8-chemo denotes CXCL8 tested on the MSD chemokine kit.

^a Number of visit-pairs contributing to interval of time between blood draws.

^b Undetectable values were assigned a value equal to one-half the plate-specific lower limit of detection.

^c Age-adjusted. Age at visit was modeled continuously and centered at the median age across all 971 HIV-uninfected person-visits.

^d One observation with an implausible value for sCD14 was excluded only from analyses of sCD14 ($n = 970$).

(≥ 0.78) up to 8 years between visits but was fair to good after 8 years, with ICCs of 0.52 (8–14.9 years) and 0.48 (≥ 15 years).

4. Discussion

Recent advances in multiplexing technologies have led to a wide array of commercially available platforms capable of quantifying an increasing number of soluble proteins in smaller sample volumes with greater sensitivity. These advancements are transforming biomarker discovery research, allowing for the exploration of the biologic underpinnings of disease pathogenesis in large population-based cohort studies. Along with unique analytic issues inherent to large-scale biomarker studies are questions of assay reliability over time. Here, we assessed several important methodological aspects of biomarker measurement including detectability, assay reliability, and within-person variability. Methodological concerns (e.g., inter- and intra-plate variation, inter- and intra-laboratory variation, assay drift over time, assay reliability across different manufactured lots) can be compounded as the number of samples and/or analytes increases, resulting in attenuated effect estimates and reduced statistical power [25].

Our results suggest acceptable reliability of the MSD platform for the chemokines, as measured by high levels of detectability and low assay CVs using the external control samples. Reliability was less consistent for the proinflammatory cytokines. IL-10, IL-12p70, IL-6, and TNF- α exhibited acceptable detectability and reliability; however, IL-2, IL-1 β , IFN- γ , and GM-CSF demonstrated inconsistent detectability. Possible reasons for this could be low sensitivity of the assay, potential degradation of the analyte due to long-term storage, or low cytokine concentrations in these control sera. Modifying the lower cutoff did not change the agreement, arguing against this last explanation. These analytes have low endogenous concentrations in the peripheral circulation, which may result from tissue utilization of cytokines or the distribution

of cytokine-expressing cells. Likewise, some analytes may be less reliable due to the antibody pairs used in the assay [26]. The occurrence of variable plate-specific CVs underscored the need to run all longitudinal samples for an individual on the same plate, as well as samples from comparative groups of interest on each plate, in order to avoid bias or the attenuation of parameter estimates [27,28].

The ICC provides another measure of assay reliability, particularly with longitudinal specimens collected closely in time, and individual biomarker tracking over longer time periods. With poor within-person correlation, a biomarker-disease association may be missed if based only a single measure, particularly if not captured in the most biologically relevant time period. While moderate to large associations may still be detected, effect sizes will be attenuated [18]. ICCs can be used to estimate the extent of measurement error associated with a single measurement and then employed to obtain corrected estimates of the “true” association, under the assumption of non-differential measurement error. This is done by multiplying the natural logarithm of the hypothesized true measure of association by the ICC and then exponentiating the result ($RR_{ob} = \exp[ICC \times \ln RR_{true}]$ where RR is the relative risk) [29,30]. ICCs > 0.80 , as we observed, would result in only a moderate downward bias of exposure-disease associations [31]. In addition, smaller sample sizes may be adequate in studies of repeated biomarker measurements of markers that have high within-person correlation [32].

Yet, while assay reliability is important, particularly for planning studies, an assay does not need to exhibit perfect reliability in order to have epidemiologic utility. Variability in biomarker concentrations should be considered in the context of the biological question. For example, significant associations may be identified in the face of large technical variability, if the biological variability is larger than the technical variability. Biomarkers with many undetectable values may still prove informative if expressed dis-

proportionately according to disease [7]. For example, prior MACS studies looking at the association between B-cell inflammatory markers and HIV-associated non-Hodgkin B-cell lymphoma identified significant biomarker-disease associations by comparing the fractions detectable versus undetectable [33,34].

Using an ICC ≥ 0.75 as a criterion for strong within-person correlation, 11 of the 22 biomarkers in our study exhibited strong reliability over the shortest time between measurements (0–1.9 years). The remaining 8 showed fair to good reliability. The short-term variability of IL-6, CXCL8, and TNF- α , which comprise the classic inflammatory cascade, has been the subject of several previous reliability analyses using other assay platforms. Studies of IL-6 have reported short-term ICCs comparable to [15,35,36] or higher [11,16,19] than what we observed. Similarly, previous studies of CXCL8 showed short-term ICCs comparable to [12,16], lower [36,11], or higher [15,19] than in the present study. The moderate ICCs observed for TNF- α were similar to some studies [16,17] but lower than others [11,12,15,19,36]. These differences may be due to differences in study populations or sample handling. The current study comprised exclusively male participants, mostly middle-aged, in a population at risk for HIV acquisition, unlike most previous studies. While uninfected with HIV, they were more likely to be exposed to HIV-related risk factors, such as chronic HCV infection and recreational drug use, potentially predisposing them to transient and variable inflammatory stimuli. In addition, we have previously reported that discrepancies may arise from different assay technologies or lack of a centralized laboratory [37].

Although the ICCs were expected to diminish with increasing time between measurements, many exhibited considerable tracking over time. With few exceptions (CXCL8, IL-6, TNF- α , and CXCL13), all demonstrated good to strong within-person tracking up to 15 years apart, with some exhibiting ICCs ≥ 0.75 even up to 8 (CCL11, CXCL10, CCL2, CCL13, CCL4, CCL17, IL-10, IL-12p70, sCD27, sIL-2R α , sTNF-R2) and 15 years (CCL11, CCL2, CCL13, sCD27) between samples. Tracking strength may serve as an important characteristic of physiologic homeostasis and be informative for both clinical assessment and pathogenesis research. Characteristics which influence these biomarkers also may be fixed within a person but very heterogeneous in the population. The men in our study were all HIV-uninfected and therefore not subjected to the impact of HIV infection on the immune system. Nevertheless, because 25% were HCV-mono-infected, we considered the potential effect of HCV on the ICCs; however, adjustment for HCV did not alter our results (data not shown).

Our study has some limitations. First, the study population to assess the ICCs consisted of young to middle-age men who have sex with men in the United States, which may limit generalizability to other populations. Second, we were unable to evaluate the ICCs of 4 cytokines (IL-2, IL-1 β , IFN- γ , and GM-CSF) because >20% of samples were undetectable. Also, the lack of a gold standard precluded assessing the validity of our biomarker measurements. In addition, concentrations in serum may not fully reflect biological processes. Moreover, because the intra-assay CVs are based on a small number of measurements per plate, estimates may not be representative of the full range of variability estimates given the relatively few data points. Finally, the estimated CVs are based on measurements from external samples and therefore may not be representative of the variability in measurements for the population used to estimate the ICCs.

Several strengths of our study should be noted. For evaluating the MSD assay reliability, we utilized the same two external control samples across 146 plates, allowing us to robustly determine assay reliability. In addition, each multiplex assay was conducted in a single laboratory with standardized protocols. The MSD testing was performed by a single technician, minimizing laboratory variability, and limited to a single assay lot. The centralized Luminex/

R&D Systems assays also were done using a single assay lot. All of the samples tested underwent only one freeze-thaw cycle, limiting variation due to specimen degradation. In addition, all longitudinal specimens for an individual were run on the same plate, thus minimizing the effects of inter-assay variability. High retention from 1984 to 2009 in a well-characterized cohort with standardized data and specimen collection permitted the assessment of biomarker reliability over increasingly long periods of time, up to 15 years, which is, to our knowledge, otherwise unaccounted for in the literature and may be useful for diseases with long etiologic windows, such as cancer, or latency periods, such as HIV. Moreover, the large sample size allowed us to estimate ICCs with greater precision than many previous studies. As far as we know, only one other analysis, from the MACS, has reported on the temporal reliability of sgp130, BAFF, and CXCL13, assessed using the Luminex platform, but this was limited to a 2-year interval between sample collections [36].

In conclusion, our data support the use of high-sensitivity multiplex assays in large-scale epidemiologic research. Despite some technical variability, most biomarkers exhibited acceptable detectability and reliability. Moreover, most yielded fair to strong within-person reliability in men over a period of up to 15 years, indicating low variability within individuals relative to between individuals and stability of these analytes over time when properly stored. Changes in these biomarkers over time may thus be efficiently studied, and evaluated for use as surrogates for endogenous and exogenous effects longitudinally. CXCL8, TNF- α , and CXCL13 may be less stable within a person over longer periods of time but were still reliable within short time periods. Our results may also be useful for statistically accounting for the attenuation of exposure-disease associations due to potential non-differential misclassification. Our findings highlight the utility of investigating inflammation and immune-related pathways through the multiplex measurement of a comprehensive panel of circulating biomarkers, which will be important for unraveling the complex profile of immune response and disease pathogenesis.

Acknowledgements

Samples and data in this manuscript were collected by the Multicenter AIDS Cohort Study (MACS) with support from an American Recovery and Reinvestment Act (ARRA) supplement with centers (Principal Investigators) at: Johns Hopkins University Bloomberg School of Public Health (Joseph Margolick), U01-AI35042; Northwestern University (Steven Wolinsky), U01-AI35039; University of California, Los Angeles (Roger Detels), U01-AI35040; University of Pittsburgh (Charles Rinaldo), U01-AI35041; the Center for Analysis and Management of MACS, Johns Hopkins University Bloomberg School of Public Health (Lisa Jacobson), UM1-AI35043. The MACS is funded primarily by the National Institute of Allergy and Infectious Diseases (NIAID), with additional co-funding from the National Cancer Institute (NCI). Website located at <http://www.statepi.jhsph.edu/mac/mac.html>. The contents of this publication are solely the responsibility of the authors and do not represent the official views of the National Institutes of Health (NIH).

Donor identification and acquisition of external control samples from Duke was supported by the NIAID DAIDS IQA Program (HHSN27220054C). Cytokine profiling of pre-shipment material was performed in the Duke Human Vaccine Institute/Regional Biocontainment Laboratory Immunology Unit (Durham, NC) under the direction of Dr. Gregory D. Sempowski. The Regional Biocontainment Laboratory at Duke received partial support for construction from the National Institutes of Health, National Institute of Allergy and Infectious Diseases (UC6-AI058607).

The authors have no conflict of interest to report.

References

- [1] S.F. Moss, M.J. Blaser, Mechanisms of disease: inflammation and the origins of cancer, *Nat. Clin. Pract. Oncol.* 2 (2) (2005) 90–97 (quiz 1 p following 113).
- [2] K. Heikkilä et al., Associations of circulating C-reactive protein and interleukin-6 with cancer risk: findings from two prospective cohorts and a meta-analysis, *Cancer Causes Control: CCC* 20 (1) (2009) 15–26.
- [3] R. Klingenberg, G.K. Hansson, Treating inflammation in atherosclerotic cardiovascular disease: emerging therapies, *Eur. Heart J.* 30 (23) (2009) 2838–2844.
- [4] D.M. Silverstein, Inflammation in chronic kidney disease: role in the progression of renal and cardiovascular disease, *Pediat. Nephrol.* 24 (8) (2009) 1445–1452.
- [5] H. Tilg, The role of cytokines in non-alcoholic fatty liver disease, *Dig. Dis.* 28 (1) (2010) 179–185.
- [6] V. Appay, D. Sauce, Immune activation and inflammation in HIV-1 infection: causes and consequences, *J. Pathol.* 214 (2) (2008) 231–241.
- [7] A.K. Chaturvedi et al., Evaluation of multiplexed cytokine and inflammation marker measurements: a methodologic study, *Cancer Epidemiol., Biomarkers Prevent.: A Publ. Am. Assoc. Cancer Res., Cosponsored Am. Soc. Prevent. Oncol.* 20 (9) (2011) 1902–1911.
- [8] I. Agalliu et al., Detectability and reproducibility of plasma levels of chemokines and soluble receptors, *Results Immunol.* 3 (2013) 79–84.
- [9] G.M. Fitzmaurice, N.M. Laird, J.H. Ware, *Applied Longitudinal Analysis*. Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, 2004. 506.
- [10] N. Galai et al., Tracking of markers and onset of disease among HIV-1 seroconverters, *Stat. Med.* 12 (22) (1993) 2133–2145.
- [11] Y. Gu et al., Reproducibility of serum cytokines and growth factors, *Cytokine* 45 (1) (2009) 44–49.
- [12] J.N. Hofmann et al., Intra-individual variability over time in serum cytokine levels among participants in the prostate, lung, colorectal, and ovarian cancer screening Trial, *Cytokine* 56 (2) (2011) 145–148.
- [13] T.V. Clendenen et al., Temporal reliability of cytokines and growth factors in EDTA plasma, *BMC Res. Notes* 3 (2010) 302.
- [14] F. Linkov et al., Reliability of tumor markers, chemokines, and metastasis-related molecules in serum, *Eur. Cytokine Netw.* 20 (1) (2009) 21–26.
- [15] S.L. Navarro et al., Reliability of serum biomarkers of inflammation from repeated measures in healthy individuals, *Cancer Epidemiol., Biomarkers Prevent.: A Publ. Am. Assoc. Cancer Res., Cosponsored Am. Soc. Prevent. Oncol.* 21 (7) (2012) 1167–1170.
- [16] S.A. Lee et al., Intra-individual variation of plasma adipokine levels and utility of single measurement of these biomarkers in population-based studies, *Cancer Epidemiol., Biomarkers Prevent.: A Publ. Am. Assoc. Cancer Res., Cosponsored Am. Soc. Prevent. Oncol.* 16 (11) (2007) 2464–2470.
- [17] R.C. Kaplan et al., Within-individual stability of obesity-related biomarkers among women, *Cancer Epidemiol., Biomarkers Prevent.: A Publ. Am. Assoc. Cancer Res., Cosponsored Am. Soc. Prevent. Oncol.* 16 (6) (2007) 1291–1293.
- [18] E.A. Platz et al., Intra-individual variation in serum C-reactive protein over 4 years: an implication for epidemiologic studies, *Cancer Causes Contr.: CCC* 21 (6) (2010) 847–851.
- [19] H.L. Wong et al., Reproducibility and correlations of multiplex cytokine levels in asymptomatic persons, *Cancer Epidemiol., Biomarkers Prevent.: A Publ. Am. Assoc. Cancer Res., Cosponsored Am. Soc. Prevent. Oncol.* 17 (12) (2008) 3450–3456.
- [20] M.M. Epstein et al., Temporal stability of serum concentrations of cytokines and soluble receptors measured across two years in low-risk HIV-seronegative men, *Cancer Epidemiol., Biomarkers Prevent.: A Publ. Am. Assoc. Cancer Res., Cosponsored Am. Soc. Prevent. Oncol.* 22 (11) (2013) 2009–2015.
- [21] R.A. Kaslow et al., Infection with the human immunodeficiency virus: clinical manifestations and their relationship to immune deficiency. A report from the Multicenter AIDS Cohort Study, *Ann. Intern. Med.* 107 (4) (1987) 474–480.
- [22] J.S. Chmiel et al., Factors associated with prevalent human immunodeficiency virus (HIV) infection in the Multicenter AIDS Cohort Study, *Am. J. Epidemiol.* 126 (4) (1987) 568–577.
- [23] J. Cohen, A coefficient of agreement for nominal scales, *Educ. Psychol. Measur.* 20 (1960) 37–46.
- [24] B.A. Rosner, *Fundamentals of Biostatistics*, vol. 6, Brooks/Cole, Thomson, 2006.
- [25] C.W. Whitney, B.K. Lind, P.W. Wahl, Quality assurance and quality control in longitudinal studies, *Epidemiol. Rev.* 20 (1) (1998) 71–80.
- [26] D. Dabito et al., Multiplex measurement of proinflammatory cytokines in human serum: comparison of the Meso Scale Discovery electrochemiluminescence assay and the Cytometric Bead Array, *J. Immunol. Meth.* 372 (1–2) (2011) 71–77.
- [27] S.S. Tworoger, S.E. Hankinson, Use of biomarkers in epidemiologic studies: minimizing the influence of measurement error in the study design and analysis, *Cancer Causes Contr.: CCC* 17 (7) (2006) 889–899.
- [28] S.S. Tworoger, S.E. Hankinson, Collection, processing, and storage of biological samples in epidemiologic studies: sex hormones, carotenoids, inflammatory markers, and proteomics as examples, *Cancer Epidemiol., Biomarkers Prevent.: A Publ. Am. Assoc. Cancer Res., Cosponsored Am. Soc. Prevent. Oncol.* 15 (9) (2006) 1578–1581.
- [29] S.E. Hankinson et al., Reproducibility of plasma hormone levels in postmenopausal women over a 2–3-year period, *Cancer Epidemiol., Biomarkers Prevent.: A Publ. Am. Assoc. Cancer Res., Cosponsored Am. Soc. Prevent. Oncol.* 4 (6) (1995) 649–654.
- [30] B. Rosner, D. Spiegelman, W.C. Willett, Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error, *Am. J. Epidemiol.* 136 (11) (1992) 1400–1413.
- [31] A.W. Hsing et al., Reproducibility of serum sex steroid assays in men by RIA and mass spectrometry, *Cancer Epidemiol., Biomarkers Prevent.: A Publ. Am. Assoc. Cancer Res., Cosponsored Am. Soc. Prevent. Oncol.* 16 (5) (2007) 1004–1008.
- [32] A.J. Kirby, N. Galai, A. Munoz, Sample size estimation using repeated measurements on biomarkers as outcomes, *Contr. Clin. Trials* 15 (3) (1994) 165–172.
- [33] E. Vendrame et al., Serum levels of cytokines, and biomarkers for inflammation and immune activation, and HIV-Associated Non-Hodgkin B cell Lymphoma Risk, *Cancer Epidemiol., Biomarkers Prevent.: A Publ. Am. Assoc. Cancer Res., Cosponsored Am. Soc. Prevent. Oncol.* (2013).
- [34] E.C. Breen et al., B-cell stimulatory cytokines and markers of immune activation are elevated several years prior to the diagnosis of systemic AIDS-associated non-Hodgkin B-cell lymphoma, *Cancer Epidemiol., Biomarkers Prevent.: A Publ. Am. Assoc. Cancer Res., Cosponsored Am. Soc. Prevent. Oncol.* 20 (7) (2011) 1303–1314.
- [35] G.Y. Ho et al., Variability of serum levels of tumor necrosis factor- α , interleukin 6, and soluble interleukin 6 receptor over 2 years in young women, *Cytokine* 30 (1) (2005) 1–6.
- [36] M.M. Epstein et al., Temporal Stability of Serum Concentrations of Cytokines and Soluble Receptors Measured Across Two Years in Low-Risk HIV-Seronegative Men, *Cancer Epidemiol., Biomarkers Prevent.: A Publ. Am. Assoc. Cancer Res., Cosponsored Am. Soc. Prevent. Oncol.* (2013).
- [37] E.C. Breen et al., Multisite comparison of high-sensitivity multiplex cytokine assays, *Clin. Vaccine Immunol.: CVI* 18 (8) (2011) 1229–1242.