

## Evaluation of Multiplexed Cytokine and Inflammation Marker Measurements: a Methodologic Study

Anil K. Chaturvedi<sup>1</sup>, Troy J. Kemp<sup>3</sup>, Ruth M. Pfeiffer<sup>1</sup>, Angelique Biancotto<sup>4</sup>, Marcus Williams<sup>3</sup>, Stella Munuo<sup>2</sup>, Mark P. Purdue<sup>1</sup>, Ann W. Hsing<sup>1</sup>, Ligia Pinto<sup>3</sup>, J. Philip McCoy<sup>4</sup>, and Allan Hildesheim<sup>1</sup>

### Abstract

**Background:** Chronic inflammation is etiologically related to several cancers. We evaluated the performance [ability to detect concentrations above the assay's lower limit of detection, coefficients of variation (CV), and intraclass correlation coefficients (ICC)] of 116 inflammation, immune, and metabolic markers across two Luminex bead-based commercial kits and three specimen types.

**Methods:** From 100 cancer-free participants in the Prostate, Lung, Colorectal, and Ovarian Cancer Trial, serum, heparin plasma, and EDTA plasma samples were utilized. We measured levels of 67 and 97 markers using Bio-Rad and Millipore kits, respectively. Reproducibility was assessed using 40 blinded duplicates (20 within-batches and 20 across-batches) for each specimen type.

**Results:** A majority of markers were detectable in more than 25% of individuals on all specimen types/kits. Of the 67 Bio-Rad markers, 51, 52, and 47 markers in serum, heparin plasma, and EDTA plasma, respectively, had across-batch CVs of less than 20%. Likewise, of 97 Millipore markers, 75, 69, and 78 markers in serum, heparin plasma, and EDTA plasma, respectively, had across-batch CVs of less than 20%. When results were combined across specimen types, 45 Bio-Rad and 71 Millipore markers had acceptable performance (>25% detectability on all three specimen types and across-batch CVs <20% on at least two of three specimen types). Median concentrations and ICCs differed to a small extent across specimen types and to a large extent between Bio-Rad and Millipore.

**Conclusions:** Inflammation and immune markers can be measured reliably in serum and plasma samples using multiplexed Luminex-based methods.

**Impact:** Multiplexed assays can be utilized for epidemiologic investigations into the role of inflammation in cancer etiology. *Cancer Epidemiol Biomarkers Prev*; 20(9); 1902–11. ©2011 AACR.

### Introduction

Chronic inflammation is now recognized as a major etiologic factor for a range of malignancies including cancers of the esophagus, stomach, gall bladder, liver, pancreas, colon and rectum, prostate, urinary bladder, and lung (1–3). Chronic inflammation in tissues arises from sustained activation of the innate immune system (neutrophils, macrophages, and fibroblasts) as well as the adaptive immune system (B and T cells; ref. 4). This

chronic inflammatory response to persistent infections or environmental insults increases cancer risk both directly, through DNA damage, and indirectly, through tissue remodeling and fibrosis (4).

One strategy to evaluate the relationship of cancer with chronic inflammation is to measure circulating levels of inflammatory markers. Most previous epidemiologic investigations of circulating inflammatory markers and cancer have included a narrow range of markers [e.g., C-reactive protein (CRP), interleukin (IL) 6, IL-10, TNF- $\alpha$ ; ref. 5]. The process of inflammation is complex and involves multiple key mediators (3) including chemokines, proinflammatory cytokines, anti-inflammatory cytokines, growth factors, angiogenesis factors, and metabolic markers. Therefore, a thorough epidemiologic characterization of inflammatory biomarkers and pathways involved in carcinogenesis requires a comprehensive evaluation of a wide range of markers.

Emerging multiplex technologies allow for the simultaneous quantification of more than 100 analytes in low specimen volumes (6, 7), underscoring their potential utility for large-scale epidemiologic investigations. Although the obvious benefits of multiplexed assays

**Authors' Affiliations:** <sup>1</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute; <sup>2</sup>Information Management Services, Rockville; <sup>3</sup>HPV Immunology Laboratory, National Cancer Institute, SAIC, Frederick; and <sup>4</sup>Center for Human Immunology, NIH, Bethesda; Maryland

**Note:** Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

**Corresponding Author:** Anil K. Chaturvedi, Infections and Immunoepidemiology Branch, Division of Cancer Epidemiology and Genetics, 6120 Executive Boulevard, EPS 7072, Rockville, MD 20852. Phone: 301-451-2495; Fax: 301-402-0817; E-mail: [chaturva@mail.nih.gov](mailto:chaturva@mail.nih.gov)

doi: 10.1158/1055-9965.EPI-11-0221

©2011 American Association for Cancer Research.

include reductions in time and specimen volume, several aspects of these assays warrant thorough evaluation and standardization including assay validity, reproducibility, stability, and appropriateness of specimen types (e.g., serum vs. plasma; ref. 6). A majority of the previous studies that have formally assessed the performance of multiplexed assays were small in size and limited in the number of markers (8–15).

In the current study, we evaluated the performance of 116 inflammation, immune, and metabolic markers across 2 Luminex bead-based commercial kits (Millipore and Bio-Rad) and 3 specimen types (serum, heparin plasma, and EDTA plasma). We specifically addressed the epidemiologic utility of these assays, as measured by their detectability in specimens from cancer-free individuals (i.e., values above the assay's lower limit of detection) and reproducibility, as measured by coefficients of variation (CV) and intraclass correlation coefficients (ICC). Our primary aim was to evaluate the performance of each marker within a specimen type and kit type. Our secondary aims were to compare assay performance across specimen types within each kit and across kits within each specimen type.

## Materials and Methods

### Study design

We conducted this study among 100 cancer-free individuals who participated in the screening arm of the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. Briefly, between 1993 and 2002, the PLCO trial recruited approximately 155,000 men and women into either the screening arm or the control arm (16). Screening arm participants provided blood specimens at the baseline visit (T0) and annually during follow-up (T1 through T5; ref. 16). All samples were processed by centrifugation at 2,400 to 3,000 rpm for 15 minutes. Specimens were frozen within 2 hours of collection and stored at  $-80^{\circ}\text{C}$  until further use. Specimens used for the current study underwent 2 thaw cycles—1 for aliquoting and 1 for laboratory testing.

We selected 100 participants with available T0 serum and T0 heparin plasma as well as EDTA plasma samples collected at the third annual visit (T3). To ensure comparability of the T3 EDTA plasma samples with another specimen type, we also included T3 heparin plasma samples from 50 of the 100 individuals. This design allowed us to compare assay performance between T0 serum versus T0 heparin plasma samples ( $n = 100$ ) as well as between T3 EDTA plasma samples versus T3 heparin plasma samples ( $n = 50$ ). The T0 and T3 heparin plasma samples were analyzed separately. To evaluate reproducibility of marker measurements, from each of the 3 specimen types (T0 serum, T0 heparin plasma, and T3 EDTA plasma), we selected 40 individuals as blinded duplicates and placed 20 pairs as within-batch duplicates and 20 as across-batch duplicates. A batch denotes 1 plate of 37 unique samples including blinded duplicates. The

subjects selected for blinded duplicates varied by specimen type but were the same across the 2 kits given a specimen type.

### Laboratory methods

We evaluated the performance of 116 inflammation, immune, and metabolic markers—67 on Bio-Rad and 97 on Millipore, with 48 markers measured on both kits. Using magnetic bead-based assays, the Bio-Rad markers were measured in 150  $\mu\text{L}$  of specimen across 4 panels: cytokine panel 1 (27 markers), cytokine panel 2 (21 markers), acute-phase protein panel (9 markers), and diabetes panel (12 markers). The Millipore kits utilized polystyrene bead-based assays to measure 97 markers in 400  $\mu\text{L}$  of specimen across 6 panels: cytokine panel 1 (39 markers), cytokine panel 2 (23 markers), cytokine panel 3 (9 markers), soluble receptor panel (13 markers), metabolic panel (10 markers), and acute-phase protein panel (3 markers). On both kits, specimens were assayed in duplicate and the duplicate measurements were averaged. On the basis of the measurements of 7 standard concentrations provided by the manufacturer, a 5-parameter standard curve was utilized to convert optical density values into concentrations (pg/mL). Using the curve-fit measurements for each standard, we also estimated CVs across unblinded duplicates as well as recovery—calculated as the ratio of the observed and expected concentrations. We note that these recoveries indicate the goodness of fit of the standard curve rather than recoveries based on known, spiked concentrations. All assays were conducted according to the manufacturer's instructions.

### Statistical analyses

For each marker, separately within each kit and specimen type, we evaluated assay performance using 3 measures: (i) detectability—the proportion of samples with values above the assay's lower limit of detection (based on the 100 unique measurements for T0 serum, T0 heparin plasma, and T3 EDTA plasma); (ii) CV for within-batch and across-batch duplicates (based on 20 pairs each for each specimen type); and (iii) ICCs, which capture the proportion of total variability in measurements that arises from interindividual variability (based on 20 pairs each of within-batch and across-batch duplicates for each specimen type). Observed concentrations of each marker were log-transformed to achieve approximate normality. CVs and ICCs were estimated using the ANOVA procedure. We considered detectability greater than 25% as acceptable, given the common use of quartiles in epidemiologic studies. CVs less than 20% were deemed acceptable.

To generalize marker performance across the 3 specimen types (T0 serum, T0 heparin plasma, and T3 EDTA plasma), we defined acceptable performance for a marker as: (i) being detectable in greater than 25% of the 100 samples on all 3 specimen types and (ii) across-batch CVs of less than 20% on at least 2 of the 3 specimen types.

These criteria allowed us to identify markers with acceptable performance within each kit across different specimen types.

We compared detectability across specimen types given a kit (T0 serum vs. T0 heparin plasma and T3 EDTA plasma vs. T3 heparin plasma) and across kits (Bio-Rad vs. Millipore) given a specimen type using the McNemar's test. Median concentrations of each marker across specimen types and kits were compared using the Wilcoxon signed-rank test. Correlations of marker measurements across specimen types and kits were estimated using the Spearman's rank correlation coefficient. Analyses for comparisons of detectability and medians and for correlation coefficients were based on the 100 unique measurements for T0 serum, T0 heparin plasma, and T3 EDTA plasma.

Head-to-head comparisons of ICCs across specimen types (given a kit) and between kits (given a specimen type) were conducted using variance components analyses. For each specimen type and kit, we estimated the ICC as the proportion of variation attributed to interindividual variation using all observations (including blinded duplicates) in mixed-effects models that included the batch number and study subjects nested within batches.

## Results

Given the large number of markers as well as the different specimen types and kit types, we present assay performance and type/kit comparisons as the number of markers with acceptable or poor performance. Detailed results for each marker, including median observed concentrations, percentage of detectability, CVs, and ICCs separately for each kit and specimen type as well as correlations of marker measurements across specimen types and kits are presented as Supplementary Material (Supplementary Tables S1 and 2).

### Bio-Rad markers

For the 67 markers measured on Bio-Rad, we initially evaluated CVs as well as recoveries on unblinded duplicates across the 7 known standard concentrations used for curve fit. Across the markers, CVs ranged from 4.3% to 27%, with only 2 markers (PCT and Ferritin) having CVs greater than 20%. Likewise, recoveries ranged from 90% to 670%, with a majority of markers (49 of 67 markers) having recoveries in the 80% to 120% range.

Using a criterion of detectable values in greater than 25 of the 100 individuals for each specimen type, a high proportion of markers were detectable (56 markers on serum, 63 markers on heparin plasma, and 64 markers on EDTA plasma; Fig. 1A and Table 1). Likewise, a high proportion of markers had CVs for across-batch duplicates less than 20% (51, 52, and 47, respectively, on serum, heparin plasma, and EDTA plasma; Fig. 2A–C and Table 1). In addition, for a majority of markers, within-batch CVs were lower than across-batch CVs on each specimen type (Fig. 2A–C).

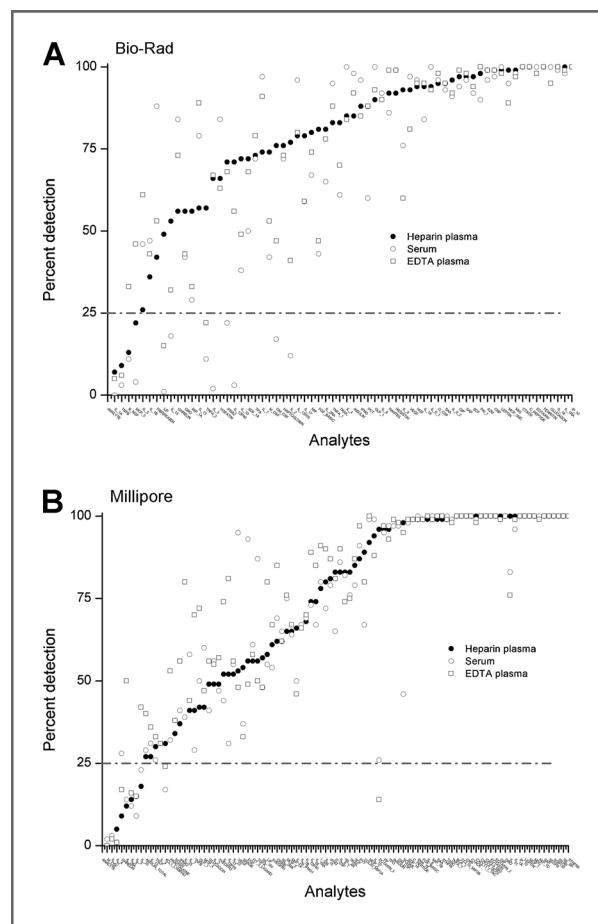


Figure 1. The proportion of samples with detectable levels for 67 markers on Bio-Rad (A) and 97 markers on Millipore (B) across 3 specimen types—serum (open circles), heparin plasma (filled circles), and EDTA plasma (open squares).

When the performance across the 3 specimen types was combined, 45 of 67 markers had acceptable performance in terms of detectability and across-batch CVs (Table 1). Markers with poor performance (<25% detection on at least 1 specimen type or across-batch CVs >20% on at least 2 specimen types) included: B-NGF, GM-CSF, G-CSF, IFNA2, IL-1A, IL-1B, IL-2, IL-3, IL-4, IL-5, IL-7, IL-10, IL-12 p40, IL-12 p70, IL-13, IL-15, IL-17, LIF, MCP-3, MIP-1A, TNF-B, and TPA (Table 2). Across the 4 panels, 12 of 27 markers on cytokine panel 1, 13 of 21 markers on cytokine panel 2, 8 of 9 markers on the acute-phase panel, and all 12 markers on the diabetes panel had acceptable performance.

On all 3 specimen types, ICCs for across-batch duplicates ranged from 0.31 to 0.99, with 23 markers on serum, 22 on heparin plasma, and 10 on EDTA plasma having ICCs greater than 0.8 (Table 1).

### Millipore markers

Across the 97 Millipore markers, CVs for the 7 standard concentrations ranged from 3.4% to 14.7% and recoveries

**Table 1.** Summary of results for evaluation of multiplexed inflammation marker assays

	Bio-Rad number of markers (%)	Millipore number of markers (%)
Total	67 (100.0)	97 (100.0)
<i>Markers with &gt;25% detection</i>		
Serum	56 (83.5)	89 (91.7)
Heparin plasma	63 (94.0)	89 (91.7)
EDTA plasma	64 (95.5)	89 (91.7)
<i>Markers with &lt;20% CV for across-batch duplicates<sup>a</sup></i>		
Serum	51 (76.2)	75 (77.3)
Heparin plasma	52 (77.6)	69 (71.1)
EDTA plasma	47 (70.1)	78 (80.4)
<i>ICCs for across-batch duplicates</i>		
Serum		
<0.50	6 (8.9)	5 (5.1)
0.50–0.80	29 (43.3)	32 (33.0)
0.80–0.90	12 (17.9)	19 (19.5)
>0.90	11 (16.4)	34 (35.1)
Heparin plasma		
<0.50	3 (4.5)	9 (9.3)
0.50–0.80	38 (56.7)	31 (31.9)
0.80–0.90	15 (22.3)	25 (25.8)
>0.90	7 (10.4)	28 (28.9)
EDTA plasma		
<0.50	5 (7.5)	7 (7.2)
0.50–0.80	49 (73.1)	25 (25.8)
0.80–0.90	5 (7.5)	24 (24.7)
>0.90	5 (7.5)	37 (38.1)
<i>Markers with acceptable performance<sup>b</sup></i>		
	45 (67.2)	71 (73.2)

<sup>a</sup>CVs and ICCs were calculated for 20 blinded duplicate samples for each specimen type that were placed across different batches.

<sup>b</sup>Acceptable performance was defined as: (i) being detectable in greater than 25% of the 100 samples on all 3 specimen types and (ii) across-batch CVs of less than 20% for blinded duplicates on at least 2 of the 3 specimen types.

ranged from 72% to 319%. A majority of markers (82 of 97 markers) had recoveries in the 80% to 120% range.

On serum, heparin plasma, and EDTA plasma samples, 89 markers each had detectable concentrations in greater than 25% of the 100 individuals (Table 1 and Fig. 1B). A high proportion of markers (75 on serum, 69 on heparin plasma, and 78 on EDTA plasma) had across-batch CVs of less than 20% (Table 1 and Fig. 2D–F). Similar to the Bio-Rad results, on each specimen type, within-batch CVs were generally less than across-batch CVs.

Combining detectability and across-batch CVs for the 3 specimen types, 71 of 97 markers had acceptable performance (Table 1). Markers with poor performance (<25% detection on at least 1 specimen type or across-batch CVs >20% on at least 2 specimen types) included: Eotaxin-3, Ghrelin, GM-CSF, IL-1B, IL-1 RA, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-9, IL-10, IL-12 p70, IL-13, IL-15, IL-17, IL-20, IL-21, IL-23, IL-28A, I-309, M-CSF, TGF- $\alpha$ , TNF-B, and CXCL1 (Table 2). Across the different Milli-

pore panels, 22 of 39 markers on cytokine panel 1, 17 of 23 markers on cytokine panel 2, 7 of 9 markers on cytokine panel 3, 9 of 10 markers on the metabolic panel, and all 3 markers on the acute-phase and all 13 markers soluble receptors panel had acceptable performance.

ICCs for across-batch duplicates ranged from 0.08 to 0.99, with 53, 53, and 61 markers on serum, heparin plasma, and EDTA plasma, respectively, having ICCs greater than 0.8 (Table 1).

#### Comparison of assay performance across specimen types and kits

We conducted comparisons of assay performance across the 3 specimen types and 2 kits for markers with acceptable performance (Table 3; 45 Bio-Rad markers, 71 Millipore markers, and 23 markers measured on both Bio-Rad and Millipore). On both Bio-Rad and Millipore, a majority of markers had similar percentage of detectability for T0 serum versus T0 heparin plasma as well as

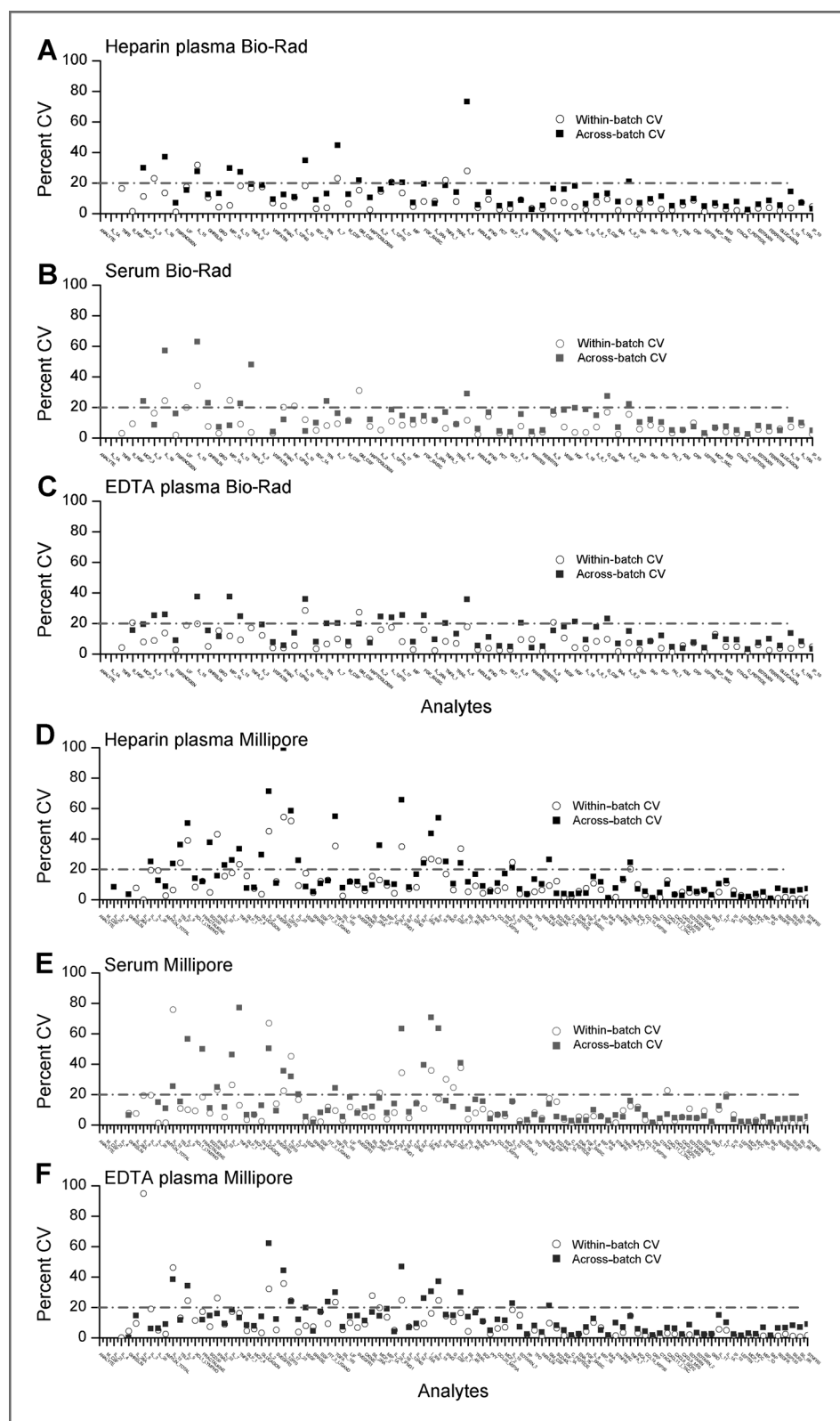


Figure 2. The CVs for 40 blinded duplicates, 20 placed within the same batch, and 20 placed across different batches. Results are shown separately for 67 Bio-Rad markers (A–C) and 97 Millipore markers (D–F) across 3 specimen types—heparin plasma, serum, and EDTA plasma.

Downloaded from <http://aacrjournals.org/cebp/article-pdf/20/9/1902/2273593/1902.pdf> by guest on 21 February 2024

for T3 EDTA plasma versus T3 heparin plasma. In contrast, for both Bio-Rad and Millipore, for a considerable number of markers, median cytokine concentrations dif-

fered between T0 serum versus T0 heparin plasma and between T3 EDTA plasma versus T3 heparin plasma (Table 3).

**Table 2.** Summary of performance of multiplexed markers on Bio-Rad and Millipore

Bio-Rad markers with acceptable performance <sup>a</sup>	Bio-Rad markers with <25% detectability <sup>b</sup>	Millipore markers with acceptable performance	Millipore markers with acceptable performance <sup>a</sup>	Millipore markers with <25% detectability <sup>b</sup>
A2M	B_NGF	Amylin_Total	MIP_1D	Eotaxin 3
CRP	GM_CSF	BCA_1	PP	Ghrelin
CTACK	IFNA2	CCL19_MIP3B	PYY	IL_20
C Peptide	IL_12P40	CCL20_MIP3A	SAA	IL_21
Eotaxin	IL_15	CKINE	SAP	IL_28a
Ferritin	IL_1A	CRP	SCD30	IL_3
FGF_Basic	IL_2	CTACK	SCD40L	IL_4
Fibrinogen	IL_3	CXCL11_I_TAC	SCF	IL_5
Ghrelin	LIF	CXCL6_GCP2	SDF_1A	M_CSF
GIP	MCP_3	CXCL9_MIG	SEGFR	XCL1_Lympho
GLP_1	TNFB	C_Peptide	SGP130	<b>Millipore markers with &gt;20% CV<sup>c</sup></b>
Glucagon		EGF	SILRII	
GRO	<b>Bio-Rad markers with &gt;20% CV<sup>c</sup></b>	ENA_78	SIL_1RI	
Haptoglobin		Eotaxin	SIL_2RA	GM_CSF
HGF		Eotaxin_2	SIL_4R	IL_10
IFNG	B_NGF	FGF_Basic	SIL_6R	IL_12P70
IL_16	GM_CSF	FIT_3_Ligand	SRAGE	IL_13
IL_18	G_CSF	Fractalkine	STNFRI	IL_15
IL 1RA	IL_10	GIP	STNFRII	IL_17
IL 2RA	IL_12P70	GLP_1	SVEGFR1	IL_1B
IL_6	IL_13	Glucagon	SVEGFR2	IL_1RA
IL_8	IL_15	GRO	SVEGFR3	IL_2
IL 9	IL_17	G_CSF	TARC	IL_21
Insulin	IL_1A	IFNA2	TNF- $\alpha$	IL_23
IP_10	IL 1B	IFNG	TPO	IL_28A
Leptin	IL_2	IL_11	TRAIL	IL_3
MCP_1MC	IL_4	IL_12P40	TSLP	IL_4
MIF	IL_5	IL_16	VEGF	IL_5
MIG	IL_7	IL_1A		IL_6
MIP_1B	LIF	IL_29_IFNG1		IL_7
M_CSF	MCP_3	IL_33		IL_9
PAL_1	MIP 1A	IL_8		I_309
PCT	TNF_B	INSULIN		M_CSF
PDGF_BB	TPA	IP_10		TGFA
Rantes		LEPTIN		TNF-B
Resistin		LIF		XCL1_Lympho
SAA		MCP_1		
SAP		MCP_2		
SCF		MCP_3		
SCGF_B		MCP_4		
SDF_1A		MDC		
TNF- $\alpha$		MIP_1A		
TRAIL		MIP_1B		
VEGF				
Visfatin				

<sup>a</sup>Acceptable performance was defined as (i) being detectable in greater than 25% of the 100 samples on all 3 specimen types and (ii) across-batch CVs of less than 20% for blinded duplicates on at least 2 of the 3 specimen types.

<sup>b</sup>Markers with less than 25% detectability on at least 1 of 3 (serum, heparin plasma, and EDTA plasma) specimen types.

<sup>c</sup>Markers with CVs for across-batch duplicates greater than 20% on 2 or more of 3 specimen types. CVs were calculated for 20 blinded duplicate samples for each specimen type that were placed across different batches.

**Table 3.** Comparison of assay performance across specimen types and kits<sup>a</sup>

Comparison	% Detectability number of markers <sup>b</sup>	Median concentration number of markers <sup>c</sup>
<i>Bio-Rad (45 markers)</i>		
T0 serum vs. T0 heparin plasma		
Significantly higher in serum	7	16
Significantly higher in plasma	10	19
Similar	21	10
Not evaluable <sup>d</sup>	7	
T3 EDTA plasma vs. T3 heparin plasma		
Significantly higher in EDTA	2	7
Significantly higher in heparin	2	12
Similar	32	26
Not evaluable	9	
<i>Millipore (71 markers)</i>		
T0 serum vs. T0 heparin plasma		
Significantly higher in serum	4	18
Significantly higher in plasma	5	13
Similar	51	40
Not evaluable	11	
T3 EDTA plasma vs. T3 heparin plasma		
Significantly higher in EDTA	5	3
Significantly higher in heparin	1	30
Similar	49	38
Not evaluable	16	
<i>Bio-Rad vs. Millipore (23 markers)</i>		
T0 Serum		
Significantly higher in Bio-Rad	2	15
Significantly higher in Millipore	2	7
Similar	10	1
Not Evaluable	9	
T0 heparin plasma		
Significantly higher in Bio-Rad	4	16
Significantly higher in Millipore	4	7
Similar	6	0
Not evaluable	9	
T3 EDTA plasma		
Significantly higher in Bio-Rad	3	15
Significantly higher in Millipore	3	8
Similar	7	0
Not evaluable	10	

<sup>a</sup>Comparisons were restricted to markers with acceptable performance—45 Bio-Rad markers, 71 Millipore markers, and 23 markers measured using both kits.

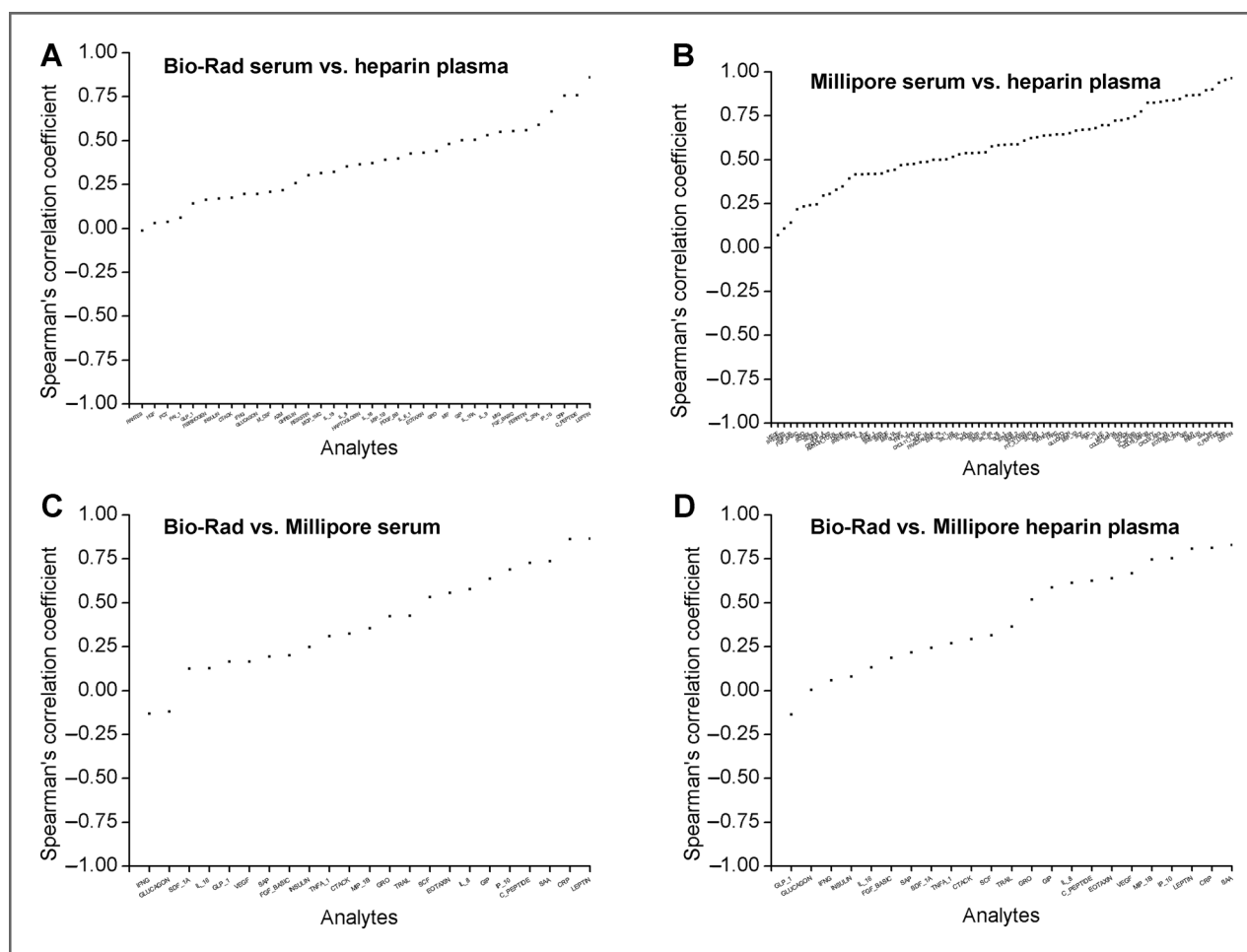
<sup>b</sup>Comparisons of percentage of detectability across specimen types and kits were conducted using the McNemar's test.

<sup>c</sup>Comparisons of median observed concentrations across specimen types and kits were conducted using the Wilcoxon signed-rank test. Observations below the assay's lower limit were excluded.

<sup>d</sup>McNemar's *P* value could not be evaluated because one of the groups had 100% detectability.

For 45 Bio-Rad markers with acceptable performance (Fig. 3A), correlation coefficients between T0 serum and T0 heparin plasma were less than 0.5 for 33 markers, 0.5 to 0.75 for 9 markers, and 0.75 or greater for 3 markers. For 71 Millipore markers with acceptable performance (Fig. 3B),

correlation coefficients between T0 serum and T0 heparin plasma were less than 0.5 for 25 markers, 0.5 to 0.75 for 31 markers, and 0.75 or greater for 15 markers. In variance components analyses, a difference less than 10% in ICCs between T0 serum and T0 heparin plasma was observed



**Figure 3.** Spearman's correlation coefficients are shown for comparisons of the rank order of marker concentrations between T0 serum and T0 heparin plasma on Bio-Rad (A) and on Millipore (B). Comparisons between Bio-Rad and Millipore for T0 serum (C) and T0 heparin plasma (D) are also shown. T0 denotes baseline visit in the PLCO study.

for 16 of 39 evaluable Bio-Rad markers with acceptable performance and for 35 of 67 evaluable Millipore markers with acceptable performance (Fig. 4A and B).

Across the 3 specimen types, percentage of detectability and median concentrations were significantly different between Bio-Rad and Millipore for a majority of the 23 markers with acceptable performance. Likewise, for all 3 specimen types (Fig. 3C and D), correlation coefficients between Bio-Rad and Millipore were low (for T0 serum:  $<0.5$  for 12 markers,  $0.5\text{--}0.75$  for 7 markers, and  $\geq 0.75$  for 4 markers; for T0 heparin plasma:  $<0.5$  for 14 markers,  $0.5\text{--}0.75$  for 7 markers, and  $\geq 0.75$  for 2 markers; for T3 EDTA plasma:  $<0.5$  for 15 markers,  $0.5\text{--}0.75$  for 4 markers, and  $\geq 0.75$  for 4 markers). In variance components analyses, ICCs differed between Bio-Rad and Millipore for a majority of markers (of 23 acceptable markers on both kits, 7 of 20 evaluable markers on T0 serum, 8 of 21 evaluable markers on T0 heparin plasma, and 7 of 19 evaluable markers on T3 EDTA plasma had  $<10\%$  difference in ICCs; Fig. 4C and D).

## Discussion

In this large methodologic study, we show that a majority of multiplexed inflammation, immune, and metabolic markers can be measured reliably in serum and plasma specimens, as evidenced by low CVs and high ICCs, on both Bio-Rad and Millipore. Median analyte concentrations and ICCs differed to a small extent across specimen types and to a large extent between Bio-Rad and Millipore. Likewise, correlations in analyte levels were moderate to high across specimen types but were low between the 2 commercial kits.

Our results underscore the utility of multiplexed technologies for large-scale investigations into the role of inflammation and immune dysregulation in the etiology of cancer and other diseases. Notably, the 45 markers on Bio-Rad and 71 markers on Millipore with good detectability and reproducibility include several components of the inflammation and immune response such as proinflammatory markers (e.g., IL-8, TNF- $\alpha$ , IFNG, GRO),



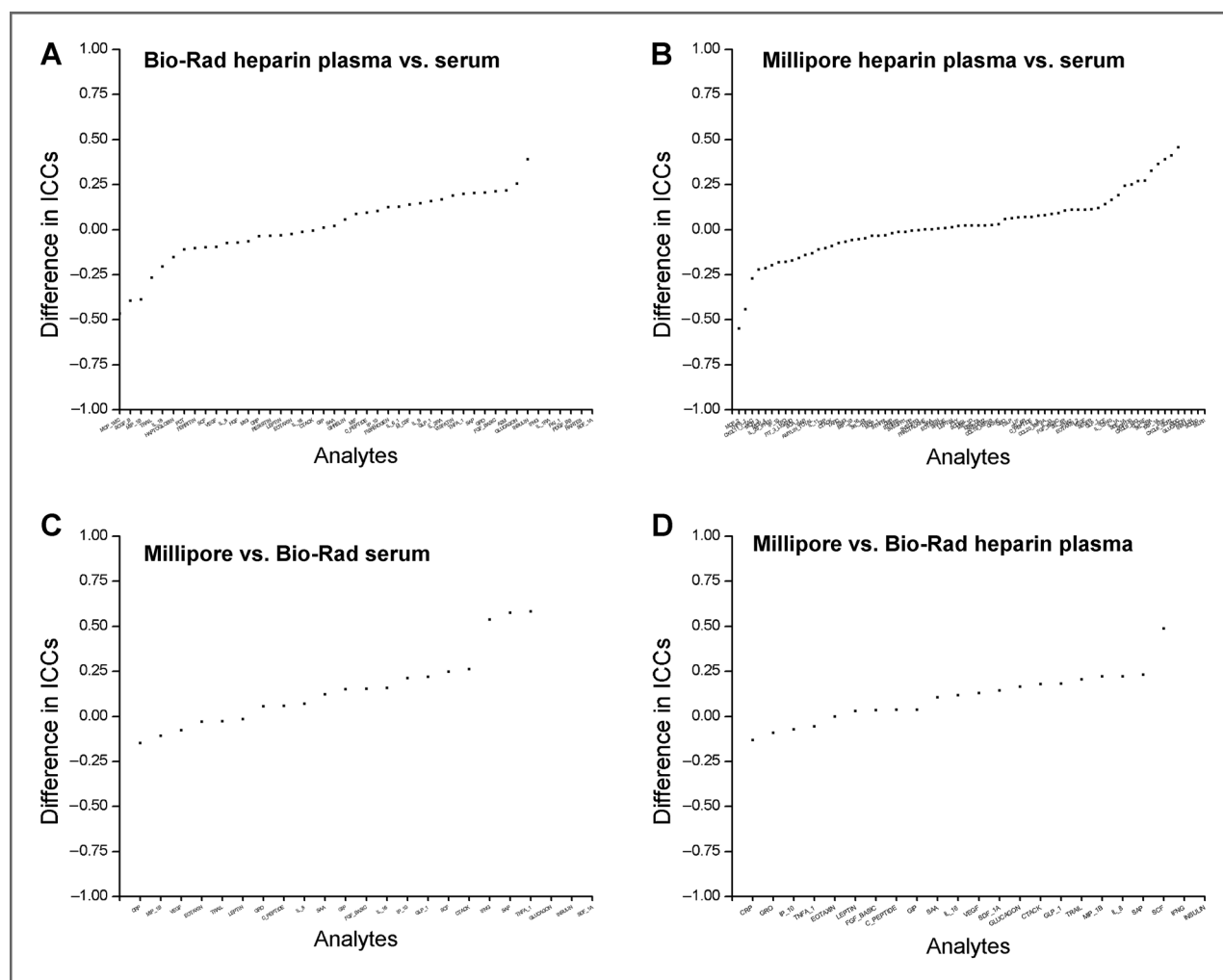


Figure 4. Differences in ICCs between T0 heparin plasma versus T0 Serum for Bio-Rad (A) and for Millipore (B). Differences in ICCs between Millipore versus Bio-Rad for T0 serum (C) and T0 heparin plasma (D) are also shown. ICCs were estimated using variance components analyses.

anti-inflammatory markers (e.g., IL-16), acute-phase proteins [e.g., CRP, serum amyloid A (SAA)], and growth and angiogenesis factors (e.g., FGF, VEGF). Reliable detection of these markers in serum and plasma samples provides the opportunity to comprehensively evaluate the role of immunity and inflammation in cancer etiology in cohort and case-control studies. Furthermore, the redundant and pleiotropic nature of most inflammation markers provides the opportunity to evaluate the association of groups of markers (defined through principal components or factor analyses) with cancer risk (9).

Despite the large number of markers with acceptable performance, classic Th1-type markers such as IL-2, IL-12, and IL-15 and Th2-type markers such as IL-4, IL-10, and IL-13 had a low proportion of samples with detectable levels, unacceptably high CVs, or low ICCs. Notably, a majority of these markers were included in panels with higher numbers of markers, and we found that assay performance decreased with increasing number of mar-

kers on a panel. For example, 17 markers (43%) on Millipore's 39-plex panel and 15 markers (55%) on Bio-Rad's 27-plex panel had poor detectability and/or reproducibility. Because markers, such as IL-2 and IL-10, from the same vendors had acceptable performance on previous studies which simultaneously measured a limited number of markers (9, 17), it is likely that interference from other markers affected the performance of these markers.

Measurement of circulating inflammation markers is potentially sensitive to several factors such as specimen types, sample handling, and processing methods (18, 19). Previous studies have reported that marker measurements are not interchangeable between serum and plasma samples (9), and these differences are believed to arise from factors such as degradation of markers during the process of clotting and degranulation of granulocytes (9). Consistent with these studies, we found that on both Bio-Rad and Millipore, for a considerable number

of markers, median analyte concentrations and ICCs differed between serum versus heparin plasma and EDTA plasma versus heparin plasma. In addition, irrespective of the specimen type, for a majority of markers, percentage of detectability, median concentrations, and ICCs differed between Bio-Rad and Millipore. Therefore, our observations indicate that results from different studies utilizing different specimen types and different multiplexed kits may not be directly comparable (20).

Circulating levels of inflammation, immune, and metabolic markers are also influenced by several demographic and behavioral characteristics such as age, sex, race, smoking, body mass index, and diet (21). Therefore, in separate studies, we are currently evaluating predictors of an inflammatory response for single markers as well as empirical groupings of markers and the temporal stability of markers with acceptable performance. The temporal stability of circulating markers is largely unknown, and single time-point measurements in prospective cohort studies could bias results to the null for unstable markers (22).

Our study has several strengths, including the standardized collection, processing, and storage of specimens in the PLCO study (16) and comprehensive evaluation of more than 100 multiplexed markers on different specimen types. We also note the limitations of our study. Importantly, our study focused on reliability, but not

validity, of marker measurements. Nevertheless, previous studies comparing the performance of multiplexed marker measurements with ELISA assays show high validity (13–15). Finally, we defined less than 25% detectability as poor performance, in part, because samples with low detection levels are generally accompanied by unacceptably high CVs and low ICCs. However, we note the possibility that some markers could be expressed only in disease conditions and therefore could be informative for disease associations.

In conclusion, our key observation was that Bio-Rad and Millipore multiplexed markers are broadly reproducible and can therefore be utilized for large-scale epidemiologic studies. Our results highlight the opportunity to comprehensively evaluate the role of a large number of circulating inflammatory markers representative of a range of immune-related processes and pathways in cancer etiology and prognosis.

### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received March 1, 2011; revised June 18, 2011; accepted June 20, 2011; published OnlineFirst June 29, 2011.

### References

- Moss SF, Blaser MJ. Mechanisms of disease: inflammation and the origins of cancer. *Nat Clin Pract Oncol* 2005;2:90–7.
- Thun MJ, Henley SJ, Gansler T. Inflammation and cancer: an epidemiological perspective. *Novartis Found Symp* 2004;256:6–21.
- Coussens LM, Werb Z. Inflammation and cancer. *Nature* 2002;420:860–7.
- de Visser KE, Coussens LM. The inflammatory tumor microenvironment and its impact on cancer development. *Contrib Microbiol* 2006;13:118–37.
- Heikkilä K, Harris R, Lowe G, Rumley A, Yarnell J, Gallacher J, 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* 2009;20:15–26.
- Eishal MF, McCoy JP. Multiplex bead array assays: performance evaluation and comparison of sensitivity to ELISA. *Methods* 2006;38:317–23.
- de JW, Rijkers GT. Solid-phase and bead-based cytokine immunoassay: a comparison. *Methods* 2006;38:294–303.
- Hosnijeh FS, Krop EJ, Portengen L, Rabkin CS, Linseisen J, Vineis P, et al. Stability and reproducibility of simultaneously detected plasma and serum cytokine levels in asymptomatic subjects. *Biomarkers* 2010;15:140–8.
- Wong HL, Pfeiffer RM, Fears TR, Vermeulen R, Ji S, Rabkin CS. Reproducibility and correlations of multiplex cytokine levels in asymptomatic persons. *Cancer Epidemiol Biomarkers Prev* 2008;17:3450–6.
- Gu Y, Zeleniuch-Jacquotte A, Linkov F, Koenig KL, Liu M, Velikhathnaya L, et al. Reproducibility of serum cytokines and growth factors. *Cytokine* 2009;45:44–9.
- Chowdhury F, Williams A, Johnson P. Validation and comparison of two multiplex technologies, Luminex(R) and Mesoscale Discovery, for human cytokine profiling. *J Immunol Methods* 2009;340:55–64.
- Khan SS, Smith MS, Reda D, Suffredini AF, McCoy JP Jr. Multiplex bead array assays for detection of soluble cytokines: comparisons of sensitivity and quantitative values among kits from multiple manufacturers. *Cytometry B Clin Cytom* 2004;61:35–9.
- Djoba Siawaya JF, Roberts T, Babb C, Black G, Golakai HJ, Stanley K, et al. An evaluation of commercial fluorescent bead-based luminex cytokine assays. *PLoS One* 2008;3:e2535.
- dupont NC, Wang K, Wadhwa PD, Culhane JF, Nelson EL. Validation and comparison of luminex multiplex cytokine analysis kits with ELISA: determinations of a panel of nine cytokines in clinical sample culture supernatants. *J Reprod Immunol* 2005;66:175–91.
- Fu Q, Zhu J, Van Eyk JE. Comparison of multiplex immunoassay platforms. *Clin Chem* 2010;56:314–8.
- Hayes RB, Sigurdson A, Moore L, Peters U, Huang WY, Pinsky P, et al. Methods for etiologic and early marker investigations in the PLCO trial. *Mutat Res* 2005;592:147–54.
- Kemp TJ, Hildesheim A, Garcia-Pineres A, Williams MC, Shearer GM, Rodriguez AC, et al. Elevated systemic levels of inflammatory cytokines in older women with persistent cervical human papillomavirus infection. *Cancer Epidemiol Biomarkers Prev* 2010;19:1954–9.
- Aziz N, Nishanian P, Mitsuyasu R, Detels R, Fahey JL. Variables that affect assays for plasma cytokines and soluble activation markers. *Clin Diagn Lab Immunol* 1999;6:89–95.
- Twoorger SS, Hankinson SE. Use of biomarkers in epidemiologic studies: minimizing the influence of measurement error in the study design and analysis. *Cancer Causes Control* 2006;17:889–99.
- Toedter G, Hayden K, Wagner C, Brodmerkel C. Simultaneous detection of eight analytes in human serum by two commercially available platforms for multiplex cytokine analysis. *Clin Vaccine Immunol* 2008;15:42–8.
- O'Connor MF, Irwin MR. Links between behavioral factors and inflammation. *Clin Pharmacol Ther* 2010;87:479–82.
- Platz EA, Sutcliffe S, De Marzo AM, Drake CG, Rifai N, Hsing AW, et al. Intra-individual variation in serum C-reactive protein over 4 years: an implication for epidemiologic studies. *Cancer Causes Control* 2010;21:847–51.