

## Need for an External Proficiency Testing Program for Cytokines, Chemokines, and Plasma Markers of Immune Activation

JOHN L. FAHEY,<sup>1\*</sup> NAJIB AZIZ,<sup>1</sup> JOHN SPRITZLER,<sup>2</sup> SUSAN PLAEGER,<sup>1†</sup> PARUNAG NISHANIAN,<sup>1</sup> JANET L. LATHEY,<sup>3†</sup> JOAN SEIGEL,<sup>4</sup> ALAN L. LANDAY,<sup>4</sup> RAKHI KILARUI,<sup>5</sup> JOHN L. SCHMITZ,<sup>5</sup> CARMEN WHITE,<sup>6†</sup> DIANE W. WARA,<sup>6†</sup> ROBERT AKRIDGE,<sup>7</sup> JOIE CUTILI,<sup>8†</sup> STEVEN D. DOUGLAS,<sup>8†</sup> JAMES REUBEN,<sup>9†</sup> WILLIAM T. SHEARER,<sup>9†</sup> MUSTAFA NOKTA,<sup>10</sup> RICHARD POLLAND,<sup>10</sup> ROBERT SCHOOLEY,<sup>11</sup> DESHRATN ASTHANA,<sup>12</sup> YAFFA MIZRACHI,<sup>13</sup> AND MYRON WAXDAL<sup>14</sup>

*University of California, Los Angeles,<sup>1</sup> University of California, San Diego,<sup>3</sup> and University of California, San Francisco,<sup>6</sup> Statistical and Data Analysis Center, Harvard School of Public Health, Boston, Massachusetts<sup>2</sup>; Rush-Presbyterian-St. Luke's Medical Center, Chicago, Illinois<sup>4</sup>; University of North Carolina, Chapel Hill, North Carolina<sup>5</sup>; University of Washington, Seattle, Washington<sup>7</sup>; Children's Hospital of Philadelphia, Philadelphia Pennsylvania<sup>8</sup>; MD Anderson Cancer Center and Baylor College of Medicine, Houston,<sup>9</sup> and University of Texas Medical Branch, Galveston,<sup>10</sup> Texas; University of Colorado Health Sciences Center, Denver, Colorado<sup>11</sup>; University of Miami School of Medicine, Miami, Florida<sup>12</sup>; Albert Einstein Medical Center New York, New York<sup>13</sup>; and FAST Systems, Gaithersburg, Maryland<sup>14</sup>*

Received 2 November 1999/Returned for modification 3 February 2000/Accepted 18 March 2000

**An external evaluation program for measuring the performance of laboratories testing for cytokines and immune activation markers in biological fluids was developed. Cytokines, chemokines, soluble cytokine receptors, and other soluble markers of immune activation (CSM) were measured in plasma from a healthy human immunodeficiency virus (HIV)-seronegative reference population and from HIV-seropositive individuals as well as in supernatant fluids from in vitro-stimulated human immune cells. The 14 components measured were tumor necrosis factor (TNF) alpha, gamma interferon, interleukin-1 (IL-1), IL-2, IL-4, IL-6, IL-10, Rantes, MIP-1a, MIP-1β, soluble TNF receptor II, soluble IL-2 receptor alpha, β<sub>2</sub>-microglobulin, and neopterin. Twelve laboratories associated with the Adult and Pediatric AIDS Clinical Trial Groups participated in the study. The performance features that were evaluated included intralaboratory variability, interlaboratory variability, comparison of reagent sources, and ability to detect CSM in the plasma of normal subjects as well as the changes occurring in disease. The principal findings were as follows: (i) on initial testing, i.e., before participating in the program, laboratories frequently differed markedly in their analytic results; (ii) the quality of testing of a CSM in individual participating laboratories could be assessed; (iii) most commercial kits allowed distinction between normal and abnormal plasma CSM levels and between supernatants of stimulated and unstimulated cells; (iv) different sources of reagents and reference standards frequently provided different absolute values; (v) inexperienced laboratories can benefit from participating in the program; (vi) laboratory performance improved during active participation in the program; and (vii) comparability between analyses conducted at different sites can be ensured by an external proficiency testing program.**

The appearance of new diseases, therapies, or technologies often leads to new clinical laboratory measurements. This is often accompanied by novel instrumentation and reagents. Under these circumstances and until laboratory procedures and reagents are standardized and laboratory leaders and technologists have been trained, laboratories may differ substantially in analytic results with comparable samples, with resultant confusion and possible misinterpretation of clinical status. Thus, external performance evaluation programs are usually introduced to achieve better performance and comparability in laboratory testing.

A good example occurred early in the spread of human immunodeficiency virus (HIV) infection and AIDS. Measurements of CD4 T-cell levels were found to be central to evalu-

ation of disease course and therapeutic decisions. However, no national proficiency testing procedures were in place. AIDS patients who were tested for CD4 levels at different locations often reported that laboratories differed substantially in their analytic results obtained by flow cytometry. Thus, investigator-initiated programs such as the Multicenter AIDS Cohort Study (MACS) (7) and the AIDS Clinical Trials Group (ACTG) (15) under National Institute of Allergy and Infectious Diseases (NIAID) auspices instituted successful external proficiency testing programs that achieved satisfactory comparability of peripheral blood lymphocyte subset measurements by flow cytometry. Subsequently, the Centers for Disease Control and Prevention and the College of American Pathologists introduced flow cytometry proficiency testing programs. More recently, an international program for quality assurance and standardization of CD4, CD8, and CD3 measurements (the QASI program) has been instituted (12). The ACTG has also instituted a performance evaluation program for quantitative HIV assays (11).

Immune system activation is increasingly recognized as a

\* Corresponding author. Mailing address: Department of Microbiology and Immunology, UCLA School of Medicine, Los Angeles, CA 90095-1747. Phone: (310) 825-6568. Fax: (310) 206-1318. E-mail: jlfahay@microimmun.medsch.ucla.edu.

† Affiliated with Pediatric ACTG Immunology Laboratories.

TABLE 1. Reporting sites<sup>a</sup>

Site no.	Site	Testing reports for study:		
		I	II	III
1	Rush Medical Center, Chicago, Ill.	—	+	+
2	UCLA	+	+	+
3	University of California, San Diego	+	+	+
4	University of Texas Medical Branch, Galveston	+	+	+
5	North Carolina Memorial Hospital	+	+	—
6	M.D. Anderson Cancer Center	+	—	+
7	Children's Hospital, Philadelphia, Pa.	+	+	+
8	University of California, San Francisco	+	—	—
9	University of Washington Fred Hutchinson Cancer Research Center	—	+	—
10	University of Colorado Health Sciences Center	—	—	+
11	University of Miami School of Medicine	+	—	—
12	Albert Einstein College of Medicine	+	—	—

<sup>a</sup> Laboratories 2, 3, 6, 7, and 8 participated in the Pediatric ACTG program. Laboratories 1, 2, 4, 5, 9, and 10 participated in the Adult ACTG program. Laboratory 2 participated in both the Adult and Pediatric ACTG programs. Laboratories 11 and 12 were added at the request of DAIDS, NIAID.

significant component of many diseases. Autoimmune disorders with activation include rheumatoid arthritis (10), inflammatory bowel disease (17), and multiple sclerosis (8). Immune activation has also been shown to be characteristic of aging, depression, and possibly some forms of chronic fatigue syndrome and fibromyalgia (9).

The role of immune activation in the pathogenesis of HIV and AIDS is receiving increasing attention. Cytokine levels in body fluids are elevated, as are soluble products of cytokine activity such as neopterin,  $\beta_2$ -microglobulin ( $\beta_2M$ ), and cytokine receptors (6). Furthermore, elevated levels of activation markers in plasma have been shown to be excellent prognostic factors in HIV infection, providing data comparable to but distinct from those provided by CD4 T-cell measurements or by viral load assays (5).

An external proficiency testing program for measurement of neopterin and  $\beta_2M$ , two important markers of immune activation, was instituted among the four centers participating in the Multicenter AIDS Cohort Study. Over a period of several years, shipments were distributed to participating laboratories, with general agreement among three laboratories for both  $\beta_2M$  and neopterin assays. One laboratory, however, was not able to master one of the assays or to obtain results that were consistent with those from the other three sites. Such discrepancies adversely affected the prognostic usefulness of such laboratory measurements (5). This prior experience emphasized the need for external proficiency testing programs to verify laboratory performance and to assess the suitability of various reagent sources to meet the needs of patients and physicians dealing with immune disorders.

In a separate series of earlier studies at the University of California at Los Angeles (UCLA) (1, 2) many factors were found to influence the outcome of assays for tumor necrosis factor alpha (TNF- $\alpha$ ), gamma interferon (IFN- $\gamma$ ), neopterin, soluble TNF receptor II (sTNF-RII), soluble interleukin-2 receptor alpha (sIL-2R), and  $\beta_2M$ . Substantial differences in apparent levels of analytes were frequently found when ELISA kits from different manufacturers were used (2). Furthermore, the analytic results from different lots of ELISA kits supplied by a single manufacturer occasionally differed by as much as 50%. In some cases, differences were found in the standards provided by separate manufacturers (2). In addition, it was demonstrated that many cytokines and products/markers of immune activation were stable on frozen storage and could be shipped to participating laboratories. Thus, batch testing of frozen stored samples is feasible. The findings indicated that

for longitudinal studies, the levels of cytokines and immune activation markers in plasma or serum should be measured using preverified reagents from one manufacturer. Furthermore, proficiency testing and external quality assurance programs can help to develop a needed consensus.

The need for uniformity in the standards for quantitative assays is clearly apparent. International reference standards are available for many cytokines (13, 14, 16) but are not available for soluble cytokine receptors or soluble activation markers. However, a 1995 report (4) noted substantial differences in terms of sensitivity and results in 11 laboratories using a variety of assays for TNF- $\alpha$ . Also, an earlier study (3) described substantial differences in commercial reagents and standards provided in ELISA kits for IL-2, IL-6, and TNF- $\alpha$  in 1992. That report (3) ended with a plea for "real standardization of immune assays for cytokine quantitation." Progress and problems in this area were reviewed in 1997 (19).

By 1996, the Adult and Pediatric ACTG programs had established more than 15 immunology laboratories to evaluate immunologic parameters relevant to HIV infection and its therapy. Recommendations were made to the ACTG and the Division of AIDS (DAIDS), NIAID, that an external proficiency testing program be tried as a quality assurance procedure for laboratory performance of cytokine and activation marker measurements. Such a program was initiated in January 1997 and terminated in January 1999. Three separate shipments of biological fluids were carried out. Plasma as well as supernatant fluids from stimulated immune cells were evaluated. Replicate samples were included to evaluate intralaboratory variability. Assays for cytokines included TNF- $\alpha$ , IFN- $\gamma$ , IL-1, IL-2, IL-4, IL-6, IL-10, and the chemokines Rantes, MIP-1 $\alpha$ , and MIP-1 $\beta$ . Assays for levels of immune activation markers in plasma included neopterin,  $\beta_2M$ , sTNF-RII, and sIL-2R. A total of 11 laboratories participated. Quite remarkable differences between laboratories became apparent in the analyses of the first shipment. Significant problems were uncovered. When these were addressed, more-uniform results were obtained. The value of an external proficiency testing program was documented.

#### MATERIALS AND METHODS

**Participation.** All of the Advance Technology Laboratories for the Adult ACTG and the Immunology Research Laboratories of the Pediatric ACTG were repeatedly invited to participate. Those that elected to participate and reported results from the Adult ACTG program are listed in Table 1. The same laboratory at UCLA (the Clinical Immunology Research Laboratory, Center for Interdis-

disciplinary Research in Immunology and Disease) participated in both the Adult ACTG and Pediatric ACTG programs. The University of Miami School of Medicine and Albert Einstein Medical Center, New York, N.Y., participated briefly at the invitation of DAIDS, NIAID. Participation was defined as contributing analytic data on one or more of the sample shipments. Conference calls were held regularly, starting in March 1997, by the ACTG Cytokine and Soluble Marker (CSM) Focus Group, which constituted an advisory group for this program.

Initial decisions were as follows. (i) EDTA plasma samples would be obtained from both normal and HIV-positive individuals. (ii) Initially, the plasma cytokines to be tested were TNF- $\alpha$  and IFN- $\gamma$  and the soluble markers of activation were neopterin, sTNF-RII,  $\beta_2$ M, and sIL-2R. Subsequently, tests for IL-1 $\alpha$ , IL-2, IL-4, IL-6, IL-10, and Rantes, MIP-1 $\alpha$ , and MIP-1 $\beta$  were added. (iii) A number of samples, including replicates, would be sent by FAST Systems, Inc. (Gaithersburg, Md.) to each participating ACTG laboratory for testing. Samples from HIV-negative and HIV-positive donors would be included, replicate samples would be randomly distributed, and sample identification would be nondescriptive. (iv) Laboratories would include these samples in their testing repertoire and were not expected to initiate new assays or do assays for all of the markers in the program but to concentrate on the ones which they were currently testing. (v) Laboratories would indicate the source of reagents and the type of methodology used for each test. (vi) Data would be reported directly to FAST Systems and transferred to the ACTG Statistical and Data Analysis Center (SDAC), Harvard School of Public Health, Boston, Mass., for evaluation of the data distribution from individual laboratories, comparison of results from various reagent sources, and consistency of individual laboratories on analyses of replicate samples. (vii) Participating laboratories would receive a comprehensive report on the analytic data obtained for each shipment. (viii) After the results of shipments were analyzed and discussed, further plans for additional testing would be formulated, appropriate samples would be obtained, and a new batch of proficiency testing samples would be distributed. (ix) Subsequently, decisions were made to add supernatants from phytohemagglutinin (PHA)- and from lipopolysaccharide (LPS)-stimulated whole blood samples and from separated peripheral blood mononuclear cells (PBMC) for testing.

**Implementation.** Large volumes of plasma were obtained by FAST Systems from a number of seropositive and seronegative individuals. Levels of selected cytokines and soluble markers were determined by the procedures and with the reagents available at UCLA so that a range of levels could be selected. The results were reported to FAST Systems, where a new coding system (known only to them and to John Spritzler, SDAC) was introduced. John Spritzler (SDAC) and Myron Waxdal (FAST Systems) determined the composition of the shipments, and 1-ml aliquots were prepared for shipment.

Individuals responsible for receiving samples at each site were identified. A preliminary notice was sent about 10 days before shipping with a request for a signed response. This is a legal requirement because of the nature of the shipment of infectious materials.

**Quantitation of plasma levels of cytokines and soluble activation markers.**  $\beta_2$ -Microglobulin was quantified using microparticle enzyme immunoassay (microparticle EIA) (Abbott Laboratories, Abbott Park, Ill.) and enzyme immunoassay (EIA) kit (Coulter, Miami, Fla.). Neopterin was measured with a competitive EIA kit (ELI test; BRAHMS, Berlin, Germany). sIL-2R was determined with EIA kits (Endogen Inc., Cambridge, Mass., and Immunotech, Marseilles, France). sTNF-RII was quantitated in plasma at a 1:20 dilution by using EIA kits (HyCult, Uden, The Netherlands; R&D Systems, Minneapolis, Minn.; and Medgenix, Fleurus, Belgium). TNF- $\alpha$  was measured with EIA kits (Medgenix; Innogenetics, Zwijndrecht, Belgium; Endogen; Biosource International; and Genzyme). IL-10 was measured with EIA kits (from Immunotech, from Endogen and from Biosource International). IFN- $\gamma$  was determined by using EIA kits with and without the CIRID at UCLA modification of the manufacturer's protocol (Immunotech, Endogen, Biosource International, Genzyme, T-Cell Diagnostics, and R&D Systems). IL-2, IL-4, IL-6, and RANTES were each measured with EIA kits from Endogen and from Biosource International. MIP-1 $\alpha$  and MIP-1 $\beta$  were measured with R&D System's Quantikine EIA kits. All assays were performed according to the manufacturer's instructions.

**Shipments.** (i) On 30 September 1997, six aliquots of plasma were sent to participating laboratories. There were three aliquots (replicates) of a single normal plasma and one sample each from three HIV-seropositive individuals. Reports were collected in October, November, and December 1997. Analyses included TNF- $\alpha$ , IFN- $\gamma$ , IL-2, IL-10,  $\beta_2$ M, neopterin, sTNF-RII, and sIL-2R. The data were analyzed at SDAC (Harvard School of Public Health) and shared with participating laboratories in the following months.

(ii) On 30 March 1998, 10 supernatant fluids from PHA-, LPS-, or mock-stimulated whole blood or separated PBMC from HIV-seropositive and HIV-seronegative individuals (which had been prepared at UCLA and transferred to FAST Systems for coding and aliquoting) were shipped to participating sites. Analyses were conducted and data were collected in April, May, June, and July 1998, evaluated, and shared with participating laboratories.

(iii) On 30 September 1998, 12 samples were shipped to participating laboratories. These included six plasma samples and six stimulated lymphoid cell supernatant fluid samples from HIV-seronegative and HIV-seropositive individuals. Three replicates were included in the stimulated cell supernatant group.

Data were collected in October, November, and December 1998 by FAST Systems and analyzed at SDAC.

**Statistical procedures.** Analytic data from participating sites were collected by FAST Systems in the months after each shipment and transferred to SDAC, Harvard School of Public Health. Results that were below the limit of detection were indicated as zero on the plots but were excluded from calculations of means and coefficients of variation.

## RESULTS

**Study I: HIV-seronegative and -seropositive plasma samples.** Results were received from nine laboratories (Table 1). Three laboratories tested six analytes, four laboratories tested three analytes, one laboratory reported two assays, and one laboratory reported only one assay. Replicate samples (three) of an HIV-seronegative plasma were included in the shipment. These results are plotted in Fig. 1 as the first three samples, but they were in fact not distributed in sequence among the six samples. Mean levels of tested analytes at different laboratories are presented in Fig. 2A, and calculated coefficients of variation are shown in Table 2. Three HIV-seropositive plasma samples with various degrees of abnormalities are included in Fig. 1 as well.

Examples of the findings are illustrated in Fig. 1. Two laboratories were able to measure plasma neopterin levels in normal samples and obtained similar data in three replicates (Fig. 1A). Also, these two laboratories determined that the HIV-positive samples had a higher neopterin content than the HIV-negative samples, and they were able to detect differences between the three HIV-positive samples. These three features are evidence of good laboratory performance. However, a striking difference between the quantitative neopterin data is evident (Fig. 1A). This could be due to reagent features or differences in the reference standards provided by the manufacturer (11). Similar findings were noted for  $\beta_2$ M when results from three laboratories were evaluated (data not shown).

Somewhat similar findings were evident in four laboratories (B, C, E, and F) testing for sTNF-RII (Fig. 1B); e.g., there were consistent values for replicate samples, higher levels in HIV-positive than in HIV-negative samples, but great differences in quantitative results between labs. Furthermore, one laboratory (F) was barely able to detect sTNF-RII in the HIV-positive samples. In the sIL-2R analyses, six of seven laboratories detected sIL-2R in normal plasma and all labs detected higher levels in HIV-positive samples, but quantitative agreement was poor (data not shown).

Problems were more evident in TNF- $\alpha$  testing (Fig. 1C), where four laboratories (A, D, F, and H) were unable to detect this analyte in the HIV-negative sample. Two laboratories (A and D) could not detect TNF- $\alpha$  in any of the HIV-positive samples, and another lab (F) could detect it in only one sample. In contrast, two laboratories (C and E) had consistent values for the replicate samples and identified appreciable elevations in the HIV-positive samples. However, one laboratory (B) did not find differences between the HIV-negative and HIV-positive samples. Similar problems were seen with IFN- $\gamma$ , where five of seven laboratories could not detect this in any samples (Fig. 2A). Several of these laboratories were just instituting these tests and had procedural or reagent problems which were identified during group discussions with staff at the more experienced sites. On further testing with improved procedures and/or with use of more appropriate reagent sources, results at these sites became comparable to those of other laboratories.

The CSM results in study I were evaluated by conference call and at a CSM study team gathering at an Advance Technology Laboratories meeting. The performance of different test re-

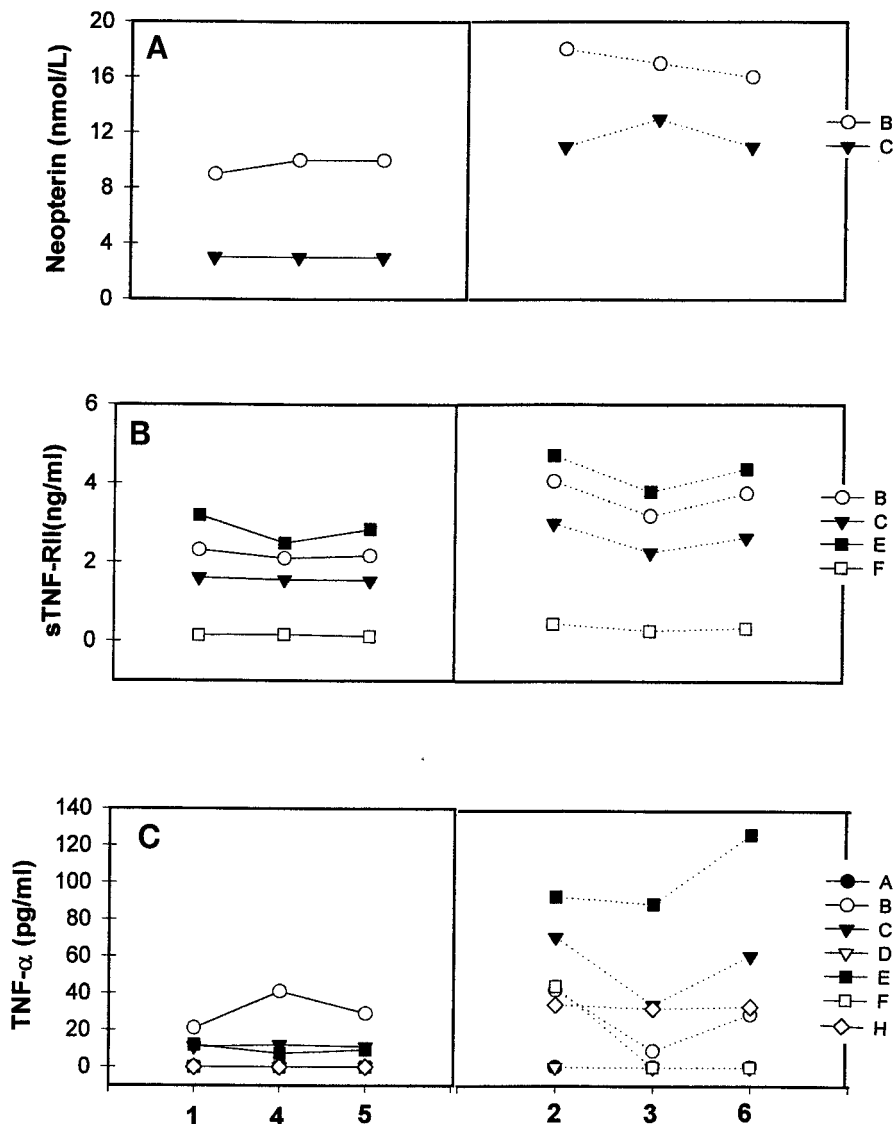


FIG. 1. Analytic data from study I for neopterin (A), sTNF-RII (B), and TNF- $\alpha$  (C) measured in several laboratories (designated by capital letters). Three aliquots (triplicates) of plasma (samples 1, 4, and 5) from a single seronegative donor were included, and the data are grouped on the left. Results for three plasma samples (samples 2, 3, and 6) from different HIV-seropositive donors are presented on the right.

agents was evaluated for sensitivity, accuracy, and reproducibility (1, 2). For instance, although the manufacturer's reported sensitivity for all reagents was at or below 5 pg/ml, significant differences were noted when the reagents were tested in this study. The following reagent sources were recommended as preferred reagent sources for plasma testing of cytokines and soluble activation markers: TNF, Medgenix; IFN- $\gamma$ , Immunotech; sTNF-RII, HyCult and Medgenix; sIL2R, Endogen and Immunotech,  $\beta_2$ M, Abbott microparticle EIA (non-ELISA) procedures; and neopterin, ELI test (BRAHMS). It must be noted that not all available reagent sources were tested or compared. Also, the more sensitive versions of TNF- $\alpha$  and IFN- $\gamma$  assays were not evaluated.

**Study II: supernatant fluids.** A total of 10 samples were shipped to seven sites. These included three replicates of a 72-h PHA-stimulated PBMC supernatant (samples 1, 4, and 8) and four nonstimulated samples (samples 2, 3, 6, and 9). One PHA-stimulated whole blood cell supernatant (sample 7) and

two LPS-stimulated cell supernatants (sample 5 from PBMC and sample 10 from whole blood) are included. A total of nine markers (IFN- $\gamma$ , TNF- $\alpha$ , IL-2, IL-4, IL-6, IL-10, MIP-1 $\alpha$ , MIP-1 $\beta$ , and Rantes) were measured using reagents from five sources. Representative results are presented in Fig. 3.

All five laboratories that reported results were consistent with the replicate analyses of IFN- $\gamma$  (Fig. 3A and Table 2). Laboratories A, C, and K were in good agreement throughout. However, lab D used a different kit and had much lower values, and lab F had at least one aberrant result.

TNF- $\alpha$  analyses were quite consistent in labs C and D. However, lab A showed variation in the replicates (Fig. 3B and Table 2) and indicated stimulation in sample 9.

IL-2 results varied (Fig. 3C and Table 2). Two labs (D and J) did not detect IL-2 in any of the samples, and lab K reported similar levels in stimulated and many nonstimulated supernatants. Laboratory A results varied substantially for the replicates as well as for the nonstimulated samples. IL-4 levels in



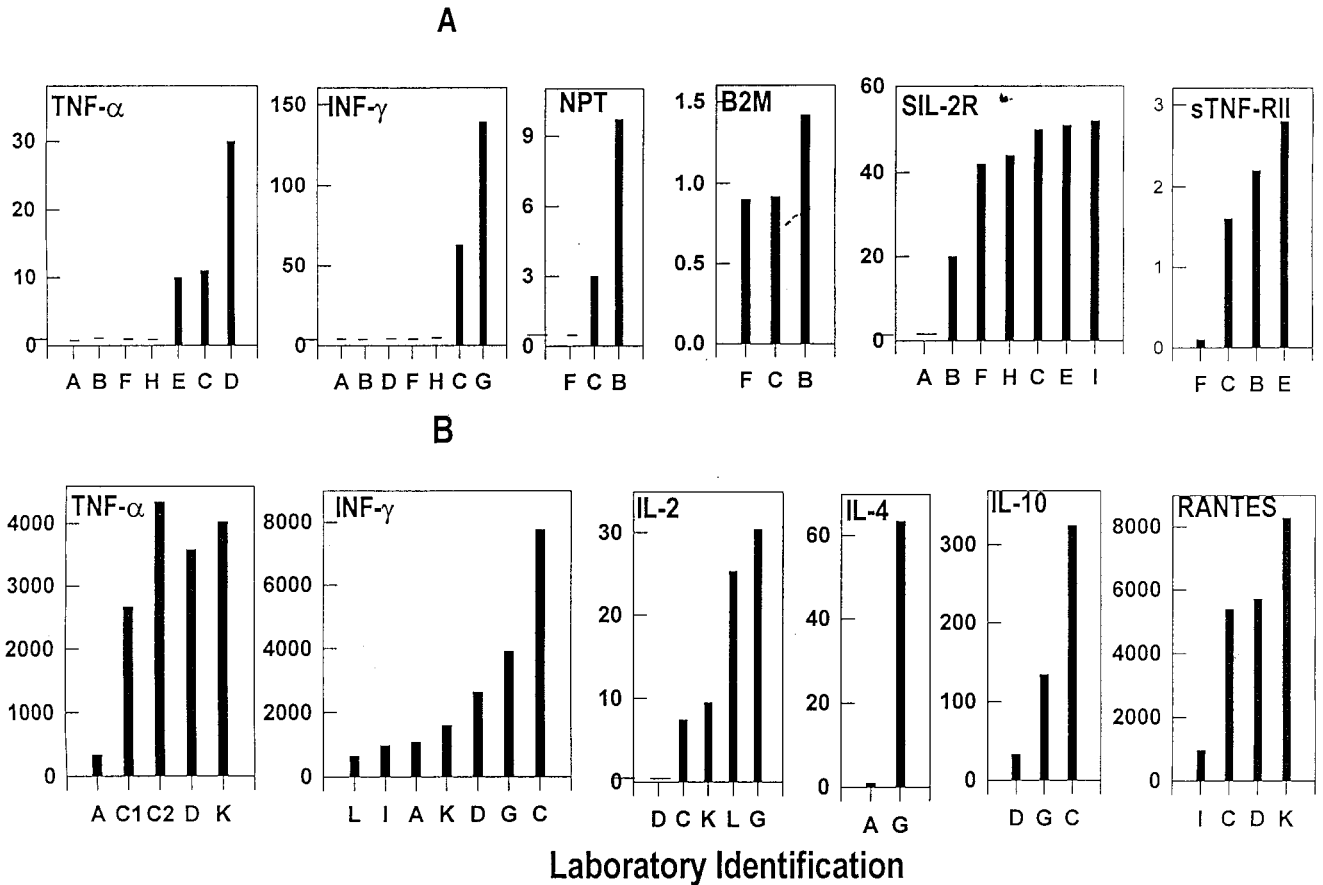


FIG. 2. Mean levels of cytokines, chemokines, and soluble products of immune activation reported by individual laboratories (designated by capital letters). Mean levels from triplicate samples of normal plasma in study I (A) and supernatant fluid from PHA-stimulated PBMC in study III (B) are presented for the laboratories that reported data. Assay values for all analytes are in picograms per milliliter.

the replicates varied in three labs as well as in the stimulated and nonstimulated samples (data not shown). There was general dissatisfaction with the IL-4 assays. Replicate agreement was good in two laboratories (A and D) reporting IL-6 analy-

ses, but some discrepancies in stimulated and nonstimulated levels were evident (data not shown).

All three labs (C, D, and J) reported elevated IL-10 levels in the three replicate and LPS-stimulated supernatants (data not shown). One lab (J) showed significant variability (Table 2). Unfortunately, there were major (but consistent) differences in the quantitative data from the three laboratories.

Rantes was tested in three laboratories (D, I, and K) using the same reagent source (Fig. 3D). Good replicate values and similar quantitative levels were reported. All performed well with replicates. However, a second assay source evaluated in lab K showed increased variability (Table 2) and failed to reveal differences between stimulated and nonstimulated samples. MIP-1 $\alpha$  and MIP-1 $\beta$  were tested in only one laboratory, where replicate samples agreed well and higher levels were found in stimulated samples than in controls (data not shown).

Overall, most laboratories performed well with the replicates for almost all cytokines and chemokines. One laboratory had some difficulties. Reagents from a single source generally gave similar results when tested in several different laboratories. In some assays, reagent sources differed substantially, indicating the need to address reference standard and/or calculation issues. IL-2 and IL-4 varied so much that levels at or below limits of detection were suspected for many samples.

TABLE 2. Variability of tests on three replicate samples included in each shipment

Laboratory	CV (%) <sup>a</sup> for assay of:				
	TNF- $\alpha$	IFN- $\gamma$	IL-2	IL-10	RANTES
A	-/41/21	-/15/8	/136/		
B	33//	-//			
C	5/11/5,8	50/3/3	//38	/10/10	//6
D	-/12/8	-/15/5	/-/-	/9/10	/2/8
E	25//				
F	-//	-/7/			
G		29//3	19//22,-	46//22	
H	-//	-//			
I		//6			/12/2
J			/-/-	/47/	
K	//12	/11/11	/25/19		/15,27/6
L		//-	//48		

<sup>a</sup> Results are presented as coefficients of variation (CV) for all three shipments (first/second/third). -, analyte was tested but not detectable. The absence of number indicates that test was not done. For example, laboratory A, did not detect TNF- $\alpha$  in the replicates in the first shipment but did in the second (with a CV of 41%) and in the third (with a CV of 21%). Laboratory B did test TNF- $\alpha$  in the first shipment but did not in the second and third shipments.

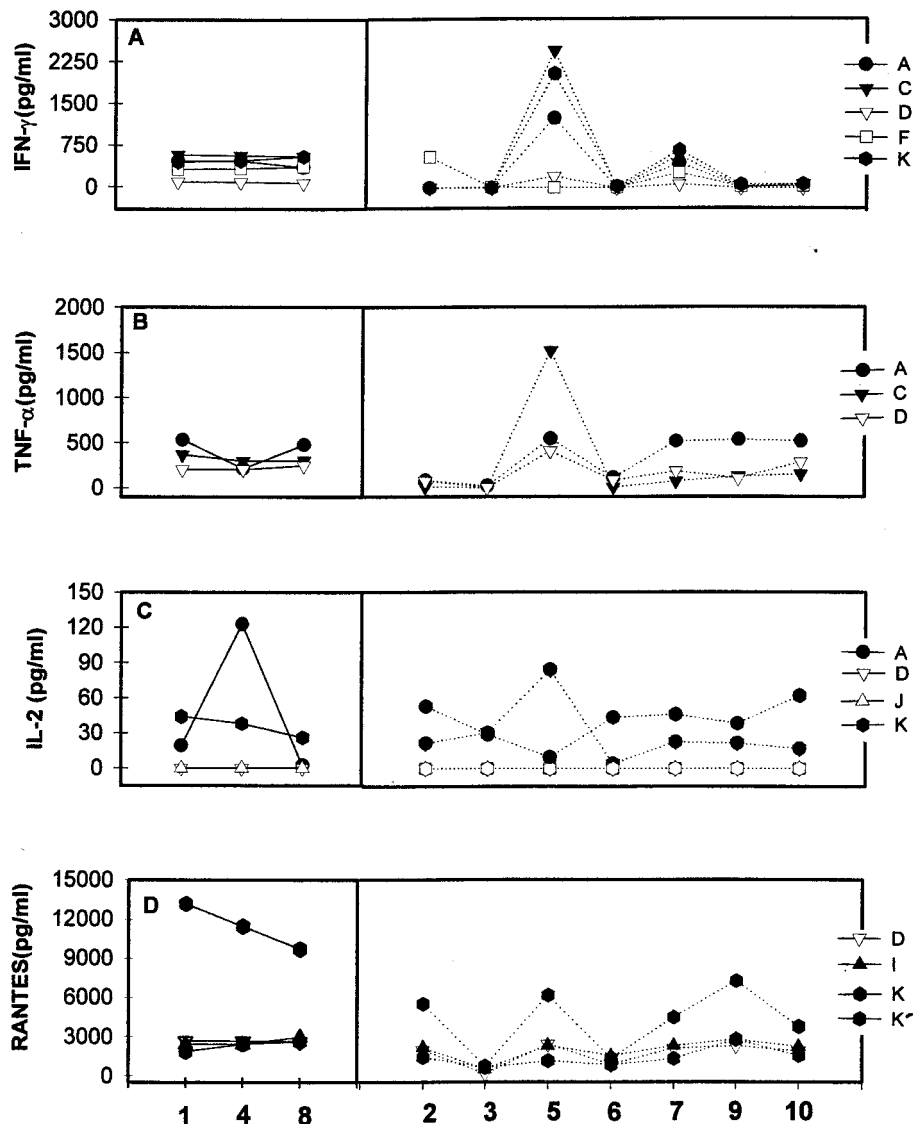


FIG. 3. Analytic data from study II for IFN- $\gamma$  (A), TNF- $\alpha$  (B), IL-2 (C), and Rantes (D) measured in several laboratories (designated by capital letters). Three replicate sample (triplicate samples 1, 4, and 8) obtained from a PHA-stimulated 72-h supernatant are grouped on the left. Four samples (samples 2, 3, 6, and 9) were not stimulated. One PHA-stimulated whole blood supernatant (sample 7) and two LPS-stimulated cell supernatants, one from PBMC (sample 5) and one from whole blood (sample 10), are included. K and K' designate results from laboratory K using two different reagent sources.

**Study III.** Six supernatant samples (1 to 6), including three replicates and six different plasma samples, were included in this study. Two or more laboratories reported analyses for TNF- $\alpha$ , IFN- $\gamma$ , IL-2, IL-4, IL-10, and Rantes. Values for IL-6, MIP-1 $\alpha$ , MIP-1 $\beta$ , sIL-2R, sTNF-RII, neopterin, and  $\beta_2$ M were reported from individual laboratories.

TNF- $\alpha$  analyses of both control and poststimulation supernatant samples are presented in Fig. 4A. Mean values of the triplicate repeats for five different labs are shown in Fig. 2B. The replicates of a supernatant were tested in three laboratories (C, D, and K) using the same reagent sources (Table 3). A second reagent source was also used in laboratory C, and laboratory A used a third source. This last source gave much lower values and missed the TNF- $\alpha$  in plasma.

Overall, the replicate values were consistent in each laboratory, and all laboratories detected the differences between stimulated and nonstimulated supernatants. Three laborato-

ries (D, K, and C) detected plasma differences between seropositive and seronegative donors, although one lab (A) failed to do so. There were some differences in absolute values, but the differences between laboratories were quite consistent.

IFN- $\gamma$  analyses (Fig. 4B) showed good replicate agreement in six laboratories (A, C, D, G, K, and I) but not in one (L) (Table 2). All laboratories distinguished stimulated from nonstimulated supernatants. One lab (C) identified elevated IFN- $\gamma$  levels in the tested plasmas of the three HIV-seropositive donors, but two labs (D and K) did not detect any IFN- $\gamma$ . Labs A and L reported detectable IFN- $\gamma$  in all plasma samples but with no difference between HIV-negative and HIV-positive samples. The three HIV-seropositive plasma samples all revealed elevated levels of  $\beta_2$ M, neopterin, sIL-2R, and sTNF-RII in comparison to the seronegative samples (data not shown).

Rantes analyses (Fig. 4C) of supernatant replicates and other supernatant samples, including a control, showed good

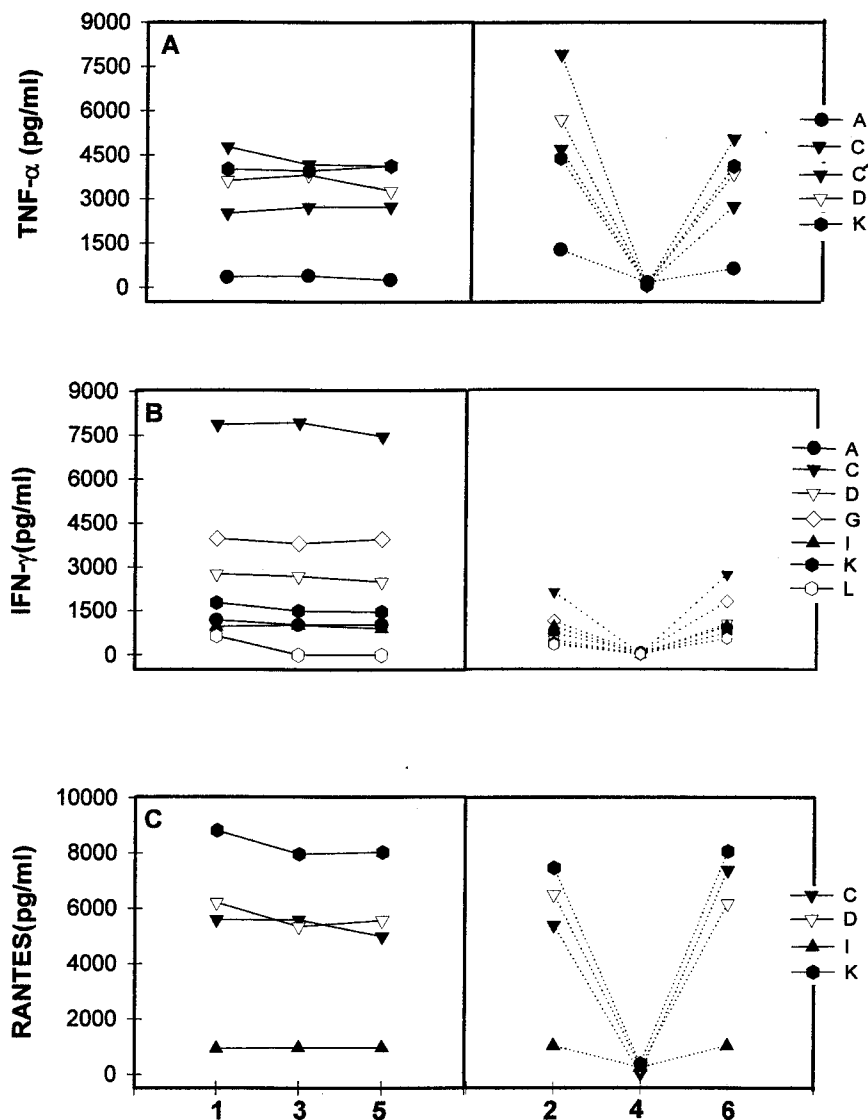


FIG. 4. Analytic data from study III for TNF-α (A), IFN-γ (B), and Rantes (C) measured in several laboratories (designated by capital letters). Three replicates of a stimulated PBMC supernatant (samples 1, 3, and 5) are grouped on the left. Three other supernatants (stimulated [samples 2 and 6] and nonstimulated [sample 4]) are presented. C and C' designate results from laboratory C using two different reagent sources.

agreement in four laboratories (C, D, K, and I) (Table 2). The quantitative differences were consistent (Fig. 2B), but the reason for the differences was not apparent because all labs used the same reagent source.

Replicate sample results (coefficients of variation) were good in almost all laboratories for all 13 components tested for each lab (good was defined as a coefficient of variation less than 20%). This indicated that laboratory performance was quite high. However, differences in quantitative values persisted.

Coefficients of variation for the tests done on replicate samples in two or more shipments are assembled in Table 2. Improvement in performance by study III is seen in most laboratories for TNF-α, IFN-γ, and Rantes. The greatest variability was seen with IL-2.

Differences in the mean levels reported by laboratories for 10 cytokines and other markers are indicated in Fig. 2. Failure

by several laboratories early in the proficiency program to detect TNF-α and IFN-γ in normal plasma is illustrated in Fig. 2A. Also, a wide difference was seen between some labs (for TNF-α and IFN-γ), but good agreement was seen between others (for sIL-2R). In Fig. 2B, however, general agreement in levels is seen with a number of assays.

The different values seen in Fig. 2 could be due to the reagent sources. This possibility was evaluated for IFN-γ and IL-2 (Table 3). For IFN-γ, Endogen kit results tended to be lower than Immunotech results. If laboratory I used the wrong decimal point, the corrected value would be 9,640 pg/ml, which would be near the level reported by laboratory C. Thus, for IFN-γ, the Immunotech kit was better than Endogen kits, with lower intralaboratory (Table 3) and interlaboratory coefficients of variation. Furthermore, in testing of the same normal plasma control samples, Endogen kits were consistently unable to detect any level of IFN-γ, while Immunotech kits detected

TABLE 3. Laboratory levels and intralaboratory coefficients of variation<sup>a</sup>

Cytokine	Reagent source and laboratory <sup>b</sup>	Mean level (pg/ml)	Coefficient of variation (%)
TNF- $\alpha$	A2	335	21.3
	C9	2,663	4.7
	D6	3,577	7.7
	K6	4,032	2.1
	C6	4,353	8.1
IFN- $\gamma$	I3	964	5.7
	A2	1,084	8.0
	K2	1,588	10.6
	D2	2,648	5.0
	C3	7,759	3.2
IL-2	C2	7.4	37.7
	K2	9.5	19.0
	L2	25.3	48.3
	G1	31.1	21.6
IL-10	D1	33	10.4
	G1	134	21.9
Rantes	I1	955	2.4
	C1	5,390	6.4
	D1	5,712	7.7
	K1	8,276	5.7

<sup>a</sup> Triplicate sample data in study III.

<sup>b</sup> Letters indicate laboratories, and numbers indicate reagent sources (1, Biosource; 2, Endogen; 3, Immunotech; 6, Medgenix; 9, Innogenetics).

an average of 14.9 pg/ml. For IL-2, the Biosource kits indicated higher levels than the Endogen data. However, substantial differences between laboratories in IL-2 mean levels were evident.

The intralaboratory variability was not unusual for TNF- $\alpha$ , IFN- $\gamma$ , and Rantes (Table 3). Laboratory performance by this criterion does not account for the differences in mean values for IFN- $\gamma$  or Rantes. Differences could be in the manner of using reference standards. However, separate from the standard issue, the tests for IL-2 showed substantially greater variability in performance than any of the other assays.

## DISCUSSION

An external evaluation program to assure quality performance of many clinically important assays is required of laboratories participating in the College of American Pathologists accreditation program. New assays, however, may not be included in these programs until a need is established, usually by extensive use in clinical practice. Proficiency testing programs for plasma cytokines, chemokines, and the soluble markers of immune activation are not in place.

The findings presented here emphasize the importance of having a well designed and critically evaluated external performance evaluation program when measurements are clinical relevant and are to be conducted at multiple sites. Levels of cytokines, chemokines and soluble products of immune activation are increased in many autoimmune and inflammatory disorders and in HIV infection and are altered by aging (18; J. L. Fahey, J. F. Schnelle, J. K. Thomas, M. E. Gorre, N. Aziz, and P. Nishanian, submitted for publication). Increasingly, immune-based therapies are designed to alter cytokine activities.

Initially, in the first study, laboratory performance was uneven as was evident with the results for replicate samples at

several sites. This proved to be due to inexperience, inadequate supervision, misunderstanding of procedures, and other reasons. However, after discussions led by experienced laboratory personnel, difficulties were addressed, and replicate values were better by the third study. However, there were laboratories where the experienced technician left to go to graduate school or other employment and the process of education and training began again.

Differences in reagents and standards between suppliers was a more resistant problem. Because international standards are available for almost all cytokines, it was surprising to find marked differences within the cytokine measurements. On the other hand, the capacity for errors in laboratory performance is almost limitless. However, there was a group of laboratories, using the same supplier, which usually had comparable results.

International standards are not yet available for the soluble receptors and other products of immune activation. Thus, it is probably advisable to identify a single source of reagents and standards for each assay after their usefulness has been verified.

We did not find one reagent supplier superior to all the others, but the study was not designed to test all suppliers or reagents kits. In general when several were compared, one or two sources appeared to be more useful. However, we have had the experience of major changes in the quality of a reagent or in a reference standard from a single reagent manufacturer (2).

The distinction between an external proficiency testing program for performance evaluation and the provision of international reference standards, such as for cytokines, is important. International standards are used for calibration of reference materials. In contrast, external proficiency testing programs assess actual assay performance and allow laboratory evaluation. Many labs assume that they are doing good work, but participation in an external proficiency testing program is a means of proving it. Also, participation in well-run external proficiency testing programs for performance evaluations can help new, inexperienced, or otherwise disadvantaged laboratories to achieve high performance levels by consultation with more experienced personnel.

The premise in undertaking this external proficiency testing program was that laboratories which were funded to support multicenter clinical therapeutic trials might not provide similar values for tests of cytokines and soluble markers of immune activation (CSM). At the time that the studies were begun, CSM testing was still evolving rapidly, with many putative suppliers of reagents and standards and a variety of techniques. No proficiency testing programs for performance evaluations were in place, and some laboratories had not established methods or quality assurance procedures. A corollary assumption was that laboratories that participated would benefit from the experience. An obverse corollary would be that laboratories that did not participate might well have unrecognized problems. Deficient laboratory performance is documented in other studies (3, 4).

The experience reported here indicates the potential value of an external proficiency testing program for CSM on a larger, conceivably national, scale. The need reaches across many areas of adult and pediatric medicine. Physicians can use immunologic tests for cytokines and for the products of cytokine activity in the evaluation of many autoimmune diseases, of infectious diseases affecting the immune system (such as HIV infection), and of other disorders characterized by changing balances within the immune system. Furthermore, quality evaluations of relevant immune activities are essential for moni-



toring therapies designed to activate or suppress immune functions.

#### ACKNOWLEDGMENTS

We appreciate the assistance of Cynthia Wilkening in data analysis and the assistance of Keri Walden in manuscript preparation.

Support from the Adult ACTG program (NIH grant AI-38858) and government contract NO—AI-45175 for Immunophenotyping Quality Assurance from DAIDS, NIAID, made this study possible.

#### REFERENCES

1. Aziz, N., P. Nishanian, and J. L. Fahey. 1998. Levels of cytokines and immune activation markers in plasma in human immunodeficiency virus infection: quality control procedures. *Clin. Diagn. Lab. Immunol.* **5**:755–761.
2. Aziz, N., P. Nishanian, R. Mitsuyasu, R. Detels, and J. L. Fahey. 1999. Variables that affect assays for plasma cytokines and soluble activation markers. *Clin. Diagn. Lab. Immunol.* **6**:89–95.
3. Bienvenu, J., L. Coulon, C. Doche, M. C. Gutowski, and G. Grau. 1993. Analytical performances of commercial ELISA kits for IL-2, IL-6 and TNF $\alpha$ . A WHO study. *Eur. Cytokine Netw.* **4**:447–451.
4. DeKossodo, S., V. Houba, G. E. Grau, and WHO Collaborative Study Group. 1995. Assaying tumor necrosis factor concentration in human serum. A WHO international collaborative study. *J. Immunol. Methods* **182**:107–114.
5. Fahey, J. L., J. M. G. Taylor, B. Manna, P. Nishanian, N. Aziz, J. V. Giorgi, and R. Detels. 1998. Prognostic significance of plasma markers of immune activation, HIV viral load and CD4 T-cell measurements. *AIDS* **12**:1581–1590.
6. Fahey, J. L. 1998. Cytokines, plasma immune activation markers, and clinically relevant surrogate markers in HIV infection. *Clin. Diagn. Lab. Immunol.* **5**:597–603.
7. Giorgi, J. V., H.-L. Cheng, J. B. Margolick, K. D. Bauer, J. Ferbas, M. Waxdal, I. Schmid, L. E. Hultin, A. L. Jackson, L. Park, J. M. G. Taylor, and the Multicenter AIDS Cohort Study Group. 1990. Quality control in the flow cytometric measurement of T-lymphocyte subsets: the Multicenter AIDS cohort study experience. *Clin. Immunol. Immunopathol.* **55**:173–186.
8. Giovannoni, G., M. Lai, D. Kidd, J. W. Thorpe, D. H. Miller, A. J. Thompson, G. Keir, M. Feldmann, and E. J. Thompson. 1997. Daily urinary neopterin excretion as an immunological marker of disease activity in multiple sclerosis. *Brain* **120**:1–13.
9. Guidi, L., A. Tricerri, D. Frasca, M. Vangeli, A. R. Errani, and C. Bartoloni. 1998. Psychoneuroimmunology and aging. *Gerontology* **44**:247–261.
10. Hahn, G., B. Stuhlmuller, N. Hain, J. R. Kalden, K. Pfizenmaier, and G. R. Burmester. 1993. Modulation of monocyte activation in patients with rheumatoid arthritis by leukapheresis therapy. *J. Clin. Invest.* **91**:862–870.
11. Jackson, J. B., J. Drew, H. J. Lin, P. Otto, J. W. Bremer, F. B. Hollinger, and S. M. Wolinsky. 1993. Establishment of a quality assurance program for human immunodeficiency virus 1 DNA polymerase chain reaction assays by the AIDS clinical trials group, ACTG PCR Working Group, and the ACTG PCR Virology Laboratories. *J. Clin. Microbiol.* **31**:3123–3128.
12. Mandy, F., J. Bradley, and J. L. Fahey. 1998. International program for quality assurance and standardization of immunological measures relevant to HIV/AIDS: QASI program for CD4 proficiency testing, p. 621–625. *In* Proceedings of the 12th World AIDS Conference, vol. 2. Monduzzi Editore, Bologna, Italy.
13. Mire-Sluis, A. R., R. Gaines-Das, and A. Padilla. 1998. WHO cytokine standardization: facilitating the development of cytokine research, diagnosis and as therapeutic agents. *J. Immunol. Methods* **216**:103–116.
14. Mire-Sluis, A. R., R. Gaines-Das, R. Thorpe, and Participants of the Collaborative Study. 1997. Implications for the assay and biological activity of interleukin-8. *J. Immunol. Methods* **200**:1–16.
15. Paxton, H., P. Kidd, A. Landay, J. Giorgi, N. Flomenberg, E. Walker, F. Valentine, J. L. Fahey, and R. Gelman. 1989. Results of the flow cytometry ACTG quality control program: analysis and findings. *Clin. Immunol. Immunopathol.* **52**:68–84.
16. Poole, S., and R. E. Gaines-Das. 1991. The international standards for IL-1 $\alpha$  and IL-1 $\beta$ -evaluation in an international collaborative study. *J. Immunol. Methods* **142**:1–13.
17. Propst, A., T. Propst, M. Herold, W. Vogel, and G. Judmaier. 1995. Interleukin-1 receptor antagonist in differential diagnosis of inflammatory bowel diseases. *Eur. J. Gastroenterol. Hepatol.* **7**:1031–1036.
18. Rea, I. M., S. E. McNerlan, and H. D. Alexander. 1999. CD69, CD25 and HLA-DR activation antigen expression on CD3 lymphocytes and relationship to serum TNF $\alpha$ , IFN $\gamma$  and sIL-2R levels in aging. *Exp. Gerontol.* **14**:79–92.
19. Wadhwa, M., R. Thorpe, and Participants of the Meeting. 1997. Standardization and calibration of cytokine immunoassays: meeting report and recommendations. *Cytokine* **9**:791–793.