

# Grading of Laboratories on CD4+ T-Lymphocyte Evaluations Based on Acceptable Data Boundaries Defined by the Measurement Error

Annalisa Kunkl,<sup>1\*</sup> Domenico Risso,<sup>2</sup> Maria Paola Terranova,<sup>1</sup> Mauro Giroto,<sup>3</sup> Bruno Brando,<sup>4</sup> Lorenzo Mortara,<sup>1</sup> and Pasquale Bruno Lantieri<sup>2</sup>

<sup>1</sup>Department of Immunology, San Martino Hospital and University of Genoa, Genoa, Italy

<sup>2</sup>Department of Health Science, Biostatistics Unit, University of Genoa, Genoa, Italy

<sup>3</sup>Immunohematology Service, Ivrea, Italy

<sup>4</sup>Transplant Immunology and Hematology Laboratory, Niguarda-Ca' Granda Hospital, Milan, Italy

**Background:** We addressed the definition of limits of error of %CD4+ and CD4+ counts (AbsCD4+) typical of laboratories of excellence, as well as the grading of laboratories based on the decision to take these limits as boundaries of unacceptable data. **Methods:** We studied the 99.9% confidence intervals of the means of 24 human immunodeficiency virus (HIV)+ and HIV- blood samples analyzed by 18 laboratories of the Liguria Region Quality Assessment Program (Liguria Region QALI). Regression equations of lower (L1) and upper (L2) confidence limits over the means of data cleared of unusual results were used to interpolate limits of error for mean values in the tested range. **Results:** L1 and L2 were symmetric around the mean and a single absolute difference (Abs Res) between the limits and the mean was found. Abs Res significantly increased over mean values ( $P = 0.0005$  for %CD4+,  $P < 0.0001$  for AbsCD4+). Limits were compatible with errors shown with blind replicates. Unacceptable results, outside the limits, accounted for 25% and 30% of %CD4+ and for 18% and 35% AbsCD4+ in the Liguria Region QALI and in the Piemonte Region QA Program, respectively. Limits interpolated over the median showed a similar grading. A comparable fraction of unacceptable data was also found with the method used in the U.K. National External Quality Assessment Scheme (NEQAS) immune monitoring scheme. **Conclusions:** We propose the general use of these regression equations to determine bounds for unacceptable data in proficiency testing and to identify laboratories of excellence. *Cytometry (Clin. Cytometry) 50:117–126, 2002. Published 2002 Wiley-Liss, Inc.*<sup>†</sup>

**Key terms:** CD4+ evaluation; measurement error; quality assessment; laboratory of excellence

Accurate and reliable measures of CD4+ T lymphocytes are essential to monitor immunodeficiency in human immunodeficiency virus (HIV)-infected subjects (1,2). Sev-

eral quality assessment (QA) programs in lymphocyte immunophenotyping have been promoted to assure clinically relevant comparisons from laboratory to labora-

Grant sponsor: Italian National Health Institute, Anti-Retroviral Therapy Project; Grant number: 880/02 1995-1998.

With the collaboration of (in alphabetical order): for the Liguria Region Quality Assessment Group: P. Bancheri, Genova; A. Campanella, Genova; P. Castagna, Massa Carrara; T. Castelli, Imperia; M. Cosso, Genova; A. De Ferrari, Savona; D. Fenoglio, Genova; O. Figari, Genova; R. Giachetti, Parma; S. Jacopini, Genova; G. Lavagna, Pietra Ligure; S. Lombardi, Massa Carrara; G. Melioli, Genova; C. Oliveri, Genova; M.R. Romano, La Spezia; L. Scarso, Genova; M. Setti, Genova; D. Tanca, Lavagna; C. Valle, Pietra Ligure; M. T. Valle, Genova. For the Piemonte Region Quality Assessment Group: P. Bernasconi, Milano; R. Borgianni, Torino; E. Camiolo, Verbania; A. Cartasegna, Novi Ligure; R. Chianese, Legnano; M.M. Ciriello, Alessandria; G. Crovetto, Gallarate; G. Delfino, Savigliano; P. Facco, Torino; F. Farinetti, Novara; C. Fiorio, Torino; A. Fruttero, Cuneo; G. Garbaccio Ioria, Biella; M. Guena, Candiolo; P. Ghio, Orbassano; A. Grillo, Torino; A. Iavarone, Torino; A. Lazzaro, Casale Monferrato; A. Livraghi, Pavia; C. Martini, Torino; M. Novelli, Torino; P. Ormede', Torino; C. Ortolani, Venezia; M. Pautasso, Orbassano; G. Pene, Torino; S. Ratibondi, Aosta; D. Rossi, Varese; T. Salvaterra, Mantova; A. Santagostino, Vercelli; M. Setti, Genova; A. Stacchini, Torino; M. Tarditi, Borgomanero; L. Vanelli, Pavia.

\*Correspondence to: Dr. Annalisa Kunkl, Department of Immunology, San Martino Hospital, Largo R. Benzi 10, 16132 Genoa, Italy.

E-mail: akunkl@smartino.ge.it

Received 16 May 2001; Accepted 8 November 2001

Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/cyto.10069

tory (3–16). Participation in QA programs with the goal of laboratory improvement to excellence is a prerequisite for accreditation by appropriate organizations (17,18). Proficiency testing (PT) became referred to as QA schemes that grade laboratories by estimating their performance (3,14,19,20).

There are two main goals of comparative studies, namely, the determination of interlaboratory variability and the evaluation of individual laboratory performance. Interlaboratory variability reflects the degree of consensus among laboratories on analyses of common samples. Normal or asymmetric distribution of data is sought and the standard deviation (SD) or the interquartile range (IQR; 5,19,21) is used as measure of spread and variability. Individual laboratory performance, on the other hand, is evaluated by accuracy, which is the degree of fit of the laboratory's value to the consensus mean (mean of data devoid of outliers) or to the median, either taken as the "true" value.

QA organizations establish their own standard for unacceptable data, based on the magnitude of the deviation of the laboratory's value from the true value. Classically, unusual results are described as being outside arbitrary boundaries, dependent on data distribution, typically defined by the consensus  $M \pm 2$  SD or by the 75th percentile plus 1.5 IQR and the 25th percentile minus 1.5 IQR in the case of asymmetric distribution (5). In an alternative approach, used in the program designed by the National Institute of Allergy and Infectious Diseases (NIAID) Division of AIDS (DAIDS) in the United States, two statistical parameters are used to define unacceptable data: the residual and the deviate (19–21). The residual is the difference between the laboratory's value and the median of all participating laboratories analyzing that specimen. It can be positive or negative, indicating overestimation or underestimation with respect to the true value, respectively. A residual equal to zero is the ultimate accuracy. The deviate is the residual divided by the variation expressed as 0.75 IQR.

The NIAID DAIDS program certifies or fails laboratories in interlaboratory performance evaluation of CD4+ T lymphocytes by grading them on the basis of acceptable limits of error for %CD4+ present in healthy individuals. It classifies as unacceptable a laboratory's analysis with an absolute residual (Abs Res) from a target of 40% CD4+ greater than or equal to 5% CD4+ and with an absolute deviate greater than or equal to 2. Moreover, it defines as unsatisfactory a level of performance characterized by at least 33% of unacceptable CD4+ analyses (19,20).

Grading is a means to reduce interlaboratory variation if criteria select out poor performers along with gross outliers. Those approaches where variability of measures on a particular specimen accounts for the determination of the boundaries for unacceptable data, such as those defined by the  $M \pm 2$  SD, do not seem to be adequate. In fact, in the case of a large variation, the wide boundaries, although excluding gross outliers, hamper the selection of exactly those incongruous data that contribute to the variability. On the contrary, in the case of a small variation,

the tight boundaries may exclude data within the test error. Most importantly, the same deviation from a certain true value is accepted or rejected on different occasions, depending on the spread of data on a particular sample.

We were interested in establishing a criterion for unacceptable data that was dependent on a single attribute of the test sample, i.e., its true value, and on the variability of measures expected in laboratories of excellence. By using experimental data obtained from laboratories participating in the Liguria Region Program for Quality Assessment of Lymphocyte Immunophenotyping by flow cytometry (Liguria Region QALD), we have tried to build an ideal population of measures devoid of odd results deemed to exceed the analytic variability in the best conditions. We defined limits of errors over a range of clinically relevant values of %CD4+ and of AbsCD4+, the crucial parameter for the management of HIV+ patients (22,23). In addition, we examined grading of laboratories in PT based on the decision to take these limits as boundaries of unacceptable data.

## MATERIALS AND METHODS

### Laboratories and Instruments

The study encompasses part of the 1994–1999 data of the Liguria Region QALI promoted by the regional AIDS Committee in 1989. Our laboratory at San Martino Hospital in Genoa (Italy) was appointed as the QA reference laboratory. Since 1993, the program includes all Ligurian laboratories involved in HIV immunophenotyping. They underwent quarterly evaluations with at least three blood samples per send-out. Eighteen laboratories were enrolled during the study. Three laboratories tested only HIV- blood samples. Three laboratories enrolled more operators, a maximum of two participating at the same time. Laboratories experiencing instrument, technical, or delivery problems were not evaluated. Flow cytometers included eight FACScan machines (Becton Dickinson [BD], San Jose, CA), three Epics XL machines (Beckman Coulter [BC], Hialeah, FL), two EPICS Elites (BC), two EPICS Profiles (BC), and three Cyturon Absolute machines (Ortho, Raritan, NJ).

### Blood Samples

HIV+ donors were referred to the Department of Infectious Disease, San Martino Hospital. HIV- volunteer donors were laboratory workers. Blood (maximum 20 ml) was collected in three 7-ml K<sub>3</sub> EDTA tubes (Vacutainer Division, BD), pooled in a 50-ml centrifuge tube (Costar, Corning, Inc., Corning, NY), mixed gently, and dispensed (1.5 ml) in 1.8-ml STORE-IT tubes (Nalge Nunc, Roskilde, Denmark). Tubes were packaged in a 50-ml plastic tube containing absorbing material, delivered by special carrier at 20–22°C, and tested on the same day of blood collection.

### Sample Preparation and Data Reporting

Laboratories used the conventional dual-platform technique to obtain absolute numbers, the two-color panel recommended by the Centers for Disease Control (CDC),

and the whole blood stain-and-lyse procedure (24). Hematological results and analyses of CD8+ T cells, CD3+, natural killer (NK), and B lymphocytes are not discussed in this article. Monoclonal antibodies (mAbs) for CD4+ analysis were purchased from BC (seven laboratories), BD (8 laboratories), Ortho (2 laboratories), and Dako (Carpinteria, CA; 1 laboratory). Lysis was performed with buffered  $\text{NH}_4\text{Cl}$  ( $1.5 \cdot 10^{-1}$  M  $\text{NH}_4\text{Cl}$ ,  $10^{-2}$  M  $\text{KHCO}_3$ ,  $10^{-4}$  M EDTA) by five laboratories, with FACS Lysing Solution (BD) by three laboratories, with Ortho Lyse (Ortho) by three laboratories, and with Q-Prep Workstation and ImmunoPrep Reagents (BC) by seven laboratories. Ten laboratories used the no-wash technique whereas the other eight washed the stained samples with  $10^{-2}$  M phosphate-buffered saline, pH 7.2 (PBS). Three out of the eight laboratories that washed their samples fixed them with freshly prepared 1% paraformaldehyde in PBS, pH 7.2. Laboratory practices were estimated by a questionnaire on shipment-specific information, such as sample conditions upon arrival, preparation, and analysis. Subset percentages, hematology data, gate purity, and lymphocyte recovery were also reported. Consistency of results was tested by internal quality controls (T sum, lymphosum; 24). An AbsCD4+ was derived by multiplying the relative frequency of CD3+CD4+ lymphocytes, obtained by flow cytometry, by the absolute lymphocyte count determined by the hematology analyzer (dual-platform method of absolute counting).

#### Study Design for Error Limits of Mean %CD4+ and AbsCD4+ Values

Our proposal was to use limits of error of target mean %CD4+ and AbsCD4+ values as boundaries for unacceptable data in PT. Therefore, the study design and the rationale of statistical analyses for the definition of the measurement error were based on the following considerations. The measurement error should reflect the unavoidable analytic variation around the mean value due to test performance in different laboratory settings. When used as a grading criterion, the measurement error should effectively select out poor performer laboratories at the same time avoiding, as far as possible, laboratory misclassification. Moreover, the measurement error should be defined for all possible mean values in the clinically relevant range. To meet these terms, data from eight send-outs for a total of 24 blood samples analyzed by 18 laboratories of the Liguria Region QALI in the period February 1994 to October 1997 were considered. For each sample, the distributions of raw data reported by laboratories for %CD4+ and AbsCD4+ were examined. Valid data were selected, characterized by being inside the interval defined by  $M \pm 2$  SD, by internal consistency (T sum and lymphosum), and by being usual by the unanimous agreement among three of the authors charged to examine results independently. The range within which the true value was almost certain to lie, i.e., the 99.9% confidence interval (CI) of the mean, was calculated for each sample. To relate all possible mean values in the clinically relevant range to the respective confidence limits, a relationship

was sought for each parameter by linear regression of the lower (L1) and the upper (L2) limits over mean values of the 24 samples. Two regression lines and equations were found for each parameter, relative to the L1 and L2 limits, respectively. The regressed limits of the CI (L1 and L2) were considered the limits of the measurement error relative to a certain mean value (x). The measurement error or Abs Res was expressed as the absolute difference between L1 or L2 and the mean.

#### Analysis of Sets of Replicates

Operators were selected that showed reliable %CD4+ and AbsCD4+ analyses within the limits of error in at least 80% of samples analyzed in three send-outs. Laboratories were sent three to five STORE-IT tubes containing 0.3-1 ml blood from one donor as blind replicates. Data from the different operators, relative to each donor's blood, were pooled, the overall mean estimated for each parameter, and the acceptable residual calculated by the regression equations. Results within each set of replicates were evaluated by two parameters: the range of results (the largest [max] value minus the smallest [min] value) and the difference between each replicate and the set's mean (intraoperator residual). Maximum values of range and of intraoperator residual distributions are reported.

#### Interlaboratory Studies With Mixtures of T-Cell Clones

Human T-cell clones were stimulated and expanded by established methods (25,26). Viable CD4+ ( $7 \cdot 10^7$ ) and CD8+ ( $1.5 \cdot 10^8$ ) cells, separated on a Ficoll/Hypaque gradient, were resuspended in PBS ( $1.2 \cdot 10^7$ /ml) and fixed by adding an equal volume of 2% freshly prepared cold paraformaldehyde (Sigma, St. Louis, MO) in PBS, pH 7.2. After 1 h of incubation on crushed ice, cells were centrifuged (300g) at 4°C and the pellet resuspended in Hank's solution containing  $2 \cdot 10^{-2}$  M L-lysine monohydrochloride (Sigma), pH 7.2, to stop fixation. Cells were washed and resuspended at  $4 \cdot 10^6$ /ml in PBS, 5% fetal calf serum (FCS), 0.01%  $\text{NaN}_3$ . Cells fixed under these conditions were shown to maintain surface antigen expression, reactivity, and specificity for anti-CD3, anti-CD4, and anti-CD8 mAbs (27,28). Different volumes of clones were mixed to obtain cellular standard samples containing CD4+ in various proportions. Mixtures were dispensed in Eppendorf tubes (100  $\mu\text{l}$ ,  $4 \cdot 10^5$  cells) and frozen at -80°C in cryo freezing containers (Nalgene, Rochester, NY). Preliminary experiments showed that cells frozen under these conditions could be stored at -80°C up to 1 week with maintenance of physical parameters and specific reactivity. For QA, a set of mixtures was thawed and shipped at 8°C to the participating laboratories for CD4+ testing on cells gated on physical parameters to exclude debris.

#### Assessment of Laboratory Performance

Limits of error were validated as a grading criterion by evaluating the performance of laboratories participating in the Liguria Region QALI from March 1998 to December 1999 and of laboratories participating in the Piemonte Region QA Program from January 1995 to April 1999. The

Piemonte Region QA Program started in 1987 under the coordination of Dr. M. Girotto at Ivrea, Italy (29,30). Forty-two laboratories were enrolled during the study period and underwent monthly evaluations with one blood sample from HIV- donors. All laboratories used the two-color CDC panel, the whole blood stain-and-lyse procedure, and the dual-platform method to generate absolute counts. Comparative analyses were performed on the same day of blood collection for the Liguria Region QALI and one day later for the Piemonte Region QA Program. For each sample, the consensus mean (A) and the overall median (B) were determined. Regressed 99.9% confidence limits for x value equal to A or to B and the corresponding Abs Res were calculated. Moreover, the difference between each laboratory's value and the consensus mean or the median (Abs Res Lab) was estimated. Abs Res Lab from the consensus mean higher than the Abs Res of the 99.9% confidence limits interpolated over the consensus mean was unacceptable by A1. Data unacceptable by A1 and with absolute deviates (Abs Res Lab / SD) equal or greater than 2 were unacceptable by A2. Abs Res Lab from the median higher than the Abs Res of the 99.9% confidence limits interpolated over the median was unacceptable by B1. Data unacceptable by B1 and with absolute deviates (Abs Res Lab / 0.75 IQR) equal or greater than 2 were unacceptable by B2. The fractions of unacceptable results found with criteria A1, A2, B1, and B2 were compared with the fractions of results outside the limits defined by the following methods: (C) 25th percentile - 1.5 IQR and the 75th percentile + 1.5 IQR (Paxton's criterion), (D) consensus  $M \pm 2$  SD (criterion D that coincided with A2), (E1) 10% trimmed  $M \pm 1$  trimmed SD, and (E2) 10% trimmed  $M \pm 2$  trimmed SD. Criteria E1 and E2, used in the U.K. National External Quality Assessment Scheme (NEQAS) immune monitoring scheme (D. Barnett personal communication), were calculated as described by Healy (31). Cohen's kappa index ( $k$ ; 32) was used to study the degree of agreement in paired comparisons. Agreement was ranked according to  $k$  values:  $>0.75$ , excellent;  $0.4-0.75$ , fair to good;  $<0.4$ , poor (33).

## RESULTS

### Definition of Measurement Errors of %CD4+ and AbsCD4+

Figure 1 shows L1 and L2 of mean %CD4+ (Fig. 1A) and AbsCD4+ (Fig. 1B) for each of the 24 samples plotted against the corresponding means, together with the regression lines and equations. Interestingly, CIs increased over mean values in particular for the AbsCD4+ parameter, as shown by the diverging trend of the regression lines relative to L1 and L2 limits. In fact, although an association between the mean and the standard error of the mean (SEM) was already evident for %CD4+ ( $r^2 = 0.43$ ,  $P = 0.0005$ ), the association was higher for AbsCD4+ ( $r^2 = 0.74$ ,  $P < 0.000$ ; data not shown).

The two regression equations, found for each parameter, were used to interpolate confidence limits for any hypothetical mean value in the tested range (Table 1). Because selected data followed a normal distribution, as

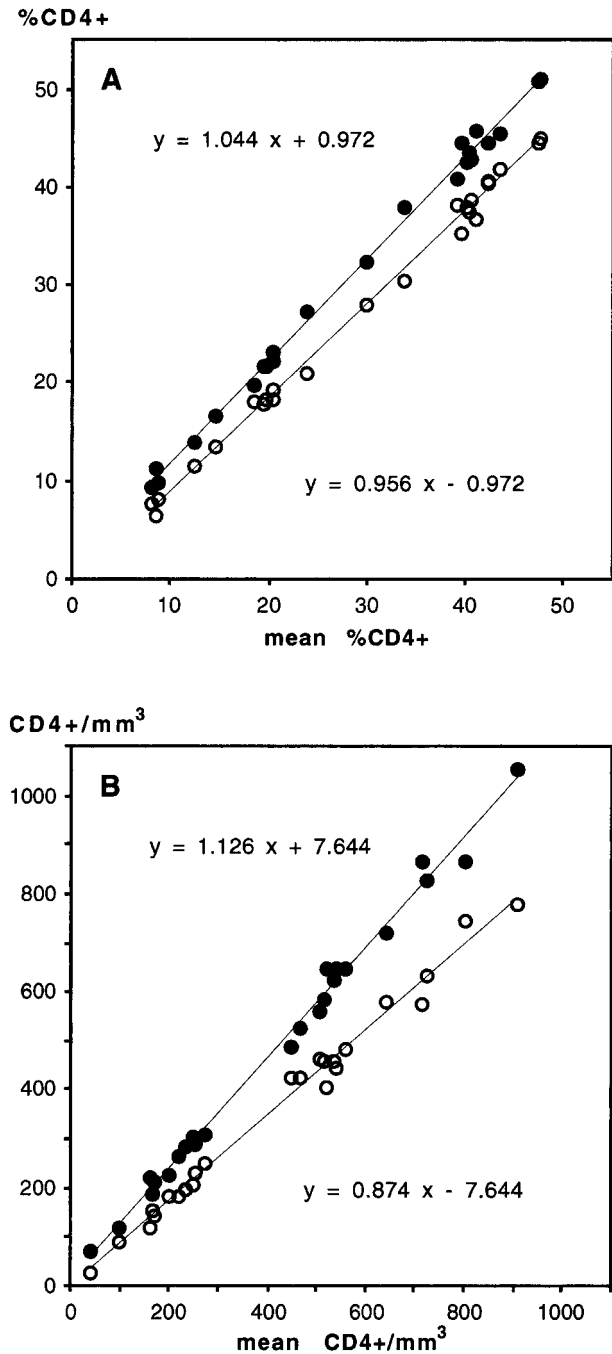


Fig. 1. Regression of lower and upper 99.9% confidence limits over mean %CD4+ and mean AbsCD4+ lymphocytes. Twenty-four blood samples (12 HIV+ and 12 HIV-) were analyzed by 18 laboratories of the Liguria Region QALI. 99.9% lower (open circles) and upper (solid circles) confidence limits of the mean of selected data of each sample were plotted against the mean %CD4+ (A) and the mean CD4+/mm<sup>3</sup> (AbsCD4+) lymphocytes (B). Regression was determined on a total of 340 valid %CD4+ data (90% of raw data) and of 274 valid AbsCD4+ data (93% of raw data). The range of mean values of the 24 samples was 8.3%–47.8% for %CD4+ and 47–913 CD4+/mm<sup>3</sup> for AbsCD4+.

suggested by the similarity of the mean and median values and by the skewness and kurtosis indices (data not shown), lower and upper limits of CI were typically sym-

Table 1  
Limits of the Measurement Error Calculated With the Regression Equations of Figure 1\*

Mean (x)	%CD4+				CD4+/mm <sup>3</sup>				
	99.9% CI		Abs Res	Relative error (%)	99.9% CI		Abs Res	Relative error (%)	
	Lower limit	Upper limit			Lower limit	Upper limit			
8.3	7.0	9.6	1.3	16	47	33	61	14	29
10.0	8.6	11.4	1.4	14	100	80	120	20	20
15.0	13.4	16.6	1.6	11	200	167	233	33	16
20.0	18.1	21.9	1.9	9	300	255	345	45	15
25.0	22.9	27.1	2.1	8	400	342	458	58	15
30.0	27.7	32.3	2.3	8	500	429	571	71	14
35.0	32.5	37.5	2.5	7	600	517	683	83	14
40.0	37.3	42.7	2.7	7	700	604	796	96	14
45.0	42.0	48.0	3.0	7	800	692	908	108	14
47.8	44.7	50.9	3.1	6	913	790	1036	123	13

\*Equations of Figure 1 were solved for y equal to the lower (L1) and the upper (L2) limits of the 99.9% CI by substituting x with the theoretical (true) values in the tested range. Abs Res was the absolute difference between L1 or L2 and the mean. The relative error was calculated as  $100 \times \text{Abs Res}/\text{mean}$ .

metric around mean values. Therefore, for each mean value, a single absolute difference (Abs Res) between the limits and the mean was found. Abs Res was considered the maximal acceptable deviation to a target mean value, the least deviation being equal to 0. For instance, a sample with a consensus mean of 300 CD4+/mm<sup>3</sup> had a 99.9% CI of 255–345/mm<sup>3</sup>. Thus, at maximum, the acceptable measurement error (Abs Res) to the mean of 300 CD4+/mm<sup>3</sup> had to be 45 CD4+/mm<sup>3</sup>, which corresponded to a relative error of 15%.

The pattern of increase of the variation with the mean showed that an increase in the mean %CD4+ from 8.3% to 47.8% CD4+ corresponded to an increase in the width of the CI such that the Abs Res of its limits varied from 1.3% to 3.1% for %CD4+. The relative error decreased from 16% to 6% accordingly. Moreover, an increase in the mean of AbsCD4+ from 47 to 913 CD4+/mm<sup>3</sup> corresponded to an increase in the Abs Res from 14 to 123 CD4+/mm<sup>3</sup> and a decrease in the error relative to the mean from 29% to 13% (Table 1).

Results thus far suggested that by substituting x with the consensus mean of data obtained in comparative studies, regression equations could be solved for y and limits of the measurement error interpolated.

#### Boundaries of Error Are Compatible With Intraoperator Variability

Measurement errors depend on the extent of interlaboratory variability, as CIs are related directly to the SEM. During the process of selecting valid data, however, we might have compressed interlaboratory variation to such an extent that errors were incompatible with the test error. We assumed that test error, due to the spread of data obtained on repetitive measures by one operator, was the least variation expected for an unknown sample.

Six samples (four HIV+ and two HIV-) were analyzed as blind triplicates (Table 2). Ranges of mean values were 6.9%–64% for %CD4+ and 162–1,238/mm<sup>3</sup> for AbsCD4+. In all samples, Abs Res from mean %CD4+ and AbsCD4+ values were higher than the Max Abs Res value of intra-

Table 2  
Validation of Measurement Errors: Analysis of Residuals From the Set's Mean\*

Sample	Lab no.	Operator no.	Data no.	%CD4+				CD4+/mm <sup>3</sup>			
				Overall data		Sets of replicates		Overall data		Sets of replicates	
				Mean <sup>a</sup>	Abs Res <sup>b</sup>	Max range <sup>c</sup>	Max Abs Res <sup>d</sup>	Mean <sup>a</sup>	Abs Res <sup>b</sup>	Max range <sup>c</sup>	Max Abs Res <sup>d</sup>
A HIV+	5	6	18	17.3	1.7	2	1.4	457	65	61	39
B HIV+	5	7	21	22.6	2.0	3	1.6	359	53	39	22
C HIV-	4	5	15	40.8	2.7	4	1.9	726	99	90	56
D HIV+	7	8	24	16.0	1.7	2	0.7	227	36	48	19
E HIV-	5	7	21	64.0	3.8	2	1.2	1238	164	121	79
F HIV+	5	7	21	6.9	1.3	1	0.8	162	28	36	20

\*Six samples (two HIV- and four HIV+) were analyzed by 15 operators in 11 labs in the first send-out (samples A, B, and C) and by 18 operators in 15 labs in the second send-out (samples D, E, and F). Each operator tested one set of blind triplicates of the same sample except for two operators who tested three and two sets of replicates, respectively, in both shipments.

<sup>a</sup>Consensus mean.

<sup>b</sup>Absolute residual calculated by the proposed equations (Fig. 1).

<sup>c</sup>Maximal range of results in sets of replicates.

<sup>d</sup>Maximal value of Abs Res distribution (intraoperator, from means of replicates).

operator residuals. As an example, in the case of sample A and for the AbsCD4+ parameter, the interpolated acceptable Abs Res for 457 CD4+/mm<sup>3</sup> was 65 CD4+/mm<sup>3</sup>, whereas the Max Abs Res was 39 CD4+/mm<sup>3</sup>, indicating that data selection compressed interlaboratory variation to a level still compatible with intraoperator variation.

However, since the Abs Res to a target of 40% CD4+ was 2.7% (Table 1), a value lower than the 5% indicated by the NIAID DAIDS program, we studied error patterns of %CD4+ evaluations in larger data sets. In the first series of studies, we shipped blind replicates of HIV- specimens showing mean %CD4+ values of 45.2% (A), 45.4% (B), and 47.8% (C; Table 3). Whereas Abs Res were 2.9%, 3%, and 3.1%, Max Abs Res were 2.5%, 2.2%, and 2.1% respectively, confirming the previous conclusion that limits of error calculated with the equations of Figure 1 fit with the test error.

In a second set of experiments, carried out in four different send-outs, residuals from the overall mean were examined in interlaboratory studies with T-cell samples consisting of mixtures of CD4+ and CD8+ T-cell clones containing predetermined %CD4+ contents in the range 5%–40% CD4+. Figure 2 showed that although the median Abs Res Lab (represented by the line within the box) increased with %CD4+ lymphocytes present in the mixtures, it was lower than 5% CD4+, even when 40% CD4+ samples were analyzed.

**Validation of Limits of Error as a Grading Criterion in PT**

Our purpose was the evaluation of the number of unacceptable data when limits of acceptable errors, calculated over the consensus mean and over the median, were used as grading criteria (A1 and B1). We studied how

Table 3  
Validation of Measurement Errors for %CD4+ Values of HIV Samples\*

Sample	Lab no.	Operator no.	Data no.	%CD4+			
				Overall data		Sets of replicates	
			Mean <sup>a</sup>	Abs Res <sup>b</sup>	Max range <sup>c</sup>	Max Abs Res <sup>d</sup>	
A	12	12	120	45.2	2.9	4.5	2.5
B	9	9	45	45.4	3.0	3.5	2.2
C	11	16	64	47.8	3.1	4.0	2.1

\*Twelve, 9, and 11 labs were selected for shipments A, B, C, respectively. Five labs participated with two operators in shipment C. Labs received 10, 4, and 5 blind replicates of the same sample in shipments A, B, and C, respectively. Evaluation of %CD4+ was performed in the context of a complete immunophenotype in B and C. In A, %CD4 was determined with a single tube (anti-CD3-FITC plus anti-CD4-PE) on lymphocytes gated on physical parameters.

<sup>a</sup>Consensus mean.

<sup>b</sup>Absolute residual calculated by the proposed equations (Fig. 1).

<sup>c</sup>Maximal range of results in sets of replicates.

<sup>d</sup>Maximal value of Abs Res distribution (intraoperator, from means of replicates).

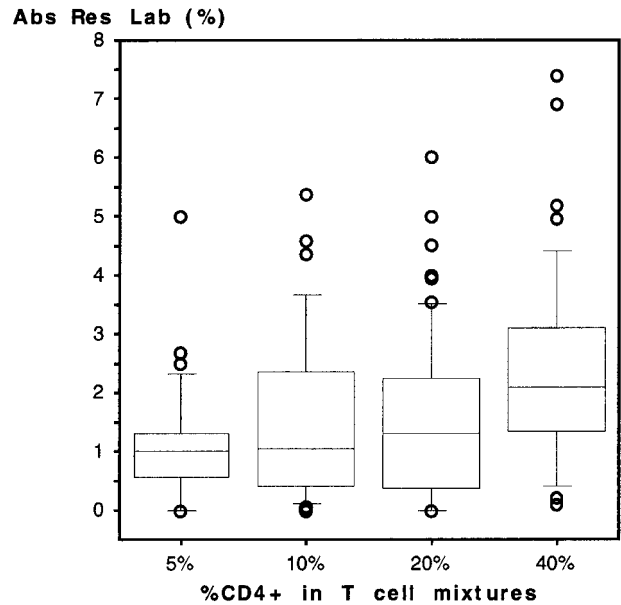


Fig. 2. Comparative studies with mixtures of T-cell clones with a predetermined CD4+ content. Results refer to four different send-outs in which different preparations of cellular standard samples were used. Tested T-cell frequencies were 5%, 10%, 20%, 40% plus a negative control (0% CD4+, 100% CD8+; not shown). Data of different experiments relative to samples containing the same CD4+ frequency were pooled and Abs Res Lab (%) from the overall mean calculated. Statistical analysis was performed on the mean of 43 data points for CD4+ analysis of the T-cell mixture. Lines on the box plot represent the 25th, the median, and the 75th percentile of Abs Res Lab (%) distribution. Circles represent data outside the 10th and the 90th percentiles.

many results unaccepted by A1 or by B1 were also unaccepted by the second criterion (absolute deviate equal or greater than 2; criteria A2 and B2). In addition, we compared our method with three other methods available for setting standards for laboratories, all based on the distribution of results obtained by laboratories on the test sample. We evaluated performance of laboratories of two external QA programs: the Liguria Region QALI, on data that followed those used to determine limits of error, and the Piemonte Region QA Program (Table 4).

The fractions of unacceptable results by criteria A1 and B1 were very similar both in the Liguria Region QALI and in the Piemonte Region QA Program sets of data. By comparing A1 versus B1, we obtained 25% versus 27% total %CD4+, 18% versus 19% total AbsCD4+ in the Liguria Region QALI; 30% versus 31% total %CD4+ and 35% versus 35% total AbsCD4+ in the Piemonte Region QA Program. The degrees of agreement of A1 with B1 for %CD4+ and AbsCD4+ were excellent (k = 0.90) in both QA programs. In the Liguria Region QALI, a slightly higher proportion of unacceptable results was found by criterion B2 compared with A2 (14% versus 10% total %CD4+, 11% versus 8% total AbsCD4+).

Comparison of our method with the other methods was performed with the Piemonte Region data set. Interestingly, our first criterion, A1, identified a similar proportion of unacceptable results as the U.K. NEQAS criterion E1

Table 4  
Validation of Measurement Errors as Boundaries for Unacceptable Performance Definition in PT Programs\*

Program	Parameter	Total data No.	Accepted data by all criteria No. %		Unacceptable data by the criteria													
					Consensus mean				Median				Paxton's criterion		U.K. NEQAS criterion			
					A1		A2 = D		B1		B2		C		E1		E2	
N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%			
Liguria Region	%CD4+	212	153	72	53	25	22	10	57	27	30	14	18	8	nd	nd	nd	nd
	CD4+/mm <sup>3</sup>	216	172	80	39	18	18	8	41	19	23	11	16	7	nd	nd	nd	nd
Piemonte Region	%CD4+	840	530	63	256	30	82	10	264	31	101	12	54	6	243	29	64	8
	CD4+/mm <sup>3</sup>	823	485	59	288	35	75	9	292	35	86	10	56	7	231	28	59	7

\*Data from five send-outs (15 samples) to 16 labs (28 operators) in the Liguria Region QALI and data from 33 send-outs (33 samples) to 42 labs in the Piemonte Region QA Program were examined by different criteria for the definition of unacceptable data: A1, Abs Res Lab from the consensus mean higher than the Abs Res of the 99.9% confidence limits interpolated over the consensus mean; A2, unacceptable results by A1 and with absolute deviates (Abs Res Lab/SD) equal or greater than 2; B1, Abs Res Lab from the median higher than the Abs Res of the 99.9% confidence limits interpolated over the median; B2, unacceptable results by A1 and with absolute deviates (Abs Res Lab/0.75 IQR) equal or greater than 2; C, results outside the 25th percentile - 1.5 IQR and the 75th percentile + 1.5 IQR (Paxton's criterion); D, results outside the consensus mean  $\pm$  2 SD (this criterion coincided with A2); E1, results outside 10% trimmed mean  $\pm$  1 trimmed SD; E2, results outside 10% trimmed mean  $\pm$  2 trimmed SD. nd, not determined because the U.K. NEQAS criterion was applicable in some cases (sample size lower than 15; 31).

(30% versus 29% total %CD4+ and 35% versus 28% total AbsCD4+, respectively). The degrees of agreement of A1 with E1 were good (%CD4+,  $k = 0.75$ ; AbsCD4+,  $k = 0.65$ ), suggesting that laboratories inside our limits of acceptable error, roughly corresponding to 70% of participating centers, can well be defined as laboratories of excellence.

It was clear that the introduction of the deviate parameter that corrects the unacceptable residual for the spread of results obtained by laboratories on that specimen made our method (criteria A2 and B2) very similar to criteria C, E2, and D (which actually coincided with A2). The agreement of A2 with C was good for %CD4+ ( $k = 0.65$ ) and excellent for AbsCD4+ ( $k = 0.77$ ), as was the agreement of A2 with E2 ( $k = 0.71$  for %CD4+ and  $k = 0.77$  for AbsCD4+). Moreover, the degrees of agreement of B2 with C and B2 with E2 were excellent for both %CD4+ and AbsCD4+ (range of  $k$  values = 0.76-0.85).

In conclusion, our grading criterion (A1 or B1), based on limits of error calculated with the regression equations, was more restrictive than methods C, D, and E2 but had a very similar impact on grading to the U.K. NEQAS first criterion (E1).

## DISCUSSION

The Liguria Region QALI includes PT as a measure of the performance of laboratories. Thus, the definition of grading criteria for PT was a pivotal and preliminary issue.

The evaluation of a laboratory's CD4+ value with a passing or a failing result, defined by the residual and the deviate statistical parameters, was adopted from the NIAID DAIDS QA program. The residual describes the deviation of the laboratory's value from a central tendency value, either the consensus mean or the median. We assumed that, for a certain mean value, the maximum of the residual should be represented by the difference be-

tween the 99.9% confidence limits of the mean and the mean in an ideal population of measures, typical of laboratories of excellence, devoid of unusual results, and normally distributed. Thus, the probability to misclassify a laboratory that reports results outside these limits as a poor performer is very low (1/1,000). Any laboratory's result included within the estimated limits of error and thus characterized by a residual between 0 (when the laboratory's value is equal to the mean or the median) and a maximal value typical of the 99.9% limits (Fig. 1, Table 1) is accepted and scored with 2. Any laboratory's value with a residual higher than that defined by the regressed limits of error, called "critical," is unacceptable by the residual, the first statistical parameter that defines unacceptable data.

We designed limits of critical error to reflect the ideal, not the actual variability of results. In fact, the final purpose of grading was the reduction of interlaboratory variability to that expected in a group of extensively standardized and quality-controlled laboratories, so-called laboratories of excellence. Although an error of 83 CD4+/mm<sup>3</sup> for a measure of 600 CD4+/mm<sup>3</sup> may not be clinically relevant, it is paramount that a CD4+ result of 200 or 300 CD4+/mm<sup>3</sup> be the same or the most alike as possible in two different laboratories. It is important to point out, however, that even considering this theoretical variability, the true value of a sample with measured 200 CD4+/mm<sup>3</sup> had 99.9% probability to be included between 167 and 233/mm<sup>3</sup>. Uncompressible variability when measurements are performed in different laboratories, technical bias over time, and within-individual variability should all be taken into consideration before staging HIV disease or before making therapeutic decisions on the basis of threshold limits of CD4+ counts (34-37).

Our data indicate that variability in results was greater in AbsCD4+ as the count increased. Moreover, due to the

association of mean values with variation, residuals increased with the increase of the mean whereas relative errors decreased (Fig. 1, Table 1). Although this pattern was expected when measures were expressed as percent values, it was more dramatic with AbsCD4+ than with %CD4+. Consistent with this observation, previous findings of a national QA program in Italy showed a higher frequency of unacceptable AbsCD4+ data on more than 300 CD4+/mm<sup>3</sup> samples than on less than 200 CD4+/mm<sup>3</sup> samples for unacceptable AbsCD4+ Abs Res equal or greater than 100 CD4+/mm<sup>3</sup> (14). The present data may also explain the significantly higher frequency of unacceptable %CD4+ data found in the Liguria Region QALI on normal versus pathological samples for %CD4+ Abs Res equal or greater than 5% CD4+ (unpublished observations). Twenty-nine unacceptable %CD4+ of a total of 167 %CD4+ results (27.3% of total data) were found in HIV- compared with 12 unacceptable %CD4+ of a total of 149 %CD4+ results (8% of total data) found in HIV+ samples ( $\chi^2$  test,  $P = 0.029$ ). The indication from these former studies that a fixed residual, irrespective of the mean CD4+ value, was unsuitable to the definition of unacceptable data in pathological samples prompted us to undertake the present study by systematically looking at the errors with samples bearing different true values. As a conclusion, we propose the general application of the regression equations, found with experimental data, to determine the boundaries for unacceptable data in relation to the true value.

Our limits of error of %CD4+ evaluations were narrower than the ones used in the NIAID DAIDS program, although they were compatible with the test error by keen operators, representing the best effort of our local laboratories (Tables 2 and 3). Various factors could have narrowed the variability of original measures (before data selection) obtained in the Liguria Region QALI. First, we tested specimens on the day they were collected, in contrast to most QA programs that use shipped samples, thus eliminating possible sources of error due to overnight delivery. We used fairly standard protocols that excluded three and four-color panels and flow cytometric-based absolute cell counting. Although multicolor panels and single-platform absolute counting should account for a lower interlaboratory variability than the two-color panel and the dual-platform absolute counting technology used throughout this study (37–40), the use of all these different methodologies at the same time might increase the extent of variability.

Second, we examined the validation of limits of error as the basis of our grading criterion. Measurement errors were defined with experimental data by a relative low number of laboratories, although they represented the majority of immunophenotyping laboratories present in our region. Therefore, we proved the feasibility of this criterion also on data of the Piemonte Region QA Program, which represents a wider variety of laboratory settings than the Liguria Region QALI.

An important issue was the similarity of impact on grading obtained with criteria A1 and B1 for x value equal

to the consensus mean and median, respectively (Table 4). The higher fraction of unacceptable data by B2 compared with A2, which was found in the Liguria Region QALI and was barely seen in the Piemonte Region QA Program data, could be ascribed to the asymmetric distributions of the Liguria Region QALI data. Indeed, a modified IQR seems to be an unbiased estimate of the true SD only if the distribution is Gaussian (41). It was interesting that for %CD4+, the agreement between B2 and C, where limits were calculated with a nonparametric approach, was excellent ( $k = 0.83$ ).

Because the overall median is calculated more readily than the consensus mean during analyses of PT data, we suggest method B (B1 and B2) as the most direct and suitable approach to evaluate laboratory performance in multisite comparison studies. A similar proportion of unacceptable %CD4+ data was found in the Liguria Region QALI and in the Piemonte Region QA Program databases (25% versus 30% by criterion A1 and 27% versus 31% by criterion B1, respectively; Table 4). On the contrary, the frequency of unacceptable AbsCD4+ was higher in the Piemonte Region QA Program than in the Liguria Region QALI (35% versus 18% with criterion A1 and 35% versus 19% with criterion B1, respectively). Therefore, the participation of a larger number of laboratories and the use of shipped specimens in the Piemonte Region QA Program seemed to have changed the dimensions of the AbsCD4+ results. Probably the sole intrinsic variability of hematological measurements in overnight samples may explain the increased variability in absolute lymphocyte counts (data not shown) and in the AbsCD4+ counts observed in the Piemonte Region QA Program. Currently, it is recognized that hematology instruments contribute more to the variation in AbsCD4+ than the flow cytometer (42). In two studies (38,40) assessing the accuracy of the single-platform measurements compared with flow cytometry plus hematology, there was a trend that showed that lymphocyte subset counts increase with the sample age when analyzed by conventional flow cytometry plus hematology. This seemed to be due to an increase in the absolute lymphocyte counts detected by some hematology instruments.

The high fraction of unacceptable data by A1 and B1 (30%–31% unacceptable %CD4+ and 35% unacceptable AbsCD4+) was similar to the fraction found with E1 (29% unacceptable %CD4+, 28% unacceptable AbsCD4+). Comparison of criterion A2 or B2 with E2 confirmed the agreement of our method with the method used in the U.K. NEQAS Scheme that is fully accredited according to the Clinical Pathology Accreditation System (EQA) UK Ltd. (D. Barnett, personal communication). The comparable impact on grading observed between our method and the U.K. NEQAS was desirable in the light of achieving a similar degree of excellence within Europe. We stress, however, the striking difference between our method and the U.K. NEQAS method. Our method allows a unique definition of a certain deviation from a true value (either acceptable or unacceptable by the limits a priori determined by the regression equations). In contrast, with the

U.K. NEQAS and with other methods (C and D), where boundaries of acceptable data are based on the variability of measures obtained by laboratories that had analyzed that specimen, the definition of an unacceptable result entirely depends on whether poor performers participated in a particular sample.

To understand the sensitivity of our criterion A1 or B1, we studied the decrease in the mean interlaboratory variability obtained over the 33 samples after removal of unacceptable data by the various methods (data not shown). Raw data showed a mean coefficient of variation (CV) of 8.4% for %CD4+ and 19.9% for AbsCD4+, figures in line with those reported in other investigations relative to the dual-platform technology (37). The mean CV after removal of unacceptable AbsCD4+ by A1 or B1 was 7%, whereas it ranged between 12% and 14% after removal of unacceptable data with C, D, or E2.

Results suggest the educational purpose of a more severe and sensitive criterion that would provide an incentive for laboratories to improve their performance with the final reduction of interlaboratory variability (CV <10% for AbsCD4+). Laboratories should be pressed to meet criterion A1 or B1, suggestive of a level of performance characterized by the production of highly accurate results. A traditional area of laboratory excellence concerns analytical procedures and results specifically (17,18), although recently customer satisfaction and the turn-around-time of tests are receiving more attention as quality measures. The fact that the Piemonte Region QA Program scheme adopted a grading criterion (consensus  $M \pm 2$  SD) not sufficiently able to select out unusual results could also explain the high fraction of unacceptable results in the Piemonte Region data found with criteria A1 and B1.

The aim of an additional looser criterion (A2 or B2) is to keep the grading system from being punitive for the laboratory. For example, the laboratory with an unacceptable residual is given another chance to move into the acceptable range and pass the test, if the variability of measures on that sample is unusually large. If the absolute deviate is lower than 2, the result is deemed acceptable and is scored 1. On the contrary, if the deviate is greater than 2, the data are definitely unacceptable and scored 0. The notification of a passing result scored 1 (result unacceptable for A1 or B1) should call the attention of the laboratory personnel to problems in the analysis procedure. A failing result scored 0 (result unacceptable for A2 or B2) should urge the laboratory to implement remedial actions. A passing result scored 2 is indicative of excellent performance.

This study made two major findings. First, we established a range of acceptable errors for different CD4+ values in the clinical relevant range. According to the intrinsic variability of measures among high-quality laboratories (laboratories of excellence), the ranges of error we report can be used as a reference term for the validation of new methods for CD4+ testing. Because the measurement error is expected to change with advances in technology, this study will be integrated with the single-

platform CD4+ analysis. Second, this grading system, based on boundaries for unacceptable data defined by the measurement error, is applicable to other PT programs for cell enumeration in clinical flow cytometry, namely, CD34+ stem cell and low-level leukocyte counting (37). Most importantly, laboratories with results within these boundaries of error can be selected as centers of excellence for institutional accreditation.

## ACKNOWLEDGMENTS

We acknowledge Myron Waxdal and Alessio Palini for useful discussions during the design of this study, Fabrizio Manca and Franco Celada for helpful criticism, and Dr. Janet Nicholson for critically reading the manuscript.

## LITERATURE CITED

1. CDC. 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescent and adults. *MMWR* 1995;41:1-35.
2. Fei DT, Paxton H, Chen AB. Difficulties in precise quantitation of CD4+ T lymphocytes for clinical trials: a review. *Biologicals* 1993; 21:221-231.
3. Nicholson JK. Quality control in immunophenotyping: U.S. efforts to establish common methodology and their impact. *American Society for Histocompatibility and Immunogenetics. Eur J Histochem* 1994;38 Suppl 1:7-12.
4. Fletcher MA, Azen SP, Adelsberg B, Gjerset G, Hassett J, Kaplan J, Niland JC, Odom-Maryon T, Parker JW, Stites DP, Mosley JW and the Transfusion Safety Study Group. Immunophenotyping in a multicenter study: the Transfusion Safety Study experience. *Clin Immunol Immunopathol* 1989;52:38-47.
5. Paxton H, Kidd P, Landay A, Giorgi J, Flomenberg N, Walker E, Valentine F, Fahey J, Gelman R. Results of the flow cytometry ACTG quality control program: analysis and findings. *Clin Immunol Immunopathol* 1989;52:68-84.
6. Rickman WJ, Waxdal MJ, Monical C, Damato JD, Burke DS. Department of Army lymphocyte immunophenotyping quality assurance program. *Clin Immunol Immunopathol.* 1989;52:85-95.
7. Martini E, D'Hautcourt JL, Brando B, Lawry J, O'Connor JE, Sansonetti F. First European Quality Control of cellular phenotyping by flow cytometry. Paris: Frison-Roche; 1990.
8. Gerber AR. Centers for Disease Control programs in flow cytometric immunophenotyping. *Ann N Y Acad Sci* 1993;677:40-42.
9. Homburger HA, Rosenstock W, Paxton H, Paton ML, Landay AL. Assessment of interlaboratory variability of immunophenotyping. Results of the College of American Pathologists flow cytometry survey. *Ann N Y Acad Sci* 1993;677:43-49.
10. Brando B, Sommaruga E. Nationwide quality control trial on lymphocyte immunophenotyping and flow cytometer performance in Italy. *Cytometry* 1993;14:294-306.
11. Barnett D, Granger V, Mayr P, Storie I, Wilson GA, Reilly JT. Evaluation of a novel stable whole blood quality control material for lymphocyte subset analysis: results from the UK NEQAS immune monitoring scheme. *Cytometry* 1996;26:216-222.
12. Vesely R, Barths J, Vanlangendonck F, Hannet I, Strauss K. Initial results of Central European Immunophenotyping Quality Control Program (CEQUAL). *Cytometry* 1996;26:108-112.
13. Gratama JW, Kraan J, Van den Beemd R, Hooibrink B, Van Bockstaele DR, Hooijkaas H. Analysis of variation in results of flow cytometric lymphocyte immunophenotyping in a multicenter study. *Cytometry* 1997;30:166-177.
14. Pandolfi F, Alario C, Girardi E, Rava L, Ippolito G, Kunkl A, Aiuti F. The Italian quality control study for evaluation of CD4 cells in centres involved in the treatment of HIV-1 patients. *Italian CD4 Quality Control Group. Clin Exp Immunol* 1998;111:564-573.
15. Bergeron M, Faucher S, Minkus T, Lacroix F, Ding T, Phaneuf S, Somorjai R, Summers R, Mandy F. Impact of unified procedures as implemented in the Canadian Quality Assurance Program for T lymphocyte subset enumeration. Participating Flow Cytometry Laboratories of the Canadian Clinical Trials Network for HIV/AIDS Therapies. *Cytometry* 1998;33:146-155.
16. Baudouin F, Sarda MN, Goguel A, Bene MC. Multicenter study of reference stabilized human blood for lymphocyte immunophenotyping.

- ing quality control in flow cytometry. *GEIL. Cytometry* 1999;38:127-132.
17. Batjer JD. The College of American Pathologists Laboratory Accreditation Programme. *Clin Lab Haematol* 1990;12 (suppl 1):135-138.
  18. CLIA Program; simplifying CLIA regulations relating to accreditation, exemption of laboratories under a state licensure program, proficiency testing and inspection—HCFA. Final rule. *Fed Regist* 1998;63:26722-26738.
  19. Calvelli T, Denny TN, Paxton H, Gelman R, Kagan J. Guideline for flow cytometric immunophenotyping: a report from the National Institute of Allergy and Infectious Diseases, Division of AIDS. *Cytometry* 1993;14:702-715.
  20. The Adult AIDS Clinical Trials Group (AACTG). Immunology Quality Assessment Program: program description and certification criteria. <http://aactg.s-3.com/iqa.htm>.
  21. Gelman R, Cheng SC, Kidd P, Waxdal M, Kagan J. Assessment of the effects of instrumentation, monoclonal antibody, and fluorochrome on flow cytometric immunophenotyping: a report based on 2 years of the NIAID DAIDS flow cytometry quality assessment program. *Clin Immunol Immunopathol* 1993;66:150-162.
  22. Mildvan D, Landay A, De Gruttola V, Machado SG, Kagan J. An approach to the validation of markers for use in AIDS clinical trials. *Clin Infect Dis* 1997;24:764-774.
  23. Li TS, Tubiana R, Katlama C, Calvez V, Ait Mohand H, Autran B. Long-lasting recovery in CD4 T-cell function and viral-load reduction after highly active antiretroviral therapy in advanced HIV-1 disease. *Lancet* 1998;351:1682-1686.
  24. 1994 revised guidelines for the performance of CD4+ T-cell determinations in persons with human immunodeficiency virus (HIV) infections. Centers for Disease Control and Prevention. *MMWR* 1994;43:1-21.
  25. Manca F, Li Pira G, Fenoglio D, Valle MT, Kunkl A, Ferraris A, Lancia F, Saverino D, Mortara L, Balderas R, Arp J, Dekaban G, Dalglish AG, Lozzi L, Theofilopoulos AN. Recognition of human T-leukemia virus (HTLV-1) envelope by human CD4+ T-cell lines from HTLV-1 seronegative individuals: specificity and clonal heterogeneity. *Blood* 1995;85:1547-1554.
  26. Kunkl A, Mortara L, Valle MT, Fenoglio D, Terranova MP, Megiovanni AM, Alessandrini A, Li Pira G, Mazzarello G, Del Bono V, Canessa A, Bassetti D, Manca F. Recognition of antigenic clusters of *Candida albicans* by T lymphocytes from human immunodeficiency virus-infected persons. *J Infect Dis* 1998;178:488-496.
  27. Kunkl A, Valle MT, Fenoglio D, Manca F. Quality control in immunophenotyping. use of human CD4+ and CD8+ T clones as standards. *Ann N Y Acad Sci* 1993;677:417-419.
  28. Kunkl A, Valle MT, Fenoglio D, Manca F. Accuracy assessment in flow cytometry quality control: use of cloned human T-lymphocytes as cellular standard. *FCS Forum* 1993;5:3-5.
  29. Giroto M, Facco G, Carubia F, Garbaccio Ioria G, Guerra MG, Lanteri M, Mari R, Saitta M, Mazzucco L. Fenotipizzazione linfocitaria: programma sperimentale di controllo di qualità interlaboratori. *La Trasfusione del Sangue* 1989;24:227-229 (Italian).
  30. Giroto M, Facco G. T-lymphocyte immunophenotyping: 6 years of quality control experience (abstract). *Eur J Histochem* 1993;37:28.
  31. Healy MJR. Outliers in clinical chemistry quality-control schemes. *Clin Chem* 1979;25:675-677.
  32. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37-46.
  33. Armitage P, Berry G. *Statistical methods in medical research*, 3<sup>rd</sup> edition. Oxford: Blackwell Scientific Publication; 1994.
  34. Hoover DR, Graham NM, Chen B, Taylor JM, Phair J, Zhou SY, Munoz A. Effect of CD4+ cell count measurement variability on staging HIV-1 infection. *J Acquir Immune Defic Syndr* 1992;5:794-802.
  35. Vedhara K, Nott KH, Richards SM. Individual variability in the reliability of CD4+ cell counts. *AIDS* 1995;9:98-99.
  36. Lyles RH, Munoz A, Xu J, Taylor JM, Chmiel JS. Adjusting for measurement error to assess health effects of variability in biomarkers. Multicenter AIDS Cohort Study. *Stat Med* 1999;18:1069-1086.
  37. Brando B, Barnett D, Janosy G, Mandy F, Autran B, Rothe G, Scarpati B, D'Avanzo GJL, Lenkei R, Schmitz G, Kunkl A, Chianese R, Papa S, Gratama JW. Cytofluorometric methods for assessing absolute numbers of cell subsets in blood. *Cytometry* 2000;42:327-346.
  38. Reimann KA, O'Gorman MR, Spritzler J, Wilkening CL, Sabath DE, Helm K, Campbell DE. Multisite comparison of CD4 and CD8 T-lymphocyte counting by single- versus multiple-platform methodologies: evaluation of Beckman Coulter flow-count fluorospheres and the tetraONE system. The NIAID DAIDS New Technologies Evaluation Group. *Clin Diagn Lab Immunol* 2000;7:344-351.
  39. Barnett D, Granger V, Whitby L, Storie I, Reilly JT. Absolute CD4+ T-lymphocyte and CD34+ stem cell counts by single-platform flow cytometry: the way forward. *Br J Haematol* 1999;106:1059-1062.
  40. Schnizlein-Bick CT, Spritzler J, Wilkening CL, Nicholson JK, O'Gorman MR. Evaluation of TruCount absolute-count tubes for determining CD4 and CD8 cell numbers in human immunodeficiency virus-positive adults. Site Investigators and The NIAID DAIDS New Technologies Evaluation Group. *Clin Diagn Lab Immunol* 2000;7:336-343.
  41. Gelman R, Wilkening C. Analyses of quality assessment studies using CD45 for gating lymphocytes for CD3+4+%. *Cytometry* 2000;42:1-4.
  42. O'Gorman MR, Nicholson JK. Adoption of single-platform technologies for enumeration of absolute T-lymphocyte subsets in peripheral blood. *Clin Diagn Lab Immunol* 2000;7:333-335.